1   **An approach for building occupancy modelling considering the urban context**

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

3   [a] *Centre for Transport Studies & Urban Systems Lab, Department of Civil and Environmental Engineering, Imperial*
4   *College London, London, SW7 2AZ, UK*

5   [b] *School of Architecture, Building, and Civil Engineering, Loughborough University, Loughborough, LE11 3TU, UK*

6   **Abstract**

7   Building occupancy, which reflects occupant presence, movements and activities within the

8   building space, is a key factor to consider in building energy modelling and simulation.

9   Characterising complex occupant behaviours and their determinants poses challenges from

10   the sensing, modelling, interpretation and prediction perspectives. Past studies typically

11   applied time-dependent models to predict regular occupancy patterns for commercial

12   buildings. However, this prevalent reliance on purely time-of-day effects is typically not

13   sufficient to accurately characterise the complex occupancy patterns as they may vary with

14   building's surrounding conditions, i.e. the urban environment. Therefore, this research

15   proposes a conceptual framework to incorporate the interactions between urban systems and

16   building occupancy. Under the framework, we propose a novel modelling methodology relying

17   on competing risk hazard formulation to analyse the occupancy of a case study building in

18   London, UK. The occupancy profiles were inferred from the Wi-Fi connection logs extracted

19   from the existing Wi-Fi infrastructure. When compared with the conventional discrete-time

20   Markov Chain Model (MCM), the hazard-based modelling approach was able to better capture

21   the duration dependent nature of the transition probabilities as well as incorporate and quantify

22   the influence of the local environment on occupancy transitions. The work has demonstrated

23   that this approach enables a convenient and flexible incorporation of urban dependencies

* 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@lboro.ac.uk (B. Howard).
Postal address: Skempton Building, South Kensington Campus, Imperial College London, London SW7 2AZ UK

leading to accurate occupancy predictions whilst providing the ability to interpret the impacts of urban systems on building occupancy.

Keywords: Urban system; Competing risk hazard model; Building occupancy simulation; Wi-Fi connection data

1. Introduction

Building occupancy is a key parameter of building energy modelling and a basic factor of modelling occupant behaviour [1]. Bahaj and James [2] discovered that different occupancy patterns lead to different energy consumptions in identical buildings. Therefore, accurately modelling and comprehensively understanding how people use building can help improve energy efficiency [3]. Although the IEA Annex 66 on Definition and Simulation of Occupant Behaviour in Buildings has given guidance on the current state of occupancy modelling, there are still unanswered questions relating to the enhancement of modelling techniques, from the sensing, interpretation and prediction perspectives [4].

Driven by multidisciplinary factors, occupancy has three features that pose challenges to occupancy modelling: stochasticity, diversity and complexity [4]. Stochasticity has been the research focus over a past decade and has enriched the methodological approaches for modelling random occupancy chain. In fact, occupancy is not purely random but could also be characterised by deterministic factors in surrounding systems. These factors could be the internal conditions of building environment and/or the conditions external to the building system. Such conditions can be natural (weather and local topology) as well as anthropogenic, related to urban systems (conditions of transport network and presence of event). Incorporating both of those conditions as potential driving factors of building occupancy into a modelling framework could be a promising way of understanding and enhancing the way we model the complexity and diversity of occupancy patterns.

Considering spatial and temporal dependencies between building occupancy and conditions on surrounding urban systems, this work proposes a modelling framework that

comprehensively and flexibly accounts for such interactions. The framework is operationalised by the means of a competing risk hazard model using implicit occupancy data inferred from wireless (Wi-Fi) networks in the case study area. This approach has apparently not been applied in this discipline so far and has capabilities exceeding those of the more conventional approaches, e.g. MCMs. In particular, it can more easily incorporate multiple exogenous variables simultaneously and allow insights into exogenous impacts on occupancy and thus energy consumption. As an additional contribution, this work also presents and applies a methodology of processing customarily available Wi-Fi data for occupancy modelling.

2. Literature Review

The review is structured to correspond to the three contributions of this work. Section 2.1 summarises studies related to interactions between human behaviour and factors originating in urban systems. Section 2.2 comparatively discusses the applications of MCMs and hazard-based models in occupancy modelling to date. Section 2.3 reviews different approaches to implicit occupancy detection, including discussion of the challenges associated with the use of Wi-Fi data for occupancy modelling in the present context. Section 2.4 summarises gaps in existing studies and states the aims of this paper.

2.1 Review of human-environment interactions in building occupancy modelling

Occupant-centric building performance is a subject of improving building design and operation through understanding how occupants interact with the building environment. Most studies have considered effects of time and building characteristics on occupant behaviour (Table 1). Studies that have incorporated the influence of external factors have been sparse, often lacking a systematic and comprehensive approach.

3

Table 1 Summary of building occupancy studies

Author(s), year	Occupancy level modelled	Empirical modelling methodology	Exogenous variables considered	Data sources considered	Model performance metrics
Hong et al. (2015) [5,6]	General energy-related occupant behaviour	Proposed DNAS framework which was tested in obXML schema	(1). Building (component, properties, and location); (2). Occupant (attributes, attitudes, location, and state); (3). Time (day, week and month); (4). Environment (climate and weather); (5). System (properties and state of building facilities).	-	-
Page et al. (2008) [7]	Presence and absence	MCM	Time of day	Movement sensors	Chi-squared test
Erickson et al. (2011) [8]	Count	MCM	Time of day	Wireless sensors and cameras	Jensen-Shannon test
Wang et al. (2011) [9]	Location and count	MCM	Time of day and schedule; Recommended but not incorporated possible influencing factors (building type and size, geographic location and climate)	Work schedules	Plots of maximum, minimum and average occupancy counts.

Reference	Type	Method	Influencing factors	Sensors	Validation metrics
Chen et al. (2015) [10]	Count	MCM	Time of day	Cameras	Normalised root-mean-square deviation (RMSD); Kullback-Leibler (K-L) divergence
Adamopoulou et al. (2016) [11]	Count	MCM; Semi-MCM	Time of day, day of week, season and zone correlations	Cameras, one acoustic sensor and one passive infrared sensor	Root-mean-square error (RMSE) and normalised RMSE
Wang et al. (2018) [12]	Count	Markov based feedback recurrent neural network algorithm	Time of day	Wi-Fi probe, cameras and environmental sensors	Mean Absolute Error (MAE); Mean Absolute Percentage Error (MAPE); RMSE; x-accuracy (one error metric intended for assessing occupancy modelling).
Wilke et al. (2013) [13]	Activity	High-order Markov Chain Survival Analysis	Time of day	Time-use survey data	Ten-fold cross validation

1   The effects of the local transport conditions have rarely been linked to occupant

2   behaviour in existing studies. The stochastic nature of an occupant transition is not only

3   naturally similar to traffic movements, but can also be similarly influenced by environmental

4   variables [14]. For example, Hong et al. [5] state that the location of the building relative to

5   busy roads is a potential factor, and should be considered to correlate with occupant behaviour

6   in future research. Oliveira-Lima et al. [15] used vehicle counts in the neighboured parking lot

7   to estimate building occupancy. While not establishing the causal relationship, the evidence

8   for correlation has pointed out that occupants used cars to travel to the building. Hence

9   disruptions on the road network, especially unexpected disruptions, could influence when or

10   even whether occupants arrived at the building, thus changing the occupancy patterns.

11   Another set of studies has linked occupant behaviour to events, including urban-scope

12   events which possibly affect occupants' transitions between buildings (e.g. sports) or local

13   events which only lead to movements within one building (e.g. meeting). Roy [16]

14   demonstrated a causal correlation between special events or exhibits in a shopping mall and

15   the frequency of shopper visits. Ab Majid et al. [17] also found that the number of occupants

16   within malls greatly increases during the periods of seasonal sales. People movements are

17   also highly attracted by sports events and are independent on the location where they are [18].

18   Similar to the effect of events, the presence of retail stores also influences occupant

19   presence and absence. Retail outlets can exist in either a multi-purpose shopping cluster or

20   in an office building (e.g. a café). Rajagopal [19] found that buildings with recreational facilities

21   and stores lead to higher frequency of visits. In spite of the increasing popularity of e-

22   commerce, Cachero-Martínez and Vázquez-Casielles [20] discovered that people are still

23   willing to visit physical stores to experience the service. As people are sensitive to price, Roy

24   [16] points out the high degree of attractiveness of discount stores to 'deal–prone' customers.

25   Occupants' behaviour and locations are significantly associated with polices and

26   regulations [21]. The longer opening times of a store in a building attracts more visits to the

27   building [17] and could influence transitions within the building. Under the indoor smoking ban,

smokers have been found to have higher frequency of indoor absence and higher rate of leaving the space [22,23]. D'Oca et al. [24] found that educating occupants on energy saving tactics can evoke the self-motivated management of their behaviours. This agrees with the study by Stephenson et al. [21] which reports that people change their behaviour to respond to the energy pricing mechanism.

Weather is another factor that has been used in some analyses of occupant behaviour [6,9,25]. Parsons [26] observed decreasing shoppers during the period of high rainfall; increasing occupants in shopping mall during the months of hot summer and cold winter. Similarly, Eliasson et al. [27] found that air temperature and wind speed have significant impacts on attendance in a place. Regarding the influence of geography, Stephenson et al. [21] emphasised the effects of culture-dependent lifestyles on energy-related behaviour in households. Fabi et al. [28] and D'Oca and Hong [29] also included habit, lifestyle, income, household composition, age and gender as the potential drivers of occupant's window opening behaviour. Motuziene and Vilutiene [30] demonstrated a causal relationship between occupants' culture and the durations of working, eating and leisure.

Regarding the links between occupant behaviour and characteristics of the building itself, space layout has been an accepted driver for modelling occupancy due to its description of functional areas [31]. Building type and locations of facilities identify the range of activities allowed in the given space and accordingly determine movement patterns [32]. Hong et al. [5] additionally point out the considerable impact of building orientation (façade exposure to solar radiation). A case study by Ab Majid et al. [17] has also revealed that architecture and interior design significantly attract the frequency of shoppers' visits.

Although the concept of an urban-level analysis has been discussed before by Zimring et al. [32], Delzendeh et al. [33] and Happle et al. [34], incorporation of multiple driving factors simultaneously in the modelling of building occupancy remains a novelty. In the previous studies, time of day is the most frequently used covariate for modelling occupancy (see Table 1), due to its good performance in capturing the regularity of occupancy patterns. However,

1   time of day is effectively a convenient proxy variable that represents broader societal dynamics

2   within urban spaces: typical work-day hours, business open hours, meal times or transport

3   network peak and off-peak periods. Hence variations in the urban environment within which

4   such dynamics exist are likely to affect building occupancy. Consequently, the existing

5   literature supports the need ready-to-operationalise modelling approaches that systematically

6   and comprehensively reflect the interactions, at different spatial and temporal scales, between

7   building occupancy and urban systems. The primary benefit of such modelling approaches

8   lies in their ability to better understand and forecast changes in the building occupancy

9   patterns (and hence energy consumption) due to changes in urban systems, including those

10   resulting from policy interventions.

11   **2.2 Review of occupancy modelling methods**

12   Various modelling approaches have been developed to simulate building occupancy

13   including MCM [7], logistic regression [35], agent-based approaches [36,37] and decision

14   trees [38]. Considering the popularity of MCM in modelling occupancy [4], it is used as the

15   benchmark in this paper. MCM is based on an assumption that the current state only depends

16   on the most recent past state. Generally, the state is defined as either the number of occupants

17   or the individual occupied location. The states can transit between the time steps, according

18   to the transition probability matrix, itself calibrated using empirical data. The earliest

19   application of generalised MCM was adapted by Page et al. [7] to model occupants' presences

20   in single-person offices, depending on the time of day. With the good prediction for the

21   stochastic occupancy process, MCM was further expanded to address more complex scenario,

22   i.e. multi-occupant in multi-zone [8–10], with various specifications for improving modelling

23   accuracy [8,11,12]. While appealing, the MCM is challenging to operationalise when a need

24   arises to include multiple covariates and interactions between them. The other limitation

25   concerns absence of the inferential statistics associated with the calibrated parameters,

26   limiting the ability to understand external impacts from a statistical perspective.

8

Another approach comprises survival analysis or Hazard-based modelling and is designed to model the expected duration until a particular event occurs. With a wide range of applications in epidemiology, engineering and economics, models of this kind have flexible and transparent ways of incorporating the effects of covariates. However, survival analysis has not yet been widely applied to model building occupancy. For example, the initial related application of survival analysis was to characterise occupants' actions with regards to opening windows [25]. Wilke et al. [13] and D'Oca et al. [39] implemented survival models with Weibull distributions to estimate the durations of intermediate absence and occupancy, while occupancy-state transition modelling still relied on the MCM. As shall be shown in section 3.2, it is possible to overcome this reliance on MCM by making use of the competing risk hazard model to achieve a joint duration- and transition-choice modelling. This methodology has been used to model the choice of activities and the duration of activity participation in transport studies [40–42]. As will be shown in this paper, these models can also underpin the operationalisation of building occupancy modelling, allowing for both interpretation of covariates effects and the simulation of occupancy patterns. To the best of our knowledge, this is the first independent application of hazard-based models (HBMs) in this discipline.

2.3 Review of implicit data usages for occupancy modelling

Implicit occupancy data is collected from existing infrastructures, which can provide information on occupancy despite not being designed to do so as the primary task [43]. Considering costs and privacy concerns, implicit occupancy sensing has seen increasing research interest in the recent decade.

In the literature, CO_2 concentration data, which indicates indoor air condition, demonstrates good performance in measuring occupancy [44–46]. However, the lower spatial and occupant resolutions of CO_2 data limit granularity in modelling outputs to room and count level. Bluetooth Low Energy (BLE), itself based on the traditional Bluetooth but consuming less energy, was adopted to detect more granular occupancy information [47–50]. Nevertheless, the data availability and quality may be affected by the occupant's device, e.g.

operating system (e.g. iOS and Android) and state of its Bluetooth (e.g. whether switched on or not). Therefore, Park, et al. [50] applied a capture and recapture methodology, using limited logs of Bluetooth connections to estimate occupancy profile for whole population in the building. Overall, both sensing techniques still need extra installations or modifications for collecting data and hence restrict their more universal application for occupancy detection.

