

# A preliminary study of representing the inter-occupant diversity in occupant modelling

William O'Brien<sup>1\*</sup>, H. Burak Gunay<sup>1</sup>, Farhang Tahmasebi<sup>2</sup>, Ardeshir Mahdavi<sup>2</sup>

\*[Liam.obrien@carleton.ca](mailto:Liam.obrien@carleton.ca), <sup>1</sup>Carleton University, Ottawa, Canada; <sup>2</sup> TU Wien, Vienna, Austria

## 1 ABSTRACT

Significant diversity between occupants and their presence and actions results in major uncertainty with regards to predicting building performance. However, current occupant modelling approaches – even stochastic ones - suppress occupant diversity by focusing on developing representative occupants. Accordingly, existing approaches tend to limit the ability of stochastic occupant models to provide probabilistic building performance distributions. Using occupancy data from 16 private offices, this paper evaluated three hypotheses: 1) occupant parameters have a continuous distribution rather than discrete; 2) modelling occupants from aggregated data suppresses diversity; and, 3) randomly selecting occupant traits exaggerates synthetic population diversity. The paper indicates that samples sizes for the studied occupants would have more appropriately been an order of magnitude higher: hundreds. This introductory paper shows that there are many future research needs with regards to modelling occupants.

## 2 INTRODUCTION

Occupant behaviour has become widely understood to be a leading cause of uncertainty in building performance. Moreover, occupants have an increasingly significant effect on building performance as buildings become more efficient (Page et al. 2008; Hoes et al. 2009). Numerous field and simulation studies have found the difference between the highest and lowest energy consuming occupants in a similar or identical set of homes or offices to be at least a factor of two (Gram-Hanssen 2010; Haldi and Robinson 2011). This uncertainty comes at a considerable cost because in most cases, engineers will size equipment or renewable energy systems (in net-zero energy buildings) to cover the highest expected loads or annual energy (Djunaedy et al. 2011). Furthermore, the predicted economic payback of various building upgrades – at design or for retrofit – can be significantly influenced by the assumptions made about occupants (Rasouli et al. 2013; Lee and Schiavon 2014). The uncertainty of occupants can be mitigated through building design and through diversity of larger buildings or district energy systems, but greater knowledge of the complex and dynamic human-building, including quantification of uncertainty and occupant diversity, is needed (O'Brien and Gunay 2015). For instance, heating equipment at the zone level should be sized in recognition that a certain fraction of occupants will prefer and choose heating setpoints that are at the upper limit of the normal range. But this does not mean that the entire building's heating plant capacity should be sized as if all occupants prefer higher temperatures. On the contrary, we expect a diversifying effect such that heating plant sizing can be achieved using the mean expected temperature setpoint. But this of course, does not relieve designers of understanding the expected occupants and their systemic characteristics (e.g., suit-wearing executives, casually-dressed students, or active laboratory workers).

It follows that the importance of occupant diversity modelling is also related to the design stage. Early in design, when building massing, window-to-wall area ratios, and other high-level decisions are being made at the building level, the importance of characterizing individual occupants is much lower than in detailed design where room features are being designed (Athienitis and O'Brien 2015).

Diversity can also provide a smoothing effect on peak loads in the temporal sense. For instance, if all occupants are modelled to arrive simultaneously and turn on lights and computers and meanwhile the controls are scheduled to activate heating or cooling simultaneously, peak building-level loads will be much higher than if arrivals are staggered.

In the past decade, the research community has been developing more advanced occupant models based on monitoring data. It has been argued that occupants should be represented and modelled stochastically rather than deterministically (Nicol 2001). That is, occupants are unlikely to respond in the same way to a given set of circumstances in a machine-like manner because there are many complexities to their decision making process. Thus, recent occupant modelling efforts have been primarily stochastic. But much of the research has focused on modelling a “typical occupant” for building simulation, rather than recognizing diversity between occupants with regards to their behaviour and presence. In essence, this results in a statistically representative occupant rather than a true representation of a population of

occupants. While a degree of uncertainty is introduced by developing agent-based stochastic models whose presence and actions are random at any given timestep, another major source of uncertainty stems from differences between occupants.

Diversity has been studied at length in other fields such as biology, psychology, sociology, organizational behaviour, and behaviour science (e.g., Gangestad and Snyder 1985; Mohammed and Ringseis 2001; Van Knippenberg and Schippers 2007), yet seldom in the context of individual occupant models for BPS. In the current context, diversity means to have inter-individual variability with respect to building occupant traits that affect building performance. The most common current reference to “diversity” in BPS applications is the derivation and utilization of schedules that are normalized to unity and multiplied by some reference value (e.g., total building occupancy capacity or lighting power density) (Abushakra, Haberl and Claridge 2004; Bourgeois, Reinhart and Macdonald 2006). There is an implicit assumption for those diversity factors that they apply to large buildings (e.g., hundreds or thousands of occupants). For estimating building-level peak cooling loads, a major driver of diversity factors and schedules, the current approach achieves its purpose (Abushakra, Haberl et al. 2004). The main limitation of these standardized profiles is that they are not specific to a given building. In contrast to the traditional diversity profiles, the current focus on diversity is on a smaller scale (room-level) where the traditional approach is limiting. For instance, in private offices it does not make sense to discuss fractions of an occupant – particularly for the purpose of agent-based modelling where the occupants are not merely sources of passive heat gains, but rather active participants in energy performance. Also in contrast to the conventional diversity factor schedules, the current paper presents a method to stochastically introduce diversity of occupants into BPS tools.

Contrary to pure laboratory-based studies, which control for contextual differences, such that personal characteristics and inter-personal traits can be quantified and controlled, there are additional layers of complexity from in-situ occupant monitoring campaigns. While, there is strong evidence to suggest that occupants have different sensitivities and preferences for comfort conditions (e.g., Nicol 2004; Boyce 2014), contextual factors further introduce noise to observed behaviours (O'Brien and Gunay 2014). The diversity in observed behaviours stems from multiple factors including sensitivity to environmental conditions, activity type, user interfaces, building controls schemes, personal health and wealth, environmental culture, job type and lifestyle, and many other contextual factors. Occupant preferences for indoor environmental conditions also depend on occupant metabolic rate, whether an occupant is reading, drawing or using a computer, and whether an occupant needs silence to concentrate. Furthermore, room and furniture layout could play a role on occupants' behaviour. For instance, an occupant who is seated at a small desk near a window would be more sensitive to daylight glare than an occupant who is seated far from a window at a large desk with flexible seating positions (Jakubiec and Reinhart 2012) – even though indoor daylight levels would be typically measured identically for both offices. Anecdotally, some occupants may even alter their departure time from an office if conditions are uncomfortable. These confounding factors would likely result in prohibitively expensive and impractical experimental and in-situ studies because of the immense number of combinations. However, the aggregate effect of contextual factors can still be captured to develop more generalized models. Most occupant modelling studies have attempted to fit models to a limited number of environmental factors (Reinhart 2004; Rijal et al. 2007) with the contextual factors accepted as random noise. A limited number of studies (Sutter, Dumortier and Fontoynt 2006; Haldi and Robinson 2010; Schweiker et al. 2012) have had the privilege of several controlled contextual factors such that their impact can be quantified. An illustrative example of the diversity between occupants with respect to manual office light controls use is shown in Figure 1. It shows that the apparent diversity between occupants' likelihood of turning on lights as a function of workplane illuminance is actually the aggregated effect of activity type, physiology, measurement method, contextual factors, and probably additional phenomena.

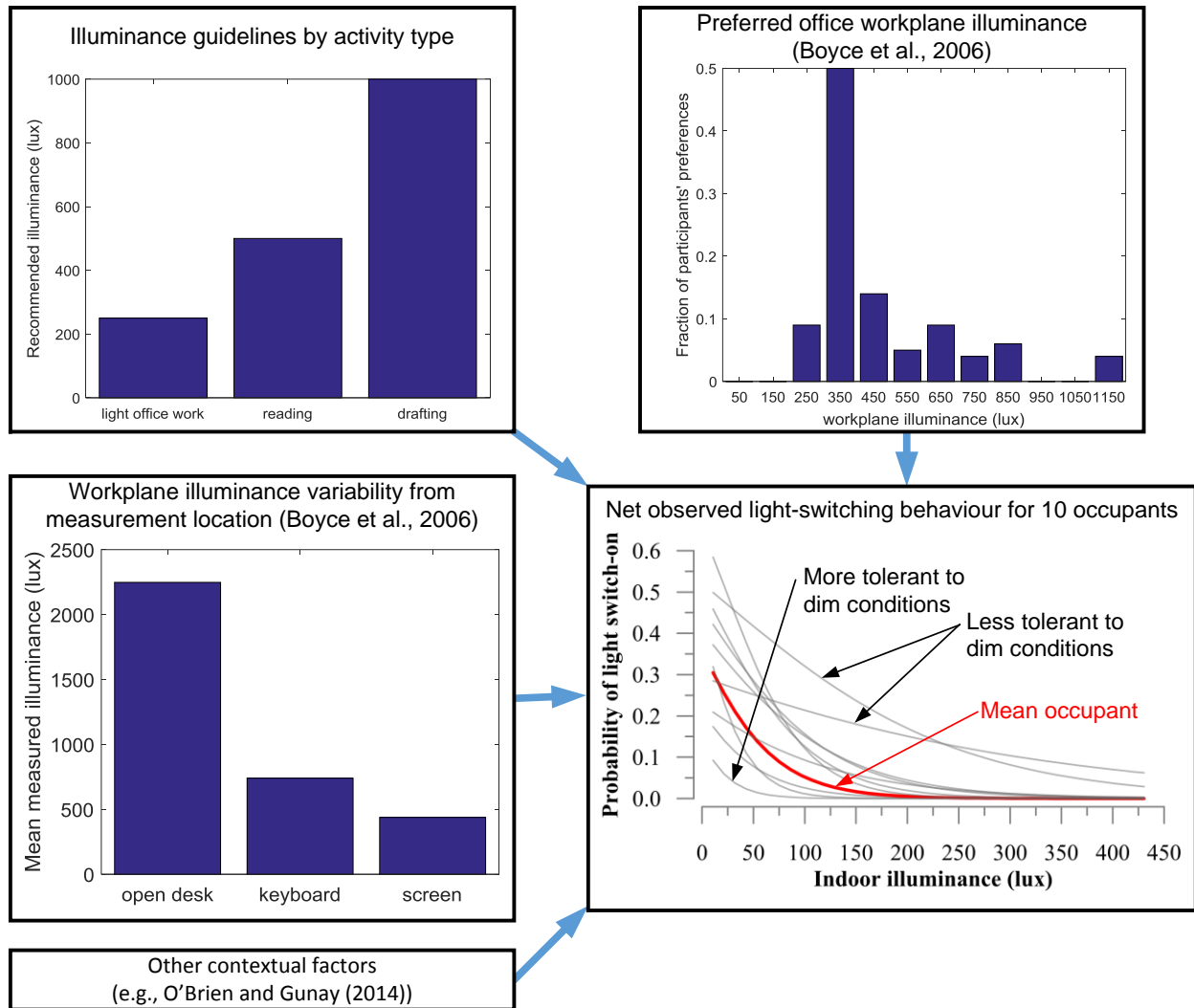


Figure 1: Illustrative example of light-switching behaviour, showing that observed occupant diversity is an aggregate of many diverse factors. The net observed light-switching behaviour plot shows great diversity between individual occupants (each of whom is shown as a separate curve). The mean occupant curve is a poor representation of the sample.

