Available online at www.sciencedirect.com
ScienceDirect
Energy and Buildings xxx (2007) xxx–xxx
ENERGY and BUILDINGS
www.elsevier.com/locate/enbuild

A generalised stochastic model for the simulation of occupant presence

J. Page *, D. Robinson, N. Morel, J.-L. Scartezzini

LESO-PB, Bâtiment LE, Station 18, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

Received 25 September 2006; received in revised form 15 January 2007; accepted 22 January 2007

Abstract

This paper describes an algorithm for the simulation of occupant presence, to be later used as an input for future occupant behaviour models within building simulation tools. By considering occupant presence as an inhomogeneous Markov chain interrupted by occasional periods of long absence, the model generates a time series of the state of presence (absent or present) of each occupant of a zone, for each zone of any number of buildings. Tested on occupancy data from private offices, the model has proven its capacity to realistically reproduce key properties of occupant presence such as times of arrival and departure, periods of intermediate absence and presence as well as periods of long absence from the zone. This model (due to related metabolic heat gains), and associated behavioural models which use occupants' presence as an input, have direct consequences for building energy consumption.

Keywords: Energy simulation; Occupants' presence model; Stochastic process; Markov chain

1. Introduction

A large number of simulation tools are available for the estimation of a building's energy consumption (such as ESP-r or ENERGY+) and very recently new tools (proposed by [1,2]) have shown their capacity to simulate urban districts comprising buildings of various sizes and uses. The latter tools emphasize the importance of implementing energy saving measures at the equipment and building levels and reveal the potentiality of district-level energy generation and distribution. These tools may produce yearly profiles of the energy needed by a building, for heating, cooling, lighting and ventilation as well as for hot water and electrical appliances. Such tools may also be used to size the associated heating, ventilation and air-conditioning plants (such as separate boilers, combined heat and power plants, district heating systems, photo-voltaic or solar thermal panels, etc.). An important issue in sizing these production systems is to estimate peaks in demand that will have to be met. It is therefore necessary to reliably estimate the fluctuations in energy consumption. Stochastic variables linked to the climate and to the behaviour of the buildings' occupants have important influences on those fluctuations. While high quality measured climate data is available at reasonably fine temporal resolution, corresponding data linked to the occupant in all its stochastic variety is scarce. However the modeling of occupant behaviour is starting to bloom.

The influence of occupants on the building they occupy can be broken down into several means of interaction (as discussed by [3]), each of which can be represented by a stochastic model as shown in Fig. 1. Being present within the building is clearly a necessary condition for being able to interact with it. Occupant presence is therefore an input to all other models and the model for occupant presence will be central to the family of other stochastic models [4]. As each human being emits heat and ''pollutants'' (such as water vapor, carbon dioxide, odours, etc.), its presence directly modifies the indoor environment. Occupants also interact with a building to enhance their personal comfort. For example they will heat, cool or ventilate their environment to improve their thermal comfort, they will adjust lighting systems or blinds to optimize their visual comfort. Finally occupants' interactions also relate to the tasks that they are accommodated to perform: in an office building occupants may use diverse electrical appliances tending to internal heat gains and the consumption of electricity. In residential buildings, household appliances can consume water (hot and cold) as well as electricity; occupants may also produce waste. A model capable of reproducing patterns of presence of occupants in a building is therefore of paramount

* Corresponding author. Tel.: +41 21 6935557; fax: +41 21 6932722.
E-mail address: jessen.page@epfl.ch (J. Page).

0378-7788/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.enbuild.2007.01.018

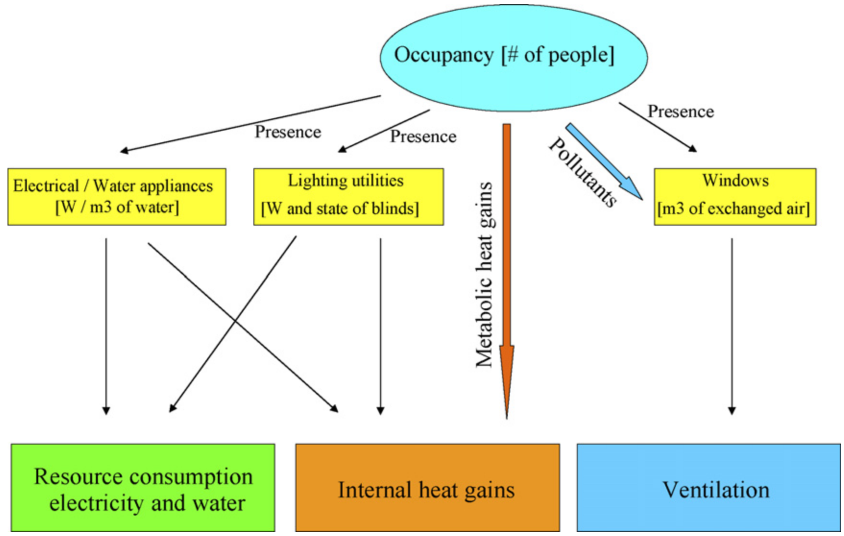

Fig. 1. Outputs of the occupancy model and their later use by stochastic models of occupants' behaviour.

importance in simulating the behaviour of occupants within a building and their effects on the buildings' demands for resources such as energy (in the form of heat, cold and electricity) or water as well as the production of waste (which may be later used to derive energy). Models for the simulation of occupant's interactions with windows ([5] and more recently [6]), lighting [7] and air-conditioning [8] have been developed and these can directly benefit from a realistic input of occupant presence. Interfaces to link models of occupant behaviour to various dynamic building simulation tools have also been developed (such as SHOCC explained in [9]). When applied to a single building simulation tool these stochastic models will help to provide information on the distribution of the demand in energy and therefore on how the production plants should be sized. When applied to a whole neighbourhood (this is possible with the software tool SUNtool [2]), several energy supply scenarios can be tested in order to choose the right mix, size and lay-out of different technologies (combined heat and power, district heating and cooling, solar or wind power, etc.).

Currently the most common way of considering occupant presence within simulation tools is by using so-called ''diversity profiles'' [10]. These are used in order to accurately estimate the impact of internal heat gains (from people, office equipment and lighting) on energy and cooling load calculations of one building. The profiles may depend on the type of building (typical categories being ''residential'' and ''commercial'') and sometimes on the type of occupants (size and composition of a family for example). Weekdays and weekends[1] are usually handled differently, especially in the case of commercial buildings. A daily profile (either for a weekday or a weekend) is composed of 24 hourly values; each of these corresponds to a fraction of a given peak load. The weekday and weekend profiles and the peak load are related to a particular category of building and type of heat gain (metabolic heat gain, receptacle load, lighting load); they may be based on data collected on a large amount of monitored buildings. Alternatively the user of the simulation tool can also enter profiles that (s)he deems best fit for the building in question. An annual load profile for each type of heat gain is constructed by repeating the weekday and weekend daily profiles and multiplying them by the peak load. The weakness of this method lies in the repetition of one or possibly two profiles and the fact that the resulting profile represents the behaviour of all the occupants of a building. The latter simplification reduces the variety of patterns of occupancy particular to each person by replacing it with an averaged behaviour. The former simplification neglects the temporal variations, such as seasonal habits, differences in behaviour between weekdays (that appear in monitored data) and atypical behaviours (early departures from the zone, weeks of intense presence and of total absence, unpredicted presence on weekends in the case of office buildings—events that all appear in monitored data).

The use of a lighting appliance, and the corresponding implications for electrical energy use, is obviously linked to the presence of its user. It is therefore of little surprise that researchers developing lighting models have been the most eager to account for the randomness of occupant presence in the most efficient way. Hunt was the first to emphasize the importance of occupant interaction with lighting appliances [11]. His work has been incorporated into simulation tools such as ESP-r or SER-Res [12]. Later on Newsham [13] and Reinhart [7] introduced a simple stochastic model of occupant presence in their work on the Lightswitch model. They were interested in reproducing more realistic times of arrival and departure of occupants to and from their offices and modified the standard profiles mentioned above to this end. Their simulated occupancy profile corresponds to working hours from 8:00 to 18:00 with a 1 h lunch break at noon and two 15 min coffee breaks in the morning at 10:00 h and in the afternoon at 15:00 h (that the occupant takes with a 50% probability). To this they added the following:

[1] Holidays are usually considered as weekends.

"All arrivals in the morning, departures in the evening and breaks are randomly scheduled in a time interval of ±15 min around their official starting time to add realism to the model". [7]

This enables them to replace the unrealistic peaks mentioned above by a more natural spread around a fixed average. Although this represents a certain progress towards a realistic simulation of occupant presence the fact that the major portion of the profile is fixed (presence of 100% during most of the working hours, presence of 0% from 18:15 to 7:45 h, repetition of the same profile for all weekdays and the assumption that the zone is unoccupied during weekends) prevents the model from reproducing the variety both in behaviours and over time of occupant presence. One important aspect of this restriction is the lack of periods of long absence (corresponding to business trips, leaves due to sickness, holidays, etc.) leading to an overestimation of the total yearly presence and associated energy consumption, as recognized by the authors. The appearance of occupants on weekends, their arrival before 7:45 h and departure after 18:15 h are phenomena that are common to the real world but are omitted by the model. Finally the absence of occupants outside of breaks is also an event that it fails to simulate.

This last deficiency was studied by Wang [14]. She examined the statistical properties of occupancy in single person offices. Based on her observations she made the hypothesis that the duration of periods of intermediate presence and absence (i.e. taking place between the first arrival of the occupant to the office and her/his last departure from the office) are exponentially distributed and that the coefficient of the exponential distribution for a single office could be treated as a constant over the day. She was able to validate these hypotheses in the case of absence but not in that of presence. To generate a simulated pattern of presence in an office she estimated the two coefficients, supposed to be constant, of the exponential distributions and generated a sequence of alternating periods of presence and absence. In addition she generated the first arrival to the office, the last departure from the office and a lunchtime break based on the assumption that these are distributed normally as Reinhart had before her. The combination of the created profiles gave her a simulated time series of presence that would vary from day to day. The model proposed is a simple and elegant one, yet it still fails to reproduce the complexity of real occupant presence. As the authors acknowledge themselves, periods of presence cannot be reproduced by an exponential distribution with a homogeneous coefficient, and times of arrival, of departure as well as absences during lunch breaks are not normally distributed. Like all its predecessors the model supposes that all weekdays are alike and that offices are always unoccupied during weekends. Periods of long absence are also neglected so that total presence is once again overestimated.

