A TOOLKIT FOR DEVELOPING DATA-DRIVEN OCCUPANT BEHAVIOUR AND PRESENCE MODELS

H. Burak Gunay[1], William O'Brien[1], Ian Beausoleil-Morrison[2]
[1]Civil and Environmental Engineering Department, Carleton University, Ottawa, Canada
[2]Mechanical and Aerospace Engineering Department, Carleton University, Ottawa, Canada

ABSTRACT

An ever-increasing amount of data from building automation systems and other indoor environmental data loggers represent an untapped opportunity to analyse occupants' presence and behaviour patterns in commercial buildings. Consequently, many emerging researchers have been entering the field of data-driven occupant modelling in buildings – particularly within IEA EBC Annex 66. Upon the methods reviewed in the literature, this paper introduces a starter toolkit for developing occupant behaviour and presence models. The toolkit contains 22 Matlab functions that take common sensory data and conduct a series of statistical analyses about the occupancy and lighting, blinds, thermostat, and plug-in equipment use patterns. The paper presents the functionalities of the toolkit through an analysis conducted upon the data gathered in an academic office building in Ottawa, Canada. We made the functions and example data files publicly available with this paper.

INTRODUCTION

Understanding occupant behaviour and presence in buildings is central to achieving better design and operation. We can tailor operating schedules specific to a building's occupancy patterns (Gunay et al. 2015a), choose better setpoints specific to the occupants' preferences (Guillemin and Molteni 2002), and predict how a building design would influence occupants' behaviour (Sanati and Utzinger 2013, O'Brien and Gunay 2015).

Availability of and demand for low-cost sensing technologies and building automation systems (BAS) in office buildings render the potential for developing data-driven occupant behaviour and presence models. These occupant models – instead of characterizing the human physiology comprehensively like ASHRAE (2004)'s PMV for thermal comfort or IES (2015)'s DGI for visual comfort – treat humans as a blackbox to seek statistically meaningful input-output relationships. These models can be implemented in building performance simulation (BPS) tools to study the impact of different design and control alternatives on occupant comfort and energy use (Gunay et al. 2015b).

Literature review: presence modelling

In the past two decades, researchers have instilled the basics of modelling human presence and behaviour in office buildings (Haldi and Robinson 2011). In presence modelling, three different methods have been used. The most common method is building weekly occupancy schedules – presenting the likelihood of presence as a function of the time of day and the day of week (Gunay et al. 2015b). The strength of this model form is that it is easy to interpret by building operators and simulationists. Building specific occupancy schedules provide valuable insights that can help operators choose operating schedules. Simulationists can incorporate them quickly in building models to represent occupancy. Recently, Mahdavi and Tahmasebi (2015) introduced a method to generate an occupancy time-series (i.e., sequential presence and absence information) from an occupancy schedule.

The second method used in occupancy modelling is the Markov chains (Page et al. 2008, Wang et al. 2011). The model predicts the likelihood of an arrival when occupants are absent and it predicts the likelihood of a departure when occupants are present. The strength of this approach is that unlike the traditional schedule-based models the likelihood of observing an arrival or departure for the rest of the day can be estimated – given current time and current state of presence. This may help making midday control decisions such as temperature setbacks when the likelihood of observing an arrival is very small for the rest of day (Gunay et al. 2015a). The Markov occupancy models are capable of creating realistic occupancy time-series which can be used in BPS models. A weakness of the Markov occupancy models is that they treat arrival and departure events independently. In reality, occupants may depart early when they arrive early or they may depart late when they arrive late (Page 2007).

Survival models (the third method) appear to be a promising alternative to tackle this limitation (Parys et al. 2011). Survival models can predict the duration of an intermediate vacancy period following a departure or they can predict the duration of an intermediate occupancy period upon an arrival (Wang et al. 2005).

Literature review: adaptive behaviour modelling

Adaptive behaviours are actions that occupants undertake to adjust their environment (e.g., blinds closing, light switch, window opening) or to adjust to their environment to improve comfort (e.g., adjust clothing) (Nicol and Humphreys 2004). In modelling adaptive behaviours, four common methods have been used. The traditional way of modelling adaptive behaviours is building schedules – e.g., presenting the ratio of the lights switched on or the mean blind occlusion rate averaged over a week or a month. This approach provides information that is easy to interpret and does not require data from indoor environmental quality (IEQ) sensors. The underlying assumption of this method is steady-periodicity. In other words, the model form predicates that the time of week or the month of year alone is adequate to make predictions for adaptive occupant behaviour. This assumption arises from the fact that indoor and outdoor environmental factors that influence adaptive behaviours tend to recur in daily or seasonal cycles. However, when a simulationist or a building operator wants to determine the outcomes of a design or a control strategy, the indoor climatic conditions that affect the occupants' behaviour will inevitably change. For example, changing the glazing material and geometry, shading material and controls, lighting fixture and controls will play a role over occupants' use of lighting. Because schedules do not incorporate indoor environmental proxies (e.g., workplane illuminance) to explain occupants' adaptive behaviours, these models may fail to mimic them under other building design and control scenarios (Hoes et al. 2009).