Many practical and theoretical questions arise for in-situ occupant monitoring and modelling campaigns, including:

- Are the differences between occupants caused purely by inter-occupant physiological differences, contextual factors, or consistency of measurement methodology – or is it some complex mixture of these?
- How should such data-driven models be converted and generalized for use in building simulation such that diversity is maintained?
- What conclusions can be drawn about adequate sample size from occupant trait diversity?

In spite of the aforementioned benefits of modelling diversity within occupant populations, this matter has been treated in relatively simple ways that leave many research questions, and in some cases could lead to significant error (Tahmasebi and Mahdavi 2016). While occupant modelling domains are plentiful and can cover adaptive and non-adaptive actions as well as occupancy (i.e., presence), this paper is focused on application of diversity modelling to a single domain: private office occupancy. But prior to focusing on this domain, the paper provides a review of occupant diversity modelling approaches and then an in-depth examination of several occupant diversity approaches. Using the current occupancy application, three hypotheses about occupant diversity modelling approaches are examined using occupancy data from 16 private offices. The Page et al. (2008) model is applied and then adapted to generate synthetic

occupants to represent population diversity. In parallel to and to support the hypotheses being tested, four diversity modelling approaches are considered.

- Cluster analysis is applied to the 16 occupants;
- The data from all occupants are aggregated to develop a generalized model;
- The 16 occupants are captured by 13 occupant traits which are randomly generated; and finally,
- The same approach as above is used by with parameter correlations maintained.

Finally, the practical, research, and monitoring campaign implications of diversity and diversity modelling are discussed.

### **3 LITERATURE REVIEW**

Where occupant diversity modelling for BPS applications is acknowledged in the literature, it is normally comprised of four possible approaches: 1) use of real measured occupant data in simulation, 2) clustering of occupant types (e.g., active and passive) and then developing models for each occupant typology, 3) developing models from aggregated data from all monitored occupants, and 4) developing models for each occupant and then developing metamodels that define the model parameter distributions (Haldi 2013). The following sections briefly summarize the methods and merits of each approach. Later in the paper, the methods are critically examined using the current dataset in order to provide stronger evidence to the literature.

#### **3.1 Mapping real data to simulation**

Perhaps the simplest way to represent occupant behaviour realistically inside BPS is to directly map observed behaviours into building models (e.g., using schedule objects) (e.g., O'Brien et al. 2010; Tahmasebi and Mahdavi 2015). Assuming the sample size is adequate, this approach incorporates diversity. However, its limitation is that it does not generalize results such that they can be applied to other buildings (e.g., unbuilt buildings) in the way that other modelling approaches can. But this approach is suitable for estimating energy savings from building retrofits or control modifications. This method is not pursued in the current paper, as it has limited applicability.

#### **3.2 Clustering**

Numerous modelling efforts have attempted to sort occupants into discrete typologies such that a small number of simulations can be run to infer a possible range of building performance levels. A particularly popular method to introduce diversity into simulation is to divide populations into passive and active users (e.g., Reinhart 2004; Rijal, Tuohy et al. 2007; Parys, Saelens and Hens 2011). In these papers, active users consistently adapt to environmental inputs while passive occupants never or rarely take actions. This approach is formed on the basis that several typologies of occupant have emerged from monitoring studies (D'Oca and Hong 2015). Similarly, Meerbeek et al. (2014) clustered all monitored occupants into groups with regards to their blind adjustment activity levels. However, it is not evident from their results that the occupants naturally cluster; the results resemble a continuous distribution.

The clustering approach has appeal because it is conceptually simple and computationally efficient (i.e., only as many simulations as occupant types need to be run). There has been a human tendency to attempt to categorize people (e.g., personality types) (Myers, McCaulley and Most 1985). However, several challenges have emerged. First, using occupant typologies require estimates of population size for each cluster (Bourgeois, Reinhart et al. 2006). Secondly, it is not evident from the data that discrete clusters occur. The few researchers who have provided individual characteristics suggest much the opposite: that the data more closely resemble a continuum (Reinhart and Voss 2003; Yun, Steemers and Baker 2008; Haldi and Robinson 2010; Meerbeek, te Kulve et al. 2014). Finally, it is unclear whether an active user of one system is also an active user of the others. One can imagine that if multiple adaptive actions are modelled, the number of distinct occupant types would quickly become unmanageable. For instance, if there are five possible actions of interest (e.g., light switching and window opening) and two user types for each, there would be 32 ( $2^5$  different occupant types). Yet after about four decades of occupant monitoring and modelling studies, only a handful of models have emerged from each common domain.

#### **3.3 Models from aggregated occupant sample data**

A prominent modelling approach in the past decade has been to develop models by aggregating monitored occupant data (e.g., window opening events and presence) and attempting to fit the data against one or more statistically significant input variables (Haldi and Robinson 2009)). Given that the equations that define the decision-making process (e.g., logistic regression curves) are unlikely to perfectly fit the data, some researchers have reported

coefficient error ranges. The error distribution can then be used to select model coefficients at the beginning of each simulation period (nominally a year). In essence, this is selecting an occupant with a different set of characteristics to be run for the entire simulation period. There are, however, several limitations to this approach. Most notably, depending on how the data from multiple occupants are combined, the models (in the case of agent-based models) are biased towards those occupants that are most active. For instance, if one occupant turns on lights ten times more than another, the net effect of combining data from the two occupants will lead to a model that has the average occupant. Furthermore, there is a tendency to reduce the variance by fitting a single equation to data generated by multiple occupants. While the resulting models are technically stochastic, they typically do not yield the large inter-occupant variations in annual energy use that are seen in reality. In brief, this is because calling such models every timestep for a year-long simulation tends to have an averaging effect.

### **3.4 Metamodels with reported coefficient distributions**

To address the bias towards certain occupants mentioned in the above method, Haldi (2013) suggested that a model be developed for each monitored occupant and then a probabilistic distribution for the model coefficients, from each occupant model, can be developed. In this way, the characteristics of each occupant are not biased through aggregation. Moreover, this approach alleviates the need to cluster occupants into discrete types. This approach has been seldom used for occupant modelling, though it is prevalent in the current paper.

### **3.5 Occupancy Modelling**

Because the current domain used to explore diversity is occupancy (presence), a brief review of occupancy modelling is provided here. Occupancy modelling for BPS originated with basic hourly schedules to reflect typical building populations. This approach offers simplicity and simulation repeatability that is important for demonstrating code compliance. But the profiles are generalized to all buildings. Moreover, they are aimed at the building-scale, where multipliers can be applied to maximum expected capacity. As discussed later in this paper, this approach is most suitable for large buildings, but less so at the room level. Notably, these mean occupancy schedules are often arguably poorly named as “diversity factors”. (At a minimum the current meaning of diversity is quite different than that of diversity factors). Finally, schedules predict presence but not arrival and departure events of individual occupants. More recent research has shown that predicting occupancy is important beyond the resulting internal heat gains and indoor air quality implications; occupancy and arrival are good predictors for lighting and equipment use (Reinhart 2004; Haldi and Robinson 2010). Accordingly, models to predict occupants’ arrival, departure, and presence began to emerge about a decade ago. Wang, Federspiel and Rubinstein (2005) proposed a non-homogeneous Poisson process model and distributions for the duration of vacancy and occupancy to model individual office occupants. Page, Robinson et al. (2008) developed an inhomogeneous Markov chain-based approach and also made the contributions of predicting long absence probability and duration and predicting a random number of daily breaks. But these models are focused on reproducing measured occupants rather than generating additional synthetic occupants for application in unbuilt buildings and they require relatively detailed inputs (e.g., mean occupancy schedule and some other abstract parameters). Andersen et al. (2014) developed a method based on inhomogeneous Markov chains and applied a hierarchical modelling strategy to distinguish between low and high presence occupants.

More recent occupancy modelling research has focused on multi-zonal buildings (Agent-based and graphical modeling of building occupancy), overtime prediction (Stochastic Modeling of Overtime Occupancy and Its Application in Building Energy Simulation and Calibration (Sun 2014)), number of occupants and location within a space (Feng, Yan and Hong 2015), and other phenomena. D’Oca and Hong (2015) applied data mining and cluster analysis to study occupancy patterns in private offices and found that four occupant types emerged from the data. Mahdavi and Tahmasebi (2015) explored the diversity between occupants’ traits and found it to be significant.

## **4 GENERAL OCCUPANT MODELLING METHODOLOGY**

This section first describes the general occupancy modelling approach and evaluates its ability to replicate real data. Following that, a dataset of 16 occupants’ presence (occupancy) – for between one and four years (depending on the office) - was examined to address three hypotheses about occupant diversity modelling.

### **4.1 Dataset description**

The 16 identical private offices are located in a modern academic building in Ottawa (Figure 2). The building automation system (BAS) logs every time that a movement is detected using commercial and integrated passive infrared (PIR) sensors. This approach is unable to determine the number of occupants, though only a single occupant is expected in the private offices for the majority of the time. It is unlikely that anyone would be in the office without

the primary occupant. The current scope is focused on predicting whether at least one person is present. The results should be taken into the context that occupants are all engineering professors, have relatively flexible hours, and are frequently absent for teaching, meetings, and travel. Thus, while their occupancy patterns may be diverse, there are no traits that inherently discretely divide them.



Figure 2: Office building where monitoring occurred.

## 4.2 Data pre-processing

The data of each office were pre-processed using custom Matlab code to remove anomalies and expected errors. First, the data was arranged into 5-minute timesteps such that if the occupant was present for any part of a 5-minute timestep, the entire timestep was assumed to be occupied. This leads to a small error and reduces the ability to study very short absences (e.g., short washroom breaks), but it is much more convenient for modelling purposes. Five minutes was selected to balance accuracy and practicality. Simulation tools typically use between 5 and 15 minute timesteps for zone and building level phenomena.

Next, occupancy events that consisted of a presence duration of 5 minutes and surrounded by an absence of 10 minutes or more, before and after, were removed. These are deemed to be janitor cleaning visits, safety inspections, or quick drop-off or pick-up events. Such events are unlikely to be associated with significant metabolic heat gains or adaptive actions (e.g., light switch events). Prior to this filtering, numerous overnight visits were observed in the data. Anecdotally, one of the authors, who is included in the dataset was not in his office between midnight and 7AM, yet numerous brief occupancy periods were detected. In previous work by Gunay et al. (2016), no false positives were detected from the PIR product used.