The latest model of occupant presence was proposed by Yamaguchi [15] in the development of a district energy system simulation model. Their aim was to simulate the "working states" (that they defined as using 1 PC, using 2 PC's, not using a PC and being out) of each occupant of a group of commercial buildings in order to derive the heat and electrical load generated from the use of energy consuming appliances. These stochastic loads combined with those resulting from non-occupant related appliances of the buildings[2] determine the electricity, heating and cooling loads to be met by a suitable energy supply system. Their model is similar to that of Wang in that it supposes that the time an occupant will spend in a working state is independent of time (in Wang's model this corresponds to the coefficients of the two exponential distributions). However it replaces the sequence of Poissonian periods of absence and presence by a mathematically equivalent (but computationally more elegant) Markov chain of working states. The transition probabilities of the Markov matrix are determined by inputs to the model and the working state of the occupant is drawn every 5 min using the "inverse function method". Moreover the times of arrival, lunch break and departure are now drawn given by empirical distributions rather than a normal distribution centered around fixed values. It is not clear in their explanation of the model in [15] whether it is being used to simulate only one repeated day of occupant activity or each and every day of the year. In the latter case it is also unclear whether weekends are treated differently or whether periods of long absence are considered; we suppose that this is not the case. The calculation of an occupant schedule for only one day, if this is the case, would be restrictive as we have argued above and the lack of long periods of absence when simulating a whole year would likewise be erroneous as we shall explain below. The hypothesis that the duration of time an occupant spends in a given working state does not depend on time (i.e. the time of day) is one that Wang proved to be wrong in at least the case of presence; dividing the state of presence into different states of working activity during presence will most probably not change that. Although their model could prove useful when only considering the use of PC's, the hypothesis of time-independence shall cause difficulties when wanting to simulate less invariable activities such as the use of lighting appliances or activities performed in residential buildings for example. We believe the coupling of an independent occupant presence model to related behavioural models to be a more general solution.

In line with this last statement, we propose in the following pages an alternative model for the simulation of occupant presence. By using a profile of probability of presence, rather than an adjusted fixed profile, as an input to a Markov chain we are able to produce intermediate periods of presence and absence distributed exponentially with a time-dependent coefficient as well as the fluctuations of arrivals, departures and typical breaks. A failed attempt to validate an earlier version of the model highlighted the importance of periods of long absence; these were included in the updated and latest version presented here. Twenty zones of an office building were monitored providing us with 2 years of data that was used for

[2] It should be noted that the loads resulting from the use of lighting and appliances used by groups of occupants are calculated based on fixed schedules.

the calibration and validation of the model. The latter was based on the analysis of indicators of importance for the stochastic models of occupant behaviour that will use the results of the model of presence as their input. Although tested with data from an office building, this model, when given the corresponding inputs, is applicable to any type of building and any pattern of occupant presence.

2. Methods

2.1. Model development

2.1.1. Aims

It is important to know what properties need to be reproduced by the model as they shall serve as a guideline for its development and as indicators to be checked during its validation. The model of occupancy is destined to deliver the metabolic heat gains and pollutants released by the occupants within the zone and to serve as an input for the use of windows, lighting appliances and other electrical and water appliances (see Fig. 1). To serve this purpose it needs to reproduce in the most reliable way properties of patterns of occupancy such as the first arrival and last departure of the occupant, the duration of the periods of intermediate presence and absence as well as of long absence and the time of intermediate arrival for each and every occupant.

Patterns of occupancy are so diverse and complex that we decided that the simplest way to develop a model capable of doing this was to build it from a priori hypotheses and check later whether the above properties are reproduced within reason.

2.1.2. Hypotheses

We are interested in simulating the presence of occupants within a specific ‘‘zone’’ of a building. This corresponds to the area occupied by a household in the case of residential buildings (typically a flat) and to a (single or multiple person) office in that of office buildings. We are not interested in simulating the movement of occupants from one zone to the other (a model for this has been proposed by [16]), but simply whether each occupant is present within the zone or not.

The hypothesis of independence allows us to model in a simple way the patterns of presence of each occupant individually. The presence of occupants sharing the same zone can then be simulated by:

(1) multiplying the obtained pattern by the total number of occupants (this case of collective behaviour would correspond to the occupancy of a meeting-room),
(2) or by simulating each occupant separately and then adding the produced patterns of presence.

We make the hypothesis that the probability of presence at a time step only depends on the state of presence at the previous time step. In other words the probability that an occupant is present now only depends on whether (s)he was present one time step ago and not on whether (s)he has been present over the past N time steps. Mathematically this statement corresponds to asserting the following property on the conditional probability:

$$P(X_{t+1} = i | X_t = j, X_{t-1} = k, \ldots, X_{t-N} = l) = P(X_{t+1} = i | X_t = j) =: T_{ij}(t) \quad (1)$$

with X_t being the random variable ‘‘state of presence at time step t’’ and i, j, k and l taking on values 0 or 1. This corresponds to considering the state of occupancy as a Markov chain with probabilities of transition $T_{ij}(t)$ (for more details on Markov Chains we direct the reader to [17]). The probability that an occupant should arrive at the office at 8:00 h or at 22:00 h are clearly not the same, therefore the values of $T_{ij}(t)$ need to be time dependent and we have the general case of an inhomogeneous Markov chain (with discrete states and discrete time steps).[3] In order to determine the time dependence of these probabilities of transition we will need the following inputs to the model: the profile of probability of presence over a typical week and a parameter of mobility that gives an idea of how much people move in and out of the zone.

2.1.3. Development

Based on our hypotheses we are looking for a model capable of generating a time series of zeros (absence) and ones (presence) that renders arrivals into and departures from the zone (typically going to work and coming from work for residential zones, arriving at work and leaving from work for office zones) as well as alternating short periods of presence and absence in between. It should not simply reproduce the pattern given as an input (the profile of probability of presence and the parameter of mobility) but create a pattern that never repeats itself while reproducing the statistics of the real world it is simulating.

To do this we have based the model on the ‘‘inverse function method’’ (IFM) that can generate a sample (in our case a time series) of events from a given probability distribution function (PDF) as shown in Fig. 2. Earlier we made the hypothesis that the value of occupancy at the next time step should only depend on the state we are in now and the probability of transition from this present state to either the same state (0 to 0; 1 to 1) or its opposite state (0 to 1; 1 to 0). These probabilities of transition T_{00}, T_{01}, T_{10}, T_{11} are therefore the PDF’s we need (in this case the values that the random variable can take are discrete). Only two of the four variables need to be known, let us say T_{01} and T_{11}, as T_{00} and T_{10} can be deduced from $T_{00} + T_{01} = 1$ and $T_{10} + T_{11} = 1$. As we have seen in previous models, the profile of probability of presence is a rather standard input for a simulation tool including occupancy and should be available to the user.[4] Having this as an input provides us with a relationship

[3] We will determine the initial state (at t_0 =00:00 h on 1 January) of the time series as ‘‘present’’ for residential buildings and ‘‘absent’’ for office buildings.
[4] In the case of users unfamiliar with such inputs, default profiles corresponding to types of occupants will have to be made available.

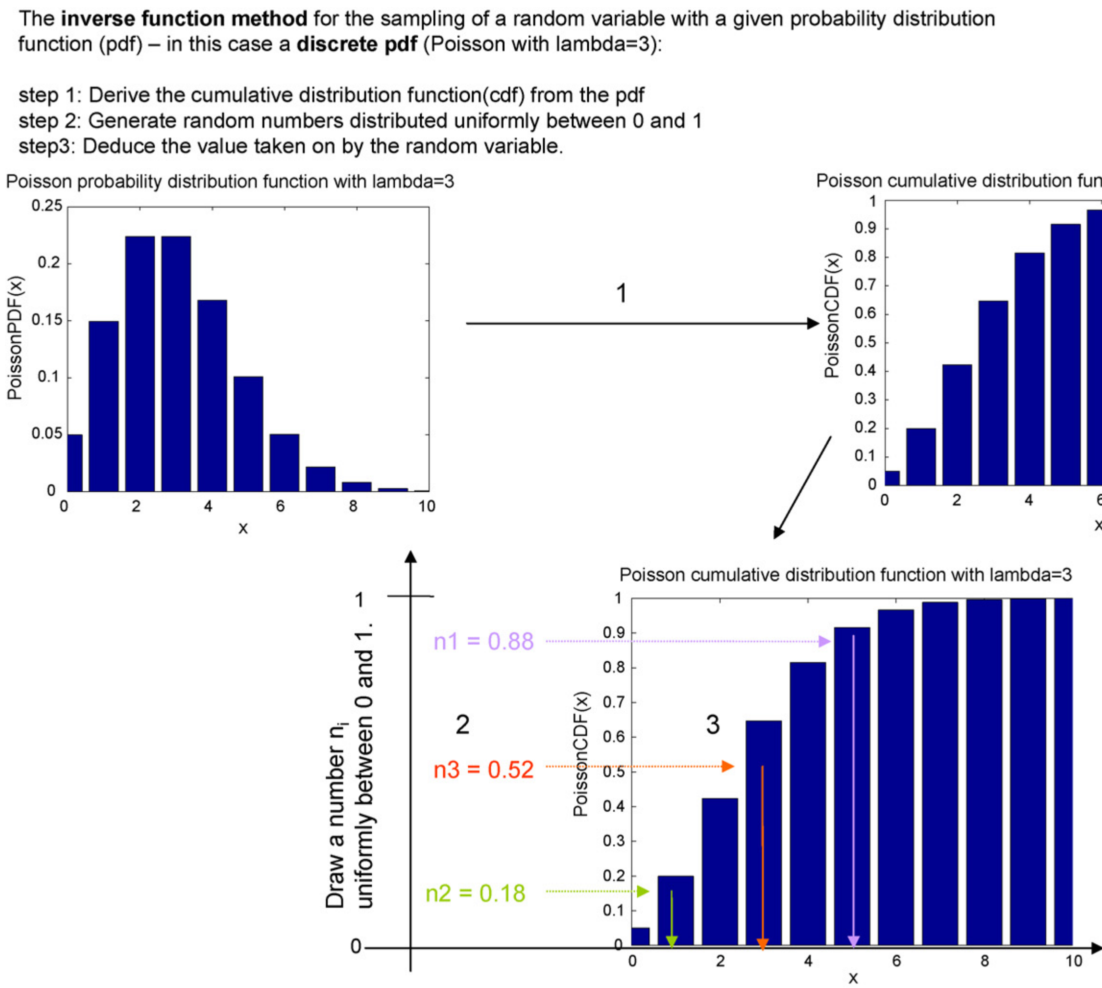

Fig. 2. Generation by the inverse function method of the series of values 5, 1, 3 based on a Poisson distribution with λ = 3.