The second method used in adaptive behaviour modelling is the Bernoulli random processes (Haldi and Robinson 2008). The Bernoulli behaviour models predict the likelihood that an occupant left a building component at a certain state (e.g., probability of the lights being on) as a function of the predictor variables (e.g., ambient illuminance). The Bernoulli model form – despite incorporating the role of environmental variables on occupant behaviours – is limited to using ambient sensory variables as predictors. This is because the outcome of an adaptive action affects the indoor conditions input to the model. For example, when a Bernoulli lighting model inputs the indoor light intensity, it provides the ratio of the time the lights are on to the entire time at a certain light level. The indoor light intensity is affected by both daylight and electric lighting. Thus, the numerator (the time the lights are on at a certain light level) will contain the effects of both daylight and electric lighting, whereas the denominator (the entire time spent at a certain light level) will be dependent on the duration of electric lighting use and daylight. Therefore, it is inappropriate to use indoor environmental variables while developing Bernoulli models. Using outdoor environmental variables as model inputs can be seen as an advantage because fewer sensors are needed – a single ambient sensor can be adequate for the entire façade. However, using ambient environmental variables as predictors limit the transferability of the models to other buildings, as it does not take into account the building design specific nuances (e.g., window material and geometry). Also, research suggests that the occupants do not actively fine-tune behaviours in response to changing environmental conditions (Gunay et al. 2013). They rather overcompensate in response to conditions of discomfort and wait much longer after the source of discomfort dissipates to reverse their action. For example, occupants tend to close their blinds upon visual discomfort and it may take days or even weeks before they reopen them (O'Brien et al. 2013). Similarly, occupants turn on their lights when it is dark and they tend to turn them off only before they leave their offices – in lieu of when there is adequate daylight (Reinhart 2004). Therefore, the Bernoulli models tend to exaggerate occupants' activeness (Haldi and Robinson 2011).

The third method used in modelling adaptive behaviours is the discrete-time Markov chains (Rijal et al. 2008, Haldi and Robinson 2009). The discrete-time Markov models predict the likelihood of undertaking an adaptive behaviour in the next timestep. They can be developed by both indoor and outdoor environmental variables, because they are derived upon conditions just before occupants undertake the behaviour. The Markov models treat adaptive actions and their reversals independently and are capable of generating realistic behaviour patterns. Note that realistic representation of the frequency of adaptive actions can act as a proxy to occupants' comfort. From thermostat overrides we can infer thermal discomfort conditions and from overrides to lighting and blinds automation systems we can infer visual discomfort conditions (Gunay et al. 2016a). A common issue regarding the discrete-time Markov models is their dependency on fixed and prescribed timesteps (Gunay et al. 2014). They only provide the likelihood of an event in the next timestep. Although mathematically it is possible to modify their prediction horizons to maintain compatibility with other building, HVAC, and occupant models that require different temporal resolutions, the procedure is undocumented and consequences over the models' predictive accuracy remain unclear. Discrete-event Markov models link the calling points of an occupant action model to an external event (Reinhart 2004, Rijal et al. 2008). For example, in Reinhart (2004)'s light switch model, simulated occupants are modelled to turn on their lights at arrivals (event). In Rijal et al. (2008) occupants were modelled to consider window opening and closing upon a change in the predicted mean vote (event) (ASHRAE 2004). The advantage of the discrete-event Markov modelling approach is that it alleviates the modellers from aforementioned limitations of the fixed and prescribed timesteps – inherent in discrete-time Markov models. However, discrete-event Markov modelling is challenged by finding an appropriate event definition to replace the timestep concept. Another limitation of this approach is that its predictive performance relies on the accuracy of the external events' predictions. For example, the predictive performance of the discrete-event Markov light switch model for arrival is subject to the occupancy models' ability to represent the frequency and timing of the arrival events accurately.

Literature review: plug-in office equipment usage

Use of plug-in office equipment is not an adaptive behaviour, because it is independent from building design and controls and it cannot be explained with occupant discomfort. However, plug-in equipment in office spaces influences both internal heat gains and electricity use – accounting for 13 to 44% of the total energy use in commercial buildings (CIBSE 2012, Menezes et al. 2012). In plug-in equipment modelling, two different methods have been used. The traditional method is building weekly plug load schedules. A data-driven plug load schedule can be a simple way to pinpoint recurring issues such as computers left on over night or weekends (Menezes 2013). When high resolution longitudinal plug load measurements are not available, weekly plug load schedules may be predicted from weekly occupancy schedules (Mahdavi et al. 2008).

The second method used in modelling plug-in equipment loads in offices is a survival model predicting a relationship between plug-in equipment loads and the duration of vacancy. Analogous to Pigg et al. (1996)'s light switch model, the plug-in equipment loads during vacancy periods can be modelled as a function of the length of absence periods (Gunay et al. 2016b) – meaning that occupants tend to turn off their equipment when they leave for a longer period of time (e.g., weekend) than when they leave for a short intermediate break.

Motivation and scope

BASs and IEQ data loggers employed in post-occupancy commissioning surveys will provide an ever-increasing amount of data for us to analyse occupant behaviour and presence patterns in buildings. Within collective research efforts for studying occupant behaviour in buildings (e.g., IEA EBC Annex 66), emerging researchers experience a steep learning curve and often redundantly spend time on the basics of human behaviour and presence modelling. Upon synthesizing the modelling methodologies in the literature, this paper presents a starter kit for developing data-driven occupant behaviour and presence models. The toolkit contains 22 Matlab functions to process common data structures used in building a large number of occupant models. The functionalities of the toolkit are presented through an analysis conducted upon the data gathered in an academic office building in Ottawa, Canada. The functions and the example data files are also made publicly available with this paper (Supplement 2016).

STRUCTURE OF THE TOOLKIT

BASs represent great potential to gather occupant behaviour and presence data. In modern buildings, most occupant actions can be registered through their interactions with control interfaces. These can include manual adjustments to the thermostat setpoints, light switches (if lighting is integrated to the BAS), and motorized blinds (if the blinds are part of the BAS). In addition, many commercial buildings are equipped with a range of sensors monitoring the indoor climate. Some of the common building sensor types include passive-infrared (PIR) motion detectors, CO_2 sensors, relative humidity sensors, photodiodes, thermistors, and current sensors. The drawbacks of relying on existing BAS infrastructure to obtain occupant data are limited flexibility on sensor type, location, and grade and access to occupied spaces to document contextual factors. Also, the ethics procedures about conducting research on already archived occupant data can be unestablished and bring about ownership issues in granting access to the data (e.g., facilities management permission vs. research ethics committee clearance).