Next, the data were rearranged in an array such that each column represented the occupancy for a day. The data revealed that mean occupancy for any given occupant does not significantly differ between weekdays for most occupants (Gunay, O'Brien and Beausoleil-Morrison 2015). Moreover, weekends were removed for simplicity. Weekend occupancy for most of the occupants was exceptionally rare.

## 4.3 Occupancy data overview

Figure 3 indicates that a wide diversity in daily arrivals and departures among the 16 occupants, with mean occupied periods (between first arrival and last departure) ranging from 2.6 to 9.3 hours.

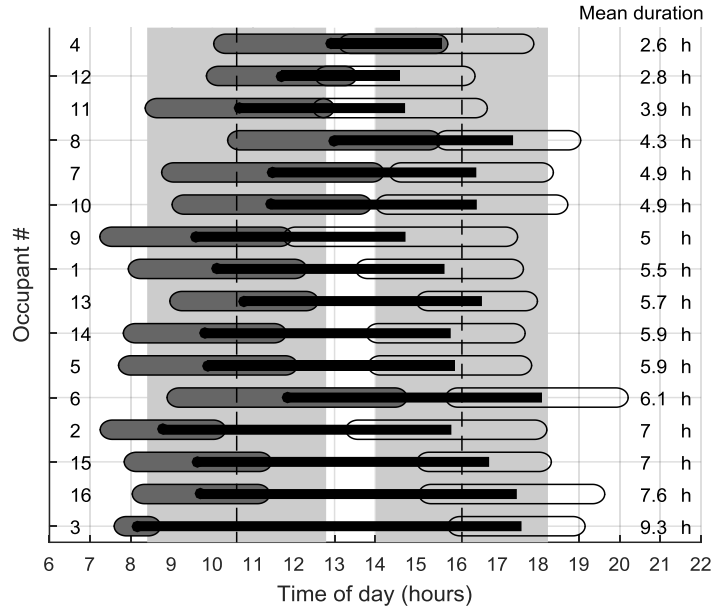


Figure 3: First arrival, last departure, and presence duration statistics for the individual occupants on occupied days only. The ellipses indicate the mean and extent of  $\pm$  one standard deviation of the arrival and departure times for each occupant. The vertical dotted lines represent the median first arrival and last departure times for the whole sample of 16 occupants, while the two gray bands represent the mean first arrival and last departure times plus or minus one standard deviation for the entire sample. The occupants are ordered from shortest to longest mean occupancy (as measured from first arrival to last departure, thus ignoring intermediate absences). The occupant numbering is consistent for the paper.

The individual occupant profiles that cover all weekdays during the monitored period (excluding absent days) are shown later in Figure 7. That figure indicates that the occupants are diverse with respect to occupancy profile shape and magnitude. Moreover, all occupants of the sample are present much less than typically assumed in offices (e.g., US DOE Reference Buildings (Deru et al. 2011)).

#### 4.4 Base occupancy modelling methodology

The Page, Robinson et al. (2008) model was implemented, as it is a commonly used algorithm due to its ability to model likelihood of state changes and modelling the frequency and duration of long absences. However, some deviations from that model were used:

- Only weekdays were modelled and only one representative weekday was modelled (not Monday through Friday separately).
- Individual occupancy profiles were fit to Gaussian mixture models for generalizability.
- The method used to describe occupant diversity was varied to address the hypotheses, as described later.

As per Page's model, there are a number of parameters that need to be extracted from the raw occupancy data, including the probability of presence as a function of time (i.e., occupancy profile), the mobility parameter, the probability of starting a long absence, and the probability distribution for the length of long absence. Readers are encouraged to refer to Page, Robinson et al. (2008) for details on the original model.

In order to improve generalizability of the occupancy profiles and regularize complex and noisy occupancy profiles, each occupancy profile was fit to a Gaussian mixture model (GMM). GMMs with between one and five modes (peaks) explored, though those with more than three modes did not offer a significantly better fit. Thus, three-mode distribution was selected to fit these occupancy profiles. This choice is specific to the current sample.

The corresponding GMM model form is:

$$occ(t) = \sum_{i=1}^3 a_i e^{-\left(\frac{x-b_i}{c_i}\right)} \quad (1)$$

where  $occ$  is the mean weekday occupancy at time of day  $t$  (in hours),  $a_i$  indicates the amplitude (probability of presence) of each mode,  $b_i$  is the mean of each mode, and  $c_i$  indicates the width of each peak. The sets of coefficients were sorted such that each of the three terms of the model are in order of time of day (e.g.,  $b_1$  indicates the timing of the first peak occupancy of the day;  $b_3$  is the last). Two representative examples are shown in Figure 4. The GMM parameter values for all 16 offices are shown in Table 1.

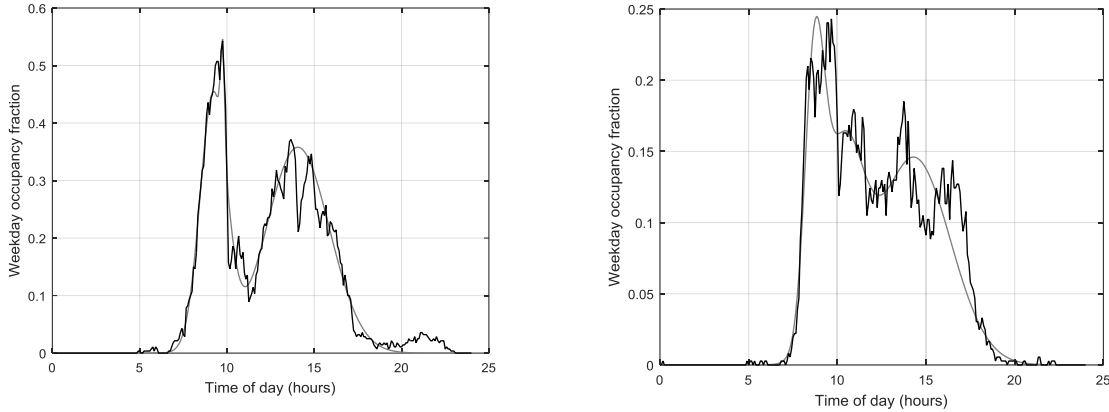


Figure 4: Example plots for two occupants' mean weekday occupancy and corresponding GMM (occupants 2 and 9).

Table 1: Nine GMM occupancy profile parameters for each occupant.

| Occupant number | a1    | b1    | c1   | a2    | b2    | c2   | a3    | b3    | c3   |
|-----------------|-------|-------|------|-------|-------|------|-------|-------|------|
| Constraints     |       | >0    |      |       | >b1   |      |       | >b2   |      |
| 1               | 0.14  | 10.07 | 1.49 | 0.35  | 12.43 | 1.22 | 0.23  | 15.54 | 1.67 |
| 2               | 0.45  | 9.22  | 1.08 | 0.17  | 9.75  | 0.16 | 0.36  | 14.06 | 2.56 |
| 3               | 0.63  | 8.93  | 0.98 | 0.57  | 12.92 | 1.92 | 0.52  | 16.72 | 1.53 |
| 4               | -0.03 | 13.61 | 1.07 | 0.11  | 14.73 | 4.25 | -0.02 | 19.79 | 1.80 |
| 5               | 0.23  | 10.18 | 1.54 | 0.27  | 14.19 | 2.17 | 0.13  | 17.06 | 0.78 |
| 6               | 0.20  | 13.53 | 1.09 | 0.14  | 15.74 | 4.27 | 0.25  | 17.83 | 0.92 |
| 7               | 0.19  | 9.53  | 0.73 | 0.21  | 13.87 | 2.65 | 0.19  | 15.98 | 0.89 |
| 8               | 0.16  | 14.25 | 2.78 | 0.19  | 16.60 | 1.00 | 0.18  | 17.94 | 1.02 |
| 9               | 0.19  | 8.72  | 0.77 | 0.14  | 10.31 | 1.60 | 0.15  | 14.32 | 2.89 |
| 10              | 0.13  | 14.11 | 3.77 | 0.04  | 15.81 | 0.40 | 0.13  | 17.36 | 0.75 |
| 11              | 0.21  | 11.26 | 2.14 | 0.19  | 13.55 | 0.44 | 0.29  | 15.19 | 1.16 |
| 12              | 0.03  | 9.20  | 0.90 | 0.26  | 12.51 | 1.44 | 0.16  | 15.14 | 1.48 |
| 13              | 0.25  | 12.96 | 0.93 | 0.27  | 13.18 | 3.24 | 0.37  | 16.44 | 0.85 |
| 14              | 0.27  | 10.65 | 2.21 | 0.15  | 13.62 | 0.73 | 0.23  | 14.98 | 2.76 |
| 15              | 0.23  | 10.63 | 2.22 | 0.15  | 13.38 | 0.37 | 0.31  | 14.73 | 2.71 |
| 16              | 0.19  | 9.65  | 1.12 | -0.26 | 14.46 | 2.09 | 0.42  | 14.51 | 3.24 |

As others (e.g., Haldi 2013) have done, the parameters were assumed to be normally distributed. However, several practical constraints or contributing factors likely cause the population parameters to not be perfectly normally distributed. Most notably, the  $b$  parameters (time of peak) are influenced by typical meeting and class schedules and typical meal times. Furthermore, the  $b$  and  $c$  parameters cannot be negative; thus the actual distribution is truncated at



zero and impossible parameter values were rejected in the Monte Carlo simulations discussed later in the paper. The authors acknowledge that this violates the Gaussian distribution.

The histograms are accompanied by best-fit normal distributions and parameters are provided in Figure 5. Moreover, the Anderson-Darling test was used to assess whether the sample originated from a normal distribution. The corresponding p-values are annotated in Figure 5. The null hypothesis ( $H_0$ ) of this test is that the data are sampled from a normally-distributed population. The interpretation of the p-value is that high values (above  $\alpha = 0.05$ ) indicate that there is insufficient evidence to disprove normality of the population (i.e., it resembles a normal distribution). While some of the distributions ( $a_1$ ,  $b_1$ ,  $c_1$ ,  $a_2$ , and  $c_3$ ) fail the normality test, the paper proceeds under the assumption that the data are normally distributed.

