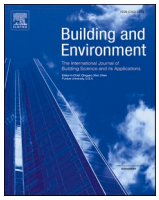
Building and Environment 225 (2022) 109661



Contents lists available at ScienceDirect

Building and Environment

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





Building occupancy modelling at the district level: A combined copula-nested hazard-based approach

Huiqiao Hou [a,*], Jacek Pawlak [a], Aruna Sivakumar [a], Bianca Howard [b]

[a] Urban Systems Lab and Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK
[b] Department of Mechanical Engineering, Columbia University, 220 S.W. Mudd Building 500 W 120th St, New York, NY, USA

ARTICLE INFO

Keywords:
Building occupancy modelling
Wi-Fi data
Copula approach
Competing risks hazard model
Urban heterogeneity

ABSTRACT

Planning and managing an energy system in a district require a comprehensive understanding and accurate modelling of people's occupancy and circulation among multiple buildings. Due to the lack of occupancy modelling tools for district scale analysis, energy models still use simplified occupancy patterns provided in building codes and standards. However, the simplified information restricts the reflection of complex occupancy patterns driven by urban heterogeneity. This paper fills this research gap and presents a hazard-based model combined with nested copula dependence to describe the complex occupants' interactions between buildings in a district, enabling the characterisation of irregular occupancy patterns in special cases. The proposed model is calibrated using Wi-Fi authentication data from the Imperial College London (UK) South Kensington campus and is validated using the following days of the same data by evaluating the performance of predicted occupancy patterns both on average and day by day. The validation results demonstrate that the model can accurately capture the effects of the urban environment on occupancy duration and choice of transition within a district. Mean Absolute Percentage Errors (MAPEs) of average-pattern predictions are between 7% and 16% for most buildings, though a bit lower in accuracy for the Library and Food Hall predictions with MAPEs of 32%–36%. We also discuss the contributions of the proposed occupancy model to potential future applications, including efficient building space use, local energy planning and management.

1. Introduction

The building sector is a substantial energy consumer giving rise to nearly 40% of global energy use [1]. An energy forecast for the period 2012–2040 suggests that global energy use for buildings will continue to increase by an average of 1.5% per year [2] because the demand for energy-consuming appliances grows and the need for heating or cooling surges because of climate change. Moreover, rules or policies to mitigate the spread of COVID-19 raise the demand for energy in the workplace. During the post-COVID-19 period, maintaining a healthy and safe built environment requires higher ventilation rates, eventually increasing the energy consumption of office buildings [3]. Heating, ventilation and air-conditioning (HVAC) is an essential service facility and dominates building energy consumption, resulting in almost 50% [4] of energy in commercial buildings. Inefficient building management due to the lack of knowledge of energy-use patterns, for example running facilities during unoccupied periods, is the main reason for high waste of energy in buildings.

Considering occupant interactions with building facilities and the built environment, and incorporating occupant behaviour in building energy analysis can help to save up to 50% of energy for single-occupancy offices [5] and broadly improve the energy performance of all types of buildings [6]. Among the multiple parameters of occupant behaviour, such as window control or air conditioning system control, building occupancy i.e. the presence or absence of different types of individuals is the basic and one of the crucial factors to consider in building energy models. It represents the occupant's presence, movement within and between buildings, arrival and departure at the disaggregated level, and can also be described as the number of occupants at the aggregated level. Moreover, building occupancy frequently varies with environmental factors to satisfy the occupants' physical, physiological, or psychological needs [7], resulting in dynamic building occupancy patterns accordingly. Both building energy models and rule-based or model-integrated HVAC control require accurate

* Corresponding author.
E-mail addresses: huiqiao.hou16@imperial.ac.uk (H. Hou), jacek.pawlak@imperial.ac.uk (J. Pawlak), a.sivakumar@imperial.ac.uk (A. Sivakumar), b.howard@columbia.edu (B. Howard).

https://doi.org/10.1016/j.buildenv.2022.109661
Received 16 July 2022; Received in revised form 13 September 2022; Accepted 29 September 2022
Available online 2 October 2022
0360-1323/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

modelling and comprehensive knowledge of building occupancy in response to urban environment heterogeneity. Therefore, enhancing the modelling of building occupancy has been a promising way to improve building energy performance [8].

Occupancy in multiple buildings at district and even urban scales has recently been the topic of interest to support urban building energy modelling. However, challenges in accessing and collecting data for the inference of the broader-scale building occupancy restrict more investigations in the context. In this situation, district-scale building energy simulation relies on building functionality-specific occupancy schedules provided in building codes and standards [9]. Nevertheless, the dynamic building occupancy is likely to be driven or influenced by the surrounding urban environment. Consequently, the average and fixed schedules in the codes and standards of most countries [10] are limited in describing the realistic occupancy patterns for dynamic energy control and analysis, all the more so when the surrounding urban environment changes significantly. For example, a one-day transport strike would impact arrival times and occupancy rate in office buildings, which could lead to shifts in occupancy schedule and even changes in occupancy demand on that day. In the existing literature, few studies have dealt with occupancy modelling beyond the scope of a single building.

1.1. Approaches to modelling building occupancy

Building occupancy modelling uses mathematical models to assist in interpreting and predicting people's stay and circulation patterns in building space. From the modelling standpoint, the occupancy question is generally addressed by three categories of model, i.e. stochastic, econometric and machine learning models. Table 1 lists and compares the features of these models, though they are currently used primarily to model a single building. The stochastic model (typically a Markov Chain approach) is the most popular family of models in the context of building occupancy, with good ability to replicate time series of random occupancy states. The econometric model, a typical example being the hazard-based model, focuses on interpreting how occupancy behaviour is affected by multiple explanatory variables. Finally, because of its high predictive performance, machine learning (ML) algorithms have seen increased use in short-term forecasting and recognising patterns in occupancy profiles. These models have their own advantages and disadvantages and are suitable for use in different scenarios as shown in the table.

Table 1
The classification of methods used to model occupancy within a single building.

Model Type	Stochastic Model	Econometric Model	Machine Learning
Motivation Primary purpose	Theory-driven Inference: To replicate how a set of occupancy locations or occupancy counts at one location changes randomly over time	Theory-driven Interpretation: To understand how one or more random variables are affected by other possibly non-random variables	Data-driven Prediction: To adaptively capture how one variable changes with other variables
Problem that fits for	In time-series	What-if analysis	Emphasis on accuracy of predictions
Distributional assumption	No need of assumptions	Relies on assumptions	No need of assumptions
Example	Markov Chain model (MCM) [11–14]	Hazard-based model [15]	Decision trees [16]
Model parameter interpretability	Easy	Easy	Difficult
Main applications in occupancy modelling and simulation	Stochastic generation of occupancy patterns	Correlation analysis and prediction for occupancy duration	Short-term forecasting or data filtering for occupancy patterns

From a geospatial standpoint, building occupancy modelling covers occupancy and circulation behaviour from within a building up to the city scale. Single building occupancy models have been developed for over two decades; hence, there is a rich body of literature and various approaches have been proposed to model it. By contrast, building occupancy analysis at the district or whole city level is an emerging field in building energy modelling [17], calling for more appropriate concepts and methods to solve the larger-scope problem. This study aims to model the broader-scale building occupancy in a strategic manner to provide building designers and managers with planning suggestions considering the urban configuration. With this perspective, an econometric model was developed to address the modelling goal.

1.2. Broader scale building occupancy studies

In light of improving the energy efficiency at the district and urban scale, modelling a single building is insufficient and fails to incorporate the interactions between buildings [18]. Urban Building Energy Modelling (UBEM), initially defined by Reinhart and Davila [19], was therefore proposed to evaluate the performance of a group of buildings in the urban context. As a growing branch of building energy modelling, UBEM scales up building modelling and simulation from a single building to a district and even an urban area [17]. Importantly, UBEM considers all the buildings in a city as an integrated system rather than simply summing the energy consumption of all urban buildings, as well intends to capture the interconnections between buildings and building interactions with the urban environment (typically with microclimate) [17].

In line with the concept of interactions between occupants and the built environment, accurate UBEM and a better understanding of building energy use in the urban context requires accurate and realistic occupancy patterns. In comparison with the rich body of literature describing isolated occupancy simulation for a single building (methods are categorised in Table 1), only few studies (listed in Table 2) have addressed the modelling of inter-building occupancy and transition at community and urban levels [20]. In existing practice, modellers and analysts still implement pre-defined deterministic occupancy profiles specified for building functionality (e.g. Ref. [21]) in UBEM to support energy-related building design, policy-making and demand control process. Although the deterministic profiles provided by building codes have been derived by statistical means from real-world data, they are by definition generic and simplified, and thus not fit to describe irregular occupancy patterns or their dynamics as a function of external factors. Therefore, modelling diversity and complexity of inter-building occupancy and circulation, possibly driven by heterogeneous urban surroundings, presents a research gap.

Accessing data that is suitable for measuring the broader-scale occupancy at high temporal and spatial resolutions has posed a challenge to modelling occupancy across multiple buildings, affecting UBEM studies. In recent years, some research has explored using travel data to model larger-scale building occupancy (see Table 2). Before the context of UBEM was proposed, Rakha et al. [18] investigated using activity-based travel survey data to generate urban-scale building occupancy schedules. Instead of travel surveys, urban mobility sensing data was also used to reflect more realistic patterns, such as positioning records on social media [25,27], mobile phone data [20,26] and location-based service data [9]. All of the studies mentioned above use the travel data as a proxy for occupancy in buildings on the broader urban scale, which reveals the explicit connection between travel and occupancy in the process of generating activity profiles. Therefore, this

Table 2
Summary of building occupancy studies at the broader scale.

Author(s), year	Case study area	Data source used	Aggregated or disaggregated analysis	Modelling work or not	Method	Implementation
Marins & Roméro (2013) [21]	District and neighbourhood levels	Occupancy rate, population density and area of building units	Aggregated	No	Deterministic model	Energy modelling for integrated urban planning involving urban morphology, urban mobility and building
Rakha et al. (2014) [18]	Urban area	Activity-based travel survey data	Disaggregated	No	K-means clustering	Generate building occupancy schedule (occupancy count)
Van Der Spek et al. (2015) [22]	Chosen buildings on campus	Wi-Fi data	Disaggregated	No	Rule-based method	Generate occupancy and movement patterns in and between campus buildings
Bon et al. (2016) [23]	Campus	Wi-Fi data	Disaggregated	No	Rule-based method	Identity features of occupancy and movement between campus buildings
Griffioen et al. (2017) [24]	Campus	Wi-Fi data	Disaggregated	No	Rule-based method	Assign circulation flows to corridors
Mohammadi & Taylor (2017) [25]	City	Positioning record on Twitter	Aggregated	No	–	Infer occupants' flows and study their relationship with the fluctuations of building energy use in the city
Barbour et al. (2019) [26]	Urban area	Mobile phone data	Aggregated	No	Probabilistic occupancy assignment	Generate urban-scale building occupancy pattern (occupancy rate)
Kang et al. (2019) [27]	Region	Active positioning requests on social network of phones	Aggregated	No	Clustering methods	Extract typical daily and weekly occupancy patterns (occupancy rate)
Wu et al. (2020) [20]	Downtown area of a city	Mobile phone data	Aggregated	No	Statistical method -Discrepancy score	Generate urban-scale building occupancy pattern (occupancy rate)
Happle et al. (2020) [9]	City	Location-based service data	Aggregated	No	–	Generate representative occupancy schedule (occupancy rate) for different building types
Wang et al. (2020) [28]	All commercial buildings in a city	Transport accessibility and population data	Aggregated	Yes	Gaussian Mixture Model	Measure and fit occupant density distributions in commercial buildings
Mosteiro-Romero et al. (2020) [29]	Campus	University register data	Disaggregated	Yes	Agent-based transportation model	Disaggregate-level occupancy schedule (occupancy count and rate) simulation

paper proposes to incorporate a travel demand model into building occupancy models to generate first arrivals[1] among multiple buildings, which also demonstrates a potential link between buildings and the transport system that could result in urban heterogeneity.