for the probability $P(t+1)$ that the occupant is present at the time step $t+1$:

$$P(t+1) = P(t)T_{11}(t) + (1-P(t))T_{01}(t) \tag{2}$$

From this we can deduce that:

$$T_{11}(t) = \frac{P(t)-1}{P(t)}T_{01}(t) + \frac{P(t+1)}{P(t)} \tag{3}$$

However, we still lack one piece of information to be able to determine uniquely the value of T_{01} and T_{11} at all times. This further input to the model should make sense to the user who will be entering it. Keeping this in mind we defined the ‘‘parameter of mobility’’ as the ratio between the probability of change of the state of presence over that of no change:

$$\mu(t) := \frac{T_{01}(t) + T_{10}(t)}{T_{00}(t) + T_{11}(t)} \tag{4}$$

To simplify the inputs to the model we consider $\mu(t)$ to be constant and to assist the user of the model we have defined numerical values to levels of ‘‘low’’, ‘‘medium’’ and ‘‘high’’ mobility. Given relationships (3) and (4) and the inputs $P(t)$ and μ we should now have a complete profile of $T_{01}(t)$ and $T_{11}(t)$:

$$T_{01}(t) = \frac{\mu-1}{\mu+1}P(t) + P(t+1) \tag{5}$$

$$T_{11}(t) = \frac{P(t)-1}{P(t)}\left[\frac{\mu-1}{\mu+1}P(t) + P(t+1)\right] + \frac{P(t+1)}{P(t)} \tag{6}$$

Unfortunately not quite. For certain values of $P(t)$, $P(t+1)$ and constant μ, the condition $0 \leq T_{ij}(t) \leq 1$ can be violated. This typically happens when $P(t)$ is far greater or smaller than $P(t+1)$. This situation corresponds to an almost deterministic change in behaviour, such as a regular time of first arrival into the zone, a regular lunch break or a regular time of last departure from the zone, rather than the random movement into and out of the zone rendered by a constant value of μ. In these cases the model would fail to reproduce such clear changes in occupancy when using the initial value of μ; to counter this the model temporarily forces μ to take on the value that fulfills the above condition and is closest to the initial constant. For an initial fixed value of the state of presence at t_0 we are now able to generate a time series of the presence of an occupant within a given zone.

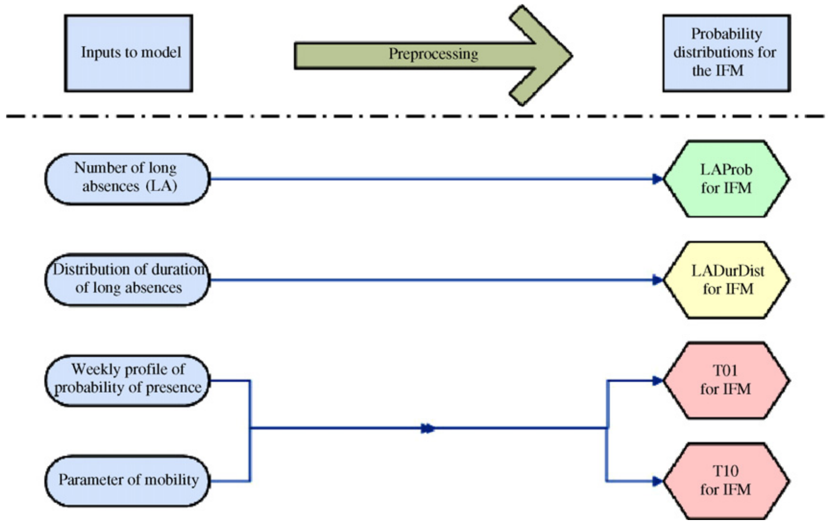

Fig. 3. Preprocessing stage: extraction from the inputs of the probability distributions needed for the inverse function method to be used in the processing stage.

This first version of the model was calibrated with data of occupancy recorded in the offices of the LESO building (for more details on the building see [18]) and a preliminary validation was made by comparing the cumulated presence over a week resulting from the original data and the simulations done with the model. The data contained a great variety of results ranging from weeks of total absence, corresponding to periods of leave due to sickness, work outside of the office or vacations, to weeks of high cumulated presence, corresponding to periods of overtime work and unusual presence over weekends. The model was only capable of producing a Gaussian distribution around the average of the empirical data. This showed that, although the Markov chain model works well at reproducing periods of short absence and presence during one day, it needs to be complemented in order for the model to generate long periods of absence. These have been included by adding to the algorithm the possibility to start, at random, a period of long absence at each time step.[5] To generate them we need to know the probability of them happening and the parameters that determine the distribution of their duration; these shall be new inputs to the model. For the validation of this improved version of the model of occupant presence the periods of long absence (lasting more than one day but not corresponding to a weekend) were extracted from the empirical data and treated to give the necessary inputs to the model. The remaining data was used to calibrate the Markov chain.

2.1.4. Algorithm

The model was implemented as a MATLAB script. The presence of each occupant in each zone was simulated independently based on the inputs related to that occupant. The profile of probability of presence and the parameter of mobility are used to determine the profile of T_{01} and T_{10} (see Fig. 3). The occupant is considered to be absent in the case of office buildings and present in that of residential buildings at t_0, i.e. 00:00 h of the 1 January. From then on the time series of presence is generated by using the IFM at each time step.

This method generally works by inverting the cumulated density function of the random variable of interest, drawing uniformly a number between 0 and 1 and using the inverted cumulated density function (CDF) to relate the drawn number to a value adopted by the random variable (see Fig. 2). In our case the CDF is a histogram of two bins (0 and 1) or more (in the case of the duration of periods of long absences). To each bin corresponds the probability that a value within that bin be chosen at random. The number drawn between 0 and 1 will determine which of the bins has been chosen, in other words what event will take place.

Fig. 4 shows how the algorithm works: given the probability of starting a period of long absence (derived from the number of long absences happening in a year, entered as an input) we first check whether the occupant starts a period of long absence or not by using the IFM, if so we determine the length of that absence given the distribution of the duration of periods of long absences (entered as an input) with the same method, during which period the occupant is considered to be absent. At her/his return, or if (s)he did not start a long period of absence, we find ourselves in the case of the Markov chain of ‘‘usual daily’’ changes in state of occupancy. The present state of occupancy will tell us which profile of probability of transition to choose between T_{01} and T_{10}; the next state of presence is determined by the use of the IFM. By doing so we are capable of generating a time series of the state of presence of a particular occupant in a particular zone. The state of presence of each occupant of one zone and the states of occupancy of different zones being considered independent it is enough to repeat this algorithm as many times as the number of total occupants, respecting, of course, the inputs particular to each occupant simulated.

[5] This means that periods of vacation will be distributed randomly over the year rather than attributed to fixed days of the year.

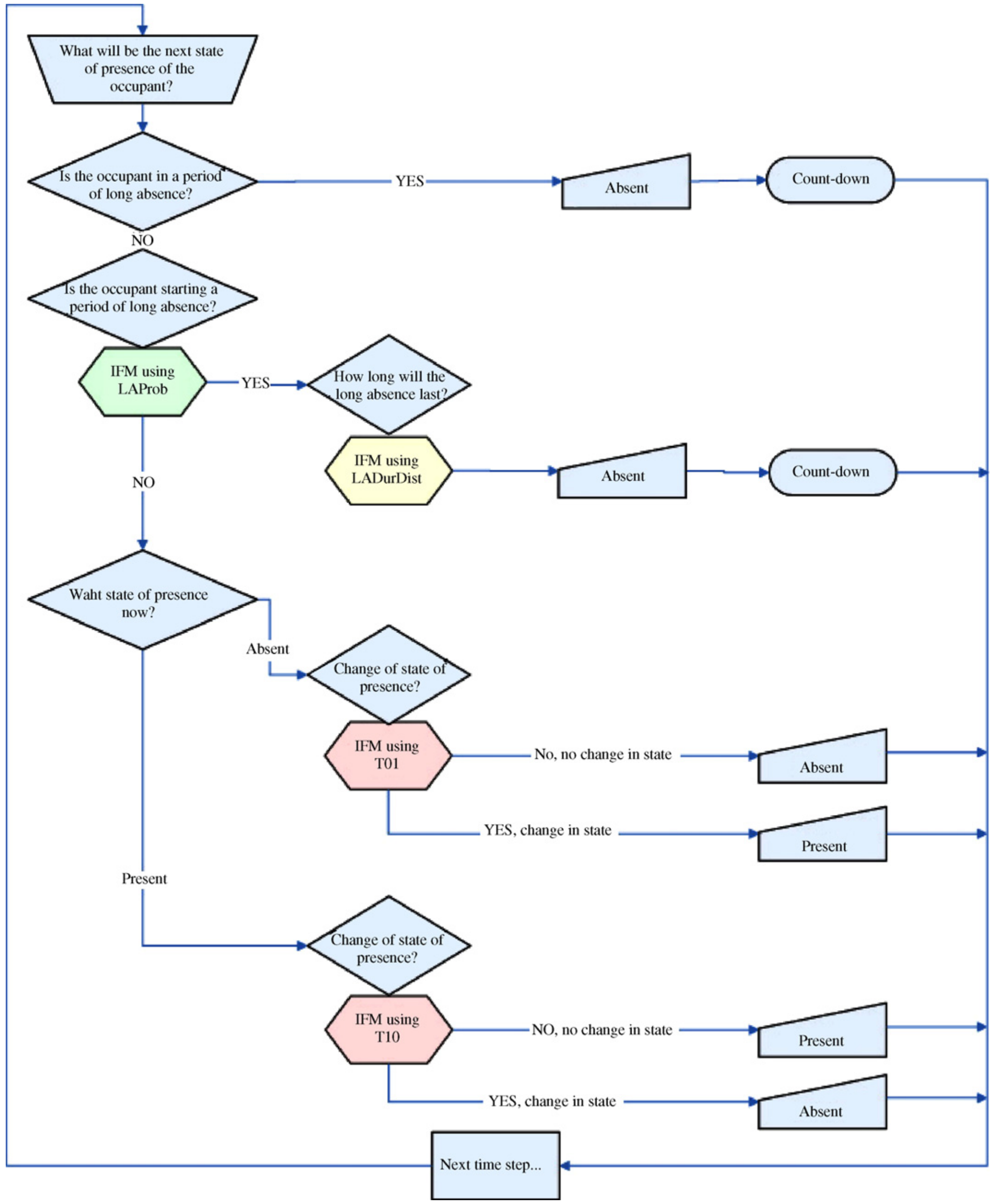

Fig. 4. Algorithm of the model (processing stage).