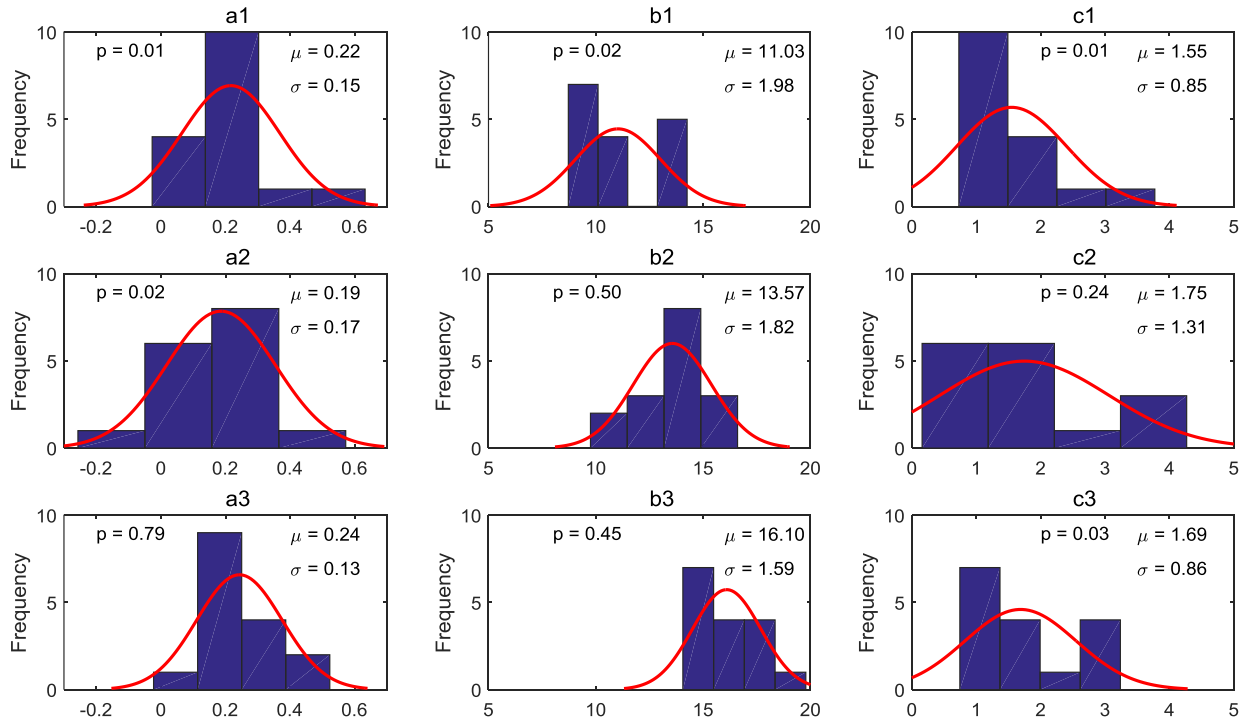


Figure 5: Frequency distributions of the nine parameters that define the occupancy profiles

Relying on these new smooth occupancy profiles would lose one vital parameter that is necessary to characterize the tendency of occupants to take intermediate breaks: the mobility parameter ( $\mu$ ). Using the raw occupancy data, the probability of transitioning from present-to-absent (T10) and absent-to-present (T01) was calculated for every timestep. The corresponding probabilities of the occupant remaining in their office (T11) or out of their office (T00) were also calculated. The resulting mobility parameter  $\mu$  for each occupant was obtained by calculating the ratio of transition events to static events  $(T10 + T01)/(T00 + T11)$ . Similarly to the conclusion of Page, Robinson et al. (2008), each of the 16 current occupants have relatively constant mobility parameters during the day, though diversity between them [0.03, 0.28] is significant. These values and the analysis that follows are based on 5-minute timesteps.

Next, the long absences were explored, which for this population is a particularly critical aspect of the model. The occupants' attendance records ranged from 23 to 91% on a daily basis. Similar to Page, Robinson et al. (2008)'s approach, the probability that an occupant embarks on a one-weekday or longer absence was calculated.

Next, the distribution of absence durations (days) was calculated for each occupant. A consistent distribution form was sought for all occupants for the sake of the approach of the current paper: to develop occupant trait distributions. Systematic testing of common probability distribution forms revealed that log-logistic performs best because of its ability to support both a peak at one day and a relatively long tail (very long absences of weeks or months). The log-logistic distribution is defined by  $\log(\text{mean})$  and  $\log(\text{shape parameter})$ . For the sample at hand, multi-day absences are not unusual since work hours are quite flexible and some occupants may travel for conferences or vacation. The distribution of long absence durations for each of the 16 occupants is shown later in Figure 11.

The mobility parameter, probability of embarking on a long absence, and the two long absence duration log-logistic are summarized in Figure 6. Again, the Anderson-Darling test for normality was applied and the results are annotated in the figure. For these parameters, the test suggests that the population resembles a normal distribution (or more specifically that the null hypothesis was not disproven).

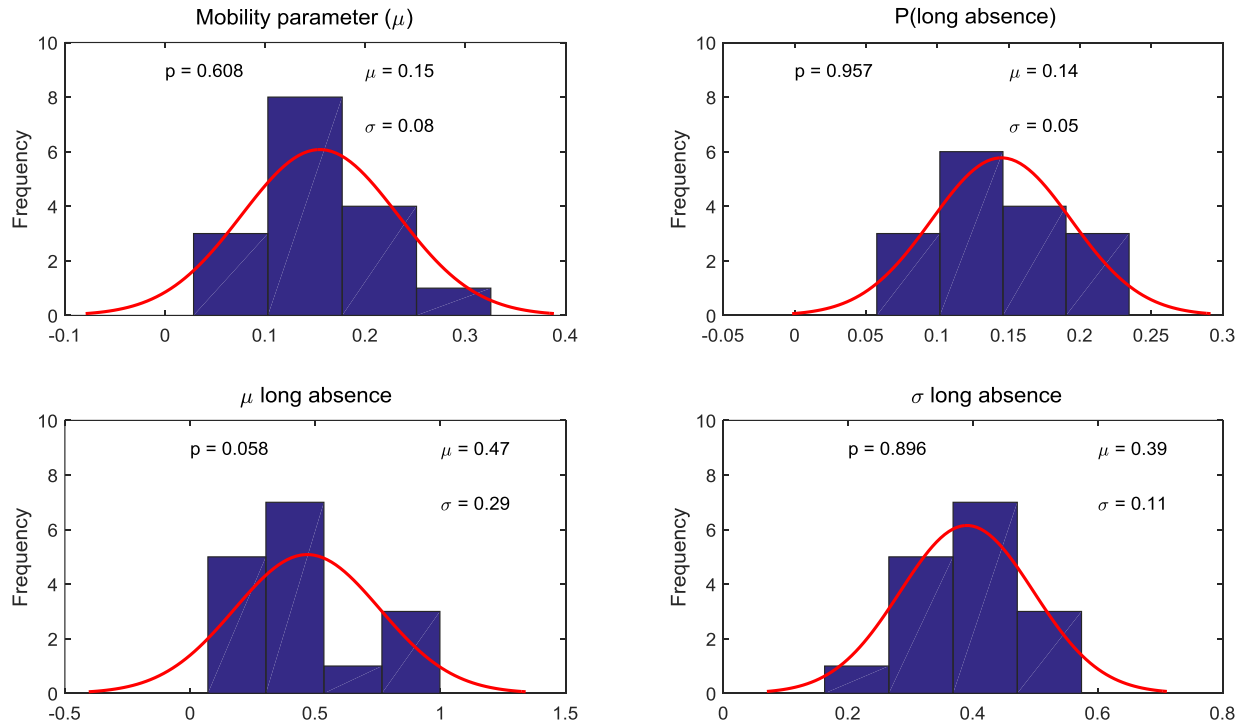


Figure 6: Frequency distributions for four occupant traits, as titled for each plot.

In all, each occupant's presence patterns were characterized by 13 parameters: nine parameters defining the three-mode GMMs, the mobility parameter, probability of embarking on a long absence and two parameters that define the log-logistic distribution of absence durations. The relationship between the parameters is further discussed in Section 5.3.

Prior to examining diversity modelling, the modelling approach was confirmed for each occupant (see Figure 7). Regardless of monitoring period, the model was run for 1000 weekdays. The corresponding mean measured and modelled occupancy (including absences) and fraction of absent days are shown. The model generally replicates measurements well, though two particular occupant characteristics were found to cause larger errors: absence duration profiles that were poorly described by the log-logistic regression (e.g., they had numerous extended absences) and occupancy profiles that were poorly described by the GMM.

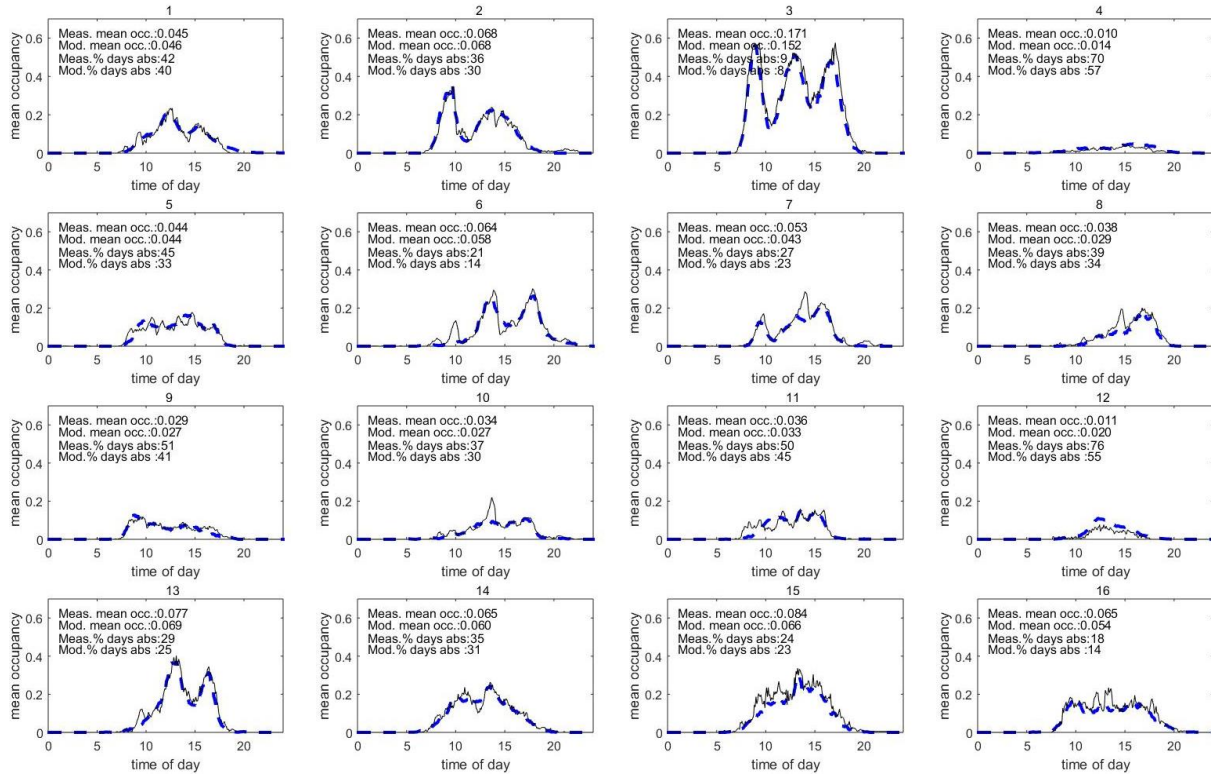


Figure 7: Comparison of modelled (blue, dotted line) and measured (black, thin line) weekday occupancy – including long absences – for all 16 occupants. Measured (meas.) versus modelled (mod.) mean daily occupancy and fraction of days absent are annotated in each plot.