Occupant data in BASs can be stored at and accessed from three different places: (1) local controllers, (2) BAS archivers, (3) BAS archivers' cloud server. Inside local controllers (where the sensors are located), the data can stored as BACnet database trendlog objects. The major limitation of this approach is that data storage capacity of local controllers are minimal. Depending on the sampling frequency and the number of trendlogs in the controller, one would need to back up data from these local controllers within weeks before newer data starts overriding the older ones. Furthermore, some occupant data need to be stored per event basis (e.g., light switch or motion detectors). Although it is possible to budget this limited data storage capacity by adjusting the sampling frequency, there is uncertainty on how fast the event-based data will overflow. With the development of building controllers with larger data storage capacities, this problem can be relieved. However, such high performance building controllers are still uncommon in modern buildings. It is no longer uncommon to see BAS data archivers in large institutional buildings. These archivers are industrial computers that scan the entire controls network, and obtain and permanently store each trendlog. The data inside BAS archivers can be accessed through a physical connection to the local controls network. The data inside BAS archivers can be sent and stored inside a cloud storage service, and the data can be accessed through its API. When BASs fail to provide high-resolution occupant data, additional data loggers can be either integrated into the BAS or set up on a separate sensor network with centralized or distributed data logging. Figure 1 illustrates the process of accessing and storing data used in occupant presence and behaviour modelling process.

The Matlab functions presented in this paper are developed to help researchers conduct basic occupant analyses. Three of the functions are intended for preprocessing of the occupant data. This includes: (1) a function for accessing, downloading, and organizing occupant data from a BAS cloud server; (2) a function using Nagy et al. (2015)'s algorithm to convert passive-infrared movement detections to occupancy time-series; (3) and a function to convert event-based data (e.g., light switch data) to time-series data. Three functions were built for developing weekly schedules for lighting, occupancy, and plug loads. Three functions develop Markov and survival occupancy models. Five functions were built to conduct lighting analysis through Bernoulli, discrete-time and discrete-event Markov models. Three functions to analyse blinds use patterns, two to conduct analyses on plug load patterns, and two functions to study thermostat use patterns were built. The toolkit also contains example datasets so that the researchers can inspect and improve the functions.

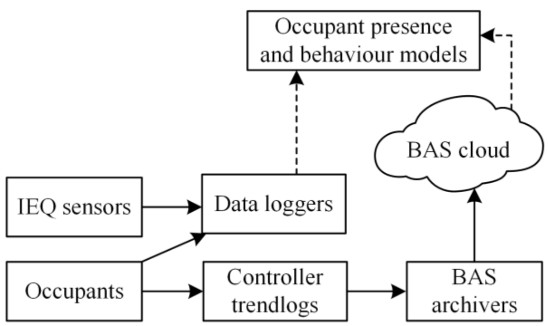

Figure 1: The process of gathering occupant behaviour and presence related data.

This paper also presents the functionalities of the toolkit through an analysis conducted upon the data gathered in an academic office building in Ottawa, Canada. The dataset contains data records for occupancy, lighting, thermostat use, plug loads, blinds, solar irradiance, indoor illuminance, and indoor temperature (see Table 1). The html scripts for the interactive versions of the figures developed upon these analyses were also included with this paper (Supplement 2016). The reader can drag and drop these scripts into an open web browser connected to the internet.

THE TOOLKIT FUNCTIONS

The first function in the toolkit is BAS2BackUp. It downloads the data of interest from the API of a BAS archiver server. The user needs to provide four inputs. Firstly, a library containing the database addresses of trendlogs (controller and trendlog IDs) of interest needs to be prepared in xls, xlsx, csv format. Table 2 presents an example library. The library should have room/space descriptors and data type descriptors. In large controls databases, it may be time consuming to download the entire library. To this end, the second user input is a string array specifying a subset of the data type descriptors (e.g., MD in Table 2). The third input is the url for the cloud server of the BAS archiver. With the permission from the building owner, it can be acquired from the buildings' controls vendor. The fourth input is the output directory where the function will create its subfolders and categorize the downloaded files.

Table 1: Characteristics of the dataset used in presenting the functionalities of the toolkit.

Data Type	Number of offices	Acquisition period	Sensor type or monitoring method	Time interval and post-processing
Occupancy	16	Jan 2014 - Dec 2015	Passive infrared-motion sensor (from BAS)	Event-based data
Lighting	9	Jan 2014 - Dec 2015	Light switch (from BAS)	Event-based data
Thermostat use	8	Sep 2015 - Dec 2015	Thermostat keypress logs (from BAS)	Event-based data
Plug loads	10	Dec 2014 - Aug 2015	Non-invasive current sensors (from data loggers)	Time-series data in 60 min timesteps
Blinds	8	Feb 2014 - Oct 2014	Time-lapse camera (manual inspection)	Time-series data in 30 min timesteps
Solar irradiance on the facade	—	Jan 2014 - Dec 2015	Pyranometer sensor (from weather station)	Time-series data in 15 min timesteps
Indoor illuminance on the ceiling	5	Mar 2015 - Dec 2015	Photodiode sensor (from BAS)	Time-series data in 15 min timesteps
Indoor air temperature	8	Jan 2015 - Dec 2015	Thermistor (from BAS)	Time-series data in 15 min timesteps

Table 2: *Example input library for BAS2BackUp*.