In comparison, Wi-Fi occupancy detection has the lowest cost as it has no additional requirements for either hardware and software installations. It is a promising single opportunity for the broad monitoring of occupancy, given its wide availability and frequent use [12,51,52]. Despite its appealing nature, there is a lack of an authoritative guideline for collecting and processing such data for modelling occupancy in either academic literature or industrial applications. Pritoni et al. [51] describe the mechanisms for extracting occupancy information in the common Wi-Fi systems and report current applications of Wi-Fi data in occupancy sensing. Although detection errors (e.g. passing by detections) have been discussed in the article, authors provide treatment strategies relying on coordination with manufactures. For the time being, the data from existing Wi-Fi infrastructure still requires substantial data cleaning methods to filter errors and obtain reliable occupancy information. However, most efforts to date have neglected, or at least not described in detail, how to process Wi-Fi data for occupancy modelling. For example, Martani et al. [53] and Balaji et al. [54] assumed Wi-Fi short-term disconnections as rare events in the future and do not show any treatments. Other studies alternatively used the active scanning technology, called Wi-Fi probe, to prevent issues due to unstable connections [55,56]. However, its application needs probes to be deployed, which is not always feasible as in our research context. More recently, Wi-Fi has been tested to act as a Doppler radar [57,58], thereby removing the usual requirement for an occupant to possess a Wi-Fi-capable device to be 'visible' in the data. Overall, it is clear that Wi-Fi as a source of data for occupancy is a promising avenue for operationalising and validating the modelling frameworks.

2.4 Aims and significance

In summary, existing studies indicate that occupancy location could be driven by surrounding urban systems and hence there is a need for an approach which integrates urban dynamics into building occupancy modelling. Extracting occupancy inputs from Wi-Fi data still lacks a standard procedure and method. Therefore, an approach capable of calibrating Wi-Fi data is also needed.

The aim of this study is to propose an urban-system level framework to comprehensively relate urban conditions with building occupancy. Under this framework, a competing risk hazard model, which can incorporate exogenous variables and filter Wi-Fi connection data, was used to model urban impacts on occupancy pattern and achieve prediction at least as well as the MCM. The means of inferring occupancy profiles from Wi-Fi data is also presented in detail in this paper.

3. Methodology

The following sections describe the urban-system level modelling framework, the HBM approach, the development of the conventional discrete-time MCM, the case study description, Wi-Fi data pre-processing and the performance metrics used to evaluate the two methods (HBM and MCM) from the estimation, interpretation and prediction perspectives.
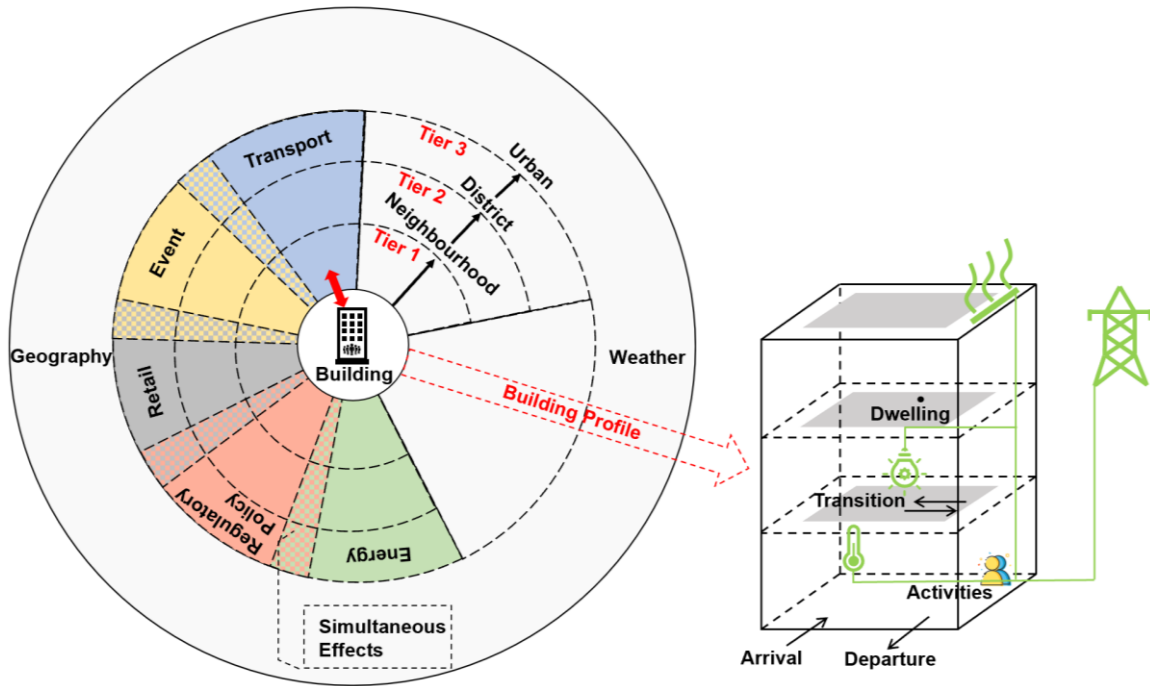
3.1 The urban-system level modelling framework

Figure 1. An urban-system level modelling framework for occupant-environment

interaction

Considering the need to represent the aforementioned interactions between building

occupancy with both internal and external factors, we propose an urban-system level

modelling framework for occupant-environment interaction (see Figure 1). The building

occupancy, located at the centre of the framework, can interact with the surrounding urban

systems: transport, events, retail spaces, regulatory policy and energy. The interactions might

be further affected by weather and geography, and take place at a local (e.g. frequent absence

due to smoke-ban), at a district (e.g. late arrivals due to regional congestions) and even at an

urban range or beyond (e.g. big events impact building uses and transition patterns).

Examples of simultaneous effects from more than one sector are visualised as overlaps in

Figure 1. Overall, the interactions in the framework can be summarised as a function (Eq. (1))

of factors from the building envelope (denoted as *BE*) and urban environment:

$$Occupancy\ Duration\ and\ Transition\ = f(TS, E, RS, RP, W, G, BE) \qquad (1)$$

where

1 TS represents the transport system, that involves the conditions of road, underground, water

2 and air traffic;

3 E represents events, that describe planned and organised occasions;

4 RS represents the retail system, reflecting retail store locations and operational details;

5 RP represents regulatory policy, i.e. local, regional and urban regulations and laws;

6 W represents external weather conditions, such as air temperature and rainfall;

7 G represents geography, including local topography, cultural habits and economic geography.

8 **3.2 Hazard-based model formulation**

9 This section describes the formulation of a HBM for building occupancy, including its

10 underlying functions and model specification, and the model estimation and simulation

11 approaches.

12 **3.2.1 Basic formulation and model specification**

13 HBM is a class of survival models, which relates the probability of state-end to its

14 duration. It uses a set of interrelated mathematical formulations to describe a random variable

15 t: probability density function (PDF, or $f(t)$), (cumulative) probability distribution ($F(t)$) and its

16 complementary survival function (SF or $S(t) = 1 - F(t)$) and hazard function (HF, or $h(t)$).

17 The HF is defined as the probability of an event taking place at time t given it has not taken

18 place before then, and is, given by,

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \qquad (2)$$

19 The probability that the state will end before time t is defined as the first integral of PDF from

20 0 to t in Eq. (3).

$$F(t) = P(T < t) = \int_0^t f(t)\, dt \qquad (3)$$

21 The SF, as a key function in the model, expresses the probability that the state has survived

22 until t,

$$S(t) = 1 - F(t) = P(T \geq t) = \int_{t}^{\infty} f(t)\, dt \qquad (4)$$
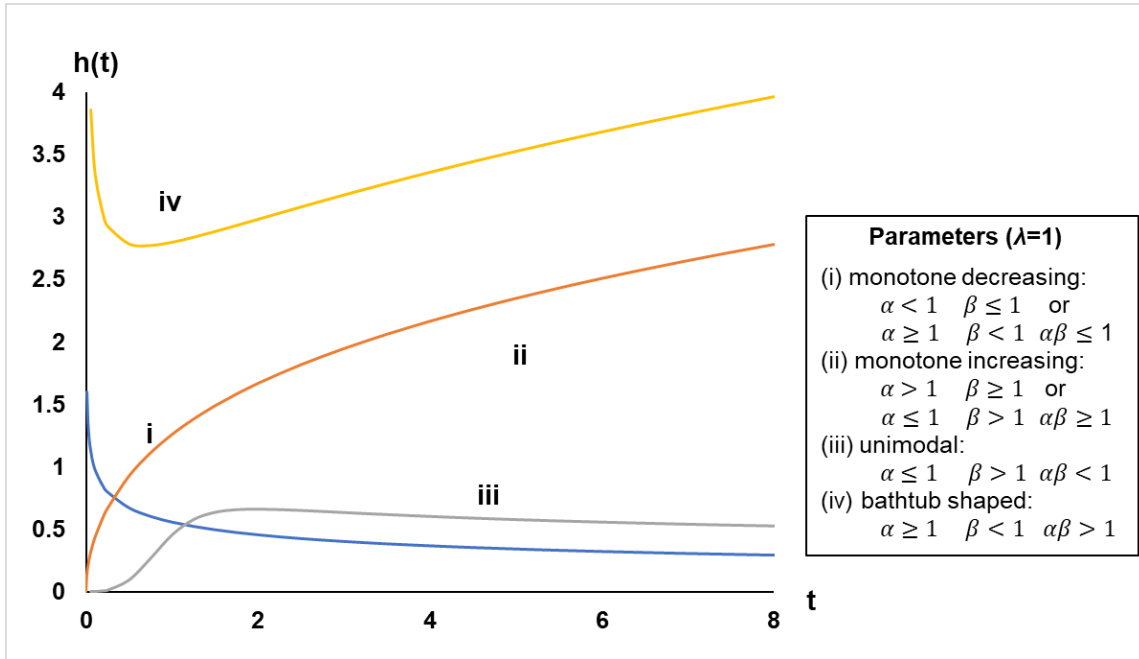

Figure 2. Shapes of parametric hazard functions [59]

1   HBMs in applied contexts commonly consist of two components: the underlying

2   probability distribution (and hence the respective PDF, SF, HF) and equations that relate the

3   distribution's parameters to exogenous covariates and their associated coefficients. Two

4   commonly used forms of HBM are the proportional hazard model and the accelerated lifetime

5   model, which both describe how hazard changes monotonically with duration. However,

6   monotonicity of hazard can be rather restrictive as an *a priori* assumption of potentially

7   complex human behaviour in the context of transitions within a building. A more flexible form

8   of HBM that can capture monotonic and non-monotonic hazard functions was proposed by

9   Dimitrakopoulou et al. [59], as shown in Figure 2. The shape and monotonicity of the hazard

10   function is controlled by two shape parameters α (>0) and β (>0), and a scale parameter λ

11   (>0), while the respective equations for the PDF, SF and HF are Eq. (5-7).

$$f(t; \alpha, \beta, \lambda) = \alpha\beta\lambda t^{\beta-1}\left(1 + \lambda t^{\beta}\right)^{\alpha-1} exp\left\{1 - \left(1 + \lambda t^{\beta}\right)^{\alpha}\right\} \qquad (5)$$

$$S(t; \alpha, \beta, \lambda) = exp\left\{1 - \left(1 + \lambda t^{\beta}\right)^{\alpha}\right\} \qquad (6)$$

$$h(t; \alpha, \beta, \lambda) = \alpha\beta\lambda t^{\beta-1}(1 + \lambda t^\beta)^{\alpha-1} \tag{7}$$

1   The formulation above considers only duration of staying in a zone. To incorporate

2   destination choice model, the model above is extended to incorporate the competing risk

3   hazard formulation [40]. In the current context, zones are modelled as if they competed against

4   each other as the next occupancy location. We assume that physical transitions from the same

5   origin are dependent (competing) and those with different origins are independent. The model

6   was specified to each transition (from zone a to zone b) and coupled with the generalised

7   formulations by Dimitrakopoulou et al. [59] to obtain the following PDF, SF and HF:

$$f_{a,b}\left(t; \alpha_{a,b}, \beta_{a,b}, \lambda_{a,b}\right)$$
$$= \alpha_{a,b}\beta_{a,b}\lambda_{a,b}t^{\beta_{a,b}-1}\left(1 + \lambda_{a,b}t^{\beta_{a,b}}\right)^{\alpha_{a,b}-1} exp\left\{1 - \left(1 + \lambda_{a,b}t^{\beta_{a,b}}\right)^{\alpha_{a,b}}\right\} \tag{8}$$

$$S_{a,b}\left(t; \alpha_{a,b}, \beta_{a,b}, \lambda_{a,b}\right) = exp\left\{1 - \left(1 + \lambda_{a,b}t^{\beta_{a,b}}\right)^{\alpha_{a,b}}\right\} \tag{9}$$

$$h_{a,b}\left(t; \alpha_{a,b}, \beta_{a,b}, \lambda_{a,b}\right) = \frac{f_{a,b}(t;\alpha_{a,b},\beta_{a,b},\lambda_{a,b})}{S_{a,b}(t;\alpha_{a,b},\beta_{a,b},\lambda_{a,b})} = \alpha_{a,b}\beta_{a,b}\lambda_{a,b}t^{\beta_{a,b}-1}\left(1 + \lambda_{a,b}t^{\beta_{a,b}}\right)^{\alpha_{a,b}-1} \tag{10}$$

8   To incorporate the urban interactions with occupancy behaviour (duration of stay in a

9   zone and transition choice probabilities), exogenous covariates X were specified as the

10   exponential (to ensure positivity) of a linear-in-parameters specification of the function α as

11   Eq. (11). Covariates in the model can be generic $X_{i,k}$ and transition-specific $X_{i,a,b,j}$, depending

12   on whether the postulated effect occurs for a particular individual across all transitions or only

13   for a particular set of zonal combinations. The other shape parameter β was specified similarly

14   though only using a constant term (Eq. (12)) to ensure covariate parameters are identifiable.