From visual inspection, generally, the model was able to reproduce the observed occupant characteristics reasonably well. Notably, the model was verified with the same data from which the model parameters were trained because the general modelling approach was previously validated by Page, Robinson et al. (2008). Further comparisons for key occupant characteristics are discussed later and provided in Figure 9.

## 5 OCCUPANT DIVERSITY MODELLING

Using the base modelling approach as a starting point, the data were analyzed and the modelling methodology varied as needed to test three diversity modelling hypotheses. This section presents both the methodologies and the results because they were developed in parallel.

### 5.1 Occupant traits are best described by a continuous function; not discrete.

As previously noted, the two prevalent schools of thought on occupant diversity are that occupants can be discretized into several typologies and that occupants cannot be categorized into several typologies because they are continuous. To assess the suitability of each approach, k-means clustering was applied to the sample dataset. In all, the following occupant traits were analyzed:

- mean first daily arrival time
- mean last daily departure time
- annual occupied duration
- mean daytime mobility parameter
- probability of starting long absence
- log-logistic parameters for long absence duration

Note that the cluster analysis was not applied to the nine GMM parameters that were used to fit the mean occupancy profiles for each occupants, as clustering 16 occupants within a nine-dimensional space is impractical. The clustering results for the six occupant traits are presented in Figure 8. The interpretation of the criterion value (“gap statistic” was used; see Tibshirani, Walther and Hastie (2001) for details) is that higher values indicate a closer resemblance to discrete clustering rather than random scatter. The Matlab algorithm seeks a compromise between the maximum criterion value and fewer clusters. That is, if the improvement to clustering is marginal for a greater number of clusters, then the algorithm indicates that the current number of clusters is optimal.

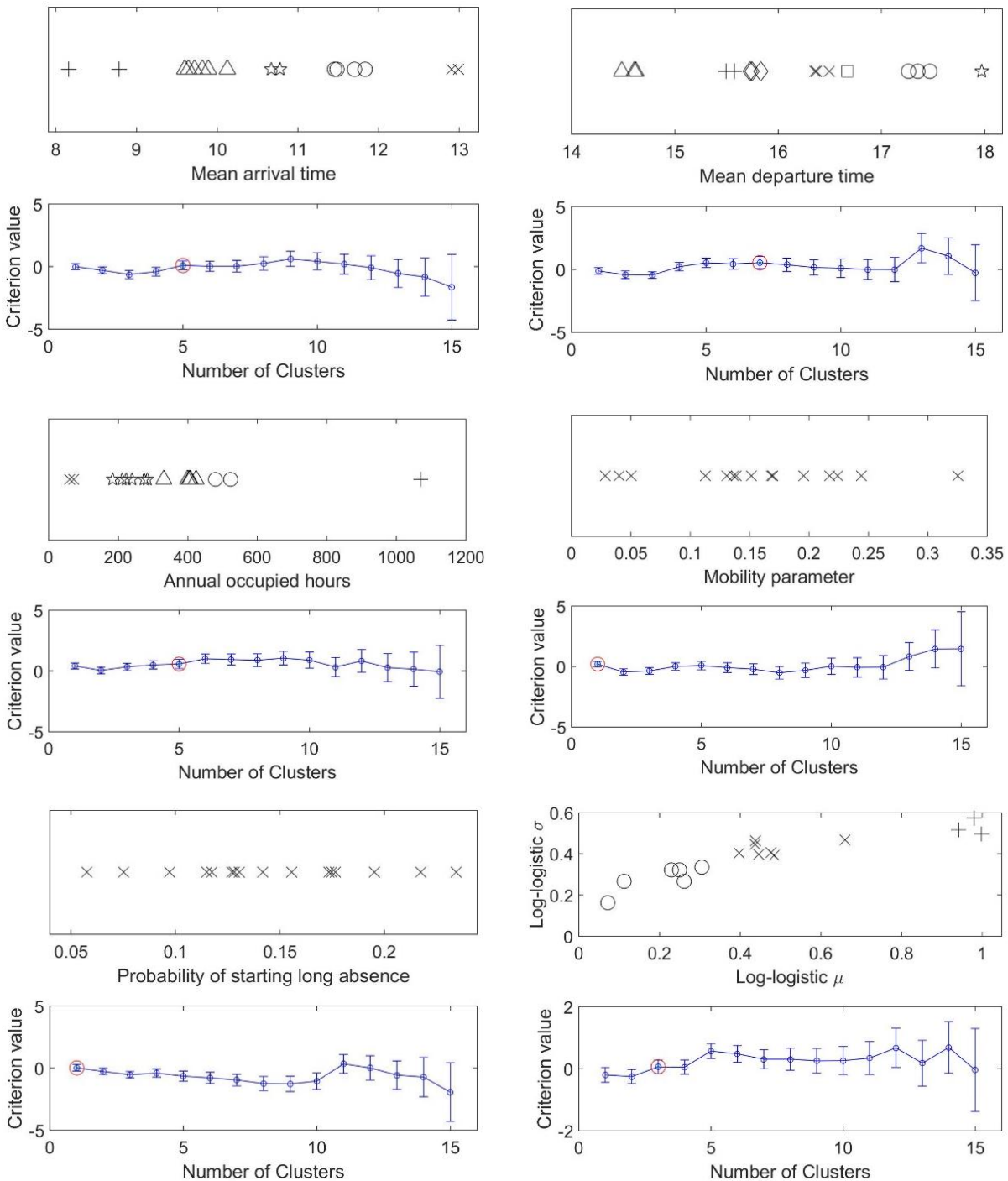


Figure 8: Results of k-means clustering of six key occupant traits. The traits are labelled in the upper graph axes. The upper plot of each pair shows the scatter of the parameters, where each point represents an occupant and the optimal clustering is indicated

by like symbols. The lower plot in each pair indicates the criterion value corresponding to the number of clusters specified. The optimal number of clusters, as determined by the algorithm, is circled.

The results of the k-means clustering analysis suggests that the occupant traits can only be weakly clustered, if at all. A visual inspection of the scatter plots shows that the found clusters are not particularly distinct. Two of the traits – mobility parameter and probability of starting a long absence – resulted in one cluster being optimal. That is, no discernable clusters emerge.

In general, the current evidence suggests that occupant traits and characteristics should not be clustered into typologies. This corroborates previous research by Haldi and Robinson (2010) and Reinhart and Voss (2003), whose plotted results show a relatively continuous distribution of occupant traits for modelling adaptive occupant behaviours. However, the current analysis must be taken in the context that the current dataset contains very similar occupants (engineering professors within the same university culture). Clusters would likely emerge if administrative staff were mixed into the current sample, given their more regular hours. Davis III and Nutter (2010) showed that single-purpose buildings reveal highly distinctive occupancy patterns with respect to each other. But the current dataset does not allow for such analysis to be performed; on the contrary, it enables the exploration of the traits of seemingly similar occupants.

## **5.2 Modelling occupants from grouped occupant data suppresses population diversity.**

A common implicit goal among previous occupant modelling approaches is to develop a single model that fits all occupants. In this way, stochasticity is still introduced through parameter uncertainty and on a time step basis, but inter-occupant diversity may be suppressed by aggregating the data from all occupants.

To explore the impact of aggregating data, the base occupancy modelling method was modified as follows:

- The mean weekday occupancy profile for all occupants combined (and equally weighted) was determined and then a distribution (Weibull) was fit to the variability of the 16 occupants with respect to the mean occupancy profile.
- The mobility parameter and probability of a long absence were fit to normal distributions as previously developed and shown in Figure 6.
- The absence duration data for all occupants was aggregated and then a log-logistic distribution was fit to it. The standard error of the two log-logistic parameters was used to introduce uncertainty into the simulation.

For each simulated occupant, the occupancy profile multiplier and four remaining parameters were independently and randomly selected. The validity of assumption of independence of these randomly sampled parameters is explored in the next hypothesis.

To establish evidence for whether modelling occupancy from aggregated sample data adequately reproduces the sample diversity, 1000 occupants were generated and simulated for 300 weekdays. Occupancy and sample characteristics were found to converge beyond these values. The resulting distribution of select occupant traits is shown in the third column of the plots in Figure 9. Comparing these distributions to that of the original sample of 16 occupants (first columns in Figure 9), it is not strongly evident that modelling occupants from aggregate sample data suppressed diversity despite the fact that 1000 occupants were modelled (vs. the 16 measured occupants). However, beyond the high-level results presented in Figure 9, the current modelling approach failed to generate occupants that vary significantly from the mean of the 16 occupants, as shown in Figure 12. For instance, distinct artefacts from the original occupants are evident in the synthetic occupants. The occupancy profiles and absence duration distributions closely resemble the shape of the sample means, despite the fact that even the sample of 16 shows considerable diversity in the shapes of these plots. Beyond the shape of the synthetically generated mean occupancy profiles, a major cause of the suppressed diversity in the current approach is that the predicted occupancy is highly-sensitive to the method for absence duration modelling because the sample of 16 occupants has high absenteeism (mean of 35% of weekdays).