Regarding the existing modelling methodologies for the broader context of building occupancy across multiple buildings, Wang et al. [28] applied a Gaussian Mixture Model using transport accessibility and population data to fit the distributions of occupant density in urban commercial buildings. Mosteiro-Romero et al. [29] adapted the modelling framework and procedure of an agent-based transportation approach, called MATSim (Multi-Agent Transport Simulation), using University registry data to simulate occupancy schedules in official and educational buildings. However, their studies suffered from two limitations. First, employee and student registry data used in the study was not always suitable for simulation of the campus buildings. For example, the COVID-19 pandemic caused significant changes in people's occupancy and circulation behaviour, resulting from the trend to mix remote and in-person work or study. But registry data only contained the arrangement of work and study activities with fixed schedules over a term, rather than the realistic and diverse patterns driven or influenced by other factors. Second, the proposed modelling approach only stochastically replicated the occupancy schedule rather than capturing causal reasons for occupancy difference resulting from urban heterogeneity, which restricts the accurate modelling for irregular occupancy patterns, typically when the surrounding context significantly changes. Therefore, understanding how occupants respond to the changes in the surrounding environment from real-world tracing data is necessary for refining occupancy simulation and enabling its reflections on different urban scenarios.

[1] The first arrival at a building represents the individual occupancy state which is detected for the first time for an individual during a day.

1.3. Objectives and scope of this study

The literature review above underpins the need for an approach that can achieve accurate and interpretable modelling for complex inter-building occupancy and circulation at the broader urban scale. Therefore, the methodology proposed in this paper will address two challenges. First, it can explain interactions between buildings and the urban environment, addressing the shortcoming of models that may provide good predictions but low interpretability of how and why such predictions arise. Second, it provides a new simulation tool using sensing data to generate occupancy patterns of multiple buildings that preserves inter-building interactions, enabling predictions of regular and irregular occupancy for strategic planning and dynamic energy controls.

Although only few studies have explored occupancy modelling beyond a single building, approaches widely used for analysis at the single building level (see Table 1) can be further adapted to fit the broader context. For example, the time-dependent MCM was converted to simulate campus-wide building occupancy schedules despite being the benchmarking approach [29]. ML methods could be applied in future to filter data streams and forecast occupancy patterns in multiple buildings relying on big data. However, these two families of approaches do not have the aforementioned properties (interpretability, flexibility in accounting for urban interactions, preserving inter-building interactions). With this perspective, an econometric model (that will be elaborated further in section 2) is proposed that meets such requirements.

In addition, we demonstrate how the proposed framework can be used to model the interactions between transport and building systems, by integrating it with an activity-based travel demand simulator [30] that generates first arrivals in multiple buildings in the same district. Finally, we discuss potential applications of the occupancy approach from multiple perspectives to improve the efficiency of facility

operations.

2. Modelling methodology

In our previous paper [15], a hazard-based model (HBM) was applied to model occupancy duration and choice of occupancy location jointly at the single-building level. Importantly, this approach provides a flexible framework for covariate specification and statistical inference, thereby revealing interactions between building occupancy and other systems, including the urban surroundings. However, the assumption of independent transitions between two locations appears too strong to describe the diversity and complexity of inter-building occupancy and transition behaviours, especially if their roles are somehow linked, e.g. being part of a campus or a shopping district. Therefore, this paper proposes to combine a copula approach [31] and HBM [15] to relax the assumption and characterise dependence between competing transitions between different locations.

2.1. Fundamentals of combined copula-nested hazard-based model

The problem and the proposed methodology are best to be described in the simplest example of two-dimensional space. An example in Fig. 1 illustrates that an individual is currently at physical location a (could be a building or indoor zone) and will leave the location at a certain time-point on the interval $[t_1, t_2]$; Meanwhile, the individual will choose location b as the next occupancy location. Based on the principle of competing risks HBM given by Ref. [32], we assumed a transition from location a to b happened at one time-point during an arbitrary small interval $[t_1, t_2]$ and the other transition from a to c had survived until t_2 (in other words, the termination of occupancy with destination c would happen at one time-point after t_2). Therefore, the probability of the happened transition equals a joint probability distribution of cumulative density function (CDF) of transition from a to b between t_1 and t_2 and survival function (SF) of transition from a to c till t_2 (see shaded regions in Fig. 1), defined by,

$$P_{a\to b}(t) = \int_{t_1}^{t_2} \int_{t_2}^{\infty} f(t_{a\to b}, t_{a\to c})\, dt_{a\to c} dt_{a\to b} \tag{1}$$

As the number of possible destinations rises, the dimensionality of the joint duration distribution will increase, which poses a difficulty with the calculation of multiple integrals. Additionally, to establish and capture dependence between transition-specific duration distributions, survival copula [33] was used to replace the two-integral function in Eq. (1) and simplify the calculation of joint distribution. Accordingly, the probability of transition from a to b at one time-point between t_1 and t_2, on conditional that transition from a to c had not happened until t_2, is given by,

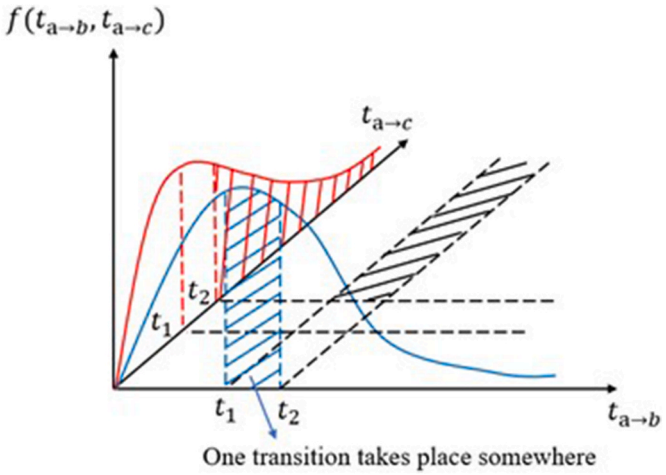


Fig. 1. Joint probability distribution for a two-dimensional case.

$$P_{a\to b}(t) = C(S_{a\to b}(t_1), S_{a\to c}(t_2)) - C(S_{a\to b}(t_2), S_{a\to c}(t_2)) \tag{2}$$

When generalising the equation to a scenario with multiple competing transitions for one occupancy state, it is necessary to consider the structure of dependence between all possible transitions. A nested copula structure was proposed to describe dependence between transitions at two hierarchical levels (see Fig. 2). The lower-level copula, C_T, establishes the correlations of competing transitions to all possible buildings except the origin. The upper-level copula, C_{TW}, considers the dependence between competing transitions inside a building and to other buildings. Incorporating the nested copula dependence, the transition probability from building a to b can be given by:

$$\begin{aligned} P_{a\to b}(t) = {} & C_{TW}(S_{a\to a}(t_2), C_T(S_{a\to b}(t_1), (S_{a\to c}(t_2))^{\delta_{access}}, \ldots, S_{a\to m}(t_2))) \\ & - C_{TW}(S_{a\to a}(t_2), C_T(S_{a\to b}(t_2), (S_{a\to c}(t_2))^{\delta_{access}}, \ldots, S_{a\to m}(t_2))) \end{aligned} \tag{3}$$

The joint survival probability at one place within building a is:

$$S_a(t) = C_{TW}(S_{a\to a}(t_2), C_T(S_{a\to b}(t_2), (S_{a\to c}(t_2))^{\delta_{access}}, \ldots, S_{a\to m}(t_2))) \tag{4}$$

where δ_{access} is an indicator for whether an individual had been authorised to access the destination building. $\delta_{access} = 1$ means the individual with the access permission to the building and hence the corresponding transition to that building is one of the competing risks in modelling. In the current context, t_1 and t_2 are times±1-min of occupancy duration. Moreover, the duration variable was converted into hours to facilitate parameter estimation.

In the copula literature, various parametric copula functions with distinct properties have been developed to describe the dependence between marginals. In this study, the multivariate Farlie-Gumbel-Morgenstern (FGM) copula was used to capture the correlation between competing transitions. Mathematically, the FGM copula describes the joint distribution of m random uniformly distributed variables as follows [31]:

$$C_{FGM}(u_1, u_2, u_3, \ldots, u_m) = \prod_{m=1}^{M} u_m \left[1 + \sum_{m=1}^{M-1} \sum_{m'=m+1}^{M} \theta_{mm'}^{FGM} (1 - u_m)(1 - u_{m'}) \right] \tag{5}$$

$$-1 \le \theta_{mm'}^{FGM} \le 1 \ \forall m, m' \in M \tag{6}$$

$$\theta_{mm'}^{FGM} = \frac{2}{1 + exp\left(\gamma_{mm'}^{FGM}\right)} - 1 \tag{7}$$

where $\theta_{mm'}^{FGM}$ is the dependence parameter, specific to the pairwise combination of variables u_m and $u_{m'}$, which can capture the correlation between two marginal distributions. The FGM copula has some properties suitable for this application over other copula functions. Firstly, compared to the Archimedean family of copulas, it has a flexible structure with the capability of specifying dependence for each pairwise combination of variables. At the same time, dependence parameters

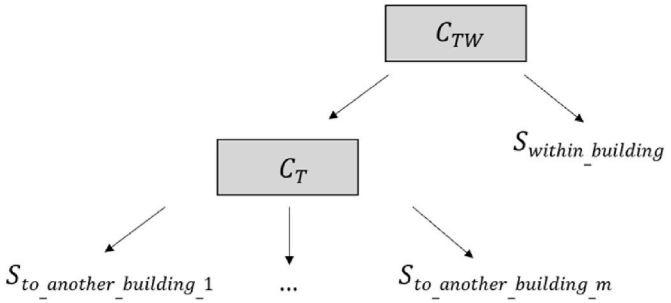


Fig. 2. Nested structure of copula dependence between competing transitions.

with values between −1 and 1 can describe correlations at two tails, though only correlations up to $\left|\frac{1}{3}\right|$ can be captured. Finally, the closed-form analytic expression of the FGM enables the use of the maximum likelihood technique for model estimation and avoids complex and computationally-heavy simulation-based estimation [34].

The dependence parameter of the FGM copula was parameterised using the linear regression similar to the specification for the shape parameter of SFs [15]. That is:

$$= \sum_i^I \sum_{a=1}^{A} \sum_{b=1}^{B} \left\{ d_{i,a,b} \left[(1 - d_{i,c}) log\left(P_{a \to b}\left(t_{i,a,b} \middle| X_{i,k}, X_{i,a,b,j} \right) \right) + d_{i,c}\ log\left(S_a\left(t_{i,a,b} \middle| X_{i,k}, X_{i,a,b,j} \right) \right) \right] \right\} \tag{9}$$

$$\gamma_{mm'}^{FGM} = \gamma_0^{FGM} + \sum_{q=1}^{Q} \gamma_q^{FGM} X_q \tag{8}$$

Specifying covariates in the copula dependence intended to capture the exogenous effects on the competing relationships between possible transitions. As the FGM copula has the limited ability to capture very strong dependence, copula constant terms were not parametrised and γ_0^{FGM} was fixed as 0, ensuring the successful parameter identification. On the other hand, this restriction meant that correlations between competing transitions exist conditional on the fact that exogenous factors impact the joint function. The exogenous effects on marginal distribution were captured by specifying covariates in single marginal functions following the application of HBM [15].