15   Specifications of α and β are largely interchangeable in this case, which means that including

16   covariates in β and constant terms in α will produce similar estimation results (though not the

17   same parameter values). The scale parameter λ is fixed as one, because it is not identifiable

18   with other parameters at the same time.

$$\alpha_{a,b} = exp(\overrightarrow{\alpha_{a,b}}X) = exp\left(\theta_{0,a,b} + \sum_{k=1}^{K}\theta_k X_{i,k} + \sum_{j=1}^{J}\theta_{(K+j),a,b}X_{i,a,b,j}\right) \quad (11)$$

$$\beta_{a,b} = exp(\overrightarrow{\beta_{a,b}}) = exp(\varphi_{0,a,b}) \quad (12)$$

1   **3.2.2 Model estimation**

2   As we mentioned, all possible destinations have a competing relationship when the

3   individual changing their state makes the decision on destination. Therefore, when one

4   transition takes place, transitions to other possible destinations must be reflected as surviving

5   during the estimation. Data noise is handled (the details of the procedure are further

6   elaborated in section 3.5) by adding weights to duplicated records. Similarly, right censoring

7   was incorporated to deal with observations for which transitions were not observed until

8   midnight, when respondent identifiers were reset (due to privacy compliance, see section 3.5).

9   Gathering the above conditions and further assuming independence between individuals and

10 the transition-specific model to be the same across the sample (to facilitate estimation), it is

11 possible to define a log-likelihood function as follows (see Table 2 for full parameter notation):

$$LL(\theta, \varphi | t, X) = \sum_{i}^{I}\sum_{a=1}^{A}\sum_{b=1}^{B} d_{i,a}\left(1 - w_i d_{i,a}\right)\Big\{\big(1 -$$

$$d_{i,c}\big)\Big[d_{i,a,b}\, log\left(f_{a,b}(t_{i,a,b}|X_{i,k},X_{i,a,b,j})\right) + (1 - d_{i,a,b})\, log\left(S_{a,b}(t_{i,a,b}|X_{i,k},X_{i,a,b,j})\right)\Big] + \quad (13)$$

$$d_{i,c}\, log\left(S_{a,b}(t_{i,a,b}|X_{i,k},X_{i,a,b,j})\right)\Big\}$$

12 which can then be used for parameter estimation using the maximum likelihood principle. The

13 estimation procedure was implemented in R and used the 'optimx' [60,61] and 'Rccp' [62]

14 packages.

15 **Table 2 Parameters used in the formulation of HBM**

Notation	Description of variable
t	The time of state transition (zone change)
a	Origin zone (occupied zone)
b	Destination zone (next zone)
I	Sample size
A	The number of origin zones

B	The number of destination zones
$\theta_{0,a,b}$	A constant term in α specific to transition from zone a to b
θ_k	The kth coefficient of a generic covariate
$\theta_{(K+j),a,b}$	The $(K+j)^{th}$ coefficient of a specific covariate
$\varphi_{0,a,b}$	A constant term in β specific to transition from zone a to b
$\alpha_{a,b}$	A shape parameter specific to transition from zone a to b
$\beta_{a,b}$	A shape parameter specific to transition from zone a to b
$\lambda_{a,b}$	A scale parameter specific to transition from zone a to b
$X_{i,k}$	The kth generic covariate
$X_{i,a,b,j}$	The jth covariate specific to transition from zone a to b, at the time when a transition happens
$f_{a,b}$	PDF specific to transition from zone a to b
$S_{a,b}$	SF specific to transition from a to b
$t_{i,a,b}$	Duration (hour) until the occupied state ends
$d_{i,a,b}$	Dummy variable that represents the combination of occupied zone a and next zone b
$d_{i,a}$	Dummy variable that indicates whether zone a is occupied for i^{th} sample
$d_{i,c}$	Dummy variable that directs whether i^{th} sample lacks destination ($d_{i,c} = 0$) or not ($d_{i,c} = 1$)
$d_{i,d}$	Dummy variable that indicates whether i^{th} sample has duplicated records
w_i	Weight to duplicated movement record $w_i = 1/\text{the number of duplication}$

1

2 **3.2.3 Simulation using HBM**

3   The HBM can be used to create a simulation process that can provide occupancy

4 distributions, given the conditions described by the exogenous variables. As the current

5 implementation of the HBM operates at the respondent level, the output of the simulation was

6 individual occupancy pattern, i.e. a set of durations of stay in zones and transitions between

7 them. The simulations for individuals were aggregated to obtain the occupancy profile as
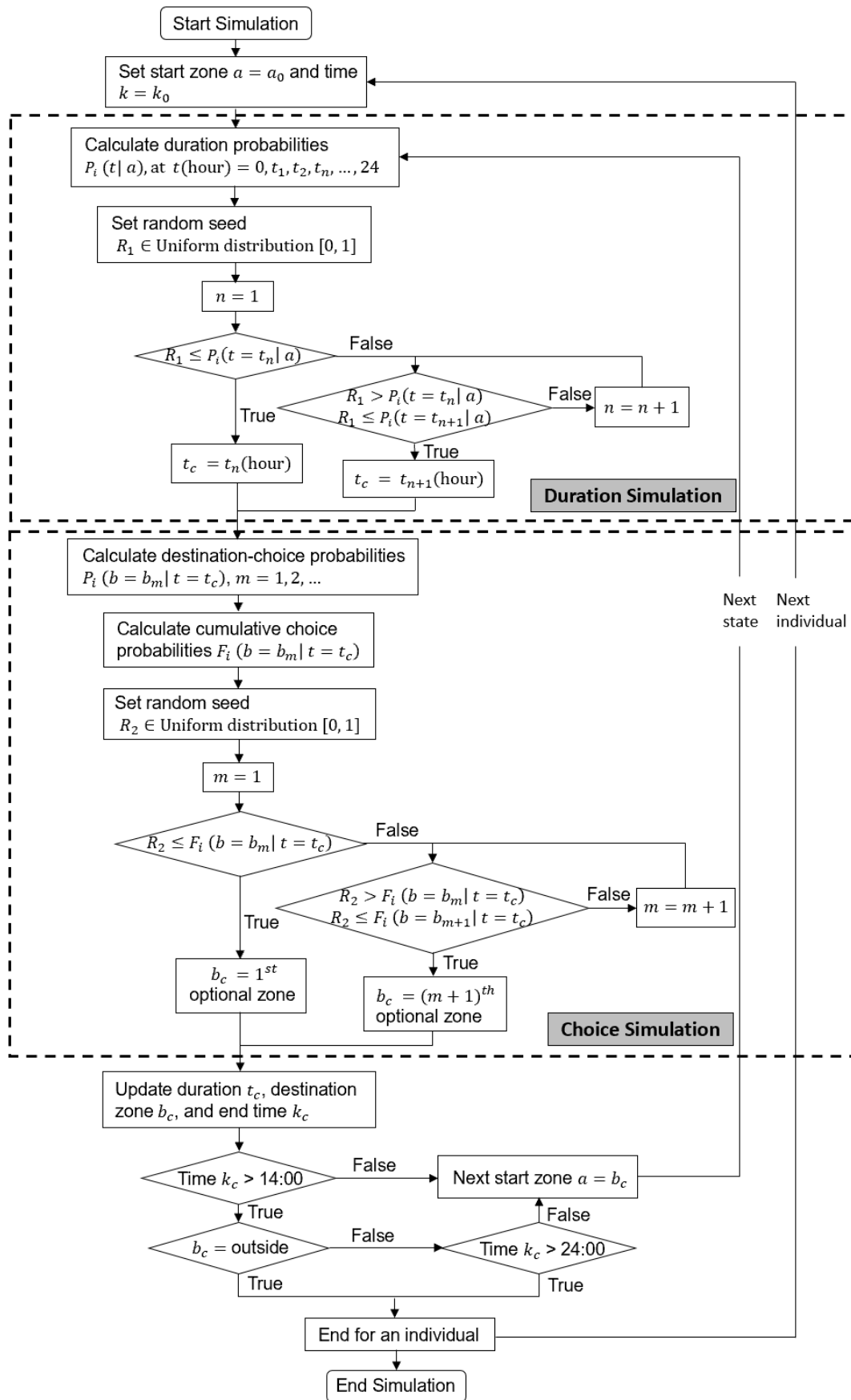
8 predicted by the HBM.

Figure 3. The workflow of HBM simulation for prediction

1   Figure 3 depicts the workflow of the HBM simulation process. The process is initiated

2   from an individual's first observed presence (obtained from empirical arrival times) to the end

3   of the day. This follows from the assumption that the initial state at the start of each individual

4   simulation can be obtained from separate arrival models, not considered in this study. The

5   simulation for each occupancy state consists of two parts: duration simulation and choice

6   simulation. As the start zone a is known, the occupancy duration can be derived from the time

7   of the next transitions out of zone a. Hence, the probability of the occupancy duration at zone

8   a is defined as Eq. (14):

$$P_i(t \mid a) = 1 - \prod_{b=1}^{B} S_{i,a,b}(t) \tag{14}$$

9   Using a random draw, the above can be used to simulate occupancy duration t_c in zone a.

10   To complement the above, the destination zone choice function was used to simulate

11   the destination. We assumed that transition was an instant process and the time when

12   occupancy ended was also the time when transition happened. From the Bayes' theorem, it

13   is possible to observe that the probability of transition from zone a to zone b given that a

14   transition occurred at time t_c (obtained from the duration model) is given by the probability of

15   transition from zone a to zone b over the past t_c hours divided by the probability of transition

16   to any zone, shown as Eq. (15).

$$P_i(b \mid t = t_c) = \frac{1 - S_{i,a,b}(t_c)}{\sum_{e=1}^{B} \left(1 - S_{i,a,e}(t_c)\right)} \tag{15}$$

17   The estimated choice probability was then cumulated over all possible destination zones,

18   generating the cumulative probability distribution of transition, defined by Eq. (16). A random

19   draw from this distribution provided the simulated next occupancy zone.

$$F_i(b = b_m \mid t = t_c) = \sum_{b \leq b_m} P_i(b \mid t = t_c) \tag{16}$$

20   Subsequently, the estimated duration and next occupancy zone were used to update

21   covariates for the simulation of the next state. The individual simulation was repeated until

1   either an individual exited the building after 14:00 (assumed to leave for home) or the time

2   crossed 24:00 with an individual still within the building. Aggregation of simulated results for

3   individuals provided occupancy profiles for the building. The simulation for one workday was

4   run 100 times consecutively and randomly.

5   **3.3 Markov Chain model formulation**

6   A first-order discrete time MCM was implemented as a reference method to compare

7   against the proposed HBM. In this paper, states during the Markov process were zone-level

8   locations. Transition probabilities between state a and b are estimated by counting the

9   frequency of transitions in historical data [8]:

$$P_{a,b} = \frac{N_{a,b}}{\sum_{e=1}^{B} N_{a,e}} \tag{17}$$

$$\sum_{b=1}^{B} P_{a,b} = 1 \tag{18}$$

10 where $N_{a,b}$ is the number of transitions from zone a to zone b in the training data. Eq. (18)

11 expresses the constraint of transition probabilities from the same origin a. Transition

12 probability can be specified to the time of day, which has been widely implemented in literature.

13 The time-specific probability is estimated by using historical data during the specific time

14 period. The estimated transition probability matrix is shown as:

$$P = \begin{bmatrix} P_{1,1} & \cdots & P_{1,b} \\ \vdots & \ddots & \vdots \\ P_{a,1} & \cdots & P_{a,b} \end{bmatrix} \tag{19}$$

15 Similar to the choice simulation process for HBM, a random seed generated from the uniform

16 distribution between 0 and 1 was compared with cumulative transition probabilities over

17 possible destinations. One difference from the HBM simulation is that, instead of using survival

18 functions to determine duration, ongoing occupancy at the next state is still characterised by

19 the transition probability $P_{a,b}$ where $a = b$.

3.4 Case study description

Our testbed, Imperial College Faculty Building, is a four-storey administrative building in South Kensington Campus, London, UK. This building is made up of open plan offices, individual offices and meeting rooms, capable of accommodating up to 350 employees. The building has two entrances: the main access gate on level 1 and a minor rear entrance under level 1 from the bike storage area.
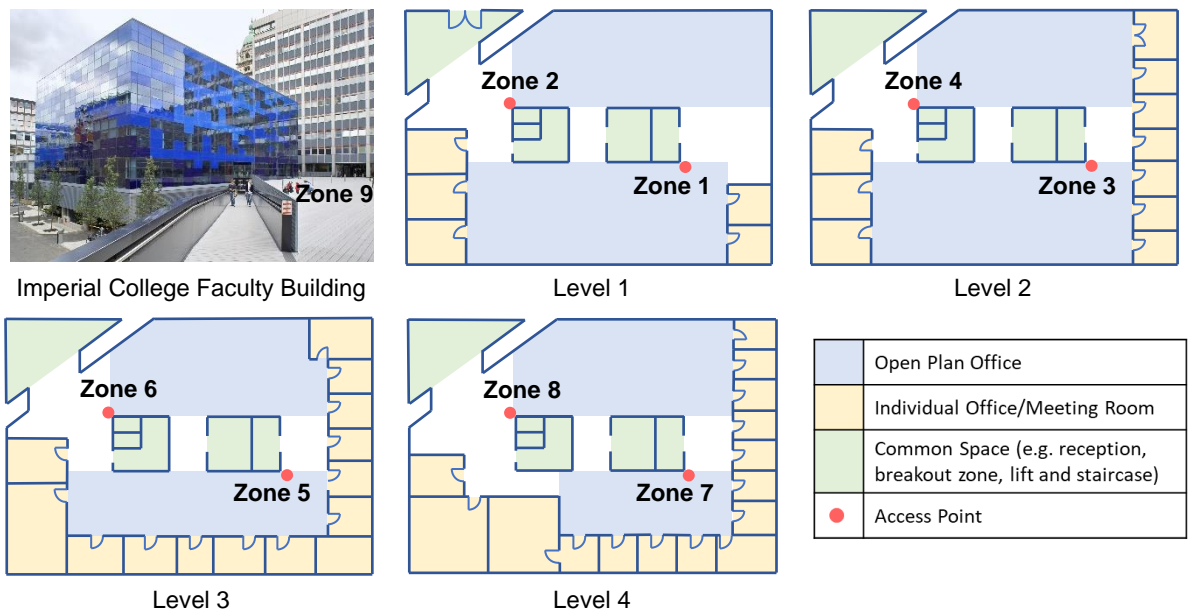