Room	Controller	TL	Description
4202	421860	32	MD
4202	421860	3	LIGHT
4202	421860	33	LUX
4205	421858	3	LIGHT
4205	421858	8	MD

The function PIR2Occ converts passive-infrared sensors' movement detections to occupancy time-series data using Nagy et al. (2015)'s adaptive time delay method. The user should specify the input directory where the data files from individual motion detectors are located. The user should ensure that the motion detector file names contain a common substring MD (e.g., RM5208_MD.csv). The user should also specify the output file directory where the occupancy time-series data created by the function will be written. The third user input is the minimum permissible timestep size – in case the adaptive time delay algorithm provides a small timestep size that is not preferred by the user.

The function Event2Time converts event-based data (e.g., light switch) to time-series data. Users need to specify the input file directory where the event-based data files are located and the output file directory where the time-series output files will be written. A common character string key needs to be provided, so that the function will find and open all the file names with the key in the input directory.

Schedule builder functions light2sch, occ2sch, plug2sch input time-series data files for lighting, presence, and plug loads from individual offices and output weekly lighting, occupancy, and plug load schedules for individual offices. Figure 2 illustrates example mean weekly occupancy and lighting schedules developed using the dataset specified in Table 1. This type of analysis could reveal that the building occupancy was low – yet it was occasionally extending over weekends. In a given building, an operator can use this information for setback scheduling of the HVAC equipment. Figure 3 illustrates the mean plug load schedule per occupant developed for the same offices. It indicates that after-hours plug-in office equipment electricity demand was more than 60% of that at peak occupancy. The facilities management can use this information to promote an energy awareness campaign or decide to invest in programmable wall plugs. In addition, a simulationist can easily input these schedules in a BPS model for post-occupancy evaluations or the design of a similar building.

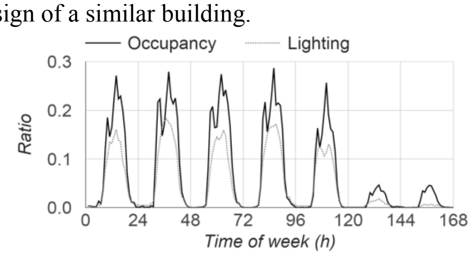

Figure 2: Mean weekly occupancy and lighting schedules (built using outputs of occ2sch and light2sch functions).

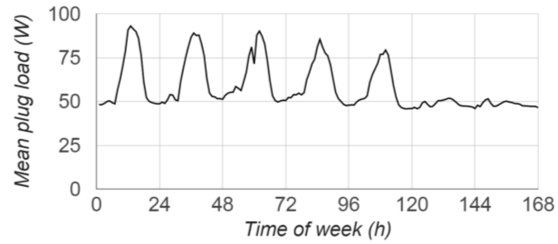

Figure 3: Mean weekly plug load schedule per occupant (built using output of plug2sch function).

Unlike occupancy, lighting, and plug loads, the blind positions do not exhibit hourly variations. Once moved, it may take days or even weeks before the blinds are adjusted again (O'Brien et al. 2013). Therefore, a schedule for the mean blind occlusion rate can be developed at a monthly time resolution. The BlindOccWrtMonths function inputs the blind position time-series from each office and outputs the mean monthly blind occlusion rate. For the offices we studied, this is shown in Figure 4. One can incorporate these results in whole-building energy and daylight simulation.

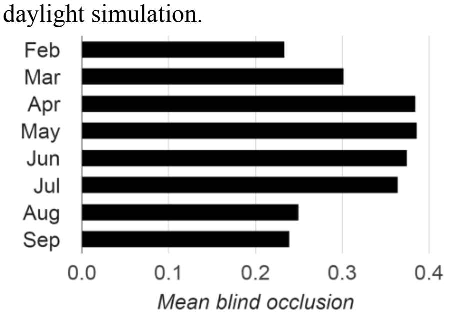

Figure 4: The mean blind occlusion rate as a function of the months (built using output of BlindOccWrtMonths).

The MarkovArrDpt is a function that calculates the likelihood of observing a weekday arrival event in the next hour when an office is unoccupied and a weekday departure event in the next hour when an office is occupied. This type of information can be used to execute midday temperature setback decisions. For example, based on the results shown Figure 5, if it is 5pm and an occupant has not arrived yet, it is very unlikely that s/he will show up for the rest of the day and a temperature setback can be initiated. Having this empirical likelihood distribution developed upon mere PIR sensor detections, one can program this control logic in a commercial building controller.

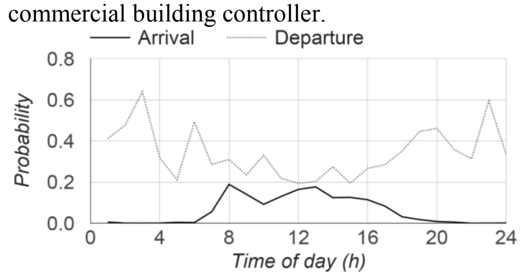

Figure 5: Discrete-time Markov occupancy model predicting the likelihood of arrival (if the space is unoccupied) and departure (if the space is occupied) in the next one hour (built using MarkovArrDpt function).

The occ2durpre and occ2durabs functions build survival models predicting the duration of intermediate occupancy and vacancy intervals. This type of information can be used in developing optimal control sequences for zone level equipment. For example, based on the results shown in Figure 6, if it is more than 3 hours since the occupant took a break, it is unlikely that s/he will show up for the rest of the day and a temperature setback can be initiated. Similarly, one can program this simple control logic by having these models for individual thermal zones.