3. Results

3.1. Data collection

The data needed for the calibration and validation of the model was collected from mid December 2001 to the beginning of January 2006 in 20 ‘‘zones’’ of the LESO-PB building at the EPFL each equipped with a movement sensor. Of these, 10 zones were offices having seen their number of occupants vary over the period of monitoring and five zones had not been constantly used as offices (printer room, conference room, classroom and workshop). The remaining five zones which had been singly occupied offices over the whole period of data acquisition, were used for model calibration and validation. The people at the LESO work mainly on research, sometimes taking or giving courses. Occupants are very mobile often leaving their office to visit other zones of the same building, such as the workshop, the library or computer-room or offices of colleagues, or to leave the building. This may make the patterns of presence not particularly representative of an office building (and even less so of a residential building!). Nevertheless, this shall not weaken the validation of the model as it has been conceived to be independent of the characteristics of the occupants to be simulated. Indeed only the inputs to the model (profile of probability of presence, parameter of mobility, distribution of periods of long absence) are related to the simulated occupants; the model itself, given the right inputs, should be applicable to any type of building and any pattern of occupant presence.

Please cite this article in press as: J. Page et al., A generalised stochastic model for the simulation of occupant presence, Energy & Buildings (2007), doi:10.1016/j.enbuild.2007.01.018

3.2. Treating data for calibration

The acquired data needed to be processed before information for the calibration and validation of the model could be extracted. Problems either with the sensor, the bus used for the transfer of monitored data or with the server used to store the data caused gaps within the acquired data, reducing the amount of usable data to approximately 2 years and the longest period of uninterrupted data acquisition to approximately 6 months. Also the acquisition system only records changes of the variable to be acquired; in the case of occupancy this means that the time and date are recorded:

(1) the first time the sensor notices motion in the office when it previously considered it to be empty—this corresponds to a switch to the state ‘‘occupied’’,
(2) when the sensor has not noticed any movement for 30 s in an office considered to be occupied—this corresponds to a switch to ‘‘vacant’’.

The sensor only recognizes two states of presence: occupied or vacant, and can therefore not distinguish whether the zone is multiply occupied or not. This however is not a problem, since the presence of guests in an office is dependent upon whether the owner of that office is present so that the owner-occupant presence is continuously accounted for. The first step in processing the raw data was to check which days of acquisition had suffered from the ‘‘technical problems’’ stated above and only conserving those that are completely intact. This data was then cleaned of all periods of absence lasting less than 2 min (this usually corresponds to a sensor that stops recording the presence of an occupant because (s)he is too still for her/his movement to be noticed). We then constructed a time series of the data with a 15 min time step by summing over each 15 min interval the duration of periods of presence and of absence and allocating to that interval the state with the longest total duration.

The treated data could then be used for the extraction of information first of all to deduce the inputs necessary to calibrate the model, then to have reference data for its validation. The first step was to check the length of both periods of presence and of absence. The periods of absence were then divided into periods of ‘‘short’’ absence (less than 24 h), of absence that could be related to weekends and periods of ‘‘long’’ absence (greater than 24 h but not taking place over a weekend).[6] Periods of long absence were studied to deduce the distribution of their duration and the average number of their occurrences in 1 year, which will both serve as inputs to the model. The long absences were removed from the time series and the remaining data was used to provide information on the day-to-day occupancy such as the profile of probability of presence and the profiles of probability of transition T_{01} and T_{10}. A profile of the parameter of mobility was deduced from the profile of T_{01} and T_{10} by using Eq. (4) and was then averaged to be used as input.

[6] Most intervals of acquired data are relatively short due to interruptions during data acquisition (intervals range from 2 days to 6 months with the average duration being 2–3 weeks). This of course limits the sample of periods of long absence available. It may also shorten periods of absence that could have lasted longer.

3.3. Validation

The occupancy model has been developed to simulate the time an occupant spends in a particular zone but above all to produce a time series of presence that will serve as an input for models capable of simulating the occupant’s behaviour. Each of these models may have different expectations of the occupancy model’s output; while a model for the opening of windows needs reliable information on the time of arrival of the occupant, a model for the use of household appliances will need to know how long the occupant’s periods of presence will last. In order to estimate the success of the model we listed the statistics that should cover each of these expectations:

- effective total amount of presence will be given by the ‘‘cumulated presence per day’’ and ‘‘cumulated presence per week’’,
- ‘‘first arrival’’ into the zone and ‘‘last departure’’ from the zone of each day of presence, the difference between these two corresponds to the duration of ‘‘daily presence’’ (in contrast to the duration of effective presence mentioned above),
- the duration of ‘‘periods of intermediate presence’’ and of ‘‘periods of intermediate absence’’,
- the ‘‘number of changes’’ of the state of presence during the same day.

We then compared the distributions of those statistics deduced from the measured data and from simulated data produced by the model.

For each of the offices of the LESO building we produced a 5-year time series based on its calibrated inputs. From these time series we calculated the profile of probability of presence, the profiles of probabilities of transition T_{01} and T_{10} and the profile of the corresponding parameter of mobility as well as the distribution of the duration of long absences, in order to make sure that the model’s output is still consistent with its inputs. While the profiles of probability of presence compare very well, the simulated values of the parameter of mobility are clearly below those entered (see Fig. 5). This implies that it is being recalculated relatively frequently so limiting its impact. The time series used to calibrate the model and those resulting from the simulations were then processed to produce the distributions of the statistics of interest for comparison.

3.4. Discussion of results

In order to restrict the number of figures produced in this article we have decided to show the results of four of our five singly occupied offices; by showing the different

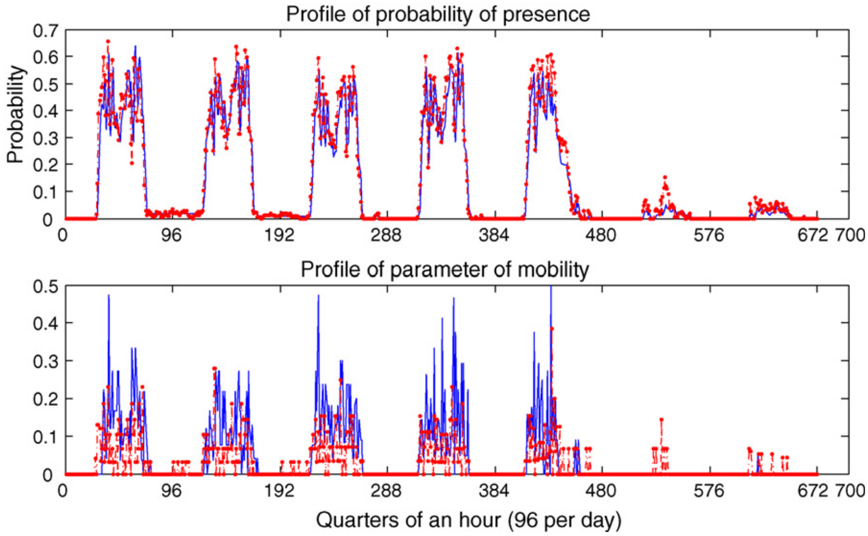

Fig. 5. Comparison, between the monitored data (solid line) and the simulated time series (dotted line), of the profiles of probability of presence (above) and of the parameter of mobility (below) for office no. 3.

behaviours of these four occupants we wish to show the reader the generality of the model. The simulation of multiply occupied offices shall nevertheless not be a problem for the future use of the model as the user will enter inputs for each occupant and each occupant will be simulated independently (unless a dependent behaviour is required, for example in the case of a meeting room).