The extended approach to accommodating the nested copula structure is called a nested hazard-based model (NHBM). Correlating the competing transitions provides the possibility of linking and interpreting the indirect effects on marginal functions. Specifically, a marginal effect on one duration distribution for one transition could impact that for another transition without the marginal effect. Therefore, this approach allows us to gain insights into the more complex dependence pattern of building occupancy on the district scale.

Table 3
Parameters used in the formulation of NHBM.

Parameter	Description
t	Occupancy duration until a transition (to another zone or building) takes place
a	Origin building (occupied building)
b	Destination building (next occupied building)
I	Sample size
A	The number of origin buildings
B	The number of destination buildings
$t_{a \to b}$	Duration (hours) until the occupied state ends and then a transition from building a to b occurs
$S_{a \to b}$	Marginal survival function specific to transition from building a to b
$P_{a \to b}$	Joint probability of transition from building a to b and survivals of competing transitions
S_a	Joint survival at building a
X_q	The q^{th} covariate specified in copula functions
$d_{i,a,b}$	Dummy variable that represents the combination of occupied building a and destination building b
$d_{i,c}$	Dummy variable that directs whether i^{th} sample lacks destination ($d_{i,c} = 1$) or not ($d_{i,c} = 0$)
$\theta_{mm'}^{FGM}$	Copula dependence parameter that describes marginal functions m and m' using FGM approach
$\gamma_{mm'}^{FGM}$	A parameter in the sigmoid restriction function for capturing copula dependence $\theta_{mm'}^{FGM}$
γ_0^{FGM}	A constant term in the linear-in-parameters function $\gamma_{mm'}^{FGM}$
γ_q^{FGM}	A coefficient of the q^{th} covariate in the linear-in-parameters function $\gamma_{mm'}^{FGM}$

2.2. Model calibration

Using the joint probability function in Eq. (3) and the joint SF in Eq. (4), the log-likelihood function with the sample of I transition observations for model estimation is given as follows (see Table 3 for a description of full parameters):

$$LL(\gamma, \omega, \varphi, \lambda \,|t, X)$$

Wi-Fi authentication logs gathered on the case study area during the whole of October 2021 and urban variables (that will be described in section 3.2) will be used to calibrate the proposed model. The parameters of marginal and copula dependence were determined based on the standard and direct maximum likelihood estimation method. The standard errors for two types of dependence parameters were also approximated by evaluating the numerically approximated Hessian matrix. Judging the dependence relationships within one SF and between SFs follows the statistical significance of associated parameters. Model estimation was undertaken in R, by running an optimisation function with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm in 'optimx'.

2.3. Simulation procedure

Unlike past studies in the literature, we generated first arrival states from an activity-based travel demand simulator developed in Ref. [30], with respect to different activities (representing people by roles, respectively). Notably, the number of people with different roles in the simulation depends on their empirical shares in the population.

The district-level occupancy simulation generally followed the microsimulation workflow given in Ref. [15] but at the building-specific locations rather than indoor zones shown in Fig. 3. Different from the previous HBM simulation, some necessary modifications to equations and rules were needed for the NHBM simulation. First, it is possible to observe occupants with different roles who take on different activities and may need the access authority to special buildings in the area. Therefore, during the simulation, only destinations accessible for the respective occupants (especially for buildings with access gate control) were incorporated as the possible choices of destinations for each transition. Moreover, the joint SF was no longer the product of marginal survivals because the copula approach was applied to capture the dependency between competing transitions. Accordingly, the probability of occupancy duration at building a in the duration simulation is given by:

$$P_i(t|\ a) = 1 - S_a(t)$$

$$P_i(t|\ a) = 1 - C_{TW}(S_{a \to a}(t_2), C_T(S_{a \to b}(t_2), (S_{a \to c}(t_2))^{\delta_{access}}, \ldots, S_{a \to m}(t_2))) \tag{10}$$

Unlike the calculation of univariate SF, the joint CDF does not equal the difference between 1 and the copula-based joint SF [33]. To incorporate the correlation between two competing hazards (i.e. transitions), the probability of transition from building a to b given that a transition occurred at time t_c in the choice simulation is:

$$P_i(b|\ t = t_c) = \frac{P_{i,a \to b}(t_c)}{\sum_{e=1}^{B} \left(P_{i,a \to e}(t_c) \right)} \tag{11}$$

The simulation for the whole population in the case study area was

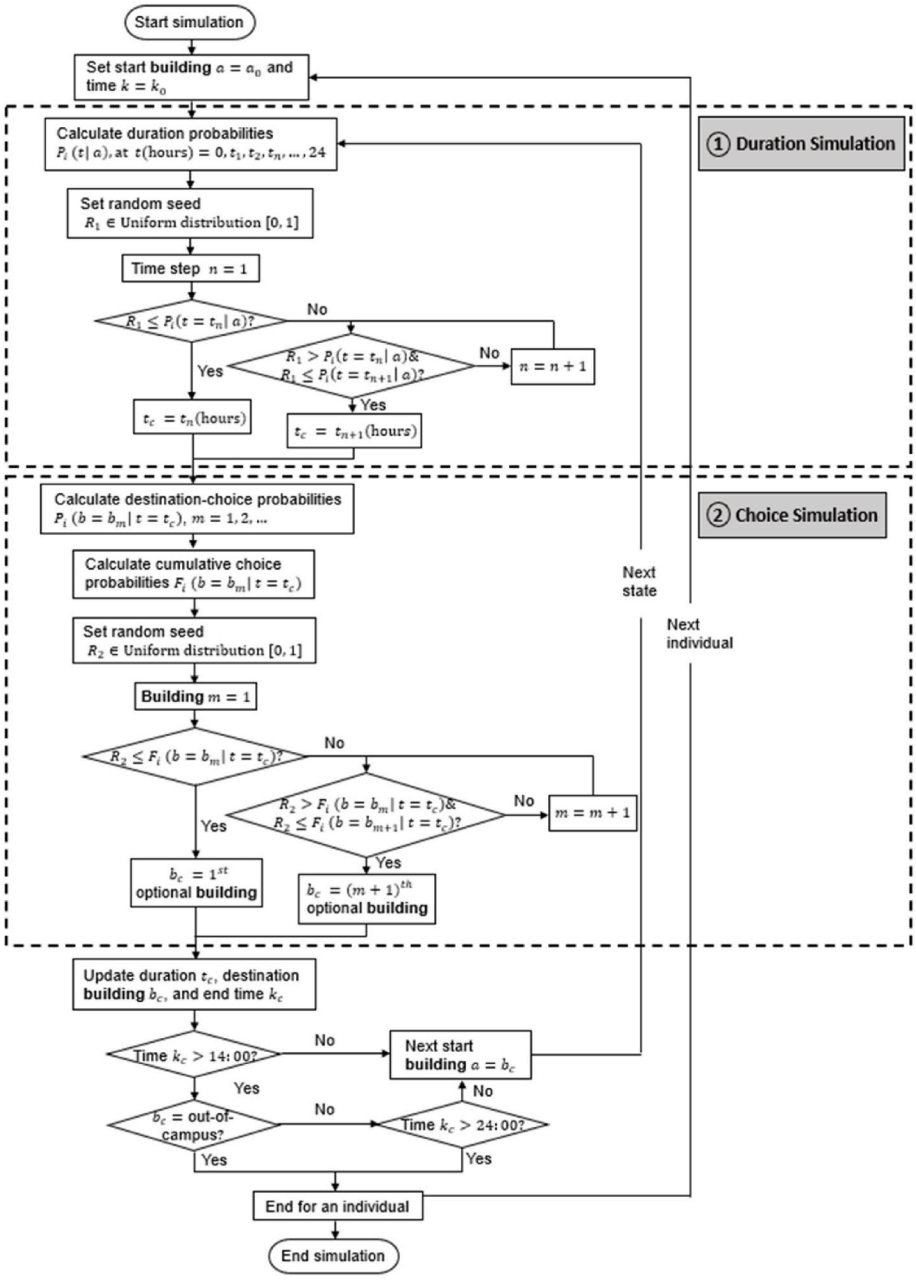


Fig. 3. The workflow of NHBM occupancy simulation.

generated 100 times. The outputs were used to validate the prediction performance of the NHBM.

3. Empirical context and validation

3.1. Case study campus and its population

South Kensington (SK) campus of Imperial College London, UK, is the case study area on which the modelling methodology was implemented and validated. The case study area was chosen for multiple reasons: 1) Wi-Fi infrastructure is broadly installed in campus buildings, which enables the availability of sensing data for measuring disaggregate building occupancy at the campus level; 2) a huge number of occupants on the campus contributes to a big sample size of occupancy states, helping us better determine the patterns; 3) the heterogeneity of occupants and building functionalities on the campus allows us to explore more knowledge of occupancy diversity and complexity; 4) the campus in central London has close connections with urban systems, which is a good fit for exploring the effects of urban heterogeneity. Overall, the case study campus can provide a representative sample for statistical analysis, due to its diversity of occupants and building functions as well as a suitably large sample size. Therefore, the findings from this study would contribute to the knowledge beneficial for wider application of the proposed methodology in urban contexts, especially campus-like contexts, including other university campuses, shopping districts or business districts.

In total, the 35 buildings on the campus can be classified into 7 types of functions and most of the buildings involve mixed uses for academic teaching and research as well as office work (see Fig. 4). The opening hours policy on campus specifies that teaching and office buildings are open to authorised people during 7:00 and midnight on weekdays. Restaurants serving lunch for college people are open between 11:45 and 14:30 and are based in a single building (called Sherfield building), while several office buildings operate cafés for longer periods. One

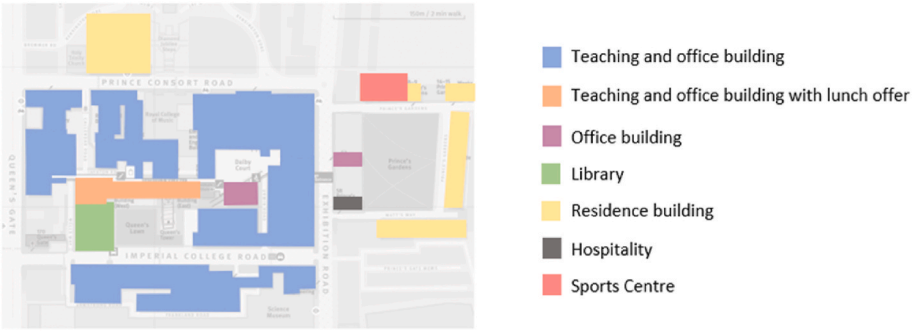


Fig. 4. Distribution of multi-type buildings on the campus map.

administrative building (with purple in Fig. 4) located in the centre of the campus is only authorised to the access of administrative staff. The central Library, the only building open for 24 hours, provides not only book loan service but also multi-types of spaces for self-and group-study. The campus also offers 8 accommodation buildings, including more than 985 beds for undergraduate and summer visiting students. All campus buildings have been covered by a Wi-Fi network via widely installed access points (APs) within buildings (more than 250 APs across the campus), allowing for the detection of continuous traces for Wi-Fi-enabled devices on campus. To facilitate the computation, these campus buildings were aggregated into 10 groups in terms of their topological connections.

The period of the case study is between October and November 2021. Since the 2021 autumn term started on 4 October, the total number of people on the SK campus approached 12,000 on Tuesday and Thursday but dropped to around 10,000 on Wednesday (in Fig. 5 (a)). As per the college timetabling policy, the more likely explanation for this is that Wednesday afternoon is usually free from teaching activities and hence fewer people go to the campus. It should be noted that the case study period is the time when the campus first reopened for in-person teaching after the COVID-19 pandemic. The hybrid model of in-person and remote teaching/working reduced occupancy count on campus. Regarding first arrivals at the campus level, Fig. 5 (b) shows the maximum arrival count at 9:00 and a slight rise at 14:00.