Figure 4. Real case study building and its layouts on each level

The occupancy data for calibrating the two models was harvested from the existing Wi-Fi infrastructure of the testbed on dates between 28 April 2017 and 18 May 2017. This Wi-Fi infrastructure only allowed for passive scanning and the default frequency of protocol scanning was 10 minutes. Access points (APs) of the infrastructure worked as detectors for recognising the existence of Wi-Fi equipped devices. Two APs were installed on each floor and totally eight APs worked to support the wireless network for the whole building (see Figure 4). Depending on the AP signal coverage, the space within the whole building was divided into 8 indoor zones and 1 'outdoor zone' representing occupant absence in the building, which is shown in Figure 4 as zone 9. The next section outlines the Wi-Fi data processing procedure.
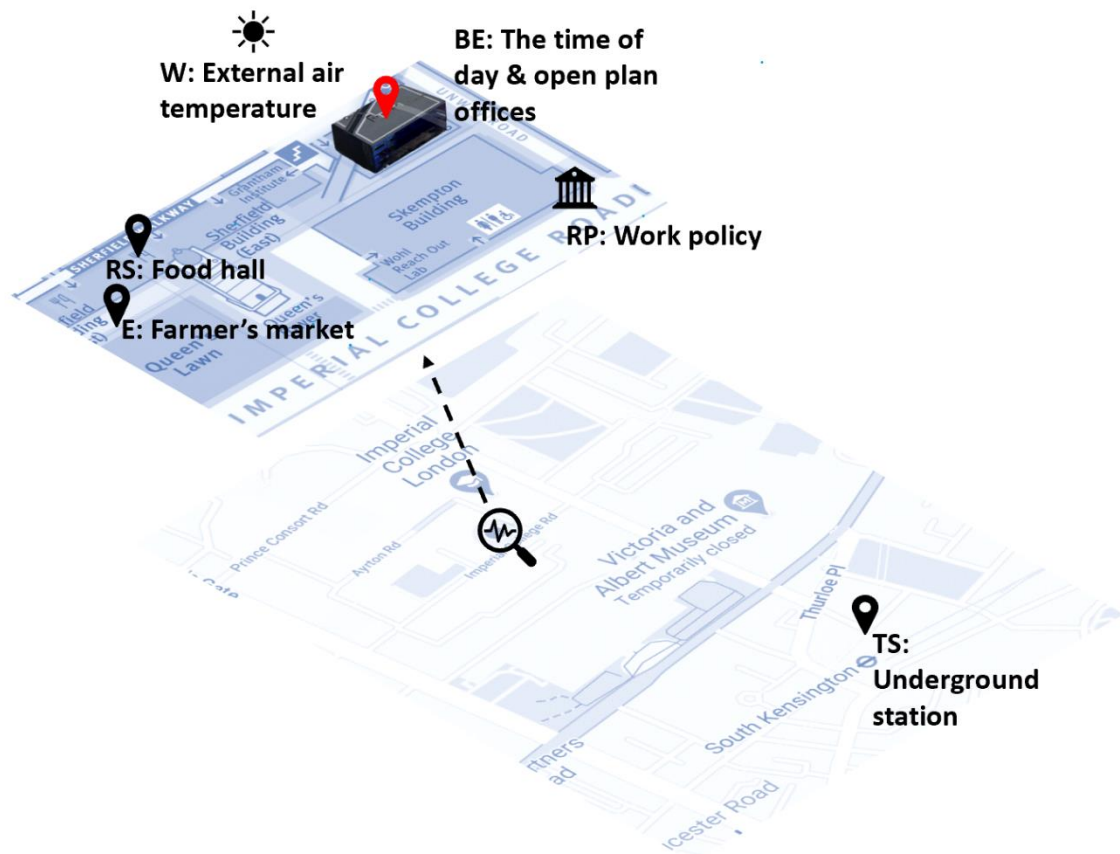
Figure 5. The spatial relationship between the case study building and the surrounding urban system

1   To complement the occupancy data, exogenous covariates which are hypothesised to be

2 correlated with the shape parameters of the occupancy distribution in Eq. (11) were gathered

3 to describe the urban system (as shown in Figure 5) as well as the building envelope, in line

4 with the conceptual model in section 3.1:

5   • *TS*: The *business of South Kensington underground station*. South Kensington station

6   is the closest underground station to the testbed. Three lines in the station serve the

7   high travel demand at South Kensington. The transport covariate used was the ratio of

8   passenger flow at a given time to the maximum passenger flow in a sample in 2017

9   [63].

- *E*: *London Farmer's Market at Imperial.* This event takes place on every Tuesday between 9:00 and 14:00, and offers fresh produce as well as takeaway lunch bites. The time of the event was used as an independent variable.

- *RS*: *Senior Common Room (SCR)* is a food retailer open to college staff and students during lunch time of every workday (11:45-14:30). Its opening time was used as an independent variable.

- *RP*: *Weekends and public holidays*. This was represented with variables for the weekends and one UK bank holiday during the data collection period.

- *W*: *External air temperature.* The impact of weather was studied by including a variable for the outdoor temperature (°C) which was extracted from Weather Underground [64].

- *BE*: *The time of day and the coverage of open plan offices.* The time of day, as the building variable in the temporal dimension, was defined in terms of work routine in this case; the proportion of open plan area in each zone was a variable to reflect the spatial impacts of building.
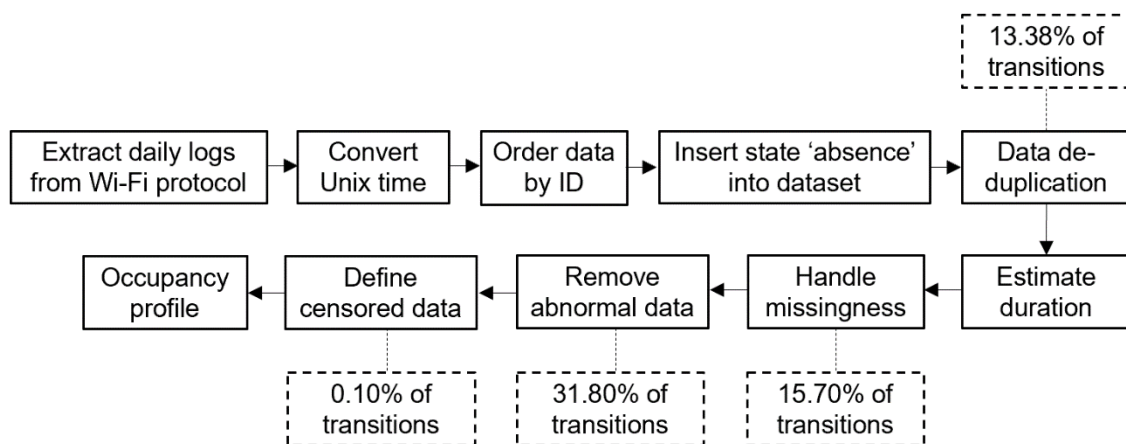
3.5 Occupancy data pre-processing

Figure 6. The flowchart of data pre-processing

1   As discussed in section 2.3, since Wi-Fi infrastructure is not designed to specifically

2   measure occupancy, it cannot be used directly for model calibration. Figure 6 presents

3   guidance showing how to process the raw Wi-Fi connection logs.

4   The raw data has been extracted using the Simple Network Management Protocol

5   (SNMP), which provided a database of devices connected to Wi-Fi networks in the building at

6   10-minute intervals, denoted by Unix date-time stamps. The initial step therefore was to

7   convert the Unix time to readable date and times and classify data according to those and the

8   anonymous IDs. For each individual, absences following early arrival in logs were considered

9   'out of building' (denoted as zone 9). After the above processing, the data clearly and

10   straightforwardly presented individual occupancy at each 10-minute time step.

11   The next steps focused on resolving record duplications that reflected as two different

12   APs (and hence zones) simultaneously associated to the same ID at a particular date-time.

13   The duplication likely resulted from either movement during the period of scanning the

14   connected devices as part of the SNMP. However, in the absence of further information

15   regarding order of visit, signal strength or physical proximity to APs, we assumed that the

16   zones had equal chance to be occupied by that respondent. In this case, the weight w_i in the

17   log-likelihood function (recall Eq. (13)) was set to 0.5 (such weights could be adjusted if the

18   aforementioned information were available). With the duplicated data resolved, occupancy

19   duration for each state could be estimated as the difference between end and start times.

20   The next step concerned dealing with missingness in the data. This could result from

21   either transition being unrecognised during the temporal gap between the consecutive

22   connections scans due to the low frequency of scanning or from short-term disconnections

23   because of signal block in lifts or staircases. The former issue cannot be observed and

24   handled based on the available logs, as it would require an increased scanning frequency. In

25   contrast, the second issue was captured by short-term absence at specific time steps in the

26   dataset. The duration of real individual absence was assumed to be longer than 10 minutes

1 while all 10-minute absences (15.70% of transitions in the dataset) were removed. All

2 absences with durations longer than 10 minutes were assigned to the out-of-building zone.

3 At the other end of the spectrum, observed abnormal durations longer than 960

4 minutes was the third issue to handle. The maximum 960-minute threshold was identified by

5 supposing that an occupant appears at the earliest time of security permission in the morning

6 and leaves at the latest permitted time in the evening. The longer durations could result from

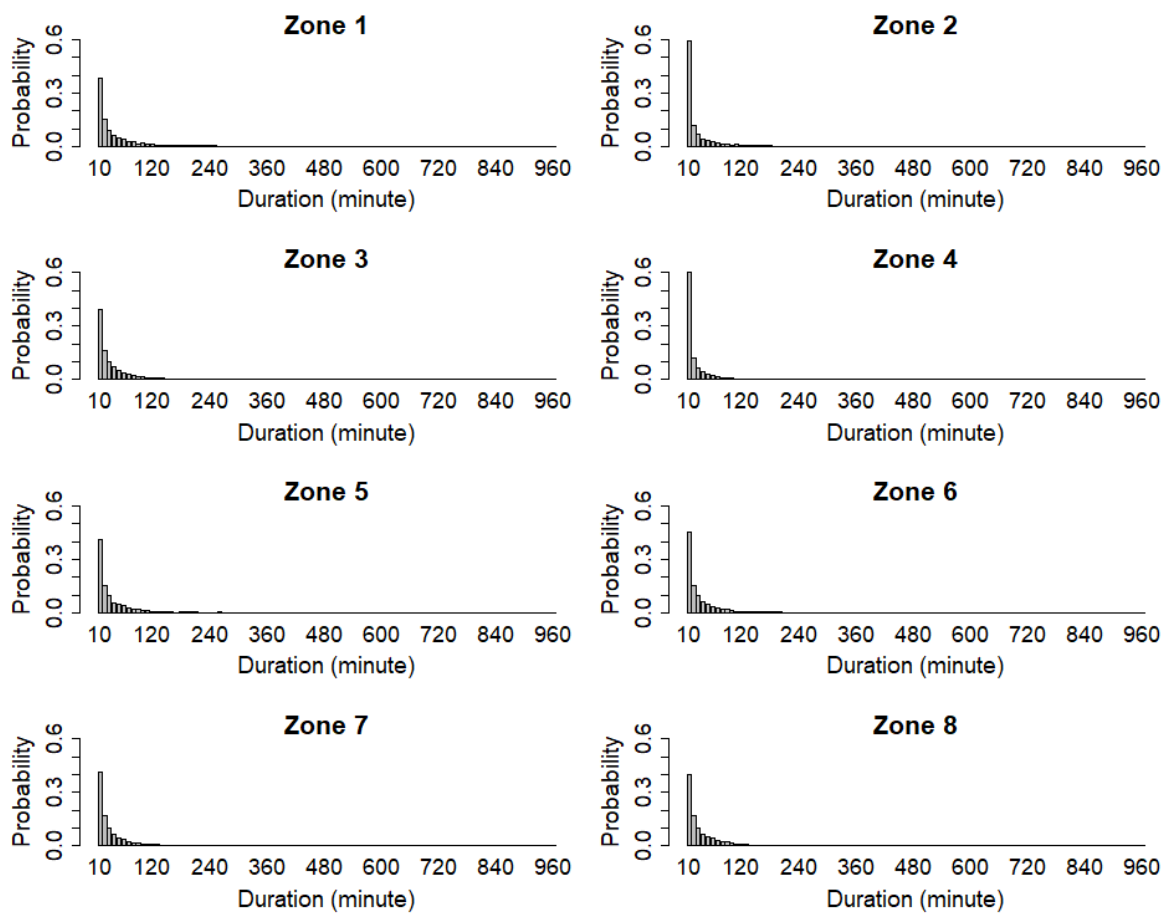7 stationary Wi-Fi items (e.g. printers) or left devices, which were deleted from the data.

Figure 7. The empirical distributions of occupancy duration in each indoor zone

8 Another abnormal duration among the observations was the 10-minute duration in the

9 building, which formed 31.8% of all occupancy states (see Figure 7). The high proportion of

10 short-term occupancy could have been caused by the detection of building passers-by. In

11 other contexts, such records could reflect real occupancy for some quick activities, such as

1 buying coffee. However, as this was unlikely in the current context, such observations were

2 removed from the dataset.

3 The last issue was about the right censoring that ID appears at a certain time and stays

4 until the end of the day. As respondent identifiers were reset every midnight for privacy

5 protection, it was unlikely to observe when and whether the censored ID left the building.

6 Therefore, we incorporated such data into survival models as shown in Eq. (13). Following the

7 data pre-processing, the final sample used for model estimation was made up of 27,286

8 transitions out of which 28 (0.10%) were censored and 3651 (13.38%) were duplicated.

3.6 Performance metrics

10 The modelling performance of the HBM and MCM were evaluated from four

11 perspectives: transition probability distribution, covariate interpretation, occupancy count and

12 duration prediction. The estimated transition probabilities by two baseline methods (with only

13 constant terms) were compared to indicate the advantages of HBM in capturing more complex

14 information with the 'new' dimension of duration. The interpretability of HBM, which is limited

15 with MCMs, enabled reflections on the external impacts on occupancy duration and transition.

16 The estimated coefficients reported the relative importance of those variables in influencing

17 occupancy patterns in the testbed. The occupancy prediction was validated in future unseen

18 data during workdays between 19 May 2017 and 25 May 2017. 100 runs of simulation were

19 done for each day using each model, using the empirical arrival data. The average of the 100

20 occupancy-count outputs was the result for accuracy validation. The precision of prediction

21 was validated through all of the simulation results.

