Journal of Cleaner Production 183 (2018) 1093–1109

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Modeling human activity in Spain for different economic sectors: The potential link between occupancy and energy usage

E.J. Palacios-García*, A. Moreno-Munoz, I. Santiago, J.M. Flores-Arias, F.J. Bellido-Outeiriño, I.M. Moreno-Garcia

Universidad de Córdoba, Departamento A.C., Electrónica y T.E., Escuela Politécnica Superior Campus de Rabanales, E-14071, Córdoba, Spain

ARTICLE INFO

Article history:
Received 6 March 2017
Received in revised form
29 January 2018
Accepted 4 February 2018
Available online 10 February 2018

Keywords:
Stochastic models
Consumer behavior
Energy management
Smart grids

ABSTRACT

Stochastic models for predicting human behavior have become an essential part of the development of demand planning strategies, as well as a high-resolution base information for building simulation software. Due to the close relationship between human presence and consumption, occupancy patterns allow for the recognition of activity peaks, and subsequently, potential maximum demand hours. This contributes to the improvement of control strategies, which combined with the active participation of consumers will drive to major energy savings. In this paper, a novel behavior model for nine economic sectors in Spain has been developed using a Markov Chain methodology that can easily be extrapolated to other locations. The model can generate daily occupancy profiles with a 10-min resolution for the selected sectors, distinguishing between the type of day and type of working hours. The results, which have been validated and compared with other works showing good accuracy, have highlighted the characteristic patterns and maximum occupancy hours of each studied sector. Furthermore, these simulated profiles have been used as input datasets for the estimation of consumption in some selected sectors, illustrating the potential link that can be established between occupancy profiles and energy usage by means of different modeling techniques.

1. Introduction

Within the total energetic consumption in developed countries, the combination of residential and office buildings comprises between 20% and 40% of the total demand. In some cases, figures higher than other sectors such as industry or transportation are reached as indicated in Pérez-Lombard et al. work, where the energy consumption in buildings was analyzed for the USA and other European countries (Pérez-Lombard et al., 2008). Specifically, in Spain, according to the latest energy report for 2015, the demand from residential and office buildings represents 31% of the aggregate energy consumption, a value close to other major consumption sectors such as industries with 24% and transportation with 42% (IDAE and Spanish Ministry of industry energy and tourism, 2015). More specifically, disaggregating the energy demand to a lower level, the electricity consumption that takes place in buildings that belong to the tertiary sector represents around 12% of the aggregate energy. This sector includes stores, hotels, and offices. The energy associated with office buildings represents half of the energy consumed by the service sector in Spain.

These energy consumption figures indicate that some new energy policies must be taken in order to decrease the energy needs of those sectors. New concepts such as Smart Grid, Smart Energy, and Smart Metering have recently arisen in this context. Therefore, as depicted in the literature analysis related to energy distribution, performed by Cardenas et al., the main area of interest nowadays is the simulation of these potential future scenarios (Cardenas et al., 2014). At the same time, a redefinition of the electrical system is needed for renewable integration, storage system, communications and new control strategies as is indicated in Strasser et al., 2015. Subsequently, the three main objectives in the distribution system are (i) an efficient distribution of energy depending on demand by means of demand side management methods (DSM) as the one proposed by Palensky and Dietrich, 2011 taking also into account the balance between robustness and user discomfort as indicated in Pournaras et al., 2014, (ii) real-time power monitoring, where the deployment of smart meters and smart metering infrastructure has

* Corresponding author.
E-mail addresses: p92pagae@uco.es (E.J. Palacios-García), el1sachi@uco.es (I. Santiago), jmflores@uco.es (J.M. Flores-Arias), fjbellido@uco.es (F.J. Bellido-Outeiriño), p92mogai@uco.es (I.M. Moreno-Garcia).

https://doi.org/10.1016/j.jclepro.2018.02.049

promoted new applications as Kabalci, and Zhou and Brown's research concluded (Kabalci, 2016; Zhou and Brown, 2016), and (iii) devices and methods that allow optimizing the energy demand from the user side both in the residential sector, as the analysis of Zhou et al. showed (Zhou et al., 2016), as well as in the industrial environment (Ding et al., 2014). All these goals, along with demand response (DR) strategies, provide an opportunity for consumers to actively participate and modify their power consumption.