The population of the SK campus includes five categories of occupants who are undergraduate (UG) students, postgraduate (PG) taught students, PG research students, staff and others (including visitors, guests, alumni etc.). Looking at the chart in Fig. 5 (c) that we plotted using the Wi-Fi logs across the SK campus, UG students are the largest group (i.e. 45.6%) of the population, whereas other categories have close shares between 11% and 16%. Moreover, 59.2% of the population (UG and PG taught students) is not arranged with either individual or shared offices; hence their occupancy locations are more flexible except for attendance to their courses. In contrast, staff and PG research students having offices are more likely to spend most of their time during the day in their office buildings, though the occupancy of some people who take responsibility for teaching is driven by course arrangement.

3.2. Data collection and processing

Increasing installations of Wi-Fi infrastructure provide an opportunity for tracking people's behaviour in both indoor and outdoor spaces, and have therefore been used in several analyses of large-scale occupancy and movement [23,35–37]. These applications demonstrate that

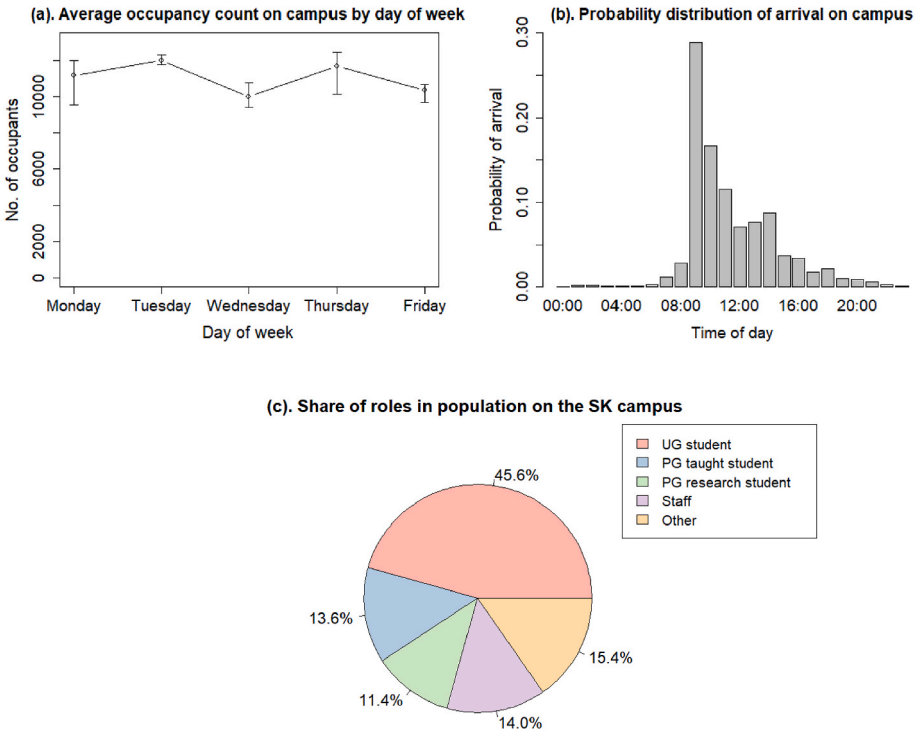


Fig. 5. Occupancy characteristics on campus.

Wi-Fi infrastructure provides a feasible low-cost way to collect individual's mobility and occupancy beyond the within-building context and even on a large city scale.

Anonymised Wi-Fi authentication data collected on the case study campus between October and November 2021 was the only data source used for calibrating and validating the proposed approach. An authentication event happens if a device needs the connectivity to an AP but has not yet been authenticated with the AP. They are recorded when the AP receives the authentication request. Although wireless clients would re-authorise every 30 min by default setting, more frequent re-authorisations take place under certain situations, especially when roaming between AP groups. In contrast, Wi-Fi association data used in the previous study [15], reflecting a set of clients under each AP, was passively collected at a fixed interval (i.e. 10 min) instead of instant update. In comparison with the Wi-Fi association logs, the authentication data was recorded irregularly with the smallest interval of 1 s. The data with such fine-grained temporal information allows us to infer and distinguish individual occupancy and passing-by behaviours on campus. Moreover, it not only gives the reason for one-to-many user-AP associations (duplication issue), which occur when using Wi-Fi association data, but can also eliminate the issue in this study.

In addition to the Wi-Fi authentication data, exogenous covariates reflecting urban heterogeneity, hypothesised to be correlated with marginal occupancy distribution and copula dependence between two marginals in the NHBM, were gathered in line with our urban interaction framework [15]. We picked up at least one typical factor to represent the exogenous impacts of each urban system on building occupancy. Most of these variables was downloaded or scraped from official websites. All exogenous factors used in the empirical analysis are summarised as follows:

- Transport system: South Kensington underground station is the closest station to the case study area, where three lines serve the high demand for travel in and out of the SK area. The transport covariate used in this study is the planned and real-time underground line service data for October 2021, which was scraped from the Transport for London website using APIs. The line service data indicates the period and scope of track affected by either severity of delays or closures.
- Events: Many academic and social events are held in multiple buildings on campus, such as colloquium, workshop, and lunch-time concert. Event venues and times were collected and incorporated to analyse their impacts on campus-wide individual circulation and occupancy patterns.
- Retail stores: Senior Common Room (SCR) restaurant is a food retailer open to college staff and students during lunchtime of every workday (11:45–14:30). Its opening time was used as an independent variable in the model.
- Regulatory policy: Opening hours and timetabling policies were included in the model, considering their significant correlations with occupancy count and possible impacts on occupancy pattern. For example, teaching activities are not arranged every Wednesday afternoon, and UG students have flexible schedules to join club activities on campus.
- Weather: External air temperature and rainfall data was gathered by connecting the API of the Met Office (UK's national weather service). In the dataset, temperature and type of weather (i.e. involving type of rainfall) were two important factors to be included in the model as independent variables.
- Building space information: Campus-wide study involves the modelling of buildings with different functionalities, of which layout covariates were gathered from Imperial building and space information database (called Pythagoras) to distinguish occupancy and energy-use patterns of buildings.

3.3. Model validation

Building occupancy models for multiple buildings in a district are still under development, and hence there is no typical method that can be used as the benchmark. In this paper, we adapted the time inhomogeneous MCM with building-level states [29] to be the benchmark for validating the proposed NHBM approach. The proposed NHBM approach is validated from the perspective of goodness-of-fit, interpretation and predictive ability as described below.

First, the full (inclusive of covariates) NHBM that captures copula dependence and marginal effects was compared with full HBM (with only marginal effects) and baseline HBM (with only constants) to demonstrate its improvement in fitting the training data. Overall, three commonly used measures of goodness-of-fit based on log-likelihood values at convergence were used to evaluate the performance of fitting the training data: Akaike information criterion (AIC), Bayesian information criterion (BIC) and likelihood-ratio (LR) test.

Second, we interpret the estimated parameters of the full NHBM from the statistical standpoint, revealing the indirect exogenous impacts on occupancy duration and transitions within the case study campus. Finally, simulation and prediction were validated in future unseen data (from the perspective of the model, which was calibrated on October 2021 data) during workdays between 1 and 5 November 2021. The prediction accuracy of NHBM and MCM for the unseen dataset was evaluated using two metrics, i.e. Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). MAE, given in Eq. (12), which measures the average of absolute errors, was used to assess the average difference between model outputs and Wi-Fi observations. MAPE, shown as Eq. (13), measures MAE normalised to observations and has intuitive interpretation in terms of relative errors. It expresses the percentage difference between model outputs and observations.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|Pred_t - Obs_t| \tag{12}$$

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{Pred_t - Obs_t}{Obs_t}\right| \tag{13}$$

4. Empirical occupancy analysis and discussion

This section will discuss the empirical and modelling results according to the model outputs. Section 4.1 evaluates how the proposed NHBM fits the calibration dataset and how the estimation results reflect the drivers of occupancy. Section 4.2 assesses the model performance in predicting the occupancy patterns averaged over the validation days. Section 4.3 evaluates how the proposed NHBM captures the changes in occupancy patterns between days and provides suggestions for model improvement in future applications.

4.1. Model fitting and interpretation of exogenous effects

Various measures of goodness-of-fit were used to evaluate the performance of model fitting for the training data, of which statistics for full NHBM, full HBM and baseline HBM are given in Table 4. The final log-likelihood values at the convergence of three models were translated into AIC, BIC and LR statistics. The decreasing AIC and BIC statistics confirm the modelling improvement for which both copula dependence and marginal external covariates were included in the model. In the LR test, a much larger statistic (17750.031) than the critical chi-square value (11.070) with 5 degrees of freedom (i.e. number of copula parameters) at the 99% confidence level is evidence to indicate a statistically and significantly better fit of the NHBM.

Exogenous covariates were specified both in marginal SFs and copula functions. Therefore, the NHBM can capture the external effects on the shape of transition (also occupancy) distribution as well as describe the effects on the dependence between distributions of two competing

Table 4
Estimated parameters of the best fitted NHBM with multiple exogenous variables and model fit statistics (sample size I = 682,719).

	Covariate	Transitions within the building	Transitions to another building	Entrance to the campus and a building	Leaving the campus
Marginal survival or hazard	The proportion of office area in a building	−7.046***	−7.046***	−7.046***	−7.046***
	The proportion of residence in a building	−3.632***	−3.632***	−3.632***	−3.632***
	External air temperature (°C)	–	−1.750***	–	−0.343***
	Rain	–	−0.007	0.167***	0.037***
	During the transport event	–	–	–	−0.024**
	00:00–07:00 (workdays; reference time)	0 (fixed)	0 (fixed)	0 (fixed)	–
	07:00–08:00 (workdays)	0.276***	−0.573***	−0.573***	–
	08:00–10:00 (workdays)	0.196***	−0.475***	−0.475***	–
	10:00–SCR lunch open (workdays)	0.328***	−0.070***	−0.070***	–
	SCR lunch open period (workdays)	0.281***	1.081*** (to college restaurant) 0.178*** (to others)	0.178***	–
	SCR lunch close–16:00 (workdays)	0.229***	0.272***	0.272***	0.426***
	16:00–18:00 (workdays)	0.324***	0.316***	0.316***	0.738***
	18:00–21:00 (workdays)	0.327***	−0.211***	−0.211***	0.629***
	21:00–24:00 (workdays)	–	–	–	0.919***

	Covariate	Transitions within buildings and to another buildings			Transitions to different buildings		
		γ_q^{FGM}	τ_θ	ρ_θ	γ_q^{FGM}	τ_θ	ρ_θ
Copula dependence	07:00–08:00 (workdays)	6.236***	−0.221	−0.332	–	–	–
	Staff	0.637***	−0.068	−0.103	–	–	–
	30 min before on-campus conferences or events	–	–	–	0.262**	−0.029	−0.043
	Wednesday timetabling policy	–	–	–	−0.400***	0.044	0.066
	16:00–18:00 (workdays)	–	–	–	−0.487***	0.053	0.080

	NHBM full covariates (with copula)	HBM full covariates (independence)	HBM baseline (only constants)
Final Log-likelihood (LL)	−3908860.824	−3917735.840	−4464758.662
Number of parameters	65	60	32
AIC	7817851.649	7835591.679	8929581.325
BIC	7818594.848	7836277.710	8929947.208

$***p \leq 0.010$ $**p \leq 0.050$ $*p \leq 0.100$

transitions. In addition, to simplify the model complexity, SF was specified to the broad categories of transitions, including transitions within a building, transitions to another building, entrance to the campus and a building, and leaving the campus. The estimated coefficients and corresponding statistical significance of covariates specified in the NHBM are shown in Table 4.