22 The accuracy of prediction was evaluated using four metrics: root-mean-square error

23 (RMSE), normalised root-mean-square error (NRMSE), Kolmogorov–Smirnov (K-S) test and

24 Kullback-Leibler (K-L) divergence. RMSE and NRMSE were used to measure differences

25 between predicted mean occupancy counts and observed occupancy counts. The RMSE and

26 NRMSE applied in the validation were defined as:

$$RMSE = \sqrt{\frac{\Sigma_{t=1}^{144}(Pred_t - Obs_t)^2}{144}} \tag{20}$$

$$NRMSE = \frac{RMSE}{Q1 - Q3} \tag{21}$$

1   where $Q1$ and $Q3$ are perspective values at the 25th and 27th percentile. The K-S and K-L tests

2   evaluated similarity of distributions of the observed and predicted durations. Lastly, we also

3   provided percentage of time periods for which observed occupancy counts belong to the

4   ranges of predictions at 95% and 99% confidence level.

5   **4. Results and Discussion**

6   **4.1 Comparison of transition probability distributions**

7   Comparison between the MCM and HBM with respect to the distribution of the

8   transition probabilities requires plotting of the underlying PDFs. However, a PDF underlying a

9   HBM parameterised using covariates will also depend on those covariates, themselves being

10 particular to the circumstances they reflect. Hence for the purpose of demonstrating how the

11 HBM captures more complex duration-dependencies in the transition distributions when

12 compared to the MCM, we use the example of a baseline HBM which includes only constant

13 terms. This ensures that the model does not vary across the sample, while still reflecting the
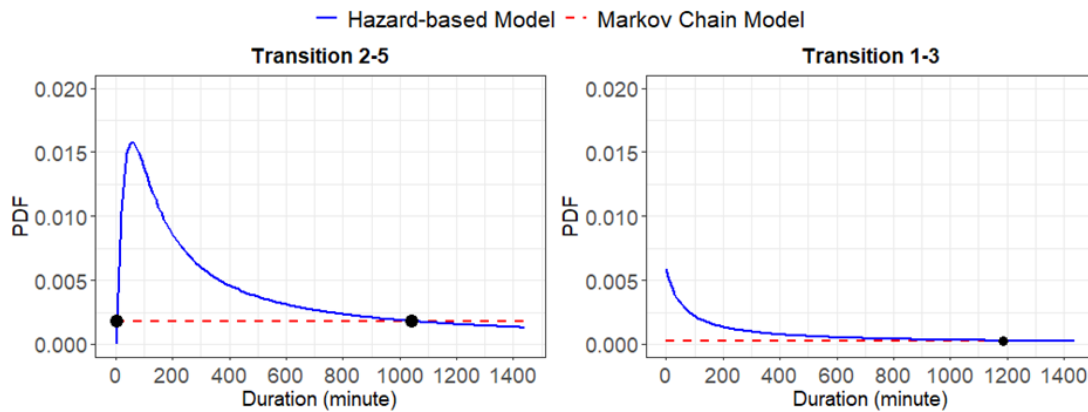
14 postulated flexibility.

Figure 8. PDFs of duration times by baseline HBM and homogenous MCM

1   Two typical distributions of occupancy duration, which is also the probability of

2   transition happening, estimated by two models are shown in Figure 8 (the estimated

3   parameters of the baseline HBM and the transition matrix for the MCM are reported in Table

4   A.1 and Table A.4). As expected, the results reflect that while the MCM assumes an underlying

5   uniform probability distribution of transitions across time (given the constant transition matrix),

6   the HBM relaxes this assumption. In other words, the inspection of estimated baseline HBM

7   parameters reveals that most transition probabilities, as computed using Eq. (8), have

8   unimodal distributions while the remaining are monotonic decreasing, which can be seen in

9   Table A.2. For the unimodal distributions, this means that probability of transition increases

10   until the peak (mode) at between 60 to 120 minutes, and falls afterwards. Inclusion of

11   covariates will alter shapes of those distributions, reflecting impacts of the circumstances on

12   occupancy behaviour.

13   **4.2 Exogenous impacts on duration and transition in the HBM**

14   Inclusion of covariates in the HBM allows us to reflect the impact of various

15   circumstances on occupancy behaviour while also providing insights into the determinants of

16   such behaviour, through parameter interpretation.

17   **Table 3 The estimation results of the HBM with urban framework variables (sample**
18   **size I=27,286)**

Covariate	Indoor zonal transitions	Leaving the building	Entering the building
TS:			
The crowding of the nearest underground station	0.226*	-0.022	-0.254***
E:			
Tuesday Farmer's Market	0.056	-0.036.	-0.092**
RS:			
Senior Common Room Lunch time	0.190*	0.047	0.376***
RP:			
Weekends and public holidays		0.010	

W:			
External air temperature (°C)	-0.110	0.825***	0.955***
BE:			
The ratio of open plan office to the total area of origin zone		-2.918***	
Time of day, reference time:			
20:00 – 07:00 (workdays)	0 (fixed)	0 (fixed)	0 (fixed)
07:00 - 09:00 (workdays)	-0.080	-0.088*	-0.006
09:00 - 11:45 (workdays)	0.213*	0.141***	0.218***
14:30 - 16:00 (workdays)	0.453***	0.507***	0.468***
16:00 - 18:00 (workdays)	0.309*	0.531***	0.194*
18:00 - 20:00 (workdays)	-0.931*	0.453***	0.331**

Goodness of fit
The log-likelihood of baseline model: -56444.884
The loglikelihood of specific covariate model: -55679.265

1 $*** \, p \leq 0.001 \;\; ** \, p \leq 0.010 \;\; * \, p \leq 0.050 \; . \, p \leq 0.100$

2 Table 3 presents the estimated coefficients of exogenous variables alongside the goodness

3 of fit. The final log-likelihood value of the models with covariates and without (baseline) are

4 respectively -55679.265 and -56444.884. Likelihood ratio test shows that the test statistic

5 (1531.238) with 29 degrees of freedom is much larger than the critical chi-squared value

6 (49.588) at the 99% confidence level. It demonstrates that the model with exogenous

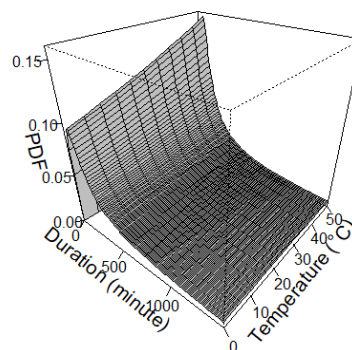7 covariates has a statistically significantly better fit.
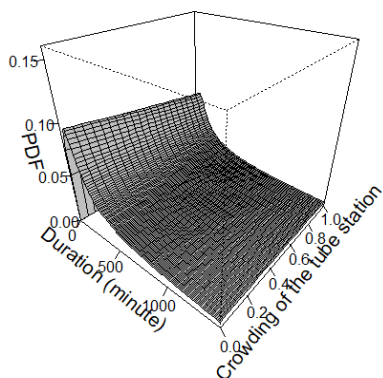
Figure 9. Example plots of transition probabilities respectively against transport and weather covariates
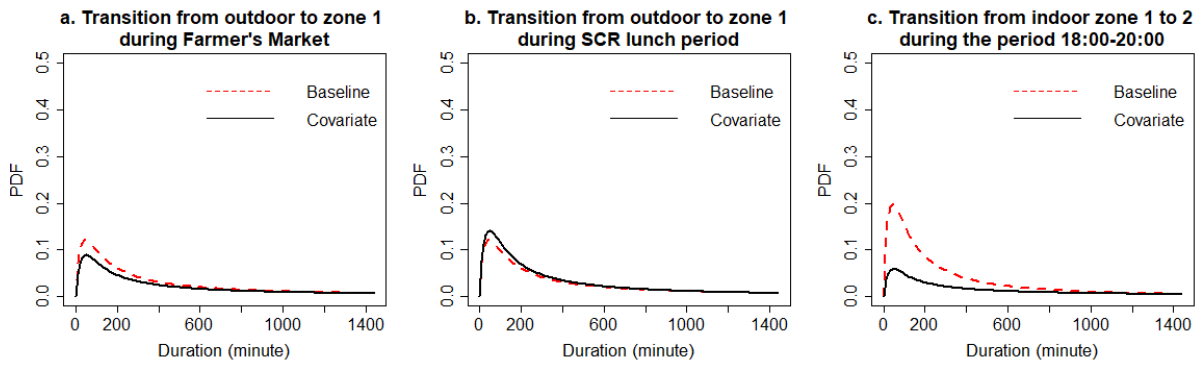
Figure 10. Example plots of transition probabilities before and after respectively adding event, retail and time covariates

1 With respect to interpreting coefficients, positive values indicate higher rate of

2 transition hence shorter expected duration of occupancy within a zone and vice versa. The

3 three-dimensional plots in Figure 9 display examples of how transition probabilities are

4 simultaneously affected by duration and exogenous continuous variables (i.e. crowding of tube

5 station and external air temperature). Figure 10 shows effects of exogenous binary variables

6 (i.e. Farmer's market, SCR lunch and time of day) on transition probability in two-dimensional

7 spaces.

8 The coefficients associated with the crowding of the nearest underground station suggest

9 a higher rate of transition within the building and a lower rate of entrance to the building as the

10 passenger flow increases. The former might indicate an increased circulation in the building

11 around peak times. The latter association could reflect the impact of station congestion on

12 arrival rate to the building. It is also possible that the coefficient captures the effect of increased

13 congestion at the station that follows end-of-day departures from the nearby buildings,

14 including the case study building. It is unlikely, however, due to the relative size of the building

15 occupancy to the station flows that there is a substantial impact of the case study building on

16 the station congestion. Nevertheless, this points out the possibility of converging transport

17 modelling and building occupancy modelling to calibrate modelling accuracies in both domains.

18 With respect to event variable, we only observed a negative impact on the transitions

19 to inside, which could indicate flatter profile of returns to the building, possibly as some

occupants could choose to visit the event. As for the SCR lunch, we observed that occupants are more likely to move within the building and enter the building when the food retail opens. This is not surprising, as at lunch time people would tend to have breaks and move around the building more, to access facilities (e.g. kitchenettes), though for shorter periods. We also observed higher rates of entering the building (see Figure 10b), which could indicate returns from lunch break but also from earlier, morning commitments with lunch time as a natural break point. This interpretation would also help explain why a similarly statistically significant effect on transitions outside was not observed. In terms of the positive coefficients associated with external temperature, we observe that occupants are more likely to make transitions between the building and the outside with warmer weather. This would be expected given the period of study (April-May) and public preference for going outside during warm weather in the UK.

When it comes to the building characteristics, the coefficient of open plan office reflects fewer transitions from the zone and longer occupancy duration with a larger ratio of open plan office. It suggests the open workplace satisfies most occupants in this case and hence there is no need for retrofitting the interior design. In contrast, the higher transition rate from the zone with the larger open area might indicate frequent movement of occupants due to noise disturbances. The open workplace is just one example of covariates representing effects of building layout on occupancy behaviour. In future implementations, more building characteristics could be calibrated in the model and then the estimated coefficients provide building designers with design and refurbishment suggestions.

These observations from estimation results highlight the need for and help further incorporation between urban-based modelling and building modelling. From the perspective of behaviour, occupancy preferences reflected by the estimation results could be used for assessing the popularity of surrounding systems and the need for developing activities within the campus. In building and energy domains, the parameterised determinants of occupancy and transition demonstrate how the use of building space changes with various urban

1 conditions and then indicate the dynamic pattern of occupancy-related energy uses (e.g. plug

2 loads) within the building. At the urban level, analysing various building occupancies under

3 different surrounding conditions could help urban planners design or retrofit cities for energy

4 efficiency.

5   With respect to the time of day dependency, we observe that as compared to the reference

6 period (between 20:00 and 7:00), the rate of transitions into and within the building between

7 7:00-9:00 is not significantly different. However, there are slightly fewer transitions out of the

8 building, likely reflecting this to be conventional arrival times. We also observe that transitions

9 within the building decrease between 18:00 and 20:00, possibly showing people's tendency

10 to stay focused to finish up their tasks. At other times, transitions are more frequent than during

11 the reference period which reflects general activity during the building working hours as

12 compared to the reference period. These temporal parameters evaluate business schedule of

13 the building, which can be used for adjusting the operation schedule of building management

14 systems (e.g. HVAC and lighting).