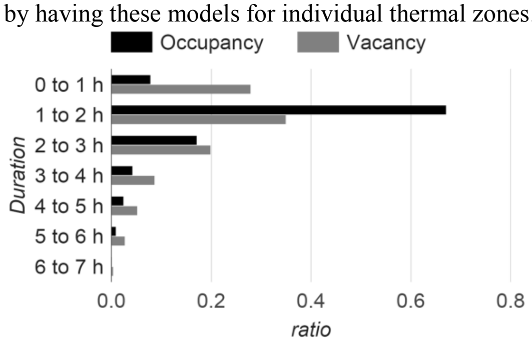

Figure 6: Survival models predicting the duration of occupancy and vacancy periods (built using outputs of occ2durpre and occ2durabs functions).

The occ2light function predicts the lighting load ratio by looking at the mean weekly occupancy schedule. The equivalent of this function for plug loads is occ2plug. This type of information can be useful to visualize the relationship between lighting/plug load and occupancy rate. For example, Figure 7 shows that the lights were on 56% of the occupied time for the example dataset we employed in this study. Figure 8 suggests that an average occupant's plug-in equipment draws about 170 W when present and 50 W when absent. Also, arguably when developing a building energy model in absence of lighting/plug use data, occupancy information may be used to estimate reasonable lighting/plug load schedules.

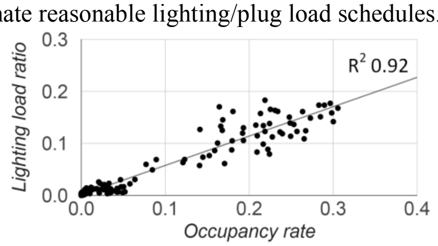

Figure 7: The relationship between occupancy and lighting load ratio (built using output of occ2light).

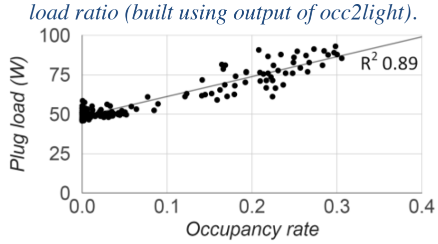

Figure 8: The relationship between occupancy and mean plug load per occupant (built using output of occ2plug).

The LightWrtSolRad function builds a Bernoulli lighting model with the predictor solar radiation on the façade. It predicts the probability of finding the lights on during occupied hours at a certain solar irradiance level. The function fits a logistic regression model between the response variable (ratio of the lights on during occupied hours) and the predictor (solar irradiance) and reports a detailed list of metrics to assess the appropriateness of the model (i.e., pseudo and ordinary R^2, Akaike and Bayesian information criterion, log-likelihood and deviance). The information from this analysis can be used to evaluate occupants' consciousness to daylight and it can be utilized to justify investments towards daylight-integrated lighting controls. For example, for the perimeter office spaces studied in this paper, the lights were on 70% of the occupied time when there was no or very limited daylight available, and they were still on 45% of the occupied time when the solar irradiance on the façade exceeded 600 W/m^2 (see Figure 9). This can be interpreted as the occupants do not actively use their lights and blinds to exploit the daylighting potential. In deep perimeter offices, this could be interpreted as daylight does not adequately reach indoors.

The equivalent of LightWrtSolRad function for blinds is BlindWrtSolRad. It predicts the blind occlusion rate during occupied hours as a function of the solar irradiance on the façade. For the perimeter office spaces we studied in this paper, the mean blind occlusion rate did not exhibit a significant relationship with the solar irradiance (see Figure 10). This confirms that occupants do not actively reposition their blinds to benefit from daylight at low solar irradiance levels.

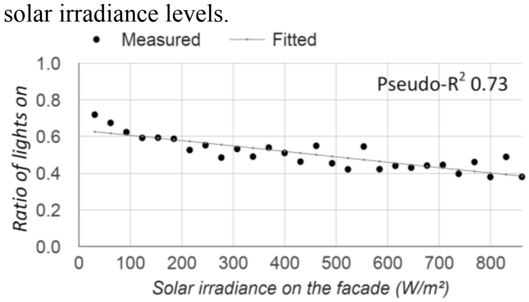

Figure 9: Bernoulli light use model (built using output of LightWrtSolRad).

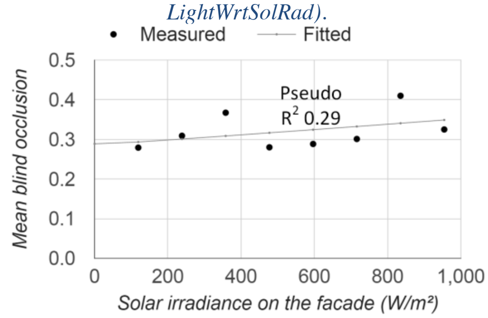

Figure 10: Bernoulli blind occlusion model (built using output of BlindWrtSolRad).

The function LightSwitchWrtSolRad builds a Markov light switch-on model upon occupancy, lighting, and solar irradiance time-series data. The model predicts the likelihood of a light switch-on action in the next timestep by looking at the solar irradiance levels. This analysis can reveal ambient solar irradiance levels triggering light switch-on actions. Results shown in Figure 11 indicate that for the studied offices the likelihood of a light switch-on action below 100 W/m^2 was about four times larger than it was above 600 W/m^2. Combined with the analysis conducted using LightWrtSolRad (see Figure 9), we can interpret this as: the occupants tend to turn on their lights when it gets dark, but they typically do not turn them off when there is adequate daylight. Given this information, a simple controls logic can be implemented to switch off lights in these perimeter offices when the solar irradiance exceeds a certain level (e.g., 400 W/m^2). This would eliminate the electric lighting use beyond that level and encourage occupants to adjust their blinds to benefit from daylight.