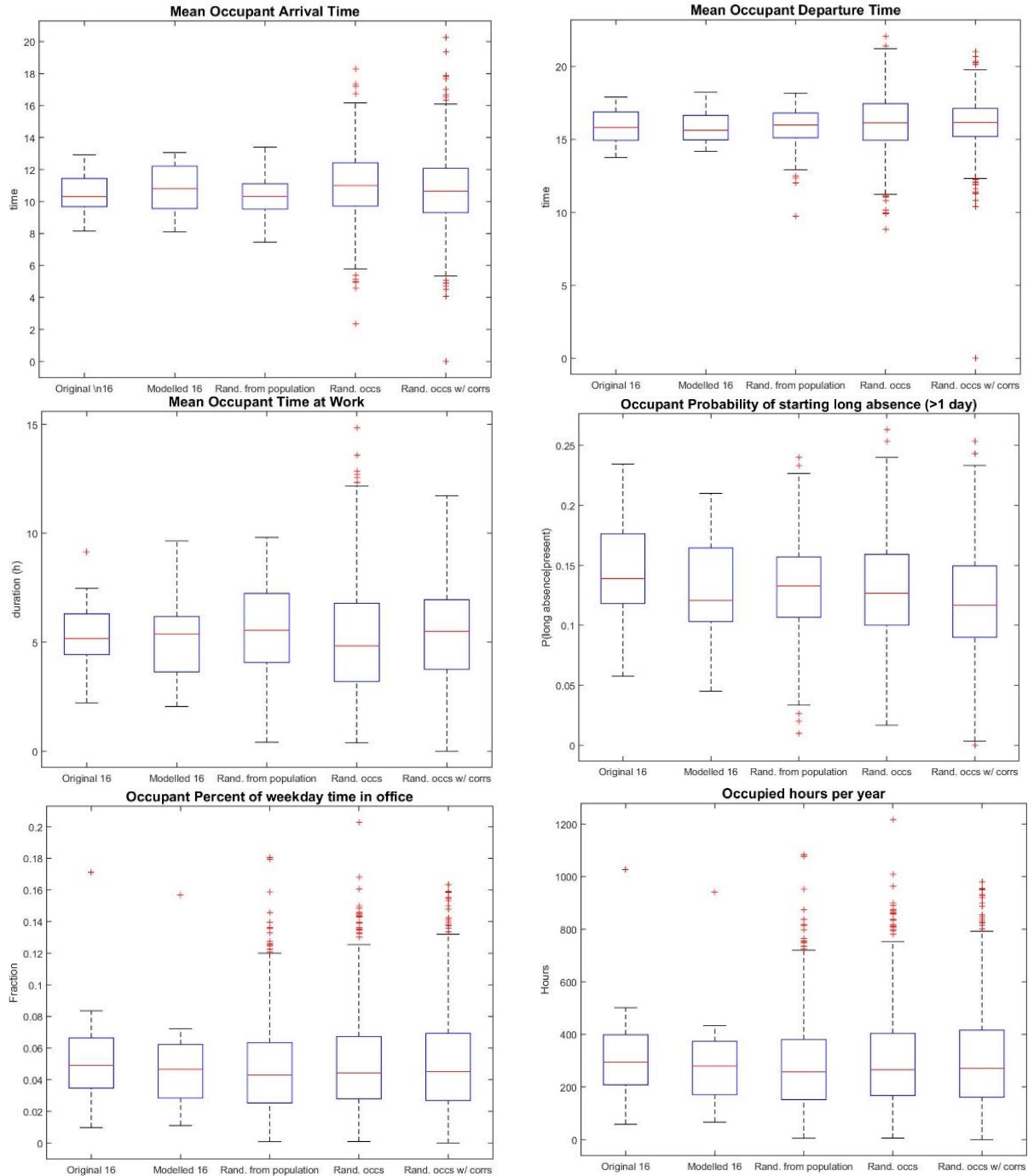


Figure 9: Comparison of key occupant traits with respect to different proposed diversity modelling methods. For each of the plots, each of the following distributions from left to right is shown: the raw data from the original monitored 16 occupants; the models derived directly from the 16 occupants; the model derived from aggregated data; the model derived from the distribution of occupant traits but without trait correlations maintained; and, the model derived from the distribution of occupant traits with trait correlations maintained. The last two modelling approaches address Hypothesis 3.

In conclusion, the modelling approach, to aggregate data initially and then develop occupant models, has a tendency to suppress occupant diversity. While the high level metrics in Figure 9 do not indicate significant suppression, the more detailed results in Figure 12 show that the current diversity modelling approach fails to provide diverse inter-occupant occupancy profiles. Thus a further conclusion from this approach is that models should be evaluated according to many metrics – particularly emphasizing those that could matter for simulation applications. An additional drawback of this method is the risk of bias towards occupants that contribute more events (e.g., long

absences) to the aggregated dataset. Care must also be taken to normalize the data such that datasets of multiple durations are not combined inappropriately.

### **5.3 Randomly selecting occupant traits to develop individual occupant models exaggerates population diversity.**

In contrast to the method above, the extracted occupancy parameters can be obtained for each occupant prior to considering inter-occupant diversity. Subsequently, metamodels can be developed from these parameter distributions. The argument for this staged approach is that preserving individuality of occupants one step further – until their parameters have been extracted from the data – better reproduces data and reduces bias by giving each occupant equal weighting.

Following the approach presented in Section 4, distributions (assumed normally distributed) were obtained after each occupant was modelled. Next, synthetic occupants are randomly created based on these parameter distributions. This section presents two variants on this metamodeling approach. The first independently randomly selects parameter values; the second randomly selects parameter values but maintains correlations between parameters.

A correlation analysis among the parameters of the 16 occupants indicates that some parameter pairs are correlated, though most are very weak. This result is specific to the current domain and possibly the current dataset; thus the conclusions cannot be generalized. A correlation plot matrix from the 16 occupants' parameters is shown in Figure 10. The resulting analysis is important for the current modelling approach but also yields some interesting qualitative conclusions:

- The timing of the three peaks in mean occupancy are mutually correlated. For instance, an occupant who tends to arrive late also tends to have later other peak periods (e.g., after lunch) in the day.
- The amplitudes of the three peaks in mean occupancy are highly correlated (Pearson's  $r = 0.80$  for the first and third peaks). The interpretation for this is that occupants who tend to be present in the morning also tend to be present in the afternoon.
- The two parameters defining the log-logistic regression of the absence duration are highly correlated (Pearson's  $r = 0.91$ ). This indicates that the probability distributions largely take on similar shapes, as later shown in the top-left plot of Figure 11. It was found that independently selecting these two parameters led to unrealistic log-logistic distributions for absence duration.
- There is an interesting negative correlation (Pearson's  $r = -0.57$ ) between the probability of presence at the first peak occupancy period and the probability of starting a long absence. In essence, occupants who are in their office less frequently in the morning are more likely to take one-day or longer breaks. A similar, but lesser effect was found for the last (third) peak of the day.
- The mobility parameter is not well correlated with the other parameters. One might have expected occupants who leave their office frequently to also have a more erratic mean occupancy profile.



Figure 10: Correlations between the 13 parameters

The first variant on metamodeling (independently random parameter selection) is shown in the fourth column of the plots in Figure 9, while the second variant (random parameter selection with correlations maintained) is in the last column of that figure. Referring to the current hypothesis, it was expected that failing to maintain variable correlations would lead to a greater spread in occupancy characteristics because the occupant parameters are less constrained and may be unrealistic. However, this effect is not evident from these results. Figure 9 shows that both variants on metamodeling predict similar levels of diversity among the six occupant characteristics that were explored. This is likely a result of the relatively low level of correlations between pairs of current parameter values.

However, for simulation applications for which individual occupants are modelled, occupancy models that merely replicate sample means are not necessarily adequate for some applications. A closer examination of the intermediate modelling results indicates that failing to maintain the correlations of the sample parameter distribution may lead to unrealistic synthetic occupants. Figure 11 shows the log-logistic fits for the original occupant absence durations and the three other methods examined in this paper. The mean profiles for all modelling methods closely resemble each other. But, when the log-logistic parameters are generated without the correlations maintained, a significant portion of the occupants are modelled as being absent most frequently for two or three days. In the original dataset, three occupants were absent nearly equally frequently for one and two-day periods, but none have a profile similar to the synthetic ones that were generated. However, when the parameter correlations are maintained, the synthetic absence duration profiles resemble the original occupants' in appearance. A similar effect is seen for mean occupancy profiles (Figure 12). The mean occupancy profile plot for which the parameter correlations were not maintained shows that many of the synthetic occupants have unusual schedules (e.g., arrivals to the office at 2 or 3AM and departures close to midnight). It is thus concluded that there is evidence to support the notion that parameter correlations need to be maintained when generating a synthetic population of occupants. It is expected that this effect is increased with stronger parameter correlations.

However, the caveat to the current conclusion that failure to maintain parameter correlations leads to unrealistic synthetic occupants needs to be further verified with a much larger sample. It is not evident at this time whether the usual synthetic occupant characteristics are merely a mathematical artefact or realistic. For instance, it is plausible that in a population of 1000 occupants there would be a small minority who would opt to work overnight and routinely work in their office one day and then work elsewhere for two to three days before returning to the office.



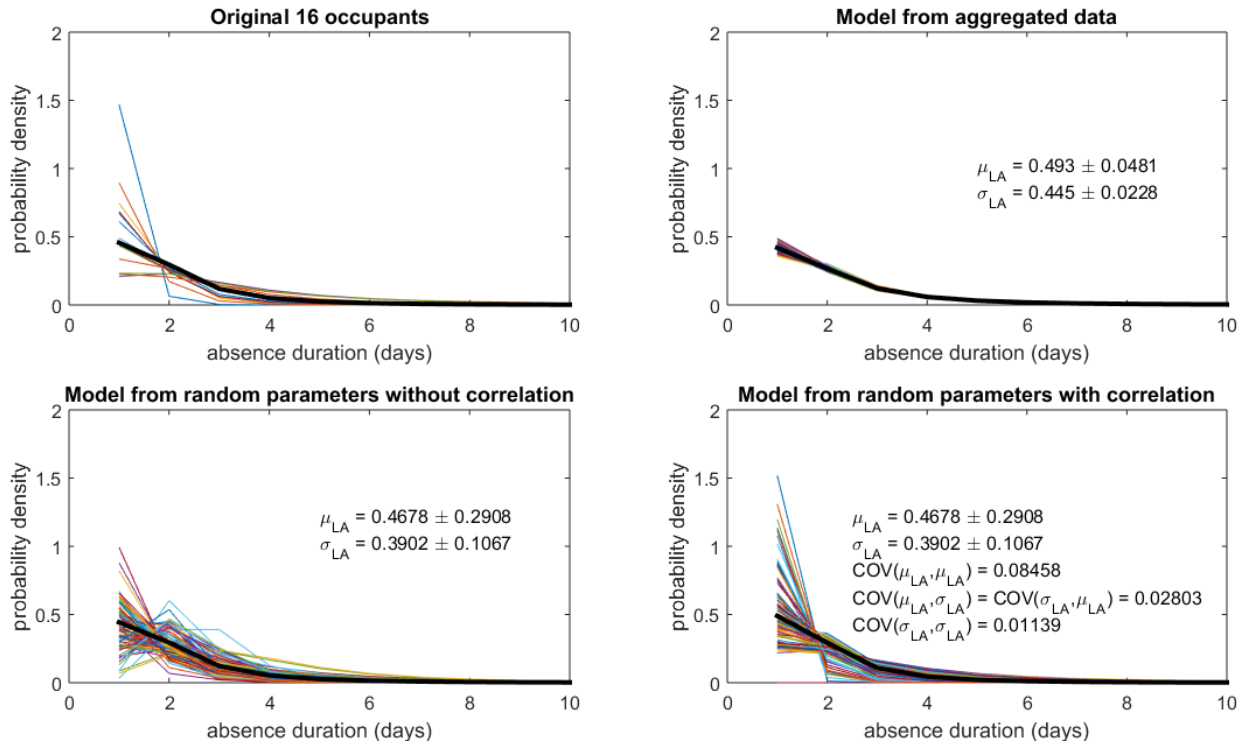


Figure 11: Impact of absence duration modelling. The top left plot shows the original occupant sample while the others are based on 100 generated synthetic occupants. 100 was used for clarity in this figure.