4.3 Comparison of full HBM and time-inhomogeneous MCM performance in
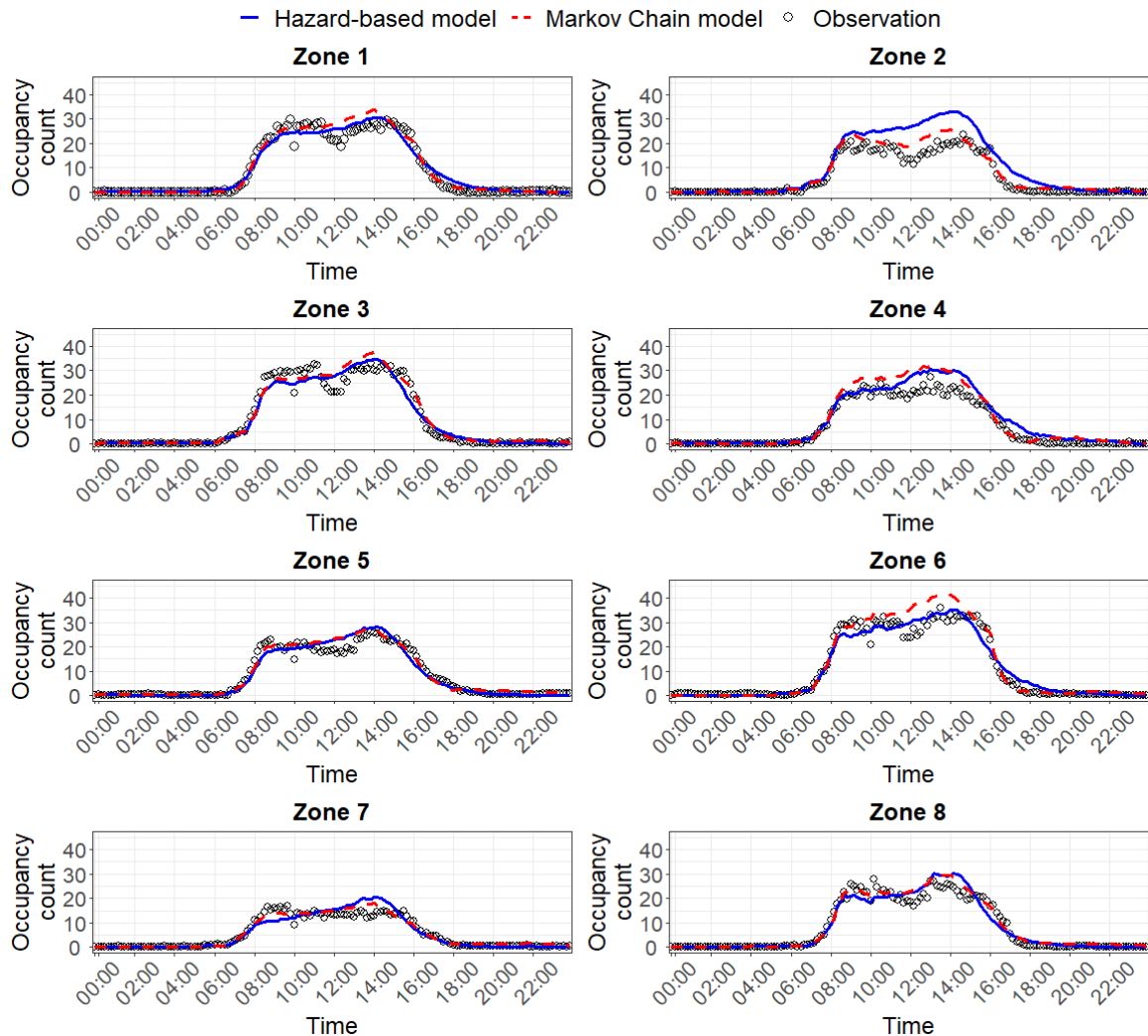
occupancy prediction

Figure 11. Mean occupancy profiles (5 days, subsequent to the estimation dates) of the full HBM (blue solid) and time-inhomogeneous MCM (red dash) for each building zone as compared to the observed Wi-Fi counts.

The mean occupancy-count profiles considering five days of validation data and for the indoor zones (zones 1-8) as predicted by the full (inclusive of covariates) HBM and the time-inhomogeneous MCM are shown in Figure 11. It can be seen from this figure that except for zone 7, all zones have similar occupancy profiles with two peaks and a local reduction around noon. The occupancy profile of zone 7 has the lowest occupancy-count and looks stable without obvious changes during the lunch time. This difference could be due to the different features of occupants: zone 7 at the top of building where most of the senior staff are

1   based. As they occupy individual offices instead of open plan workplaces, the number of

2   occupants is relatively low as compared to other zones. Occupants of those offices might also

3   have a substantial number of external commitments, which reduces their presence and hence

4   the lunch-time patterns observed elsewhere.

5   Considering matches between predicted counts and observed Wi-Fi counts, both

6   models do well to capture the increasing occupant numbers during the morning arrival;

7   predictions of the MCM for the decreasing trend during the period of afternoon departure are

8   fairly closer to the observed Wi-Fi counts. Whereas, the full HBM makes better predictions as

9   compared to the time-inhomogeneous MCM for zones 1, 3, 4 and 6 regarding the occupancy

10   between arrival and departure times; while both methods perform similarly for zones 5 and 8.

11   The HBM performs worse in zone 2, possibly due to the high proportion of noise caused by

12   the presence of the main building entrance. Specifically, the relatively large reception area in

13   zone 2 may attract lots of passers-by detections and hence short-time occupancy for waiting

14   or guest logging. The removal of the 10-minute occupancy observations as part of the data

15   pre-processing may have caused the transition probability distribution to skew towards the

16   longer duration.

17   **Table 4 Performance metrics for evaluating the accuracy of two methods**

Zone	RMSE			NRMSE			K-S statistic		K-L divergence	
	Baseline HBM	HBM	MCM	Baseline HBM	HBM	MCM	HBM	MCM	HBM	MCM
1	3.328	**2.196**	2.405	0.316	**0.227**	0.239	**0.210**	0.386	**0.272**	0.689
2	6.092	6.158	**2.760**	0.565	0.590	**0.343**	**0.209**	0.297	**0.301**	0.505
3	3.473	**2.589**	2.703	0.307	**0.251**	0.264	**0.212**	0.403	**0.273**	0.653
4	4.753	**3.416**	3.852	0.442	**0.350**	0.389	**0.219**	0.346	**0.335**	0.525
5	2.342	2.082	**1.948**	0.257	0.251	**0.220**	**0.190**	0.372	**0.280**	0.620
6	4.178	**2.471**	3.727	0.335	**0.222**	0.296	**0.189**	0.355	**0.291**	0.594
7	2.335	2.305	**1.526**	0.386	0.413	**0.264**	**0.213**	0.385	**0.258**	0.698
8	2.560	2.500	**2.023**	0.285	0.297	**0.226**	**0.196**	0.357	**0.233**	0.547

18   The results of formal performance evaluation for the HBM and the MCM are shown in

19   Table 4. In line with observations from Figure 11, the RMSEs and NRMSEs of the two models

20   indicate that the full HBM does well to capture occupancy counts in zone 1, 3, 4 and 6, and

the time-inhomogeneous MCM outperforms the full HBM for the rest of the zones. The improved RMSEs and NRMSEs of HBM when adding urban covariates demonstrates the usefulness of integrating urban systems impacts into building occupancy models. We also observe that urban covariates greatly help improve occupancy-count predictions for zones in which the full HBM outperforms MCM. These findings might suggest that occupants on the lower floor of the building are more sensitive to impacts reflected in the covariates. This could be related with more flexible work schedules allowing them to more easily move around the building and outside as compared to occupants of zones 5, 7 and 8, which host offices of more senior staff with possibly less flexible but also more diverse agendas. However, in absence of further data covering these aspects, exploration of these factors is suggested as a direction of future study.
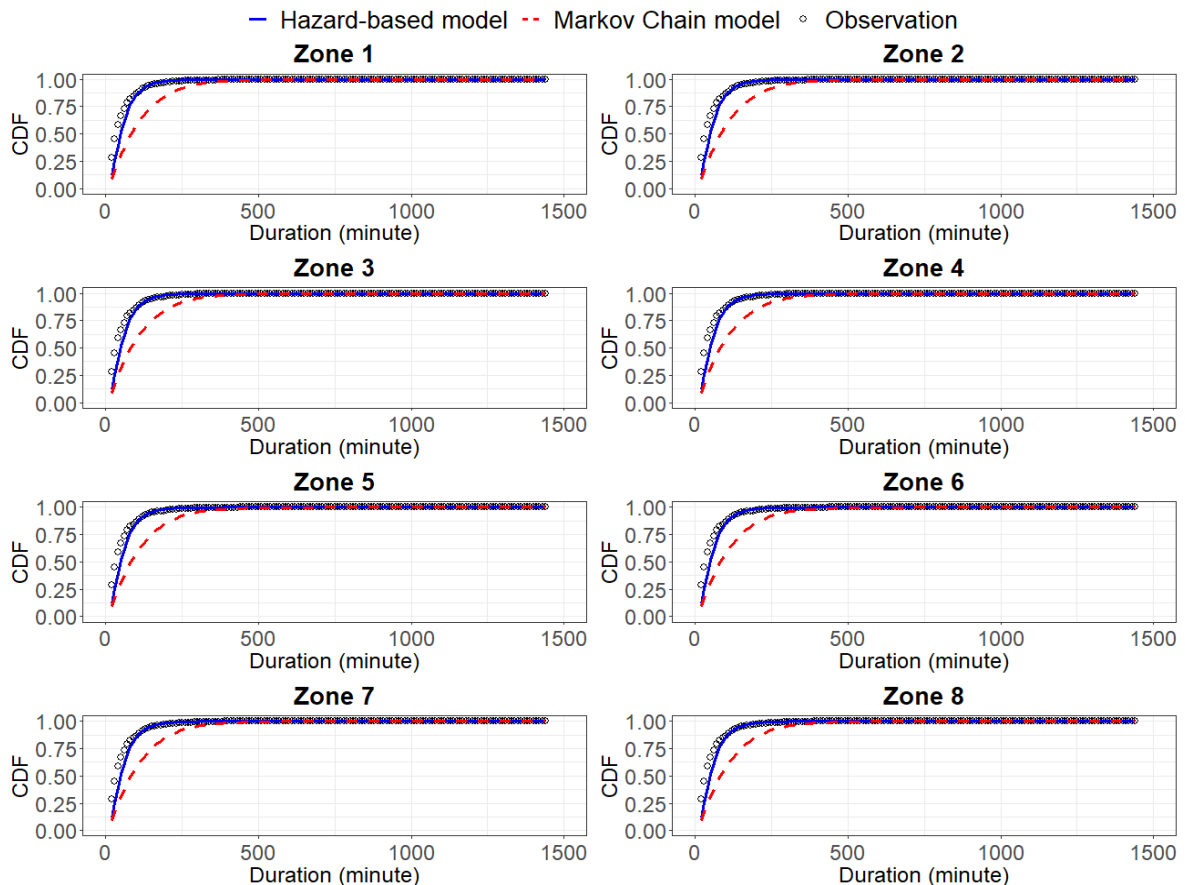
Figure 12. The CDFs of duration estimated from Wi-Fi connection logs, modelled by the full HBM (blue solid) and the time-dependent MCM (red dash)

The results of the K-S tests indicate that from the statistical standpoint, all simulated duration distributions differ from the observed distributions. However, those tests tend to be very rigorous and might be too conservative for evaluating the performance of models of complex behaviour. Looking comparatively at the K-S statistics for the HBM and MCM in Table 4, we observe that in general the HBM outperforms the MCM in reflecting the occupancy duration distributions. This is also confirmed in Figure 12 as well as in the K-L divergence tests which quantify the distance between simulated and observed distributions. It is also found that divergence of the HBM for zones 2 and 4 are higher than that for other zones, which can be the reason for the significant overestimation of occupancy counts in Figure 11. Recalling the empirical distributions in Figure 7, those two zones have the highest proportion of 10-minute occupancy durations. Removing these data might have resulted in missing some real 10-minute occupancy states biasing the duration and occupancy-count predictions. This finding suggests the urgent need for developing methods of more effective filtering of passers-by observations while retaining the real occupant observations.

One potential limitation of the continuous-time full HBM in duration prediction is the treatment of changes in exogenous variables after the prediction process for each occupancy state. In its conventional form, as presented in the current paper, the expected occupancy duration is predicted by considering exogenous factors *at the start* of the occupancy. In reality, the occupant's decision can be influenced by the environment at any moment. For instance, the HBM simulation may predict a long duration outside given the high external temperature. But it would not take into account changes in the temperature *during* that stay, instead considering the temperature again only at the point when the expected outside duration ends. This issue could be particularly severe for long episodes and rapidly changing covariates. In this study, relatively high proportions of shorter durations in the sample (see Figure 7) and slow changes of exogenous variables over time jointly lead to fewer biases in duration predictions. However, this approach should be cautiously applied to cases with a sudden change in the surrounding environment, e.g. sudden congestion due to an unexpected traffic

1   incident. In future implementations, upper limits might be incorporated to restrict the

2   overestimation of duration. Alternatively, dynamic HBMs [65] may be used at the cost of

3   additional complexity.
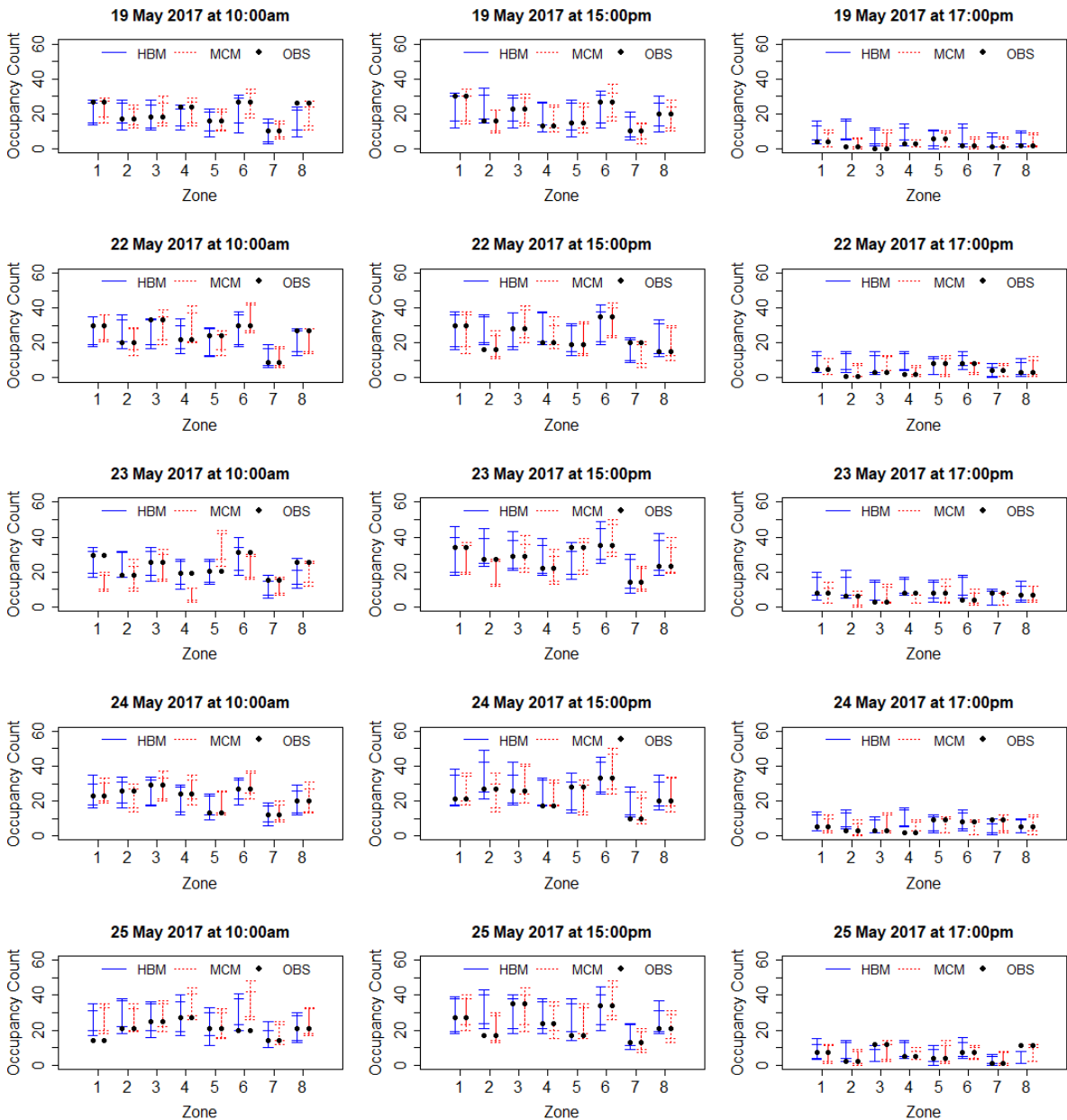
Figure 13. The intervals of predicted occupancy count at three specific time points as compared to the observed Wi-Fi counts (OBS)

4   To complement the performance accuracy analysis above, we also analysed precision

5   of the simulated results. In particular we looked at how narrow the occupancy-count profiles

6   were in 100 simulations of each workday. To that end, Figure 13 displays ranges of predicted

7   occupancy for eight zones at 10:00, 15:00 and 17:00. In the time slice, the outer horizontal

lines and inner horizontal lines respectively represent the boundary of occupancy interval at 99% confidence and 95% confidence. Overall, most slices are able to capture the observed Wi-Fi counts but MCM generally produces the narrower range of occupancy counts for most zones at those three times. HBM performs better at 10:00 as 95% of observations fall in their prediction intervals at 99% confidence as compared to 83% by MCM. At 15:00, both methods do well to capture the observed Wi-Fi counts at 99% and 95% confidence, though MCM appears to slightly outperform the HBM. At 17:00, both models perform weaker than the earlier predictions. Although most observations are within the predicted intervals, they are skewed towards the boundary.