However, in order to apply all these concepts and reduce the energy consumption within these sectors, a better understanding of the demand is essential. For these aim, many works focused on the consumption forecasting are being carried out, however, as concluded in Arguira et al. work, where various energy prediction methods were analyzed, the consumer's behavior strongly influence this consumption (Arghira et al., 2012). This idea is backed up by Torriti, who after analyzing the occupancy patterns in 15 countries concluded that DR and demand side management (DSM) techniques require a high level of granularity between 5 and 30 min for accurately taking the control decisions (Torriti, 2012). Therefore, and due to the absence of real measured data, stochastic models for predicting the behavior of users in commercial and industrial sectors are becoming more popular, since occupancy is directly related to energy consumption. In this field, various studies that apply this relationship between occupancy and consumption can be found such as Jia et al. (2017), where the acquisition technologies, as well as the modeling techniques for building occupancy assessing were addressed, and Kim and Srebric (2017) that proved the existence of a high correlation between occupancy and electrical consumption, especially that generated by plug loads.

An overview of modeling techniques for energy consumption is addressed in Swan and Ugursal, 2009 and Grandjean et al., 2012. These reviews depict that, from the wide range of possible types of models, the so-called bottom-up approach is the most suitable for implementing and simulating new control strategies and technologies. These models base their predictions on variables that are in a level below energy consumption such as occupancy profiles, activities schedule, and appliances technologies.

Focusing the interest on the bottom-up stochastic models, previous works have been developed for the residential sector. Richardson et al. modeled the occupancy profiles in the UK (Richardson et al., 2008) and used this data as the basic input for a global energy consumption model that estimates the domestic lighting consumption (Richardson et al., 2009), as well as the appliance energy demand (Richardson et al., 2010), having published recently an enhanced version of this model (McKenna and Thomson, 2016). In the case of Sweden, similar works were carried out by Widén et al. relating activity patterns with energy in households (Widén and Wäckelgård, 2010).

In the Spanish residential sector, a stochastic model was implemented and validated by Lopez et al., which modeled occupancy patterns in Spanish households showing the close relation between occupants' activity and energy consumption (Lopez et al., 2013). The stochastic model allowed for the calculation of daily active occupancy profiles of Spanish households with 10-min resolution and distinguished between the number of residents, location and the type of day, which are the input parameters required for the simulation. These occupancy profiles were finally used in a higher level algorithm for estimating the electrical consumption of some appliances (Santiago et al., 2013). In addition, this occupancy model was also used as the base input for a recently developed lighting demand estimation system in dwellings (Palacios-Garcia et al., 2015).

All these energy consumption models have common sources of information from which the occupancy profiles were extracted. These sources are the time use surveys (TUS), several statistical studies carried out in most European countries, and whose most interesting characteristic is that the interviewees write down their daily activity in 10-min intervals. That confers this model one main feature, which is the simulation of the energetic consumption with a high temporal resolution and detailed enough so that energy management strategies or policies can be derived from the application of these models.

Nevertheless, the situation in the commercial, and industrial sectors or other economic activities is significantly different. In Europe, only projects regarding residential demand were developed, such as the REMODECE (de Almeida et al., 2011), but neither data nor models are available for the other sectors. The situation is slightly different in the United States, where some general studies are periodically performed by the U.S. Energy Information Administration (U. S Energy Information Administration, 2012). However, Spanish daily schedules differ from those found in U.S. or even within Europe, so a detailed model is needed in this country.

Regarding this situation, and the lack of high-resolution consumption models for the main economic activities, the goal of this paper is to develop a stochastic model that aims to simulate the users' behavior in the framework of these sectors. This model will provide output data for the daily occupancy profiles for different kinds of buildings (offices, retail stores, industries, etc.). This prediction will be based on a set of input parameters, which are the number of workers, the division of working hours during the day and whether it is a weekday or the weekend.