By observing the effects on marginal hazard or survival functions (see Table 4), a majority of coefficients are found to be statistically significant, apart from rain impact on transitions to another building, though the effects of transport events are accepted at only 95% confidence level. These coefficients reveal how the corresponding covariates impact the shape of transition-specific occupancy probability density function (PDF), individual occupancy duration and choice of transition accordingly. Specifically, a positive coefficient suggests the shorter occupancy duration at the given origin and the higher rate of transition from the origin to a destination for an individual, and vice versa.

The covariates of copula dependence can reveal how external variables impact the correlation between two marginal functions and further drive the change in joint distribution. As mentioned in section 2.1, the nested structure of copula dependence was introduced to describe the correlations between transitions at different scopes, i.e. within a building and between two buildings. Table 4 shows statistically significant parameters and the converted correlation coefficients, Kendall's tau (τ_θ) and Spearman's rho (ρ_θ) [32], in copula functions. A positive parameter

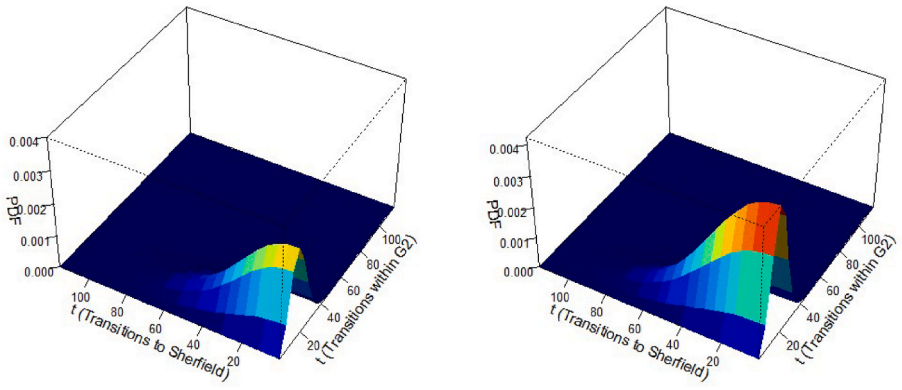


Fig. 6. Joint distribution of one happened transition and one survived transition for staff and other types of occupants.

(also negative copula dependence) suggests that the higher probability of a transition-specific SF is associated with the lower probability of that for the other competing transition and verse vice.

The period of 07:00–08:00 and being a staff member both drive the negative copula dependence between transitions within the building and to another building. Fig. 6 gives an example to show how the role of an individual impacts the joint distribution of their occupancy timings specific to transitions within Group 2 buildings and transitions to the Sherfield building (with lunch offers). Compared to other occupants, the staff appears to have a significantly higher joint PDF of transition within Group 2 buildings when it competes against the transition to Sherfield. In other words, the negative copula dependence for staff increases the difference between two competing strengths. At the same time, incorporating such copula dependence with the effects on marginal survivals allows us to interpret the indirect effects of those exogenous variables on an unspecified survival of transition. For example, the marginal effects of transport events reveal that departure from the campus is directly discouraged, while incorporating the negative copula dependence for staff indicates more frequent movements within the occupied buildings during the transport event. Similarly, with external air temperature rises in the autumn, the lower rate of transitions to another building (also the longer occupancy duration in a building) for staff is associated with a higher rate of transitions inside the occupied building.

Concerning the dependence between transitions to different buildings, the model captures the effects of on-campus events, Wednesday timetabling policy and peak period of departure. Such a level of dependence helps to describe the relationships between choices of transitions to different buildings that the generic variables in marginal functions cannot capture. During 30 min ahead of the events, transitions to the buildings where the events happen have negative correlations with transitions to other buildings, which reflects that people are attracted by these on-campus events. According to the college timetabling policy, every Wednesday afternoon is free from teaching activities, during which transitions to different buildings have positive correlations. The positive correlation reveals that college people make frequent movements between buildings under the policy due to more Student Union activities for UG students and seminars attracting research students. During the peak time of departures from the campus, positive correlations between transitions to different buildings might indicate frequent transitions across the campus often, but it appears to be fewer transitions between buildings when a transport event happens. It is reasonable to hypothesise that people often find their friends in another building and then leave together, therefore pointing out the possibility of dependence between individuals for departure.

4.2. Performance of prediction on average occupancy patterns

To assess the prediction performance, we compared the occupancy patterns on average over a week (i.e. between 1 and 5 November 2021), subsequent to the period from which data was used to calibrate the model, generated by Wi-Fi observations, and predicted by full NHBM and time inhomogeneous MCM. This section aims to evaluate how well the two models describe the observed average patterns. We deliver discussions on prediction performance in terms of observational patterns and numerical results of performance metrics. In the discussions, two models are cross compared to suggest which model fits for which scenario and potential further improvements are recommended.

Fig. 7 illustrates the variation of occupancy rate[2] on average during a week against the time of day for different groups of buildings, in which patterns from UK codes and surveys, Wi-Fi observations (OBS), full NHBM and time inhomogeneous MCM are given. To complement the visual comparison, Table 5 presents the numerical results of two performance metrics to assess and compare the overall prediction performance of the two models.

The Wi-Fi inferred occupancy patterns show that building occupancy varies with building functionality, for which residential buildings (i.e. Groups 4 and 7), the building with lunch offer (i.e. Sherfield) and Library have quite different patterns from others (i.e. teaching and office buildings). The observed occupancy patterns for teaching and office buildings except Group 5 are similarly distributed with a bimodal shape and a local valley during lunchtime. Occupancy in Group 5 has a relatively flat part during lunchtime and follows an afternoon peak. This difference could result from the movement of the chemistry department to another campus and mainly lab activities continuing in the old buildings. Although Groups 4 and 7 both have residential buildings, occupancy patterns during working hours are considerably diverse due to the different proportion of residence area over buildings. Sherfield and Library are unimodally distributed, but different slopes of morning increases and afternoon decreases are presented. In addition, as compared to standard patterns in the UK's building code, the observed patterns in Groups 1–3, 6 and Faculty generally follow the variations in the morning, lunchtime and afternoon, whereas other buildings are significantly distinct from the code. This finding demonstrates the limitation of the code-given pattern in accurately describing multifunctional buildings in this case.

As for prediction performance for different buildings, comparing plots of two models suggests full NHBM and time inhomogeneous MCM have broadly similar performance in describing the variation of average occupancy patterns over a day for multiple non-residence buildings. Most predictions of the two models fall within the 95% confidence intervals of observations, which confirms their satisfactory abilities for multi-building applications. However, the consistently lower values of MAE and MAPE (in Table 5) specify that the full NHBM greatly outperforms on predicting building groups except Groups 3, 6 and Library. It also can be seen that full NHBM has relatively more accurate predictions on occupancy for two groups with residential spaces, which verifies the importance of including space functionality covariate. The outperformance of full NHBM can point towards occupancy models that can embed a wider range of covariates once these data are accessible.

A slightly weaker performance of the full NHBM for two buildings with events held sometimes (i.e. Groups 3 and 6) could result from the biased characterisation of the impact of events. This result suggests that careful consideration is necessary for specifying the impact of events on occupancy because of the diverse preferences for event attendance within the university population. It may be concluded that adding a generic event-related covariate to occupancy modelling is likely to cause biased simulation outputs if such preference information is unknown. Additionally, MCM can replicate more reasonable and accurate occupancy patterns for Library, whereas the full NHBM prediction is a bit away from observations. The possible reason for the discrepancy is that generic time parameters were included to capture Library occupancy though it has a different nature from other buildings.

In addition, MAPE expresses how many occupants are overpredicted or underpredicted by models. For example, the bias of NHBM prediction is 170 occupants for Group 1 buildings with a maximum capacity of 1300 on average. According to MAPE values in the relevant literature [39–42], MAPEs of their occupancy predictions are commonly in the range of 15%–60%. As compared to the range, two models in this application have shown acceptable prediction accuracy for all buildings and even higher-level performance for teaching, office and residential buildings.

By comparing the prediction performance for different times of day, both models demonstrate better fitting for afternoon decreases than morning increases. Although two models describe diverse patterns for Group 4 after 16:00, full NHBM acceptably picks up the decreasing trend of the observations. However, neither of the models succeed in capturing valleys during lunchtime for teaching and office buildings. The possible reason might be imprecisely defining the time interval of lunch which

[2] Occupancy rate is the ratio of occupancy count at a given time over the maximum occupancy count during a day.

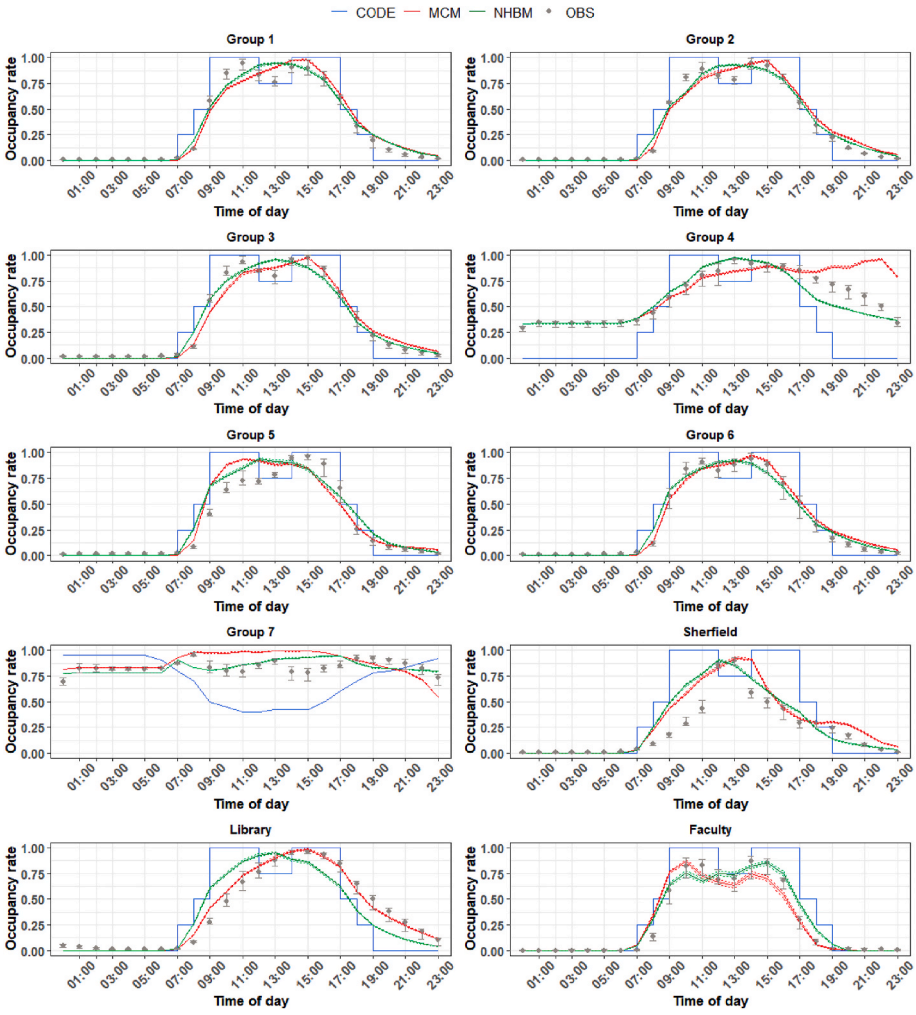


Fig. 7. Average occupancy patterns (with 95% confidence intervals) at an hourly scale, produced using the UK building code for office [10] and Time Use survey for residence [38] (blue line), Wi-Fi observations (grey error-bar plots), time inhomogeneous MCM (red line) and full NHBM (green line) for 10 building groups. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 5
Performance metrics for measuring and comparing the prediction accuracy of full NHBM and time inhomogeneous MCM.