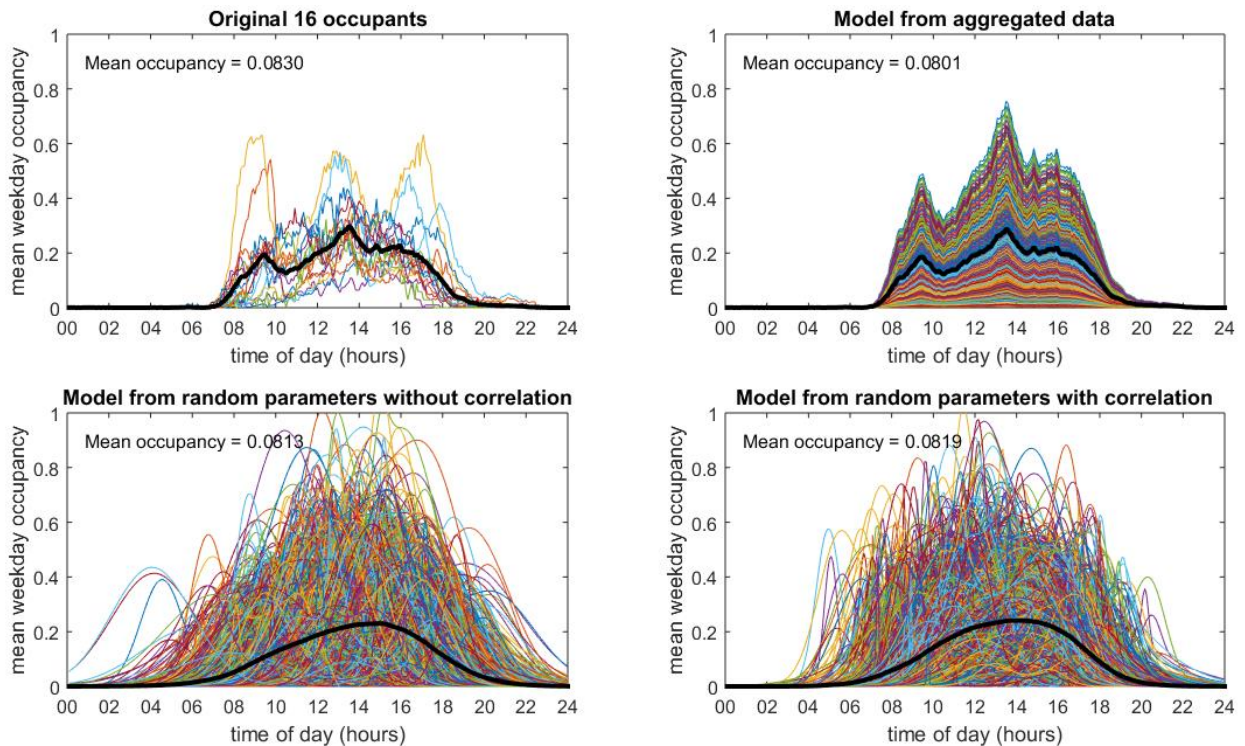


Figure 12: Impact of mean weekday occupancy profile modelling. The top left plot shows the original occupant sample while the others are based on 1000 generated synthetic occupants according to the plot titles.

The current diversity modelling methods (pertaining to Hypothesis 3) yielded somewhat higher diversity predictions than the original sample of 16 occupants and significantly higher than that in the aggregated modelling approach (Hypothesis 2). The expected diversity of the fictional sample is at least as high as that of the measured 16. Ideally, a very large sample would be available to compare the diversity of the measured sample to modelled sample. This is left as future work.

## 6 DISCUSSION

This section provides some final discussion on monitoring and simulation implications of occupant diversity.

### 6.1 Sample size

A recurring acknowledgement in the text above is that the current sample size – 16 – limits the confidence of the conclusions drawn with respect to the hypotheses. In fact, the majority of comprehensive agent-based occupant modelling studies use a similar sample size.

One of the objectives of this research was to use the knowledge of diversity between occupants to estimate the required sample size of monitored occupants required to properly represent a population. Occupant monitoring campaigns can be very costly, both with regards to time and money. There is also significant effort required to comply with ethical constraints, recruit participants, install and maintain sensors and logging systems, and to process and analyze the data. Moreover, studies may be limited by the number of consenting occupants in a particular building. Thus, there is tremendous value to researchers in minimizing the sample size while ensuring adequate representation.

Drawing from statistics theory, we can estimate the required sample size required to estimate the population parameter means (National Institute of Standards and Technology (NIST) 2016). Assuming that the occupant traits resemble a normal distribution (as previously discussed), the sample size required is approximately:

$$n \approx \left( \frac{z_{\alpha}^2}{\delta^2} \right) \left( \frac{\sigma^2}{\mu^2} \right) \quad (2)$$

where  $n$  is sample size,  $z_{\alpha}^2$  is the z-score corresponding to the significance level  $\alpha$ ,  $\delta$  is required precision of the estimate (relative to  $\mu$ ),  $\sigma^2$  is the population variance, and  $\mu$  is the population mean. In the current case, the mean and variance of the population are unknown, thus they are approximated as being equal to those of the sample.

Some occupant traits may be quite homogeneous, while others could have considerable diversity. If we wish to accurately capture the diversity of all important occupant traits, we must ensure that the sample size is adequate to quantify the most diverse trait. Applying Equation 2 to the 13 traits in the current model formulation, we can estimate the appropriate minimum sample size, recognizing the current assumptions and limitations.

Table 2: Assessment of minimum sample size for each of the 13 occupant traits for margin of error = 10% and 95% confidence.

|                          | a1    | b1     | c1    | a2    | b2     | c2    | a3    | b3     | c3    | $\mu$ | p(LA) | $\mu_{LA}$ | $\sigma_{LA}$ |
|--------------------------|-------|--------|-------|-------|--------|-------|-------|--------|-------|-------|-------|------------|---------------|
| sample mean, $\mu$       | 0.216 | 11.032 | 1.551 | 0.186 | 13.566 | 1.747 | 0.243 | 16.098 | 1.689 | 0.155 | 0.145 | 0.468      | 0.390         |
| sample std. dev., $s$    | 0.152 | 1.983  | 0.854 | 0.169 | 1.822  | 1.315 | 0.132 | 1.594  | 0.865 | 0.078 | 0.049 | 0.291      | 0.107         |
| minimum sample size, $n$ | 189   | 13     | 117   | 316   | 7      | 218   | 114   | 4      | 101   | 98    | 44    | 149        | 29            |

The results indicate that a sample size of at least 316 (or 20 times more than currently available) would be necessary to accurately estimate the population means within 10% error margin with 95% confidence. It should be noted that this is a rather simplistic approach to estimating appropriate sample size that neglects the relative impact of the parameters on the simulation results. For instance, it is important to accurately characterize parameters that have a significant impact on occupancy even if the variance of that parameter is not dominant.

The above illustrative example suggests that sample sizes on the order of hundreds are required to represent the total occupant population. This is typically an order of magnitude greater than what is used in the state of the art monitoring

campaigns: 10 to 15. However, further research is required to explore acceptable model error in various applications (e.g., early building design and controls applications).

## 6.2 Occupant uncertainty in simulation

As discussed in the introduction, uncertainty for occupants and other sources are costly for building design and operation. But the real implications for occupant uncertainty are highly dependent on scale, as quantified by O'Brien and Gunay (2015). The law of large numbers applied to occupancy modelling means that for groups of occupants, the mean occupancy will approach the population mean. Thus, the importance of simulating distinct occupants is high for small spaces (e.g., single offices) and lower for large office buildings, communities, and cities. Put differently, the implications of this paper are greatest for zone-level design (e.g., terminal HVAC units) and least for urban designers.

Despite the cost of occupant uncertainty, diversity offers some important benefits. Most notably, diversity reduces peak demand (space, power, water, and so on). With good knowledge of occupant diversity, building designers can take calculated risks to minimize the frequency of under-sizing systems while also minimizing costs. As a brief illustrative example, the employer of the 16 occupants in the current paper could have saved considerable cost on building construction if they had considered occupant diversity and the occupants were willing to give up dedicated offices (i.e., hotelling style office management). Figure 13 shows the time distribution for number of simultaneous occupants present in their offices. For instance, 99.8% of the time, the six offices would have been sufficient (vs. 16).

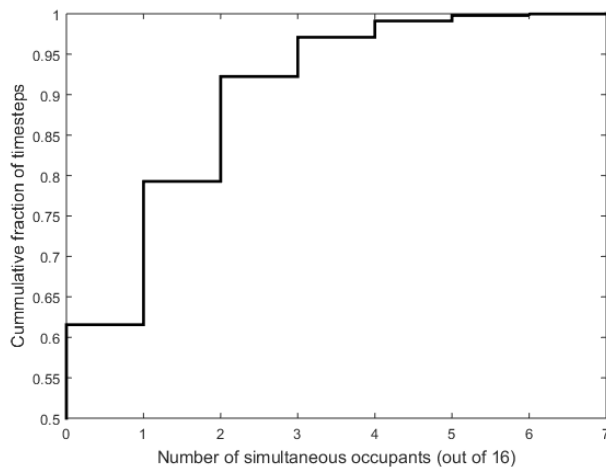


Figure 13: Cumulative distribution of time as a function of number of simultaneous occupants present (out of the sample of 16)

## 7 CONCLUSION

This paper set out to explore a seldom-discussed issue in the field of occupant behaviour modelling for building simulation applications: diversity between occupants. Prior modelling efforts have typically focused on attempting to fit data to one or a small number of occupant archetypes without attempting to characterize the diversity of occupants and their behaviour. The objective of this paper was to provide a better understanding of occupant diversity modelling and propose new methods for attacking this age old research problem. While, the dataset is limited to 16 occupants, the results provide evidence for the appropriateness of the diversity modelling approaches and the methodology lays the groundwork for future studies. The key conclusions, as per the hypotheses tested, are:

1. Occupant traits for an occupant sample more closely resemble a continuum than a discrete distribution.
2. Developing occupant models from aggregated occupant data significantly suppresses diversity relative to observations.
3. Developing multi-parameter metamodels from individual occupants in a sample preserves occupant diversity. But in generating synthetic occupants, maintaining parameter correlations should be preserved to prevent unrealistic occupants from being simulated.