The equivalent of LightSwitchWrtSolRad function for blinds is BlindClosingWrtSolRad. The function inputs solar irradiance intensity on the façade during occupied hours when the shade was at least half-open to predict the likelihood of closing them in the next timestep. Analogous to the lighting use behaviour, occupants tend to adjust their blinds when it gets bright (see Figure 12), but they infrequently adjust them again and consequently the blinds remain in that position regardless of the solar irradiance (see Figure 10).

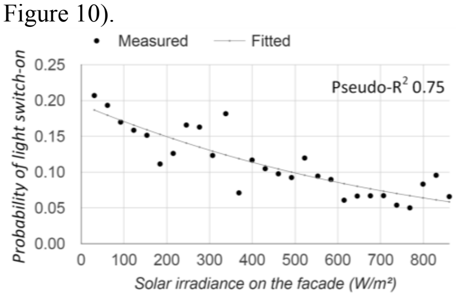

Figure 11: Discrete-time Markov light switch-on model using solar irradiance as the predictor (built using output of LightSwitchWrtSolRad) (timestep size 15 min).

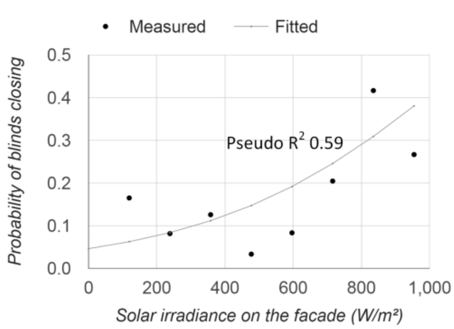

Figure 12: Discrete-time Markov blinds closing model using solar irradiance as the predictor (built using output of BlindClosingWrtSolRad) (timestep size 30 min).

The function LightSwitchWrtIndLux is identical with the LightSwitchWrtSolRad with the exception that it employs indoor illuminance measurements from individual offices – rather than the ambient solar irradiance measurements. As it takes into account the indoor nuances such as the interior blind position, it is more suitable to make decisions for lighting automation. For the studied offices, the probability that an occupant turns on the lights when the illuminance on the ceiling exceeds 150 lux was very small (see Figure 13). By having Markov light switch-on models for individual spaces, one can implement a controls logic that automatically turns off the lights beyond appropriate indoor light intensities.

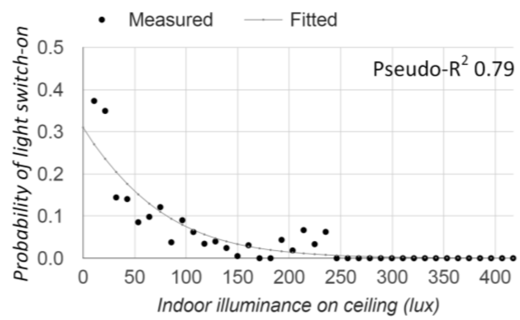

Figure 13: Discrete-time Markov light switch-on model using indoor illuminance as the predictor (built using output of LightSwitchWrtIndLux) (timestep size 15 min).

The LightSwitchWrtIndLuxArr function builds a discrete-event Markov model that predicts the likelihood of observing a light switch-on action at arrival (both intermediate arrivals and first arrival of the day) and the LightSwitchWrtIndLuxInt builds a discrete-time Markov model that predicts the likelihood of observing a light switch-on action in the next timestep during intermediate occupancy. This analysis reveals whether or not occupants' presence state plays a role in their activeness in using light switches. For the offices we studied, occupants tend to undertake light switch-on actions as they arrive to their offices (see Figure 12). This underlines the importance of the accessibility of control interfaces to the seated occupants. Also, if the model is intended to be used within BPS, the discrete-event model form will alleviate the simulationists from choosing a fixed timestep – which was a limitation of the discrete-time Markov models. However, the quality of the occupancy data becomes more sensitive with this model form as we need to predict the intermediate vacancy periods accurately.

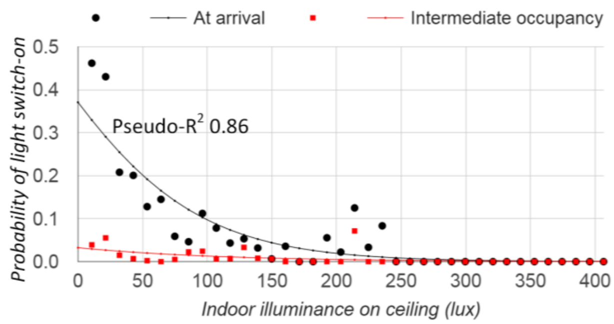

Figure 14: Discrete-event Markov light switch-on models (built from LightSwitchWrtIndLuxArr and LightSwitchWrtIndLuxInt). The scatter points represent likelihood weights and straight lines are the logistic regression fit.

The function PlugWrtAbsDur investigates how mean plug load values change at different occupancy periods. It inputs the plug load and occupancy time-series data from individual offices. It outputs the mean plug load per occupant at five different periods: (a) during occupancy, (b) during intermediate breaks (absence periods less than 12 h), (c) after-hours on weekdays (absence periods between 12 and 24 h), (d) on weekends (absence periods between 24 and 72 h), and (e) on vacations (absence periods longer than 3 days). Figure 13 illustrates that the mean plug load at different periods for the studied offices. Results indicate that the plug-in office equipment load during intermediate breaks was almost as high as it was during occupancy. Even during absence periods longer than three days, the mean plug load remained more than 50% of the mean plug load during occupancy periods. This analysis can be used to promote energy awareness campaigns and to justify the investments such as programmable wall plugs. We can use the plug load models from individual offices – in tandem with an occupancy model – as inputs to model-based predictive control algorithms or with BPS models.