	MAE		MAPE		Occupancy size
Group	**NHBM**	**MCM**	**NHBM**	**MCM**	
G1	**0.044**	0.052	**13.13%**	15.30%	1312
G2	**0.042**	0.047	**12.56%**	13.90%	1039
G3	0.045	**0.041**	12.65%	**11.30%**	968
G4	**0.066**	0.087	**11.22%**	14.80%	554
G5	**0.080**	0.081	**25.60%**	25.70%	488
G6	0.038	**0.035**	11.60%	**10.50%**	729
G7	**0.061**	0.091	**7.26%**	11.00%	664
Sherfield	**0.081**	0.084	**35.62%**	36.90%	670
Library	0.126	**0.041**	33.35%	**10.80%**	962
Faculty	**0.044**	0.050	**15.87%**	17.90%	41

counteracts leaving for lunch and coming back. Such bias could be addressed by splitting the long lunch period (11:45–14:30) into earlier and later parts for both models. On the contrary, the Faculty building with purely staff occupancy shows better simulations for the lunch valley. This result suggests that the spread of lunch time for occupants in other buildings is quite different compared to the 'pure office' building. As the NHBM is a parametric model, specifying interactive variables for capturing both building and temporal diversities in the model would appropriately enhance its prediction performance.

4.3. Performance of prediction on occupancy day by day

This section intends to validate how well the proposed NHBM describes the changes in occupancy patterns between days, using the same validation dataset as section 4.2. MCM is still the benchmark in this comparison and validation. The changes in patterns, no matter whether the NHBM succeeds in capturing them or not, can indicate some essential considerations for developing future occupancy models. The following subsections provide strategic suggestions on incorporating exogenous variables with occupancy modelling.

Fig. 8 shows occupancy patterns predicted by the two models and observed from Wi-Fi data over time of day for five future workdays. In terms of the observations, it appears remarkably different patterns on 3 November (Wednesday) for Groups 1, 2, and 6, and on 1 (Monday) and 5 November (Friday) for Faculty, while other patterns for the same group have minor differences. It is also observed that buildings with different functionalities have considerably different occupancy patterns, such as Group 7 and others. These observations indicate a need for a model having the ability to capture the variations between days and between buildings. Although the prediction pattern shows that the time inhomogeneous MCM fails to capture the changes in patterns for different days, additionally specifying week-day dependence in the estimation for transition matrices could unlock the limitation in principle. However, the inclusion of more dependence in MCM naturally requires hosting more specific transition matrices. Comparatively, the proposed NHBM

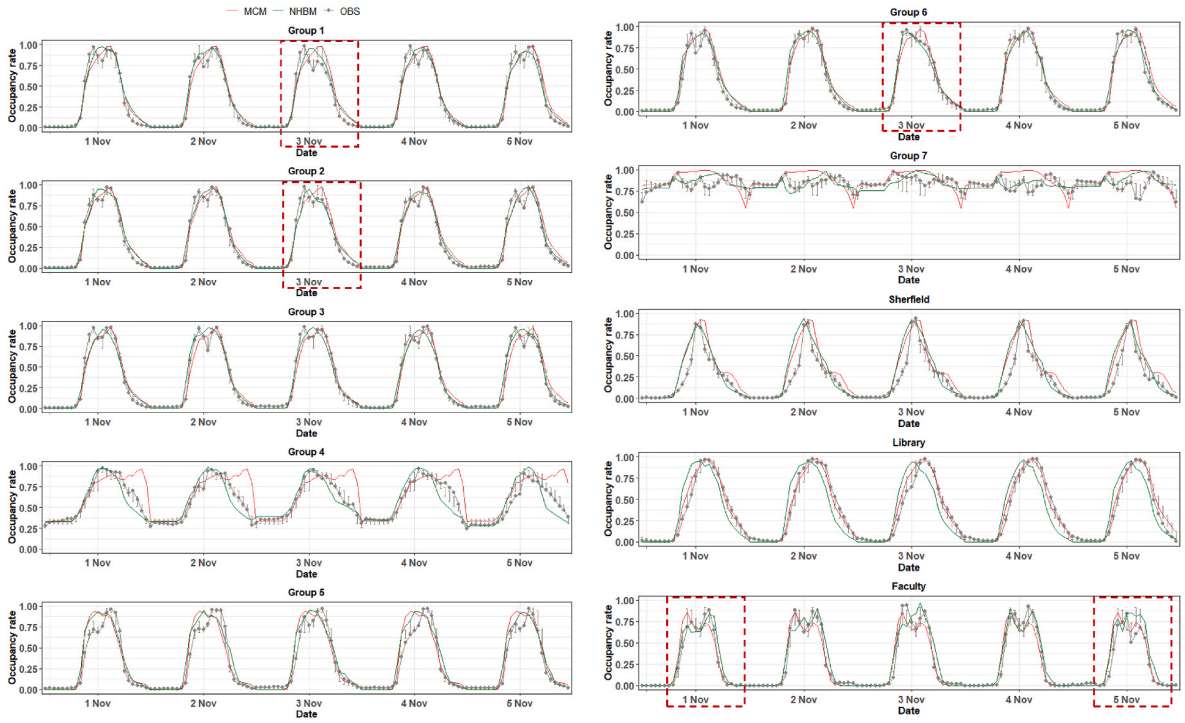


Fig. 8. The hourly-scale building occupancy patterns predicted by time-inhomogeneous MCM (red line) and full NHBM (green line) and extracted from Wi-Fi observations (grey error-bar plot) for five future workdays (from Monday to Friday). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

describes the occupancy patterns in a flexible way of incorporating causal reasons that were assumed to be constantly varied urban systems and in a much more 'compact' way to adjust model specifications by only flipping the parameters. As a result, the NHBM applied to this case can generate occupancy outputs specified for the validation day by incorporating urban variables on that day.

4.3.1. Simulations on irregular teaching days

Fig. 8 indicates the occupancy patterns on Wednesday (3 November) distinct from other workdays for Groups 1, 2 and 6 (teaching and office buildings). In these patterns, the global peak of occupancy appears in the morning, likely due to the effects of the Wednesday timetabling policy imposed on campus. As mentioned in section 3.2, Wednesday afternoon, free from teaching activities, is an important covariate in the NHBM calibration. The plots for simulation results show that full NHBM outperforms in capturing the trend of observations on Wednesday. This outperformance strongly confirms that the policy variable specified in the copula dependence of NHBM plays a role in distinguishing patterns on Wednesday and other days. Specifically, the estimated coefficient of the policy variable in the copula functions correctly describes the positive correlation between transitions as well as frequent circulations during that period. Comparatively, MCM fails to describe such irregular occupancy patterns because it only accounts for the variation that most frequently happened in the sample.

This finding reveals that the timetabling policy drives the changes in occupancy patterns in case study buildings, which confirms the hypothesis of an interaction between policy and building occupancy in the proposed framework [15]. Consequently, this dependence is worth being involved in other short-term occupancy forecasting models to enhance their accuracy. Although the same policy might not be implemented in other cases, it is essential for users to consider the incorporation of other specific policies into building occupancy modelling.

Another possible determinant of the occupancy pattern is the dependence between days because today's behaviour could be associated with yesterday's behaviour for the same individual. An example could be the social-distance rules and remote work schemes, limiting the presence to only certain days of a week due to preference to work from the office only on some days. However, this relationship is not incorporated in this study due to limitations on the data and model structure. Incorporating dependence between aggregated occupancy at days means a more complex log-likelihood function for model estimation to replace the simple summation of the log-likelihood across population and days like in Eq. (9). Therefore, future studies to describe such dependence should consider an approach (e.g. conditional probability defined in Eq. 14–17) to link likelihoods between two adjacent days.

$$LL = log(L(Tue|Mon)) + log(L(Wed|Tue)) + log(L(Thu|Wed)) + log(L(Fri|Thu)) + log(L(Mon)) \tag{14}$$

$$L(Tue|Mon) = \frac{L(Mon \cap Tue)}{L(Mon)} \tag{15}$$

$$L(Mon) = \prod_{i=1}^{I_{(Mon)}} \prod_{a=1}^{A} \prod_{b=1}^{B} \left(P_{a\to b}\left(t_{i,a,b}(Mon)|X\right)^{\left(1-d_{i,c}\right)} S_a\left(t_{i,a,b}(Mon)|X\right)^{d_{i,c}} \right)^{d_{i,a,b}} \tag{16}$$

$$L(Mon \cap Tue) = \prod_{i=1}^{I_{(Mon \cap Tue)}} \prod_{a=1}^{A} \prod_{b=1}^{B} \left(P_{a \to b}\left(t_{i,a,b}(Mon \cap Tue) \middle| X\right)^{\left(1-d_{i,c}\right)} S_a\left(t_{i,a,b}(Mon \cap Tue) \middle| X\right)^{d_{i,c}} \right)^{d_{i,a,b}} \quad (17)$$

4.3.2. Simulations on different types of spaces

Comparing occupancy patterns by buildings reveals considerably distinct patterns of Groups 4 and 7 to others, because these groups consist of all residential buildings on campus. For these buildings, we observe almost constant occupancy during out-of-work hours and dynamic occupancy during working hours. Consistent with findings in section 4.2, occupancy patterns of the two groups are generally better predicted by full NHBM over five validation days, though uncertainties for the afternoon phases are less captured. This outperformance confirms the value of incorporating a covariate that describes the ratio of residential space over the total area of buildings. In line with the same principle, a covariate representing office ratio was specified to distinguish occupancy features in different buildings, though they accommodate the same activities. However, the figure shows that occupancy of Group 5 is not constantly well predicted over days as occupancy of other buildings, for which the possible reason might be either the lack of building-specific time dependence or the use of old layout data before the building refurbishment due to the unavailability of the new layout. Such imperfection can be corrected only if the stored university building layouts are up to date.

As compared to MCM, NHBM provides the flexible mechanism for simultaneously specifying multiple independent variables, thereby allowing to capture occupancy differences flexibly. At the same time, the correct specification of covariates in the model leads to accurate predictions. However, it should be noted that specifying office area and the number of offices could obtain different estimated coefficients and prediction performance, as the size of each office depends on the role of occupants and space arrangement rather than the deterministic occupancy density given in building codes. Therefore, it is worth exploring the effects of the number of offices alongside the roles of their occupants and corresponding prediction performance in future applications. In summary, space use is an important covariate to capture differences in occupancy between buildings, which needs to be considered not only in occupancy simulation but also in short-term occupancy forecasting.

4.3.3. Importance of long-term observations

It is observed that NHBM performs much more accurately on 1 and 4 November for the Faculty building, occupied by staff only. Investigating covariates of the model indicates the surprisingly lower external air temperature during 07:00–10:00 on 2, 3 and 5 November, out of the range of air temperature in the training dataset. At the same time, the positive coefficient of staff suggests the negative correlations between transitions within buildings and to another building. Thus, incorporating the effects of lower temperature and staff role would skew the occupancy distribution to an extreme scenario. In other words, a higher proportion of staff would choose to transit out of Faculty to another building during the period of low temperature. This explains the NHBM's underpredictions for occupancy rates during those intervals.