The current findings and methods have widespread implications for the BPS researcher and user community. It raises questions about adequate experimental sample sizes, the importance of reporting all factors leading to monitored occupant diversity, occupant modelling approaches, and strategies to simulate and exploit occupant diversity and uncertainty. The current dataset suggests that it would have been more appropriate to have a sample of hundreds rather

than 16. But notably most current occupant monitoring and modelling studies are in the range of 10-15; some even smaller. However, as outlined in the discussion, the necessary sample size is greatly dependent on the diversity of the occupants. Inherent uniformity of a population will require a smaller sample size (e.g., school children who are at a similar age and have a fixed schedule).

This paper also discussed the importance of modelling occupant diversity for different scales of simulation (room to community) and design stage. The authors argue that it is most important at small scales where assumptions about individuals are critical to estimating energy and comfort performance and for sizing equipment. For large scales where the most important performance metrics are aggregated (e.g., impact of occupants on the electrical grid or district-scale co-generation sizing), the impact of individuals is much less important than the aggregated behaviour of all occupants. But for rooms and other small scales, understanding the uncertainty of individuals is of much greater importance (e.g., for sizing radiator capacity). This, however, does not mean that occupant models that are aimed at larger scale simulation applications can be constructed from small sample sizes. To the contrary, it is even more critical for such applications that the population is properly characterized. For if a sample from a population is poorly chosen, extrapolated results could cause costly errors.

While this paper focused on modelling occupant diversity with regards to occupancy patterns, the authors speculate that similar approaches could be taken for adaptive and non-adaptive occupant actions and states. Moreover, similar theoretical research is required to support the suitability for modelling the inter-occupant diversity for these other domains.

The subject of occupant diversity modelling is in its infancy and is rich with opportunities. Important future work to extend the current paper includes:

- Repeating the analysis in this paper for samples of hundreds or thousands to verify the conclusions drawn from the current sample;
- Exploring diversity and diversity modelling of both occupancy and occupant actions;
- Quantifying temporal diversity of occupants (e.g., at which frequencies do occupancy and occupant actions become periodic, if any?); and,
- Exploring the relationship between occupant model parameters for multiple domains (e.g., occupant use of lights, blinds, windows, etc.) to determine if there are correlations. For instance, is a diligent light user also energy conserving with regard efficiently controlling window openings?

## 8 ACKNOWLEDGEMENTS

The financial support from the Natural Sciences and Engineering Research Council (NSERC) Canada is gratefully acknowledged. This paper benefited greatly from discussion with members of IEA EBC Annex 66.

## 9 REFERENCES

Abushakra, B., J. S. Haberl and D. E. Claridge. 2004. "Overview of Existing Literature on Diversity Factors and Schedules for Energy and Cooling Load Calculations." *ASHRAE Transactions* 110(1).

Andersen, P. D., A. Iversen, H. Madsen and C. Rode. 2014. "Dynamic modeling of presence of occupants using inhomogeneous Markov chains." *Energy and Buildings* 69: 213-223.

Athienitis, A. K. and W. O'Brien (2015). *Modeling, design, and optimization of net-zero energy buildings*, Ernst & Sohn.

Bourgeois, D., C. Reinhart and I. Macdonald. 2006. "Adding advanced behavioural models in whole building energy simulation: A study on the total energy impact of manual and automated lighting control." *Energy and Buildings* 38(7): 814-823.

Boyce, P. R. (2014). *Human Factors in Lighting, Third Edition*, Taylor & Francis.

D'Oca, S. and T. Hong. 2015. "Occupancy schedules learning process through a data mining framework." *Energy and Buildings* 88(0): 395-408.

- Davis III, J. A. and D. W. Nutter. 2010. "Occupancy diversity factors for common university building types." *Energy and Buildings* 42(9): 1543-1551.
- Deru, M., K. Field, D. Studer, K. Benne, B. Griffith, P. Torcellini, B. Liu, M. Halverson, D. Winiarski and M. Rosenberg (2011). Commercial Reference Building Models of the National Building Stock, US Department of Energy.
- Djunaedy, E., K. Van den Wymelenberg, B. Acker and H. Thimmana. 2011. "Oversizing of HVAC system: Signatures and penalties." *Energy and Buildings* 43(2): 468-475.
- Feng, X., D. Yan and T. Hong. 2015. "Simulation of occupancy in buildings." *Energy and Buildings* 87: 348-359.
- Gangestad, S. and M. Snyder. 1985. "' To carve nature at its joints": On the existence of discrete classes in personality." *Psychological Review* 92(3): 317.
- Gram-Hanssen, K. 2010. "Residential heat comfort practices: understanding users." *Building Research & Information* 38(2): 175-186.
- Gunay, H. B., A. Fuller, W. O'Brien and I. Beausoleil- Morrison. 2016. "Detecting Occupants' presence in office spaces: a case study". eSim 2016, Hamilton, Ontario
- Gunay, H. B., W. O'Brien and I. Beausoleil-Morrison. 2015. "Development of an occupancy learning algorithm for terminal heating and cooling units." *Building and Environment* 93, Part 2: 71-85.
- Haldi, F. 2013. "A probabilistic model to predict building occupants' diversity towards their interactions with the building envelope". Building Simulation Chambery, France Aug. 26-29
- Haldi, F. and D. Robinson. 2009. "Interactions with window openings by office occupants." *Building and environment* 44(12): 2378-2395.
- Haldi, F. and D. Robinson. 2010. "Adaptive actions on shading devices in response to local visual stimuli." *Journal of Building Performance Simulation* 3(2): 135-153.
- Haldi, F. and D. Robinson. 2011. "The impact of occupants' behaviour on building energy demand." *Journal of Building Performance Simulation* 4(4): 323-338.
- Hoes, P., J. L. M. Hensen, M. G. L. C. Loomans, B. de Vries and D. Bourgeois. 2009. "User behavior in whole building simulation." *Energy and Buildings* 41(3): 295-302.
- Jakubiec, J. A. and C. F. Reinhart. 2012. "The 'adaptive zone'—A concept for assessing discomfort glare throughout daylight spaces." *Lighting Research and Technology* 44(2): 149-170.
- Lee, K. H. and S. Schiavon. 2014. "Influence of three dynamic predictive clothing insulation models on building energy use, HVAC sizing and thermal comfort." *Energies* 7(4): 1917-1934.
- Mahdavi, A. and F. Tahmasebi. 2015. "The inter-individual variance of the defining markers of occupancy patterns in office buildings: a case study". 14th International Conference of IBPSA, Hyderabad, India, December 7-9, 2015
- Meerbeek, B., M. te Kulve, T. Gritti, M. Aarts, E. van Loenen and E. Aarts. 2014. "Building automation and perceived control: A field study on motorized exterior blinds in Dutch offices." *Building and Environment* 79: 66-77.
- Mohammed, S. and E. Ringseis. 2001. "Cognitive Diversity and Consensus in Group Decision Making: The Role of Inputs, Processes, and Outcomes." *Organizational Behavior and Human Decision Processes* 85(2): 310-335.

Myers, I. B., M. H. McCaulley and R. Most (1985). MBTI Manual: A guide to the development and use of the Myers-Briggs type indicator, Consulting Psychologists Press.

National Institute of Standards and Technology (NIST). (2016). "Selecting Sample Sizes." Retrieved February 5, 2016, 2016, from <http://www.itl.nist.gov/div898/handbook/ppc/section3/ppc333.htm>.

Nicol, F. 2004. "Adaptive thermal comfort standards in the hot-humid tropics." *Energy and Buildings* 36(7): 628-637.

Nicol, J. F. 2001. "Characterising occupant behaviour in buildings: towards a stochastic model of occupant use of windows, lights, blinds, heaters and fans". Building Simulation, Rio de Janeiro, August 13-15

O'Brien, W. and H. B. Gunay. 2014. "The contextual factors contributing to occupants' adaptive comfort behaviors in offices—A review and proposed modeling framework." *Building and Environment* 77: 77-87.

O'Brien, W., K. Kapsis, A. K. Athienitis and T. Kesik. 2010. "Methodology for quantifying the performance implications of intelligent shade control in existing buildings in an urban context". SimBuild 2010, New York City, August 11-13, 2010

O'Brien, W. and H. B. Gunay. 2015. "Mitigating office performance uncertainty of occupant use of window blinds and lighting using robust design." *Building Simulation* 8(6): 621-636.

Page, J., D. Robinson, N. Morel and J.-L. Scartezzini. 2008. "A generalised stochastic model for the simulation of occupant presence." *Energy and Buildings* 40(2): 83-98.

Parys, W., D. Saelens and H. Hens. 2011. "Coupling of dynamic building simulation with stochastic modelling of occupant behaviour in offices—a review-based integrated methodology." *Journal of Building Performance Simulation* 4(4): 339-358.

Rasouli, M., G. Ge, C. J. Simonson and R. W. Besant. 2013. "Uncertainties in energy and economic performance of HVAC systems and energy recovery ventilators due to uncertainties in building and HVAC parameters." *Applied Thermal Engineering* 50(1): 732-742.

Reinhart, C. F. 2004. "Lightswitch-2002: a model for manual and automated control of electric lighting and blinds." *Solar Energy* 77(1): 15-28.

Reinhart, C. F. and K. Voss. 2003. "Monitoring manual control of electric lighting and blinds." *Lighting Research & Technology* 35(3): 243-260.

Rijal, H., P. Tuohy, M. Humphreys, J. Nicol, A. Samuel and J. Clarke. 2007. "Using results from field surveys to predict the effect of open windows on thermal comfort and energy use in buildings." *Energy and Buildings* 39(7): 823-836.

Schweiker, M., F. Haldi, M. Shukuya and D. Robinson. 2012. "Verification of stochastic models of window opening behaviour for residential buildings." *Journal of Building Performance Simulation* 5(1): 55-74.

Sun, K. (2014). Stochastic Modeling of Overtime Occupancy and Its Application in Building Energy Simulation and Calibration.

Sutter, Y., D. Dumortier and M. Fontoynt. 2006. "The use of shading systems in VDU task offices: A pilot study." *Energy and Buildings* 38(7): 780-789.

Tahmasebi, F. and A. Mahdavi. 2015. "The sensitivity of building performance simulation results to the choice of occupants' presence models: a case study." *Journal of Building Performance Simulation*: 1-11.

Tahmasebi, F. and A. Mahdavi. 2016. "An inquiry into the reliability of window operation models in building performance simulation." *Building and Environment* 105: 343-357.

Tibshirani, R., G. Walther and T. Hastie. 2001. "Estimating the number of clusters in a data set via the gap statistic." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(2): 411-423.

Van Knippenberg, D. and M. C. Schippers. 2007. "Work group diversity." *Annu. Rev. Psychol.* 58: 515-541.

Wang, D., C. C. Federspiel and F. Rubinstein. 2005. "Modeling occupancy in single person offices." *Energy and Buildings* 37(2): 121-126.

Yun, G. Y., K. Steemers and N. Baker. 2008. "Natural ventilation in practice: linking facade design, thermal performance, occupant perception and control." *Building Research & Information* 36(6): 608-624.