Table 5 The percentage of time slice which contains the observed value at 95% confidence and 99% confidence

Zone	95% confidence		99% confidence	
	HBM	**MCM**	**HBM**	**MCM**
1	**90%**	**90%**	**94%**	93%
2	**81%**	80%	**88%**	84%
3	**86%**	85%	**90%**	**90%**
4	**76%**	75%	**83%**	82%
5	84%	**85%**	88%	**89%**
6	**80%**	74%	**87%**	79%
7	**67%**	**67%**	72%	**74%**
8	84%	**88%**	89%	**91%**

Finally, we also looked at how many of the simulated confidence intervals for the 144 time points per day captured the observed Wi-Fi counts in the validation dataset (see Table 5). The results in this table suggest that the HBM is more likely to predict more true values across the entire week for zones 1-4 and 6. With respect to zones 5, 7 and 8, though the MCM performs better, both methods have similarly good performance with a difference of no more than 4%.

4.4 Summary of findings

As discussed above, the HBM has remarkable advantages over the MCM in capturing duration dependence and exogenous impacts, helping gain insights into the interactions between building occupancy and surrounding conditions. The estimation results of the HBM

confirms the significance of time variables which have been widely used in past studies, as well demonstrates that proposed urban variables (i.e. transport, retail, events, building and weather) make statistically significant contributions to fitting occupancy models.

The significant covariates show that building occupancy is linked to the activity of local urban environment such as different retail stores, the transport network and microclimate. As occupancy is a significant driver of building energy use, the incorporation of an urban-aware occupancy model allows designers to consider how the location of a building in the urban environment can affect its energy usage. This understanding could translate into more energy efficient design and operation by accounting for how occupants will use the space considering the context of their urban environment.

The prediction difference between the two methods is small for this case but it could be much larger for buildings with more dynamic occupancy. The reason for this could be that office buildings have fairly stagnant occupancy profiles. Even though they have similar performance, we observed that the HBM produces relatively better predictions for zones which have more occupants with flexible work agendas, who could be significantly driven by exogenous factors. In contrast, the MCM slightly outperforms in simulating occupancy with fixed work agendas.

5. Conclusions

In this paper, we propose an urban-system level framework conceptualising the interactions between building occupancy and urban systems. We demonstrate how the framework can be operationalised using a competing risk hazard model, incorporating Wi-Fi connection logs, to analyse the occupancy of Imperial College Faculty Building in April-May 2017. A time-inhomogeneous discrete-time MCM was also implemented as the reference method. Comparing the performance of the two models, we identify three key findings. Firstly, the proposed HBM model, which combines duration modelling and transition-choice modelling, is capable of capturing the observed transition probability distribution in the duration dimension

as opposed to the stationary distribution assumed in the Markov Chain approach. Secondly, the proposed model operationalises the framework variables in a convenient and flexible way. The estimated coefficients support the strong associations between time of day and occupancy, as well reflect the importance of external covariates in influencing occupancy duration and transition. Thirdly, the proposed model performs better in predicting occupancy patterns which are sensitive to external impacts. Overall, the satisfactory performance of the proposed model suggests that it is a useful tool as an alternative and even an upgrade of the conventional MCM.

Whilst this work is a significant contribution towards independently using the competing risk hazard model to address building occupancy modelling, some issues still exist for further implementations and improvements. The imperfect match between individual trajectory and device trajectory is an issue to handle, for developing an extra value of Wi-Fi logs in occupancy inference. Especially, it was found that a certain proportion of passers-by data in this case could affect the prediction performance. Thus, there remains the need for studying how to effectively filter Wi-Fi data for measuring occupancy. Moreover, this paper only explores the impacts of several urban covariates as a preliminary test of external importance. Future work should, therefore, include more external variables to broadly evaluate urban influences on building occupancy and enhance accuracy and precision of occupancy prediction. The next intended step is also to link the proposed model to building energy models, in order to evaluate how it can improve modelling and possibly reduce building energy consumption.

Acknowledgements

6. Appendices

Table A.1 Estimated constants for α and β parameters in the baseline HBM

θ	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8	Zone 9
Zone 1	-	-1.418***	-5.139***	-3.765***	-4.624***	-3.863***	-4.476***	-3.875***	-0.591***
Zone 2	-1.238***	-	-4.373***	-2.791***	-4.101***	-3.588***	-5.056***	-3.285***	-0.447***
Zone 3	-4.555***	-3.291***	-	-1.879***	-3.294***	-2.948***	-4.891***	-3.670***	-0.436***
Zone 4	-3.430***	-2.481***	-1.380***	-	-3.785***	-1.732***	-6.098***	-3.262***	-0.569***
Zone 5	-4.243***	-3.589***	-3.205***	-3.411***	-	-1.761***	-5.015***	-3.604***	-0.496***
Zone 6	-4.244***	-2.874***	-3.056***	-2.153***	-1.767***	-	-4.046***	-2.095***	-0.553***
Zone 7	-4.165***	-3.291***	-4.256***	-4.538***	-4.310***	-3.111***	-	-1.438***	-0.408***
Zone 8	-3.790***	-2.684***	-3.777***	-3.527***	-3.928***	-1.874***	-1.755***	-	-0.404***
Zone 9	-1.952***	-1.843***	-1.582***	-2.053***	-1.964***	-1.861***	-2.047***	-1.882***	-
φ	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8	Zone 9
Zone 1	-	0.462***	0.281	0.725***	0.709***	0.435**	0.600**	0.701***	0.386***
Zone 2	0.455***	-	0.323	0.602***	0.645***	0.773***	0.921**	0.654***	0.421***
Zone 3	0.866***	0.868***	-	0.586***	0.575***	0.636***	0.883***	0.546***	0.290***
Zone 4	0.486***	0.618***	0.427***	-	0.581***	0.604***	0.939.	0.552***	0.424***
Zone 5	0.621***	0.770***	0.498***	0.742***	-	0.474***	0.661*	0.574***	0.286***
Zone 6	0.407*	0.589***	0.556***	0.504***	0.529***	-	0.530***	0.524***	0.417***
Zone 7	0.598**	0.611***	0.700***	0.776**	0.484*	0.540***	-	0.427***	0.338***
Zone 8	0.829***	0.856***	0.908***	0.782***	0.878***	0.625***	0.538***	-	0.416***
Zone 9	0.513***	0.623***	0.473***	0.654***	0.507***	0.530***	0.512***	0.477***	-

$*** \, p \leq 0.001 \;\; ** \, p \leq 0.010 \;\; * \, p \leq 0.050 \, . \, p \leq 0.100$

Table A.2 Estimated distribution shapes for transition-specific baseline HBMs

	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8	Zone 9
Zone 1		unimodal	monotone decreasing	unimodal	unimodal	unimodal	unimodal	unimodal	unimodal
Zone 2	unimodal		monotone decreasing	unimodal	unimodal	unimodal	unimodal	unimodal	unimodal
Zone 3	unimodal	unimodal		unimodal	unimodal	unimodal	unimodal	unimodal	unimodal
Zone 4	unimodal	unimodal	unimodal		unimodal	unimodal	unimodal	unimodal	unimodal
Zone 5	unimodal	unimodal	unimodal	unimodal		unimodal	unimodal	unimodal	unimodal
Zone 6	unimodal	unimodal	unimodal	unimodal	unimodal		unimodal	unimodal	unimodal
Zone 7	unimodal	unimodal	unimodal	unimodal	unimodal	unimodal		unimodal	unimodal
Zone 8	unimodal	unimodal	unimodal	unimodal	unimodal	unimodal	unimodal		unimodal
Zone 9	unimodal	unimodal	unimodal	unimodal	unimodal	unimodal	unimodal	unimodal	

Table A.3 Estimated constants for α and β parameters in the full (inclusive of covariates) HBM

θ_0	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8	Zone 9
Zone 1	-	0 (fixed)	-3.723***	-2.344***	-3.204***	-2.446***	-3.058***	-2.455***	0.890***
Zone 2	-0.730***	-	-3.864***	-2.280***	-3.591***	-3.077***	-4.544***	-2.776***	0.101
Zone 3	-3.346***	-2.077***	-	-0.671***	-2.088***	-1.741***	-3.682***	-2.466***	0.819***
Zone 4	-2.902***	-1.947***	-0.846***	-	-3.256***	-1.200***	-5.563***	-2.734***	0.032
Zone 5	-3.470***	-2.812***	-2.434***	-2.637***	-	-0.987***	-4.241***	-2.833***	0.319***
Zone 6	-3.902***	-2.529***	-2.711***	-1.809***	-1.421***	-	-3.702***	-1.751***	-0.143***
Zone 7	-3.542***	-2.667***	-3.633***	-3.914***	-3.690***	-2.489***	-	-0.810***	0.254***
Zone 8	-3.756***	-2.648***	-3.742***	-3.493***	-3.894***	-1.839***	-1.721***	-	-0.336***
Zone 9	-2.184***	-2.076***	-1.814***	-2.282***	-2.197***	-2.092***	-2.280***	-2.115***	-
φ	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8	Zone 9
Zone 1	-	0.479***	0.294	0.739***	0.723***	0.449**	0.614**	0.716***	0.428***
Zone 2	0.464***	-	0.329	0.610***	0.652***	0.780***	0.928**	0.661***	0.451***
Zone 3	0.882***	0.886***	-	0.607***	0.592***	0.653***	0.900***	0.562***	0.340***
Zone 4	0.507***	0.641***	0.452***	-	0.603***	0.630***	0.961.	0.573***	0.463***
Zone 5	0.640***	0.789***	0.517***	0.762***	-	0.496***	0.680**	0.593***	0.326***
Zone 6	0.423**	0.605***	0.572***	0.521***	0.548***	-	0.546***	0.542***	0.453***
Zone 7	0.616**	0.629***	0.719***	0.793***	0.501*	0.558***	-	0.452***	0.394***
Zone 8	0.842***	0.870***	0.921***	0.796***	0.892***	0.641***	0.554***	-	0.447***
Zone 9	0.534***	0.645***	0.495***	0.675***	0.529***	0.552***	0.532***	0.498***	-

$*** \, p \le 0.001 \;\; ** \, p \le 0.010 \;\; * \, p \le 0.050 \;. \, p \le 0.100$

Table A.4 Estimated transition probability matrix in the MCM

	Zone 1	Zone 2	Zone 3	Zone 4	Zone 5	Zone 6	Zone 7	Zone 8	Zone 9
Zone 1	0.927	0.013	<0.001	0.001	<0.001	<0.001	<0.001	0.002	0.054
Zone 2	0.022	0.901	<0.001	0.004	0.002	0.002	<0.001	0.003	0.066
Zone 3	<0.001	<0.001	0.918	0.009	0.003	0.003	<0.001	0.002	0.063
Zone 4	0.002	0.004	0.020	0.902	0.002	0.015	<0.001	0.003	0.052
Zone 5	<0.001	<0.001	0.004	0.001	0.927	0.011	<0.001	0.002	0.053
Zone 6	0.001	0.002	0.004	0.008	0.010	0.913	<0.001	0.010	0.052
Zone 7	0.001	<0.001	0.001	<0.001	<0.001	0.002	0.918	0.015	0.061
Zone 8	0.002	0.004	0.002	0.001	0.002	0.011	0.012	0.905	0.062
Zone 9	0.010	0.009	0.014	0.007	0.009	0.009	0.008	0.010	0.924

References

[1]   X. Feng, D. Yan, T. Hong, Simulation of occupancy in buildings, Energy Build. 87 (2015) 348–359. https://doi.org/10.1016/j.enbuild.2014.11.067.

[2]   A.S. Bahaj, P.A.B. James, Urban energy generation: The added value of photovoltaics in social housing, Renew. Sustain. Energy Rev. 11 (2007) 2121–2136. https://doi.org/10.1016/j.rser.2006.03.007.

[3]   B. Dong, D. Yan, Z. Li, Y. Jin, X. Feng, H. Fontenot, Modeling occupancy and behavior for better building design and operation—A critical review, Build. Simul. 11 (2018) 899–921. https://doi.org/10.1007/s12273-018-0452-x.

[4]   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.

[5]   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.