The green solid lines correspond to data from the monitored offices, the red dotted lines are the results from the simulations. We have shown both the PDF and the CDF of the statistics. For comparison we have added to the CDF's the histogram(s) equivalent to a standard deterministic representation of occupant presence used in dynamic thermal simulation programs of buildings: 100% presence on weekdays from

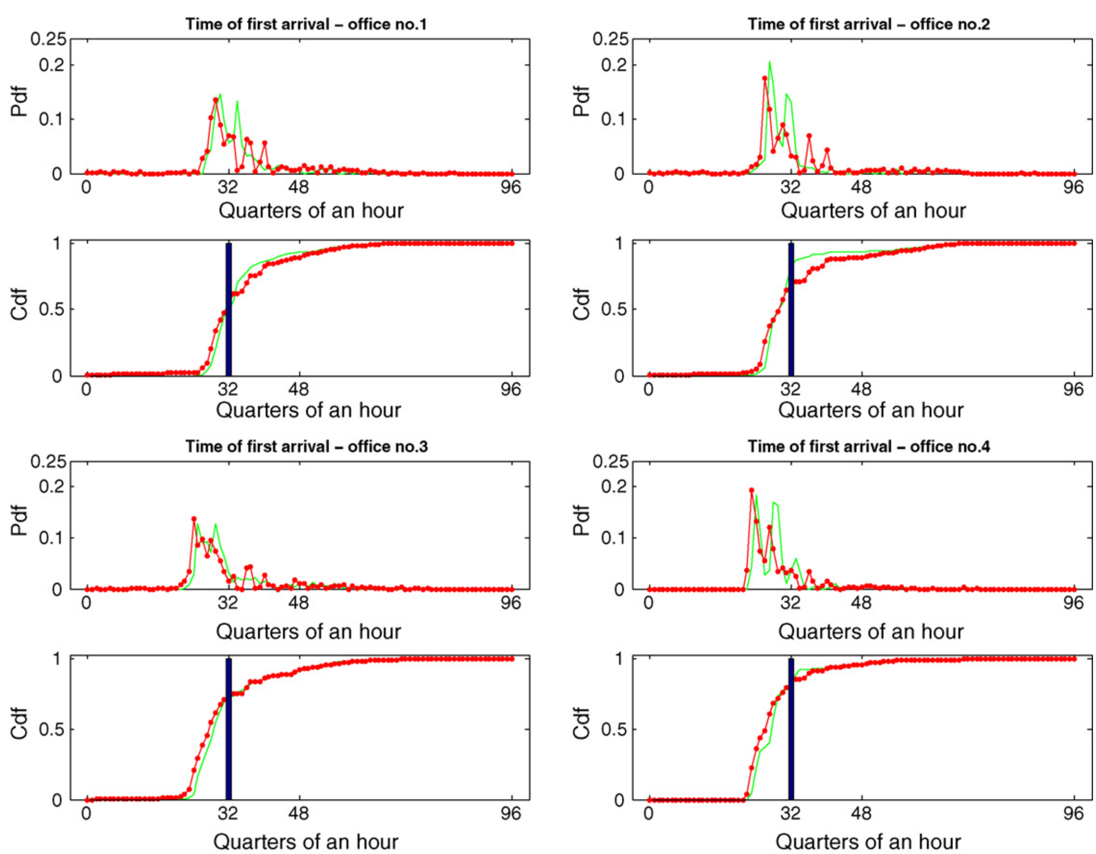

Fig. 6. Comparison, between the monitored data (solid line) and the simulated time series (dotted line), of the PDFs and CDFs of the time of ''first arrival'' into the zone for four private offices. The blue histograms correspond to the repeated use of a standard fixed profile.

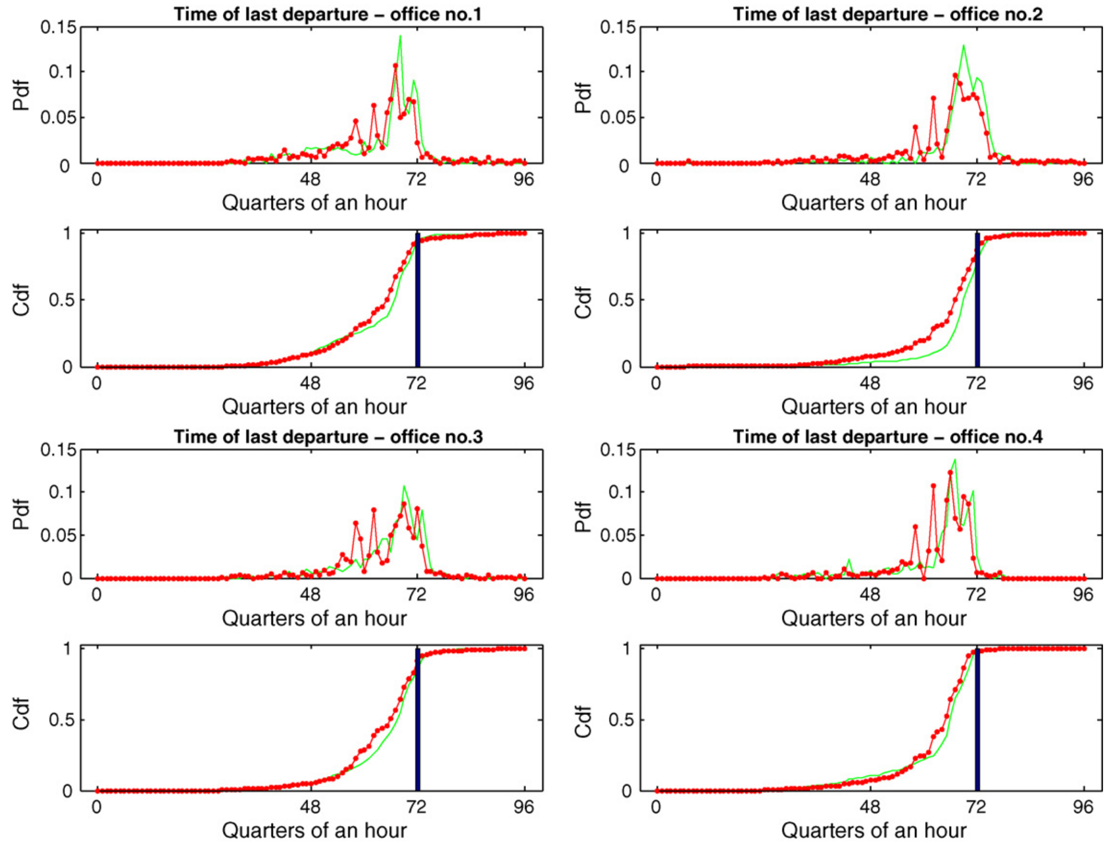

Fig. 7. Comparison, between the monitored data (solid line) and the simulated time series (dotted line), of the PDFs and CDFs of the time of ''last departure'' from the zone for four private offices. The blue histograms correspond to the repeated use of a standard fixed profile.

8:00 to 12:00 h and from 14:00 to 18:00 h. The results from the monitored data provide us with a valuable insight into what the statistics really look like and in understanding the influence of occupants on the building. We will now discuss the different categories of statistics and observe how well the model reproduces what happens in reality.

3.4.1. Arrivals into and departures from the zone

The first category of interest is that of the first arrival of the occupant into the zone and her/his last departure from the zone (shown in Figs. 6 and 7). The difference between the two, the ''daily presence'' (shown in Fig. 8), gives us an idea of how long the occupant could have interacted with the zone, although (s)he might not have always been present during that interval. The behaviour of occupants is usually very different at their first arrival and last departure than during any intermediate arrivals and departures. The first arrival of the occupant usually corresponds to the setting by the occupant of her/his environment to her/his favourite configuration, for example the setting of the state of the blinds, the state of the lights and appliances, the set-point of the heating system or the opening status of windows. These might often stay unchanged until the last departure, during which the occupant returns the zone to its unoccupied state (with, for example, lights and appliances being switched off, windows closed, etc.), knowing (s)he will not be back until the next day or beyond.

Although most values of arrival and departure correspond approximately to values one would expect (arrival around $8:00 \pm 1$ h —corresponding to 32 ± 4 quarters of an hour, departure around $18:00 \pm 1$ h—corresponding to 72 ± 4 quarters of an hour) and that are adopted by other models of occupancy, the figures show that the times of arrival and of departure are particular to the occupant and that these times can depend on the day of the week simulated, explaining the lesser peaks. Although the results from the model do sometimes differ from the original data it has clearly picked up these trends. Values might be off by a time step or two (15–30 min) and the peaks from the original data might be spread out a little but the model captures quite well the different characteristics, recognising the main peak while also reproducing the later arrivals, earlier and later departures as well as the days of longer or shorter daily presence. It should be pointed out that the occasional very early arrivals that appear in the simulations are the result of the model reproducing the non-zero probability of the occupant being present overnight that can be seen in the profile of probability of presence entered as input (Fig. 5). One can also notice in Fig. 8 the stochastic nature of the model: while some occupants will depart exactly 10 h after a first arrival that might fluctuate around an average (offices 1, 2 and 4), the simulated occupant might arrive and leave a bit earlier or later without the two being strongly correlated (just like the behaviour of the occupant of office no. 3).

Please cite this article in press as: J. Page et al., A generalised stochastic model for the simulation of occupant presence, Energy & Buildings (2007), doi:10.1016/j.enbuild.2007.01.018

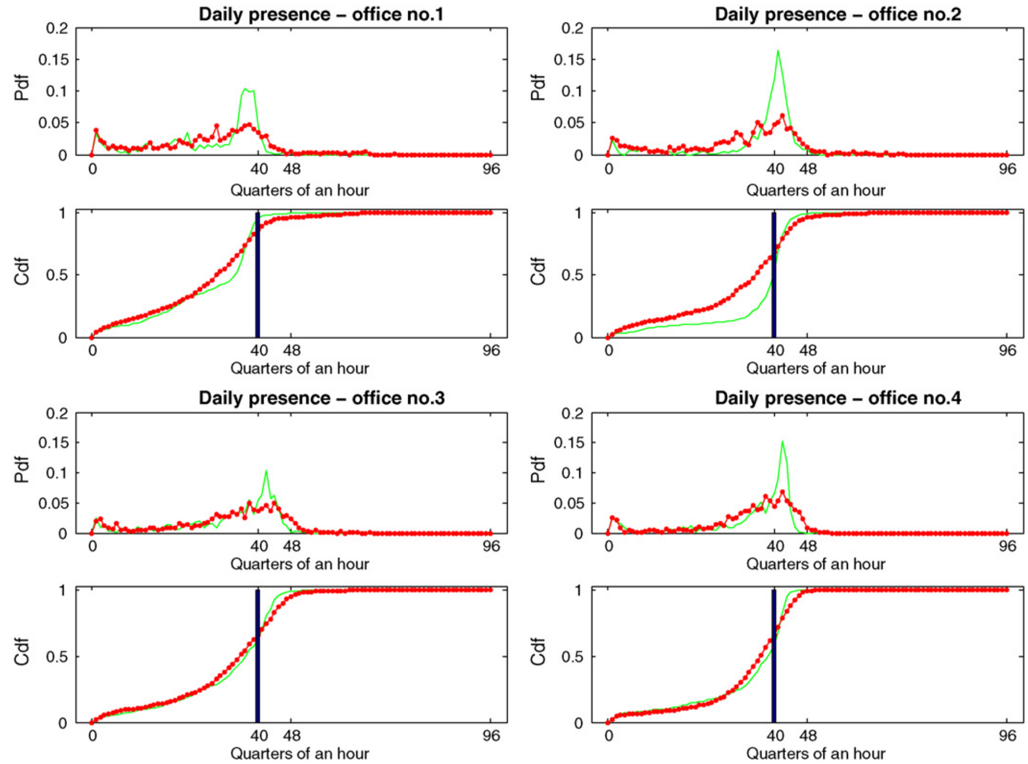

Fig. 8. Comparison, between the monitored data (solid line) and the simulated time series (dotted line), of the PDFs and CDFs of the ‘‘daily presence’’ within the zone for four private offices. The blue histograms correspond to the repeated use of a standard fixed profile.