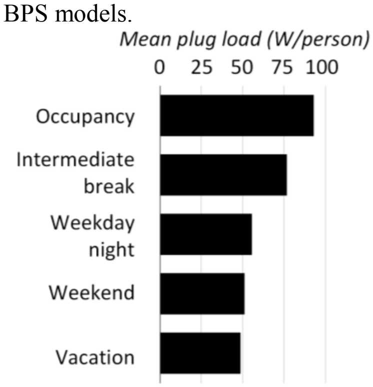

Figure 15: The mean plug load at different occupancy periods (built using output of PlugWrtAbsDur).

The function TstatUseWrtIat groups the indoor temperature measurements at time-instances occupants increase or decrease the temperature setpoints. This type of analysis can reveal the indoor temperatures typically disliked by the occupants. In addition, by conducting this analysis in individual offices the operators can choose more personal indoor temperature setpoints – which inherently takes into account the differences in sensors' calibration and positioning and individual preferences. For the studied offices, the median indoor temperature at the setpoint increase instances was 19.8°C, and it was 24.7°C at the setpoint decrease instances.

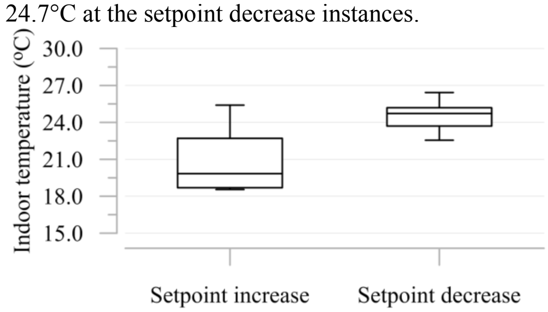

Figure 16: The indoor temperature distribution at the thermostat setpoint increase and decrease instances (built using output of TstatUseWrtIat).

The MarkovTstat function is a discrete-time Markov thermostat use model. It predicts the likelihood of observing a setpoint increase or decrease event as a function of the indoor temperature. The model learns from occupants' behaviours in two different ways: (1) when occupants accept the indoor conditions and do not change the temperature setpoint and (2) when occupants change the temperature setpoint. Upon these observations, the model develops discrete empirical likelihoods representing the ratio of number of setpoint changes registered to the number of timesteps spent at a certain temperature. Then, the function fits logistic regression models to these discrete likelihood distributions. For the studied offices, Figure 17 presents these models predicting the likelihood of a setpoint change in the next 30 minutes. This type of analysis can be useful in selecting temperature setpoints and the models generated by the function can be used within BPS models to represent manual thermostat control.

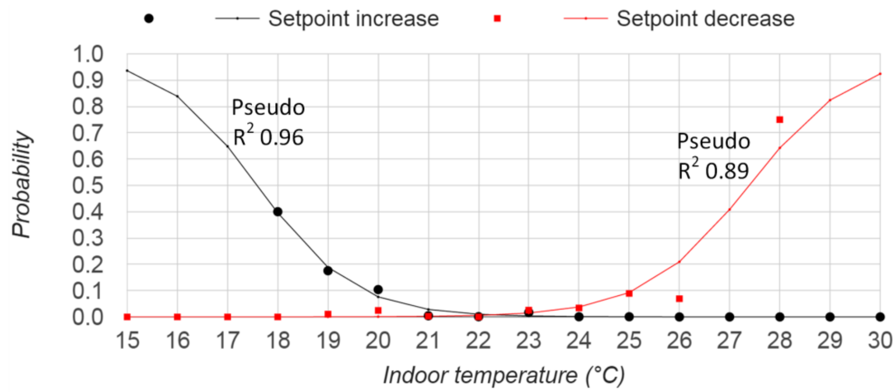

Figure 17: Discrete-time Markov thermostat use models (built using output of MarkovTstat). The scatter points represent discrete likelihood weights and the straight lines are the logistic regression fit.

CONCLUSIONS AND FUTURE WORK

In an effort to facilitate the occupant behaviour and presence model development process, a starter toolkit was prepared. The toolkit contains 22 Matlab functions that can be used in conducting statistical analyses on occupancy and lighting, blinds, thermostat, and plug-in equipment use patterns. The functionalities of the toolkit are presented through an analysis conducted upon the data gathered in an academic office building in Ottawa, Canada. The functions and the example data files are also made publicly available with this paper (Supplement 2016). Future work is planned to expand the functions library in collaboration with other Annex 66 researchers, demonstrate the importance of these occupant-modelling exercises in BPS-based design, and move the functions to other programming environments such as Python and R-programming.

ACKNOWLEDGEMENTS

The research presented in this paper benefited from the authors' participation in the ongoing efforts of the IEA-EBC Annex 66 (Definition and Simulation of Occupant Behavior in Buildings) and the associated discussions. Financial support from Delta Controls, Regulvar, Natural Sciences and Engineering Research Council of Canada (NSERC), Natural Resources Canada (NRCan) through research contracts is gratefully acknowledged.

REFERENCES

ASHRAE, A. (2004), "Standard 55-2004, Thermal Environmental Conditions for Human Occupancy," *American Society of Heating, Refrigerating and Air-Conditioning Engineering, Atlanta, GA*.

CIBSE, G. F. (2012), "Energy Efficiency in Buildings," *Chartered Institution of Building Services Engineers*.

Guillemin, A., and Molteni, S. (2002), "An Energy-Efficient Controller for Shading Devices Self-Adapting to the User Wishes," *Building and Environment*, 37, 1091-1097.

Gunay, H. B., O'Brien, W., and Beausoleil-Morrison, I. (2013), "A Critical Review of Observation Studies, Modeling, and Simulation of Adaptive Occupant Behaviors in Offices," *Building and Environment*, 70, 31-47.