[6]   T. Hong, S. D'Oca, S.C. Taylor-Lange, W.J.N. Turner, Y. Chen, S.P. Corgnati, An ontology to represent energy-related occupant behavior in buildings. Part II: Implementation of the DNAS framework using an XML schema, Build. Environ. 94 (2015) 196–205. https://doi.org/10.1016/j.buildenv.2015.08.006.

[7]   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.

[8]   V.L. Erickson, M.Á. Carreira-Perpiñán, A.E. Cerpa, OBSERVE: Occupancy-based system for efficient reduction of HVAC energy, in: Proc. 10th ACM/IEEE Int. Conf. Inf. Process. Sens. Networks, IPSN'11, 2011: pp. 258–269. https://ieeexplore.ieee.org/document/5779043.

[9]   C. Wang, D. Yan, Y. Jiang, A novel approach for building occupancy simulation, Build. Simul. 40 (2011) 149–167. https://doi.org/10.1007/s12273-011-0044-5.

[10]   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.

[11]   A.A. Adamopoulou, A.M. Tryferidis, D.K. Tzovaras, A context-aware method for building occupancy prediction, Energy Build. 110 (2016) 229–244. https://doi.org/10.1016/j.enbuild.2015.10.003.

[12]   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.

[13]   U. Wilke, F. Haldi, J.L. Scartezzini, D. Robinson, A bottom-up stochastic model to predict building occupants' time-dependent activities, Build. Environ. 60 (2013) 254–264. https://doi.org/10.1016/j.buildenv.2012.10.021.

[14]   D. Wang, C.C. Federspiel, F. Rubinstein, Modeling occupancy in single person offices, Energy Build. 37 (2005) 121–126. https://doi.org/10.1016/j.enbuild.2004.06.015.

[15]   J.A. Oliveira-Lima, R. Morais, J.F. Martins, A. Florea, C. Lima, Load forecast on intelligent buildings based on temporary occupancy monitoring, Energy Build. 116 (2016) 512–521. https://doi.org/10.1016/j.enbuild.2016.01.028.

[16]   A. Roy, Correlates of mall visit frequency, J. Retail. 70 (1994) 139–161. https://doi.org/10.1016/0022-4359(94)90012-4.

[17]   R. Ab Majid, A. Nor Zalika, R. Said, O.O. Olanrele, Drivers Influencing Customers' Visit to Shopping Centres in Malaysia, Int. Surv. Res. J. 5 (2015) 43–58.

[18]   H.-J. Ottevanger, A. Portegies, Sport Tourism: Factors of influence on sport event visit motivation, Master of Arts in European Tourism Management, Bournemouth University (UK), 2007, pp. 7–28.

[19]   Rajagopal, Growing shopping malls and behaviour of urban shoppers, J. Retail Leis. Prop. 8 (2009) 99–118. https://doi.org/10.1057/rlp.2009.3.

[20]   S. Cachero-Martínez, R. Vázquez-Casielles, Developing the Marketing Experience to Increase

Shopping Time: The Moderating Effect of Visit Frequency, Adm. Sci. 8 (2018) 77.

https://doi.org/10.3390/admsci8040077.

[21]   J. Stephenson, B. Barton, G. Carrington, D. Gnoth, R. Lawson, P. Thorsnes, Energy cultures:

A framework for understanding energy behaviours, Energy Policy. 38 (2010) 6120–6129.

https://doi.org/10.1016/j.enpol.2010.05.069.

[22]   R.C. Brownson, D.P. Hopkins, M.A. Wakefield, Effects of Smoking Restrictions in the

Workplace, Annu. Rev. Public Health. 23 (2002) 333–348.

https://doi.org/10.1146/annurev.publhealth.23.100901.140551.

[23]   A.B. Albers, M. Siegel, D.M. Cheng, L. Biener, N.A. Rigotti, Effect of smoking regulations in

local restaurants on smokers' anti-smoking attitudes and quitting behaviours, Tob. Control. 16

(2007) 101–106. https://doi.org/10.1136/tc.2006.017426.

[24]   S. D'Oca, T. Hong, J. Langevin, The human dimensions of energy use in buildings: A review,

Renew. Sustain. Energy Rev. 81 (2018) 731–742. https://doi.org/10.1016/j.rser.2017.08.019.

[25]   F. Haldi, D. Robinson, Interactions with window openings by office occupants, Build. Environ.

44 (2009) 2378–2395. https://doi.org/10.1016/j.buildenv.2009.03.025.

[26]   A.G. Parsons, The Association Between Daily Weather and Daily Shopping Patterns,

Australas. Mark. J. 9 (2001) 78–84. https://doi.org/10.1016/s1441-3582(01)70177-2.

[27]   I. Eliasson, I. Knez, U. Westerberg, S. Thorsson, F. Lindberg, Climate and behaviour in a

Nordic city, Landsc. Urban Plan. 82 (2007) 72–84.

https://doi.org/10.1016/j.landurbplan.2007.01.020.

[28]   V. Fabi, R.V. Andersen, S. Corgnati, B.W. Olesen, Occupants' window opening behaviour: A

literature review of factors influencing occupant behaviour and models, Build. Environ. 58

(2012) 188–198. https://doi.org/10.1016/j.buildenv.2012.07.009.

[29]   S. D'Oca, T. Hong, A data-mining approach to discover patterns of window opening and

closing behavior in offices, Build. Environ. 82 (2014) 726–739.

https://doi.org/10.1016/j.buildenv.2014.10.021.

[30]   V. Motuziene, T. Vilutiene, Modelling the effect of the domestic occupancy profiles on

predicted energy demand of the energy efficient house, Procedia Eng. 57 (2013) 798–807. https://doi.org/10.1016/j.proeng.2013.04.101.

[31]   R. Goldstein, A. Tessier, A. Khan, Space layout in occupant behavior simulation, in: Proc. Build. Simul. 2011 12th Conf. Int. Build. Perform. Simul. Assoc., 2011: pp. 1073–1080.

[32]   C. Zimring, A. Joseph, G.L. Nicoll, S. Tsepas, Influences of building design and site design on physical activity: Research and intervention opportunities, Am. J. Prev. Med. 28 (2005) 186–193. https://doi.org/10.1016/j.amepre.2004.10.025.

[33]   E. Delzendeh, S. Wu, A. Lee, Y. Zhou, The impact of occupants' behaviours on building energy analysis: A research review, Renew. Sustain. Energy Rev. 80 (2017) 1061–1071. https://doi.org/10.1016/j.rser.2017.05.264.

[34]   G. Happle, J.A. Fonseca, A. Schlueter, A review on occupant behavior in urban building energy models, Energy Build. 174 (2018) 276–292. https://doi.org/10.1016/j.enbuild.2018.06.030.

[35]   P.D. Andersen, A. Iversen, H. Madsen, C. Rode, Dynamic modeling of presence of occupants using inhomogeneous Markov chains, Energy Build. 69 (2014) 213–223. https://doi.org/10.1016/j.enbuild.2013.10.001.

[36]   C. Liao, Y. Lin, P. Barooah, Agent-based and graphical modelling of building occupancy, J. Build. Perform. Simul. 5 (2012) 5–25. https://doi.org/10.1080/19401493.2010.531143.

[37]   Y. Chen, T. Hong, X. Luo, An agent-based stochastic Occupancy Simulator, Build. Simul. 11 (2018) 37–49. https://doi.org/10.1007/s12273-017-0379-7.

[38]   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.

[39]   S. D'Oca, H.B. Gunay, S. Gilani, W. O'Brien, Critical review and illustrative examples of office occupant modelling formalisms, Build. Serv. Eng. Res. Technol. 40 (2019) 732–757. https://doi.org/10.1177/0143624419827468.

[40]   D. Ettema, A. Borgers, H. Timmermans, Competing risk hazard model of activity choice, timing, sequencing, and duration, Transp. Res. Rec. 1493 (1995) 101–109.

https://pure.tue.nl/ws/files/1872442/589548.pdf.

[41]   C.R. Bhat, A hazard-based duration model of shopping activity with nonparametric baseline specification and nonparametric control for unobserved heterogeneity, Transp. Res. Part B Methodol. 30 (1996) 189–207. https://doi.org/10.1016/0191-2615(95)00029-1.

[42]   J. Pawlak, J.W. Polak, A. Sivakumar, A framework for joint modelling of activity choice, duration, and productivity while travelling, Transp. Res. Part B Methodol. 106 (2017) 153–172. https://doi.org/10.1016/j.trb.2017.10.010.

[43]   K. Christensen, R. Melfi, B. Nordman, B. Rosenblum, R. Viera, Using existing network infrastructure to estimate building occupancy and control plugged-in devices in user workspaces, Int. J. Commun. Networks Distrib. Syst. 12 (2014) 4–29. https://doi.org/10.1504/IJCNDS.2014.057985.

[44]   R. Zhang, K.P. Lam, Y.S. Chiou, B. Dong, Information-theoretic environment features selection for occupancy detection in open office spaces, Build. Simul. 5 (2012) 179–188. https://doi.org/10.1007/s12273-012-0075-6.

[45]   C. Jiang, M.K. Masood, Y.C. Soh, H. Li, Indoor occupancy estimation from carbon dioxide concentration, Energy Build. 131 (2016) 132–141. https://doi.org/10.1016/j.enbuild.2016.09.002.

[46]   S.H. Ryu, H.J. Moon, Development of an occupancy prediction model using indoor environmental data based on machine learning techniques, Build. Environ. 107 (2016) 1–9. https://doi.org/10.1016/j.buildenv.2016.06.039.

[47]   G. Conte, M. De Marchi, A.A. Nacci, V. Rana, D. Sciuto, BlueSentinel: A first approach using iBeacon for an energy efficient occupancy detection system, in: BuildSys 2014 - Proc. 1st ACM Conf. Embed. Syst. Energy-Efficient Build., 2014: pp. 11–19. https://doi.org/10.1145/2676061.2674078.

[48]   A. Corna, L. Fontana, A.A. Nacci, D. Sciuto, Occupancy detection via iBeacon on Android devices for smart building management, in: Proc. -Design, Autom. Test Eur. DATE, 2015: pp. 629–632. https://doi.org/10.7873/date.2015.0753.

[49]   W. Wang, J. Chen, T. Hong, Modeling occupancy distribution in large spaces with multi-feature classification algorithm, Build. Environ. 137 (2018) 108–117. https://doi.org/10.1016/j.buildenv.2018.04.002.

[50]   J.Y. Park, E. Mbata, Z. Nagy, Good to see you again: Capture and recapture method on mobile devices to estimate occupancy profiles, in: BuildSys 2019 - Proc. 6th ACM Int. Conf. Syst. Energy-Efficient Build. Cities, Transp., 2019: pp. 145–148. https://doi.org/10.1145/3360322.3360869.

[51]   M. Pritoni, B. Nordman, M.A. Piette, Accessing Wi-Fi Data for Occupancy Sensing, LBNL-2001053. (2017). https://eta-publications.lbl.gov/sites/default/files/accessing_wi-fi_data_final.pdf.

[52]   Z. Chen, C. Jiang, L. Xie, Building occupancy estimation and detection: A review, Energy Build. 169 (2018) 260–270. https://doi.org/10.1016/j.enbuild.2018.03.084.

[53]   C. Martani, D. Lee, P. Robinson, R. Britter, C. Ratti, ENERNET: Studying the dynamic relationship between building occupancy and energy consumption, Energy Build. 47 (2012) 584–591. https://doi.org/10.1016/j.enbuild.2011.12.037.

[54]   B. Balaji, J. Xu, A. Nwokafor, R. Gupta, Y. Agarwal, Sentinel: Occupancy based HVAC actuation using existing wifi infrastructure within commercial buildings, in: SenSys 2013 - Proc. 11th ACM Conf. Embed. Networked Sens. Syst., 2013: pp. 1–14. https://doi.org/10.1145/2517351.2517370.

[55]   E. Vattapparamban, B.S. Çiftler, I. Güvenç, K. Akkaya, A. Kadri, Indoor occupancy tracking in smart buildings using passive sniffing of probe requests, in: 2016 IEEE Int. Conf. Commun. Work. ICC 2016, 2016: pp. 38–44. https://doi.org/10.1109/ICCW.2016.7503761.

[56]   W. Wang, J. Chen, X. Song, Modeling and predicting occupancy profile in office space with a Wi-Fi probe-based Dynamic Markov Time-Window Inference approach, Build. Environ. 124 (2017) 130–142. https://doi.org/10.1016/j.buildenv.2017.08.003.

[57]   Q. Chen, B. Tan, K. Woodbridge, K. Chetty, Indoor target tracking using high doppler resolution passive Wi-Fi radar, in: ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., 2015: pp. 5565–5569. https://doi.org/10.1109/ICASSP.2015.7179036.

[58] X. Li, D. Zhang, Q. Lv, J. Xiong, S. Li, Y. Zhang, H. Mei, IndoTrack: Device-Free Indoor Human Tracking with Commodity Wi-Fi, Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol. 1 (2017) 1–22. https://doi.org/10.1145/3130940.

[59] T. Dimitrakopoulou, K. Adamidis, S. Loukas, A lifetime distribution with an upside-down bathtub-shaped hazard function, IEEE Trans. Reliab. 56 (2007) 308–311. https://doi.org/10.1109/TR.2007.895304.

[60] J.C. Nash, R. Varadhan, Unifying optimization algorithms to aid software system users: Optimx for R, J. Stat. Softw. 43 (2011) 1–14. https://doi.org/10.18637/jss.v043.i09.

[61] J.C. Nash, On best practice optimization methods in R, J. Stat. Softw. 60 (2014) 1–14. https://doi.org/10.18637/jss.v060.i02.

[62] D. Eddelbuettel, Seamless R and C++ integration with Rcpp, 2013 ed., Springer, New York, 2013. https://doi.org/10.1007/978-1-4614-6868-4.

[dataset] [63] Transport for London, Station Passenger Link Flows. TfL data, 2017. http://crowding.data.tfl.gov.uk.

[dataset] [64] Weather Underground, Daily Observations: Kensington, London, England, United Kingdom Weather History, 2017. https://www.wunderground.com/history/daily/gb/london.

[65] D. Gamerman, M. West, An Application of Dynamic Survival Models in Unemployment Studies, Stat. 36 (1987) 269–274. https://doi.org/10.2307/2348523.