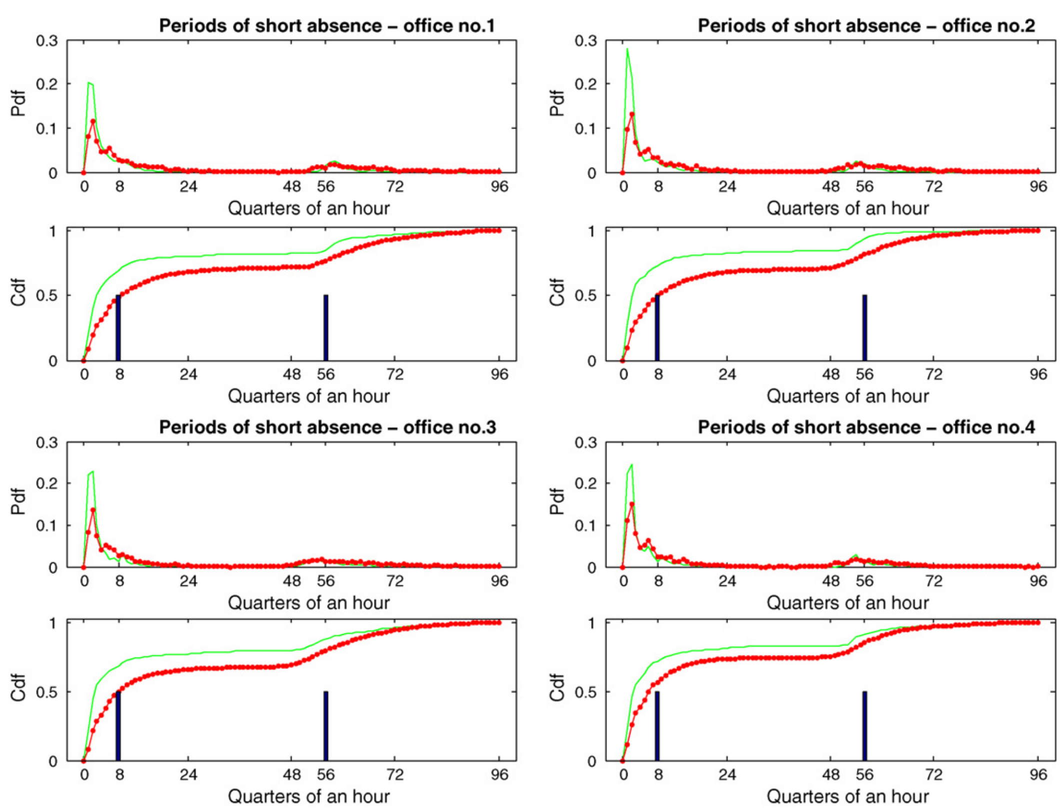

Fig. 9. Comparison, between the monitored data (solid line) and the simulated time series (dotted line), of the PDFs and CDFs of ‘‘periods of short absence’’ for four private offices. The blue histograms correspond to the repeated use of a standard fixed profile.

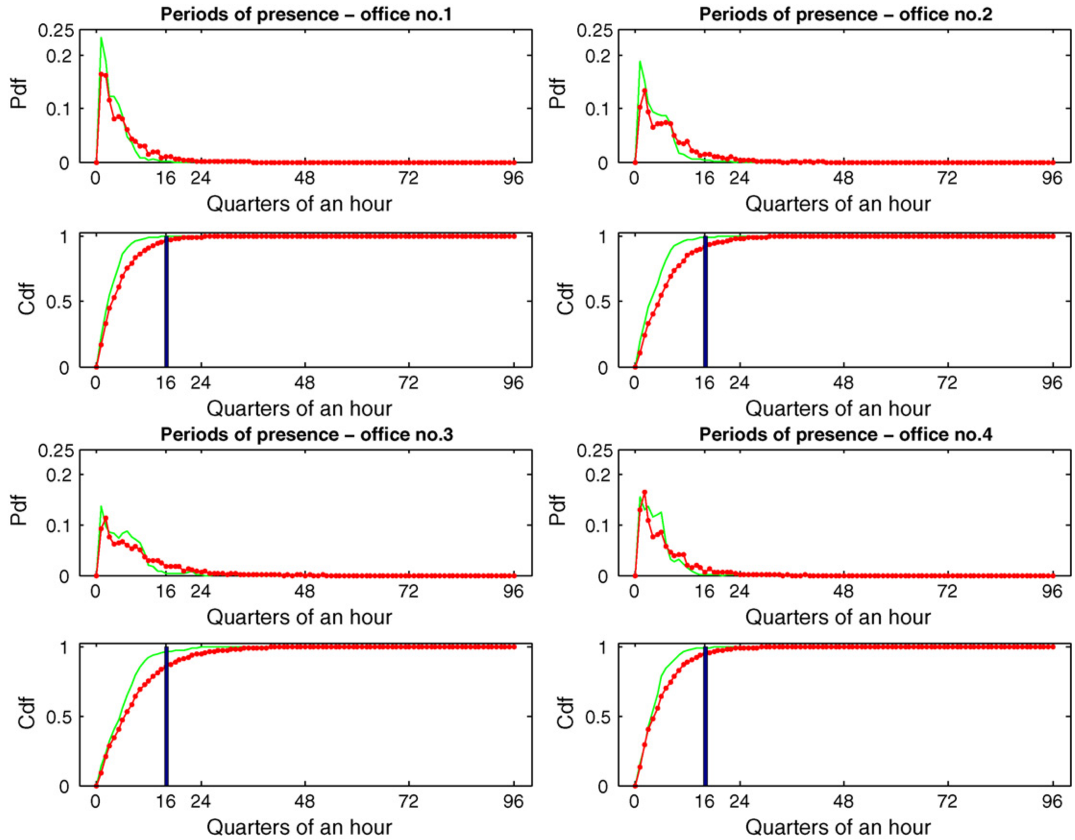

Fig. 10. Comparison, between the monitored data (solid line) and the simulated time series (dotted line), of the PDFs and CDFs of ''periods of presence'' for four private offices. The blue histograms correspond to the repeated use of a standard fixed profile.

3.4.2. Periods of intermediate presence and absence

Figs. 9 and 10 show the distribution of periods of presence and of short absence (less than 24 h). So far the models that have tried to reproduce periods of intermediate presence and absence[7] have done so by assuming their duration is distributed exponentially and is independent of time. Standard profiles of occupancy propose the histograms shown with the CDFs; they correspond to two periods of 4 h of presence separated by a 2 h lunch break and the 14 h of absence between the last departure (at 18:00 h) of one day and the first arrival (at 8:00 h) of the next. The periods of short absence simulated by the model can be split into periods of such absence between workdays (the lower peak at the right of the figure) and periods of intermediate absence (to the left and smaller than, let us say, 48 quarters of an hour). Periods of very short presence and absence (15–30 min) are clearly underestimated in the case of absence and only slightly in that of presence. Nevertheless the model confirms that presence does not follow an exponential curve and that each occupant has her/his own behaviour, which the model picks up rather well. The lack of very short periods favours the occurrence of longer periods; a confirmation of this can be seen in the distributions of the number of changes per day of Fig. 11. These appear in pairs as none of the occupants ever stayed present overnight; each arrival is therefore followed by a departure. The model suffers from two flaws:

(1) it underestimates the number of days of total absence,
(2) it underestimates the number of changes by approximately one pair, suggesting less intermediate periods of absence and presence than seem to take place.

This last discrepancy is linked to the drop in value of the parameter of mobility mentioned earlier; and, as occupants move less than in reality but their daily presence is realistically reproduced, longer periods of intermediate presence and absence will be slightly favoured as we have observed.

3.4.3. Effective time spent in the zone

So far we have discussed:

(1) the times of arrival and departure of the occupant, stressing that these are the instants of a day when the occupant is most likely to interact with her/his environment (as observed for example by [11]), as well as the daily presence that gives an idea of how long the occupant will actively (when present) or passively (when temporarily absent) interact with the zone,
(2) the number of changes of the state of occupancy and the durations of periods of presence and short absence that take

[7] Wang had some success in the case of periods of absence, while Yamaguchi has, to our knowledge, not published any validation of their model.