The reason for such a bias could be either non-linear effects of the external temperature or the limited temporal scope of observations that exclude days with much lower temperature, rather than methodological limitations. This problem is universal for all predictions incorporating the effects of external air temperature. It suggests the importance for users to attempt transformations for the temperature variable and to monitor the long-term observations for calibrating the detailed occupancy model (at fine-grained intervals) if the effects of air temperature are considered. However, the longer term of observations does not mean higher modelling accuracy because some changes (e.g. building refurbishment) could happen during a long period, which could lead to great changes in occupancy patterns. The reasonable length of observations depends on the optimal coverage being a function of the stability of the surrounding environment, considering supply (e.g. layout) and demand (e.g. occupant behaviour).

5. Conclusions and limitations

5.1. Summary

This paper presents a novel combined copula-nested hazard-based approach (i.e. NHBM), calibrated and validated using Wi-Fi authentication data and multiple covariates describing the surrounding urban environment. The approach addresses the need to model occupancy of multiple inter-dependent buildings on campus that have different functions and accommodate different roles of occupants. Incorporating copula dependence in the NHBM enables an additional interpretation for potential correlations between competing choices of destinations (for a transition) at different spatial scales, i.e. within a building and to another building. The proposed NHBM is also capable of reflecting changes in occupancy with the variations in the surrounding urban environment, thereby providing occupancy profiles for energy analysis and dynamic control of building facilities taking into account not only what takes place inside but also in the vicinity of the building. For example, the interpretation for transport effects indicated that, as compared to other roles, staff in the case study area were sensitive to transport events and their circulations within the occupied building were indirectly driven by the timing of the transport events. This finding supports the need for a future research stream and a real application towards deeply integrating building and transport systems.

5.2. Potential applications of the occupancy model

The model paves the way also towards using its outputs for evaluating re-design and refurbishment options. For example, we observe that layout coefficients reveal that increasing office proportion leads to higher occupancy-related energy consumption and vice versa. The results suggest that encouraging the shared office and allocating space to other activities could help to save building energy. In addition, we show how the policy (namely campus timetabling policy) and events cause frequent circulations between buildings and possible higher energy demand, which could in turn be used for scheduling electric vehicle charging times on campus, especially when taking into account local generation. Secondly, the occupancy simulations for extreme irregular scenarios (e.g. heatwaves and transport delays) can suggest shifts in HVAC operation times and changes in ventilation rates. Since the proposed model allows 'days ahead' predictions for irregular occupancy patterns, it can be further integrated with the model predictive control of the HVAC system and power grid management to aid variations of occupancy-driven energy demand in campus energy management.

Our model has shown that occupancy is a function of urban variables from transport, retail, events, policy and weather systems and building physics, which are not captured in other occupancy model formulations. This can be useful in strategic planning to help designers and managers to allocate resources with minimum cost and greenhouse gas emissions.

In the context of evaluating changes in building design and layout,

our model enables a designer to assess not only the change in energy consumption from physical dimension changes but also the changes in how people will circulate through the space with respect to the connections to the urban environment. In the context of evaluating policy impact on district-level energy, being able to predict dynamic energy demand in buildings resulting from timetable-driven factors can be very useful for planning and scheduling neighboured electric vehicle charging infrastructure. It also helps with planning optimal battery size for the microgrid operation to respond to dynamic pricing schemes in the energy market. In the context of developing evacuation plans, our model can simulate circulation patterns driven by events and schedules, thereby suggesting evacuation strategies for special periods. In the context of managing HVAC operations, our model can simulate occupancy patterns driven by the dynamic urban environment for extreme irregular scenarios (e.g. heatwaves and transport delays). The occupancy patterns generated can be used to determine the start and stop times and ventilation rates for dynamic HVAC control.

5.3. Limitations and future work

Overall, the proposed occupancy model is not limited to an application to the campus context but can be adapted to other areas, akin to those of a university campus, such as business districts or shopping districts, especially under common operations management. Such areas are increasingly instrumented with video surveillance, environmental sensors (e.g. temperature, rain), people counting, car park occupancy sensors. Such sources can be used as data feeds for models such as the one proposed in the current paper, providing ways to interpret dynamic interactions between case study buildings and their surroundings. Such a situation opens interesting questions concerning data sharing agreements, whereby an incentive is created to exchange data concerning buildings, e.g. between building owners, due to the mutual benefit of being able to operationalise the occupancy model and benefit from energy savings.

Naturally, the abundance of data sources raises the question of how to select the right site of independent variables to obtain a good predictive power, model transparency and but also avoid over-fitting or excessive computational cost (itself possibly leading to an excessive energy consumption).

As for the limitations of the methodology, we observed that the FGM copula formulation may lead to model identification issues when two marginals have an extreme dependence. However, the FGM function is one of only two closed-form copula functions which can be generalised to describe the bidirectional (positive and negative) correlations. The other one with the two-tail dependence is Ali–Mikhail–Haq Archimedean copula, but it has not been applied to compute the joint distribution of multivariate marginals in the literature. In order to improve the model robustness, it is possible to use a stable one-tail copula function (like the study in Ref. [34]) and collaborate with the direction of dependence evidenced by the FGM estimation. In addition, the prediction results indicate an imperfect description for library that is a representative building with multi-functions. This finding recommends future studies carefully consider including parameters specific to this type of buildings rather than generic ones, though at the cost of heavy computation.

Concerning the limitation of data source, the proposed approach was operationalised by Wi-Fi data independently, but it can be adapted to use multiple sensing data sources (through data fusion) in applications to other buildings and districts. The use of Wi-Fi data relies on an assumption that a device-detected location is the real physical occupancy location for an individual. However, because of radio signal bouncing between locations and signal absorption, using Wi-Fi data solely could lead to biased occupancy measurements and combining other data with Wi-Fi data would help to improve the measurement close to the real behaviour. In addition, this approach was validated by using Wi-Fi data on days subsequent to the period of calibration dataset. Therefore, it demonstrates the ability to statistically learn occupancy patterns and suggest strategic planning for energy management. In future studies, ground-true data (e.g. camera) is needed to validate the operational ability of the proposed approach.

CRediT authorship contribution statement

Huiqiao Hou: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Jacek Pawlak:** Writing – review & editing, Supervision. **Aruna Sivakumar:** Writing – review & editing, Supervision. **Bianca Howard:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

Acknowledgments

This research was financially supported by the China Scholarship Council and Imperial College London. The authors would like to appreciate Mr. Matt Hubbard for providing Wi-Fi authentication data. The authors acknowledge the support from the Integrated Development of Low-Carbon Energy Systems (IDLES) research programme at Imperial College London funded by the Engineering and Physical Sciences Research Council (EPSRC) Programme Grants [grant number EP/R045518/1].

Appendix. Estimation results of full NHBM

	Parameter	Value	Standard Error	t-value	p-value	p-stars
Copula	γ_w_b_7:00_8:00	6.236	1.883	3.311	0.001	***
	γ_w_b_staff	0.637	0.041	15.546	0.000	***
	γ_b_event	0.262	0.123	2.126	0.034	**
	γ_b_16:00_18:00	−0.487	0.068	7.208	0.000	***
	γ_b_Wed	−0.400	0.034	11.738	0.000	***
Marginal	$\omega_{0,1}$	0.000 (fixed)				
	$\omega_{0,2}$	−0.330	0.009	35.406	0.000	***
	$\omega_{0,3}$	−0.234	0.009	26.730	0.000	***
	$\omega_{0,4}$	−0.800	0.010	76.780	0.000	***
	$\omega_{0,5}$	−0.633	0.013	46.919	0.000	***

(*continued on next page*)

(*continued*)

Parameter	Value	Standard Error	t-value	p-value	p-stars
$\omega_{0,6}$	−0.364	0.008	47.519	0.000	***
$\omega_{0,7}$	−0.411	0.011	36.727	0.000	***
$\omega_{0,sh}$	−0.279	0.007	38.853	0.000	***
$\omega_{0,lib}$	−0.287	0.007	43.562	0.000	***
$\omega_{0,fac}$	−0.660	0.027	24.346	0.000	***
$\omega_{0,out}$	−0.981	0.008	118.906	0.000	***
$\varphi_{0,1}$	0.149	0.002	61.678	0.000	***
$\varphi_{0,2}$	0.574	0.008	72.307	0.000	***
$\varphi_{0,3}$	0.395	0.006	62.423	0.000	***
$\varphi_{0,4}$	1.058	0.009	112.275	0.000	***
$\varphi_{0,5}$	0.344	0.009	37.779	0.000	***
$\varphi_{0,6}$	0.210	0.005	44.735	0.000	***
$\varphi_{0,7}$	0.962	0.010	96.874	0.000	***
$\varphi_{0,sh}$	0.423	0.005	85.162	0.000	***
$\varphi_{0,lib}$	0.169	0.004	43.694	0.000	***
$\varphi_{0,fac}$	0.941	0.025	37.811	0.000	***
$\varphi_{0,out}$	0.934	0.005	173.352	0.000	***
$\lambda_{0,1}$	0.936	0.008	119.083	0.000	***
$\lambda_{0,2}$	1.812	0.009	192.191	0.000	***
$\lambda_{0,3}$	1.552	0.011	135.150	0.000	***
$\lambda_{0,4}$	2.247	0.003	666.106	0.000	***
$\lambda_{0,5}$	1.468	0.016	89.984	0.000	***
$\lambda_{0,6}$	1.155	0.011	101.387	0.000	***
$\lambda_{0,7}$	2.176	0.005	434.303	0.000	***
$\lambda_{0,sh}$	1.592	0.008	194.304	0.000	***
$\lambda_{0,lib}$	0.963	0.011	86.121	0.000	***
$\lambda_{0,fac}$	2.105	0.016	131.472	0.000	***
$\lambda_{0,out}$	0.351	0.004	90.799	0.000	***
ω_ratio_office	−7.046	0.038	183.549	0.000	***
ω_ratio_residence	−3.632	0.015	245.383	0.000	***
ω_temp_to_another	−1.750	0.004	397.996	0.000	***
ω_temp_exit_campus	−0.343	0.007	47.393	0.000	***
ω_rain_to_another	−0.007	0.014	0.509	0.611	
ω_rain_to_entre_campus	0.167	0.018	9.519	0.000	***
ω_rain_exit_campus	0.037	0.010	3.541	0.000	***
ω_transport_exit_campus	−0.024	0.009	2.499	0.012	**
ω_7:00_8:00_within	0.276	0.012	23.744	0.000	***
ω_7:00_8:00_to_another	−0.573	0.021	27.539	0.000	***
ω_7:00_8:00_entre_campus	−0.573	0.021	27.539	0.000	***
ω_8:00_10:00_within	0.196	0.005	39.257	0.000	***
ω_8:00_10:00_to_another	−0.475	0.007	72.362	0.000	***
ω_8:00_10:00_entre_campus	−0.475	0.007	72.362	0.000	***
ω_10:00_lunch_within	0.328	0.005	68.724	0.000	***
ω_10:00_lunch_to_another	−0.070	0.006	11.085	0.000	***
ω_10:00_lunch_entre_campus	−0.070	0.006	11.085	0.000	***
ω_lunch_period_within	0.281	0.004	71.368	0.000	***
ω_lunch_period_to_another	0.178	0.006	29.967	0.000	***
ω_lunch_period_to_sh	1.081	0.008	130.435	0.000	***
ω_lunch_period_entre_campus	0.178	0.006	29.967	0.000	***
ω_lunch_16:00_within	0.229	0.004	55.759	0.000	***
ω_lunch_16:00_to_another	0.272	0.009	31.474	0.000	***
ω_lunch_16:00_entre_campus	0.272	0.009	31.474	0.000	***
ω_lunch_16:00_exit_campus	0.426	0.008	53.513	0.000	***
ω_16:00_18:00_within	0.324	0.005	62.840	0.000	***
ω_16:00_18:00_to_another	0.316	0.011	27.640	0.000	***
ω_16:00_18:00_exit_campus	0.738	0.008	87.314	0.000	***
ω_18:00_21:00_within	0.327	0.005	61.380	0.000	***
ω_18:00_21:00_to_another	−0.211	0.010	20.929	0.000	***
ω_18:00_21:00_exit_campus	0.629	0.008	75.950	0.000	***
ω_21:00_24:00_exit_campus	0.919	0.010	90.204	0.000	***

References

[1] U. Ali, M.H. Shamsi, C. Hoare, E. Mangina, J. O'Donnell, Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis, Energy Build. 246 (2021), 111073, https://doi.org/10.1016/j.enbuild.2021.111073.