Gunay, H. B., O'Brien, W., and Beausoleil-Morrison, I. (2015a), "Development of an Occupancy Learning Algorithm for Terminal Heating and Cooling Units," *Building and Environment*, 93, Part 2, 71-85.

Gunay, H. B., O'Brien, W., and Beausoleil-Morrison, I. (2015b), "Implementation and Comparison of Existing Occupant Behaviour Models in Energyplus," *Journal of Building Performance Simulation*, 1-46.

Gunay, H. B., O'Brien, W., Beausoleil-Morrison, I., and Bursill, J. (2016a), "Implementation of an Adaptive Occupancy and Building Learning Temperature Setback Algorithm," *Ashrae Transactions*, 122.

Gunay, H. B., O'Brien, W., Beausoleil-Morrison, I., Goldstein, R., Breslav, S., and Khan, A. (2014), "Coupling Stochastic Occupant Models to Building Performance Simulation Using the Discrete Event System Specification Formalism," *Journal of Building Performance Simulation*, 7, 457-478.

Gunay, H. B., O'Brien, W., Beausoleil-Morrison, I., and Gilani, S. (2016b), "Modelling Plug-in Equipment Load Patterns in Private Office Spaces," *Energy and Buildings*.

Haldi, F., and Robinson, D. (2008), "On the Behaviour and Adaptation of Office Occupants," *Building and Environment*, 43, 2163-2177.

Haldi, F., and Robinson, D. (2009), "Interactions with Window Openings by Office Occupants," *Building and Environment*, 44, 2378-2395.

Haldi, F., and Robinson, D. (2011), "The Impact of Occupants' Behaviour on Building Energy Demand," *Journal of Building Performance Simulation*, 4, 323-338.

Hoes, P., Hensen, J. L. M., Loomans, M. G. L. C., de Vries, B., and Bourgeois, D. (2009), "User Behavior in Whole Building Simulation," *Energy and Buildings*, 41, 295-302.

IES. (2015), "The Lighting Handbook 10th Edition."

Mahdavi, A., Mohammadi, A., Kabir, E., and Lambeva, L. (2008), "Occupants' Operation of Lighting and Shading Systems in Office Buildings," *Journal of Building Performance Simulation*, 1, 57-65.

Mahdavi, A., and Tahmasebi, F. (2015), "Predicting People's Presence in Buildings: An Empirically Based Model Performance Analysis," *Energy and Buildings*, 86, 349-355.

Menezes, A. C., Cripps, A., Buswell, R. A., and Bouchlaghem, D. (2012), "Benchmarking Small Power Energy Consumption in the United Kingdom: A Review of Data Published in Cibse Guide F," *Building Services Engineering Research and Technology*, 0143624412465092.

Menezes, A. C. K. d. (2013), "*Improving Predictions of Operational Energy Performance through Better Estimates of Small Power Consumption*," © Anna Carolina Kossmann de Menezes.

Nagy, Z., Yong, F. Y., Frei, M., and Schlueter, A. (2015), "Occupant Centered Lighting Control for Comfort and Energy Efficient Building Operation," *Energy and Buildings*, 94, 100-108.

Nicol, J. F., and Humphreys, M. A. (2004), "A Stochastic Approach to Thermal Comfort--Occupant Behavior and Energy Use in Buildings," *ASHRAE transactions*, 110.

O'Brien, W., Kapsis, K., and Athienitis, A. K. (2013), "Manually-Operated Window Shade Patterns in Office Buildings: A Critical Review," *Building and Environment*, 60, 319-338.

O'Brien, W., and Gunay, H. B. (2015), "Mitigating Office Performance Uncertainty of Occupant Use of Window Blinds and Lighting Using Robust Design," *Building Simulation*, 8, 621-636.

Page, J. (2007), "*Simulating Occupant Presence and Behaviour in Buildings*," EPFL.

Page, J., Robinson, D., Morel, N., and Scartezzini, J. L. (2008), "A Generalised Stochastic Model for the Simulation of Occupant Presence," *Energy and Buildings*, 40, 83-98.

Parys, W., Saelens, D., and Hens, H. (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, 339-358.

Pigg, S., Eilers, M., and Reed, J. (1996), "Behavioral Aspects of Lighting and Occupancy Sensors in Private Offices: A Case Study of a University Office Building," *ACEEE 1996 Summer Study on Energy Efficiency in Buildings*.

Reinhart, C. F. (2004), "Lightswitch-2002: A Model for Manual and Automated Control of Electric Lighting and Blinds," *Solar Energy*, 77, 15-28.

Rijal, H. B., Tuohy, P., Nicol, F., Humphreys, M. A., Samuel, A., and Clarke, J. (2008), "Development of an Adaptive Window-Opening Algorithm to Predict the Thermal Comfort, Energy Use and Overheating in Buildings," *Journal of Building Performance Simulation*, 1, 17-30.

Sanati, L., and Utzinger, M. (2013), "The Effect of Window Shading Design on Occupant Use of Blinds and Electric Lighting," *Building and Environment*, 64, 67-76.

Supplement. (2016), "Functions and Interactive Figures Used in the Toolkit for Developing Data-Driven Occupant Behaviour and Presence Models," https://drive.google.com/folderview?id=0B6FnY3Z5MEfeRjBUUDFCS1ZYUHc&usp=sharing.

Wang, C., Yan, D., and Jiang, Y. (2011), "A Novel Approach for Building Occupancy Simulation," *Building Simulation*, 4, 149-167.

Wang, D., Federspiel, C. C., and Rubinstein, F. (2005), "Modeling Occupancy in Single Person Offices," *Energy and Buildings*, 37, 121-126.