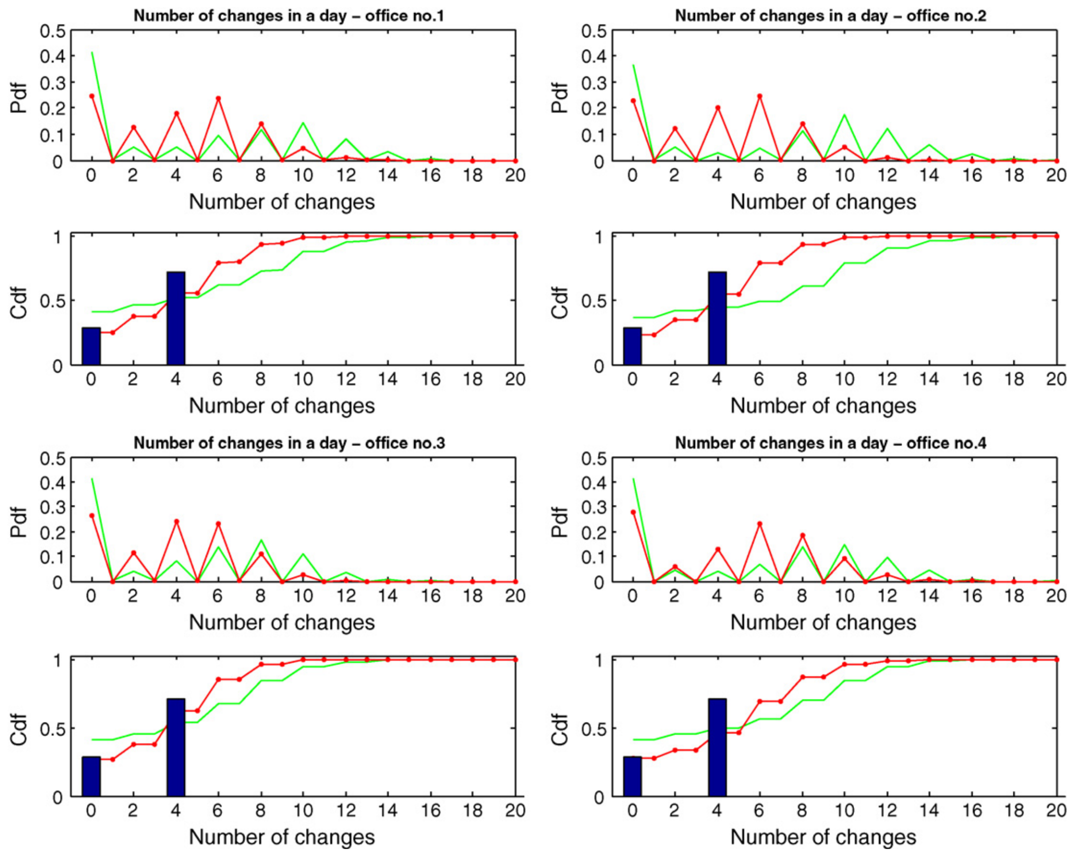

Fig. 11. Comparison, between the monitored data (solid line) and the simulated time series (dotted line), of the PDFs and CDFs of the number of changes for four private offices. The blue histograms correspond to the repeated use of a standard fixed profile.

place during one day, that give an idea of how often the occupant might interact with her/his surroundings at intermediate arrivals and departures.

What we need to know now is how much time the occupant effectively spends in the zone during a day, a week or the whole year. This will govern what heat gains and pollutants each occupant will emit as well as how much total time the occupant has to affect her/his zone of occupancy. This can be deduced from the presence cumulated (i.e. total number of 15 min time steps) over one day or over a whole week.

Fig. 12 shows the total number of quarter hours of presence during a whole week. As we can see from the monitored data, although the occupants' duration of daily presence is typically greater than 12 h (48 quarters of an hour per week), her/his effective cumulated presence over 1 week averages to around 24 h. This is explained by the great movement of occupants and the work time they spend outside of their office. It can also be explained by days of total absence from the zone that called for the revision of the model we mentioned earlier. By adding periods of prolonged absence we have been able to adapt our model to weeks of total absence and weeks of overtime giving us a similar spread distribution as with the collected data. A χ^2-test with $\alpha = 0.95$ confirms that both histograms could be the result of the same distribution for these four offices.

Even though the addition of periods of long absence to the model has drastically improved its performance we still seem to underestimate the number of days of complete absence. This can be seen from the value of the CDF at 0 quarters of an hour in Fig. 13, that teaches us that approximately 35–40% of days include no period of presence (2 days of absence per week, such as a weekend, would correspond to 28.5%) while the model only predicts an absence of 23–28%. The top part of Fig. 13 shows the whole PDF of this distribution. By subtracting the bin of zero presence we get an idea of what the distribution of cumulated presence looks like for days when the occupant appears for at least 15 min (see bottom line of plots in Fig. 13). This shows us how well the model reproduces the statistic and how it covers very closely the whole span of the distribution.

4. Discussion

To fully grasp the contribution of the model of presence it is important to compare it with models already available. Those recently developed by Wang [14] and Yamaguchi [15] are the first capable of simulating realistic periods of presence and absence between an occupant's arrival and departure from an office. However they are based on the fact that the duration of periods of presence is time-independent and omit the inclusion of periods of long absence. Both Wang's analysis and the validation of our model (see Fig. 10) have shown that the former hypothesis is wrong whereas the validation of a previous version of the model presented in this article highlighted how important it is to consider periods of long absence when generating a time

Please cite this article in press as: J. Page et al., A generalised stochastic model for the simulation of occupant presence, Energy & Buildings (2007), doi:10.1016/j.enbuild.2007.01.018

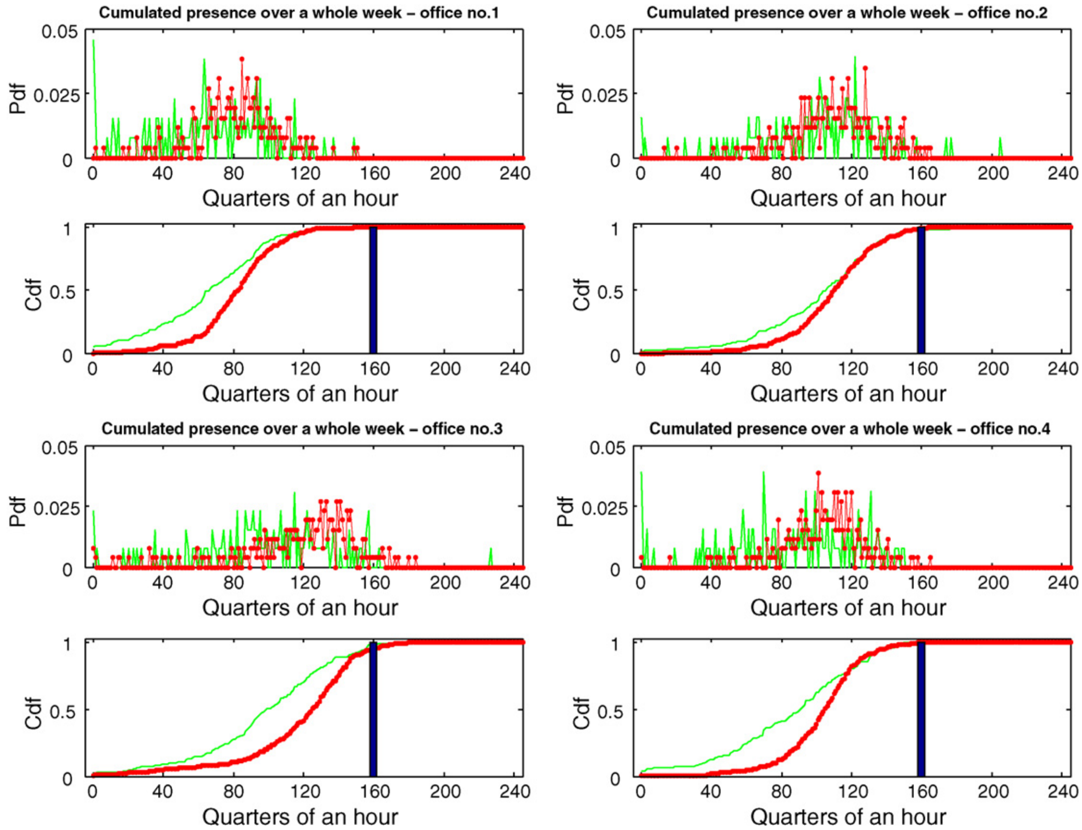

Fig. 12. Comparison, between the monitored data (solid line) and the simulated time series (dotted line), of the PDFs and CDFs of the ''cumulated presence'' over one full week for four private offices. The blue histograms correspond to the repeated use of a standard fixed profile.

series of occupant presence. All other methods used to model occupant presence can be summarised by the repetition of a standard or averaged fixed profile, with, in the best of cases, the spreading of times of arrival and departure using a Gaussian distribution in order to avoid strong peaks [7]. The results above have shown that our model, while having simple inputs, is capable of producing a non-repeating time series of any length, including essential periods of long absence and otherwise reasonable movements to and from the zone resulting in an excellent estimate of the total time an occupant really spends within the zone simulated. It is true that the model underestimates the amount of days of total absence as well as the amplitude of real movement into and out of the offices. The former must be due to an underestimation of the number or duration of periods of long absence.[8] The latter is probably due to the recalculation of the parameter of mobility. These two aspects need to be better understood and improved. Nevertheless the model has still proven itself capable of simultaneously:

- reproducing periods of absence,
- picking up the trends of periods of presence that cannot be simply modeled by an exponential distribution,
- smoothing the peaks of times of arrival and departure,
- while also considering days of atypical presence or of total absence.

Although the benefits of the model will become apparent in terms of numbers only once it is coupled to models of occupant behaviour, one can already assert from the distributions shown in Figs. 5–13 that:

- Fig. 5: people working during weekends will need power and maybe heating or cooling that will be predicted neither by the standard model nor by Wang's model, but will be by our model.
- Fig. 6: likewise, arrival earlier than that predicted by other models will correspond to 1 h of extra-lighting during the darkest days of the year, just as later arrival can correspond to saved lighting.
- Fig. 7: early departure could correspond to saved hours of lighting.
- Figs. 9 and 10: there is potential for saving electricity by implementing a smart switch-off option for appliances and lights left on when occupants are not present during periods of intermediate absence.
- The clear difference between Figs. 8 and 13 shows that, although occupants might be ''at work'' approximately 10 h a day, they only spend about half that time in their office. This corresponds to a decrease of 50% of the predicted metabolic heat gains (within that zone) as well as a potential decrease in electricity consumption linked to the non-use of lights and appliances.

[8] The probability that the Markov chain alone might produce a work day of complete absence is extremely small as it will try at each time step to direct the simulated profile of occupant presence towards the profile of probability of presence it is given as an input.