[2] Y. Zhang, X. Bai, F.P. Mills, J.C.V. Pezzey, Rethinking the role of occupant behavior in building energy performance: a review, Energy Build. 172 (2018) 279–294, https://doi.org/10.1016/j.enbuild.2018.05.017.

[3] N.D. Cortiços, C.C. Duarte, COVID-19: the impact in US high-rise office buildings energy efficiency, Energy Build. 249 (2021), https://doi.org/10.1016/j.enbuild.2021.111180.

[4] H. Huang, L. Chen, E. Hu, A new model predictive control scheme for energy and cost savings in commercial buildings: an airport terminal building case study, Build. Environ. 89 (2015) 203–216, https://doi.org/10.1016/j.buildenv.2015.01.037.

[5] T. Hong, H.-W. Lin, Occupant behavior: impact on energy use of private offices, ASim 2012 - 1st asia conf, Int. Build. Perform. Simul. Assoc. (2013) 12.

[6] D. Yan, T. Hong, B. Dong, A. Mahdavi, S. D'Oca, I. Gaetani, X. Feng, IEA EBC Annex 66: definition and simulation of occupant behavior in buildings, Energy Build. 156 (2017) 258–270, https://doi.org/10.1016/j.enbuild.2017.09.084.

[7] T. Hong, S. D'Oca, W.J.N. Turner, S.C. Taylor-Lange, An ontology to represent energy-related occupant behavior in buildings. Part I: introduction to the DNAs

framework, Build. Environ. 92 (2015) 764–777, https://doi.org/10.1016/j.buildenv.2015.02.019.
[8] W. O'Brien, A. Wagner, M. Schweiker, A. Mahdavi, J. Day, M.B. Kjærgaard, S. Carlucci, B. Dong, F. Tahmasebi, D. Yan, T. Hong, H.B. Gunay, Z. Nagy, C. Miller, C. Berger, Introducing IEA EBC annex 79: key challenges and opportunities in the field of occupant-centric building design and operation, Build. Environ. 178 (2020), 106738, https://doi.org/10.1016/j.buildenv.2020.106738.
[9] G. Happle, J.A. Fonseca, A. Schlueter, Context-specific urban occupancy modeling using location-based services data, Build. Environ. 175 (2020), 106803, https://doi.org/10.1016/j.buildenv.2020.106803.
[10] W. O'Brien, F. Tahmasebi, R.K. Andersen, E. Azar, V. Barthelmes, Z.D. Belafi, C. Berger, D. Chen, M. De Simone, Simona d'Oca, T. Hong, Q. Jin, D. Khovalyg, R. Lamberts, V. Novakovic, J.Y. Park, M. Plagmann, V.S. Rajus, M. Vellei, S. Verbruggen, A. Wagner, E. Willems, D. Yan, J. Zhou, An international review of occupant-related aspects of building energy codes and standards, Build, Environ. Times 179 (2020), 106906, https://doi.org/10.1016/j.buildenv.2020.106906.
[11] J. Page, D. Robinson, N. Morel, J.L. Scartezzini, A generalised stochastic model for the simulation of occupant presence, Energy Build. 40 (2008) 83–98, https://doi.org/10.1016/j.enbuild.2007.01.018.
[12] C. Wang, D. Yan, Y. Jiang, A novel approach for building occupancy simulation, Build. Simulat. (2011), https://doi.org/10.1007/s12273-011-0044-5.
[13] Z. Chen, J. Xu, Y.C. Soh, Modeling regular occupancy in commercial buildings using stochastic models, Energy Build. 103 (2015) 216–223, https://doi.org/10.1016/j.enbuild.2015.06.009.
[14] A.A. Adamopoulou, A.M. Tryferidis, D.K. Tzovaras, A context-aware method for building occupancy prediction, Energy Build. (2016), https://doi.org/10.1016/j.enbuild.2015.10.003.
[15] H. Hou, J. Pawlak, A. Sivakumar, B. Howard, J. Polak, An approach for building occupancy modelling considering the urban context, Build. Environ. 183 (2020), 107126, https://doi.org/10.1016/j.buildenv.2020.107126.
[16] S. D'Oca, T. Hong, Occupancy schedules learning process through a data mining framework, Energy Build. 88 (2015) 395–408, https://doi.org/10.1016/j.enbuild.2014.11.065.
[17] T. Hong, Y. Chen, X. Luo, N. Luo, S.H. Lee, Ten questions on urban building energy modeling, Build. Environ. 168 (2020), 106508, https://doi.org/10.1016/j.buildenv.2019.106508.
[18] T. Rakha, C.M. Rose, C.F. Reinhart, A framework for modeling occupancy schedules and local trips based on activity based surveys, 2014 ASHRAE/IBPSA-USA Build. Simul. Conf. (2014) 433–440.
[19] C.F. Reinhart, C. Cerezo Davila, Urban building energy modeling - a review of a nascent field, Build. Environ. 97 (2016) 196–202, https://doi.org/10.1016/j.buildenv.2015.12.001.
[20] W. Wu, B. Dong, Q. (Ryan), M. Kong Wang, D. Yan, J. An, Y. Liu, A novel mobility-based approach to derive urban-scale building occupant profiles and analyze impacts on building energy consumption, Appl. Energy 278 (2020), 115656, https://doi.org/10.1016/j.apenergy.2020.115656.
[21] K.R. de Casas Castro Marins, M. de Andrade Roméro, Urban and energy assessment from a systemic approach of urban morphology, urban mobility, and buildings: case study of agua branca in sao paulo, J. Urban Plann. Dev. 139 (2013) 280–291, https://doi.org/10.1061/(asce)up.1943-5444.0000149.
[22] S. Van Der Spek, E. Kalogianni, R. Šilerytė, M. Lam, E. Verbree, Passive WiFi monitoring of the rhythm of the campus, Proc. 18th Agil. Int. Conf. Geogr. Inf. Sci. (2015) 9–16.
[23] M. Bon, X. den Duijn, B. Duka, S. Griffioen, Y. Kang, M. Vermeer, Identifying Movement patterns from large scale wifi based location data, Geomatics Synth. Gr. Proj. (2016).
[24] S. Griffioen, M. Vermeer, B. Dukai, S. Van Der Spek, E. Verbree, Exploring indoor movement patterns through eduroam connected wireless devices, Proc. 20th Agil. Int. Conf. Geogr. Inf. Sci. (2017).
[25] N. Mohammadi, J.E. Taylor, Urban energy flux: spatiotemporal fluctuations of building energy consumption and human mobility-driven prediction, Appl. Energy 195 (2017) 810–818, https://doi.org/10.1016/j.apenergy.2017.03.044.
[26] E. Barbour, C.C. Davila, S. Gupta, C. Reinhart, J. Kaur, M.C. González, Planning for sustainable cities by estimating building occupancy with mobile phones, Nat. Commun. 10 (2019), https://doi.org/10.1038/s41467-019-11685-w.
[27] X. Kang, D. Yan, H. Sun, Y. Jin, P. Xu, An approach for obtaining and extracting occupancy patterns in buildings based on mobile positioning data, Build. Simul. Conf. Proc. 4 (2019) 2364–2370, https://doi.org/10.26868/25222708.2019.211151.
[28] C. Wang, Y. Wu, X. Shi, Y. Li, S. Zhu, X. Jin, X. Zhou, Dynamic occupant density models of commercial buildings for urban energy simulation, Build. Environ. 169 (2020), 106549, https://doi.org/10.1016/j.buildenv.2019.106549.
[29] M. Mosteiro-Romero, I. Hischier, J.A. Fonseca, A. Schlueter, A novel population-based occupancy modeling approach for district-scale simulations compared to standard-based methods, Build. Environ. 181 (2020), https://doi.org/10.1016/j.buildenv.2020.107084.
[30] J. Pawlak, A. Faghih Imani, A. Sivakumar, How do household activities drive electricity demand? Applying activity-based modelling in the context of the United Kingdom, Energy Res. Social Sci. 82 (2021), 102318, https://doi.org/10.1016/j.erss.2021.102318.
[31] R. Nelsen, An Introduction to Copulas, Springer Science & Business Media, 2006.
[32] T. Hossein Rashidi, K. Mohammadian, Application of a nested trivariate copula structure in a competing duration hazard-based vehicle transaction decision model, Transp. A Transp. Sci. 12 (2016) 550–567, https://doi.org/10.1080/23249935.2016.1162873.
[33] C. Cameron, P. Trivedi, Microeconometrics: Methods and Applications, Cambridge university press, 2005.
[34] C.R. Bhat, I.N. Sener, A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units, J. Geogr. Syst. 11 (2009) 243–272, https://doi.org/10.1007/s10109-009-0077-9.
[35] F. Meneses, A. Moreira, Large scale movement analysis from WiFi based location data, 2012 Int. Conf. Indoor Position. Indoor Navig. IPIN 2012 - Conf. Proc. (2012), https://doi.org/10.1109/IPIN.2012.6418885.
[36] A.J. Ruiz-Ruiz, H. Blunck, T.S. Prentow, A. Stisen, M.B. Kjaergaard, Analysis methods for extracting knowledge from large-scale WiFi monitoring to inform building facility planning, 2014 IEEE Int. Conf. Pervasive Comput. Commun. PerCom 2014 (2014) 130–138, https://doi.org/10.1109/PerCom.2014.6813953.
[37] M. Ribeiro, N. Nunes, V. Nisi, J. Schöning, Passive Wi-Fi monitoring in the wild: a long-term study across multiple location typologies, Personal Ubiquitous Comput. (2020), https://doi.org/10.1007/s00779-020-01441-z.
[38] V. Aragon, S. Gauthier, P. Warren, P.A.B. James, B. Anderson, Developing English domestic occupancy profiles, Build. Res. Inf. 47 (2019) 375–393, https://doi.org/10.1080/09613218.2017.1399719.
[39] J. Massana, C. Pous, L. Burgas, J. Melendez, J. Colomer, Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes, Energy Build. 130 (2016) 519–531, https://doi.org/10.1016/j.enbuild.2016.08.081.
[40] W. Wang, J. Chen, T. Hong, Occupancy prediction through machine learning and data fusion of environmental sensing and Wi-Fi sensing in buildings, Autom. ConStruct. 94 (2018) 233–243, https://doi.org/10.1016/j.autcon.2018.07.007.
[41] W. Wang, J. Chen, T. Hong, N. Zhu, Occupancy prediction through Markov based feedback recurrent neural network (M-FRNN) algorithm with WiFi probe technology, Build. Environ. 138 (2018) 160–170, https://doi.org/10.1016/j.buildenv.2018.04.034.
[42] S. Kim, S. Kang, K.R. Ryu, G. Song, Real-time occupancy prediction in a large exhibition hall using deep learning approach, Energy Build. 199 (2019) 216–222, https://doi.org/10.1016/j.enbuild.2019.06.043.