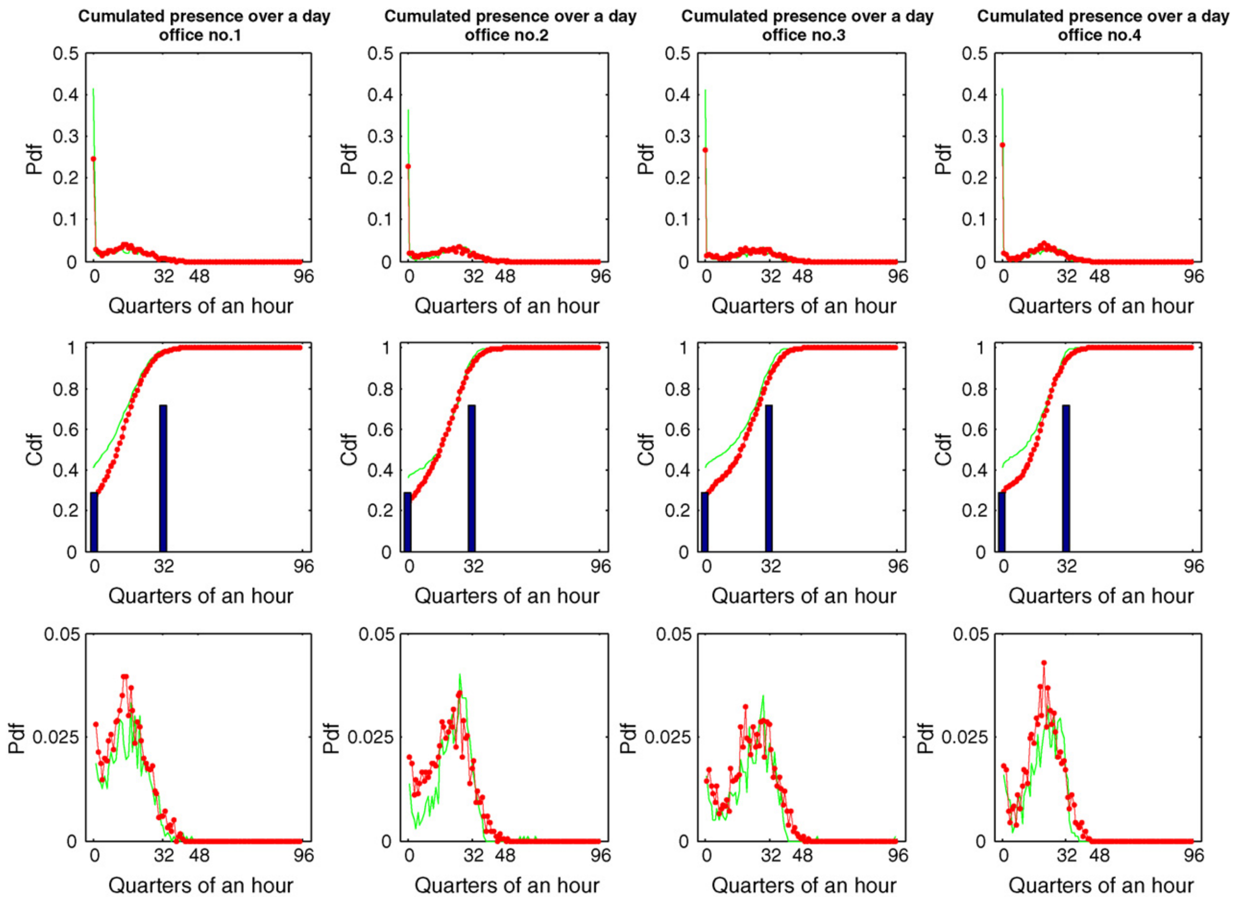

Fig. 13. Comparison, between the monitored data (solid line) and the simulated time series (dotted line), of the PDFs and CDFs of the ‘‘cumulated presence’’ over one day for four private offices. The blue histograms correspond to the repeated use of a standard fixed profile. The bottom figure corresponds to the same PDF as above without a value of cumulated presence equal to zero.

While there is still room for its improvement the model already produces a realistic picture of occupant presence within zones of a building, the basis for related models of occupant interactions with their environment.

The model has been conceived in such a way that it can simulate any pattern of occupancy of any type of building when given the corresponding inputs. This general nature of the model has allowed us to validate it with data from an office building and still claim that it will be useful for the simulation of any other type of building, in particular residential buildings. The data used for the calibration of the model being more detailed than normally available, the authors are now working on the simplification of the profiles of probability of presence and their temporal resolution and observing the effect of this simplification on the model's accuracy, as well as a preparation of guidance for the best use of the model (estimation of μ, entering of profiles of probability of presence and their temporal resolution, entering of inputs related to periods of long absence).

5. Conclusion

The better we build our buildings, the more impact the people living in them will have on their consumption of resources. While simulation tools reproduce the deterministic physical behaviour of buildings with ever greater detail, the behaviour of their inhabitants has so far been represented by repeating standard patterns of occupant presence and occupants' use of elements such as lights and appliances. These assumptions lead to considerable errors in predictions of the peak demand in resources of a building, which in turn will strongly influence the choice and sizing of the means (HVAC systems, supplies for power and water - hot and cold) used to cover that demand. Yet considering occupant behaviour is not only important for the prediction of peak demand. In [9] Bourgeois has shown that ‘‘building occupants that actively seek daylighting rather than systematically relying on artificial lighting can reduce overall primary energy expenditure by more than 40%, when compared to occupants who rely on constant artificial lighting’’. Thus by integrating into building simulation tools the variety of ways people occupy a building and interact with it one can conceive and assess new ways to save energy and enhance occupants' comfort within buildings.

In this article we have proposed an important step towards integrating the effect of occupant behaviour into modern building simulation tools, namely the modeling of occupant presence. Although relatively simple it has proven itself capable of reproducing important characteristics such as the times of first arrival and of last departure, typical long absences and the effective time of presence of the occupant within the zone of simulation using only a small set of simple inputs. The model produces a time series of the state of presence that includes the typical randomness of human behaviour: each person arrives into and departs from the zone they occupy at different times; people tend to enter and leave the zone several times during a typical period of occupancy, reducing the amount of total time spent in the zone and increasing the number of departures from and arrivals into the zone; people may be absent from the zone during long periods of time. This gives a more realistic picture of the time the occupant spends in the zone and how often (s)he might interact with her/his indoor environment; it also avoids the unnatural peaks that arise from repeating the same pattern for each occupant. The model does not simulate the displacement of occupants from one zone to another.

Our future efforts will be dedicated to making the model easy for the potential user to handle. For this purpose we will work on simplifying the profile of probability of presence and helping the user in her/his choice of parameters of mobility and periods of long absence.

Acknowledgements

The funding for this work by the European Commission's DG TREN is gratefully acknowledged, as well as the funding of the LESO-PB by the Swiss Federal Office of Education and Science. We are also grateful to the Swiss National Fund for their support within the National Research Programme NRP 54.

We wish to thank David Lindelöf (LESO-PB/EPFL) for his essential work in the collection of data within the LESO building and for his helpful comments during the development of the model and the review of this paper.

References

[1] Y. Yamaguchi, Y. Shimoda, M. Mizuno, Transition to a sustainable urban energy system from a long-term perspective: case study in a Japanese business district, Energy and Buildings 39 (2007) 1–12.

[2] D. Robinson, S. Stankovic, J. Page, N. Morel, F. Deque, M. Rylatt, K. Kabele, E. Manolakaki, N. Jyri, Integrated resource flow modelling of urban neighbourhoods: projet suntool, in: Buiding Simulation, Eindhoven, The Netherlands, 2003, pp. 375–380.

[3] D. Robinson, Some trends and research needs in energy and comfort prediction, in: Comfort and Energy Use in Buildings, Windsor, United Kingdom, 2006.

[4] J. Page, N. Morel, D. Robinson, J.-L. Scartezzini, Simulating stochastic demand of resources in an urban neighbourhood, in: CISBAT 2005, Lausanne, Switzerland, 2005, pp. 509–513.

[5] R. Fritsch, A. Kohler, M. Nygard-Ferguson, J.-L. Scartezzini, A stochastic model of user behaviour regarding ventilation, Building and Environment 25 (2) (1990) 173–181.

[6] S. Herkel, U. Knapp, J. Pfafferott, A preliminary model of user behaviour regarding the manual control of windows in office buildings, in: Proceedings of the Ninth International IBPSA Conference, Montral, Canada, 2005.

[7] C. Reinhart, Lightswitch-2002: a model for manual and automated control of electric lighting and blinds, Solar Energy 77 (2004) 15–28.

[8] J. Tanimoto, A. Hagishima, H. Sagara, Cooling load simulation considering actual variation of inhabitants' behaviour for accurate estimation of urban maximum energy requirement, in: Proceedings of the Ninth International IBPSA Conference, Montral, Canada, 2005.

[9] D. Bourgeois, C. Reinhart, I. Macdonald, 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 (2006) 814–823.

[10] B. Abushakra, A. Sreshthaputra, J. Haberl, D. Claridge, Compilation of diversity factors and schedules for energy and cooling load calculations, Tech. rep., Energy Systems Laboratory, Texas A and M University (2001).

[11] D.R.G. Hunt, Predicting artificial lighting use—a method based upon observed patterns of behaviour, Lighting Research and Technology 12 (1) (1980) 7–14.

[12] P. Haves, P. Littlefair, Daylight in dynamic thermal modelling programs: case study, Building Services Engineering Research and Technology 9 (4) (1982) 183–188.

[13] G. Newsham, A. Mahdavi, I. Beausoleil-Morrison, Lightswitch: a stochastic model for predicting office lighting energy consumption, in: Right Light Three, Proceedings of the Third European Conference on Energy Efficient Lighting, Newcastle-upon-Tyne, UK, (1995), pp. 60–66.

[14] D. Wang, C. Federspiel, F. Rubinstein, Modeling occupancy in single person offices, Energy and Buildings 37 (2005) 121–126.

[15] Y. Yamaguchi, Y. Shimoda, M. Mizuno, Development of district energy system simulation model based on detailed energy demand model, in: Proceedings of the Eighth International IBPSA Conference, Eindhoven, The Netherlands, 2003.

[16] V. Tabak, B. de Vries, J. Dijkstra, User behaviour modelling applied in the context of space utilisation, in: Design and Decision Support Systems in Architecture and Urban Planning, Eindhoven University of Technology, Eindhoven, The Netherlands, 2004, pp. 141–156.

[17] W. Feller (Ed.), An Introduction to Probability Theory and Its Applications, vol. 1 of Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, 1968.

[18] C.-A. Roulet, Leso building - report on monitoring, Tech. rep., EPFL/LESO (1999).