The article is structured as follows. In Section 2, the methodology followed in the development of the model is addressed. Section 3 provides and discusses the simulation results. Subsequently, the model validation for the studied sectors is exposed in Section 4, whereas Section 5 compares the results with previous works. In Section 6, the link between occupancy and energy intensity is presented with two examples of widely used modeling methodologies. Finally, Section 7 summarizes the main ideas presented in the paper and related future works.

2. Methods

The development of the model presented in this article is divided into three different methodological stages. The first step was the selection of the input data necessary for constructing the models. Subsequently, those data were processed and analyzed to construct the model. Finally, due to the stochastic methodology that was applied to obtain the model, a simulation mechanism was to be implemented. These three stages are addressed in detail in the following subsections.

2.1. Time use survey

The base information employed in the development of this model was collected by the time use survey (TUS). Those surveys have been carried out in many European countries and their results, as well as their microdata, are usually available. Moreover, the statistical office of the European Union (Eurostat) together with some national statistical institutes have joined efforts to establish several guidelines under the harmonized European time use surveys (HETUS) database to make the results of these works comparable between each other.

In the case of Spain, the last TUS was conducted in 2010 with 19,295 people who were at least 10 years old and lived in a total of 9,541 households. In the TUS, the interviewees wrote down in a diary, in 10-min intervals, data about the activities performed during the 24 h of one random day, the place where the activities took place and whether someone accompanied them (National Statistics Institute of Spain. Ministry of Economy and

Competitiveness, 2010).

Moreover, regarding work, the survey also records those people who develop a paid employment and the economic activity associated with it using a normalized classification. Therefore, as it is addressed in the next subsection, the behavior and daily patterns for the different economic sectors can be studied. Furthermore, other variables can be included in the study such as the type of day (weekday or weekend) and the configuration of the working hours or schedules (continuous or split shifts).

2.1.1. Economic activities classification

The survey contains a specific field where the economic activity is encoded using the definitions established by the national codes for economic activities (CNAE) in its 2009 version (Official Bulletin of Spanish Government (BOE) No.102, 2009). In the basic version, the codes are composed of a two-digit number related unequivocally to a normalized economic activity and are grouped into different sectors. However, not only is the CNAE encoding system regulated in Spain, but it also compliances with the statistical classification of economic activities in the European Community (NACE Rev. 2) (Eurostat, 2008), which provides a correspondence with the international standard industrial classification of all economic activities (ISIC Rev. 4) (United Nations, 2008). Therefore, the same sectors studied in Spain can be extrapolated to any country following the same methodology.

The CNAE codes range from 1 to 99 for the different economic activities. Nevertheless, these codes were not individually considered in the study, but two criteria were followed to select and group each one. First, since the goal of this study is to relate the daily human activity profiles with the energy consumption in various sectors, only those that represent an important part of the demand has been considered. On the other hand, due to the relatively reduced number of samples in the survey and with the aim to maintain a statistically significant population, no individual codes were considered, but the smallest division unit was the groups defined in the CNAE. In addition, some groups were studied together because of their close relationship as far as energy consumption is regarded. The selected groups are listed below.

- Group A (01–03): Agriculture, forestry, and fishing.
- Group C (10–33): Manufacturing.
- Group F (41–43): Construction.
- Group G (45–47): Wholesale and retail trade.
- Group H (49–53): Transportation and storage.
- Group I (55–56): Accommodation and food service activities.
- Group J, K, L, M, N and O (58–84): Information and communication; Financial and insurance activities; Real estate activities; Professional, scientific and technical activities; Administrative and support service activities, Public administration and defense; compulsory social security.
- Group P (85): Education
- Group Q (86–88): Human health and social work activities.

As it can be observed, one of the divisions includes several groups (from group J to O) with different economic activities. However, despite the clear difference in their final service, they all have one characteristic in common, which is being carried out in an office or laboratory. Subsequently, from the energetic point of view, the human activity in all these groups means the occupancy of a space in a building and, therefore, the usage at least of lighting, as well as heating, ventilation and cooling (HVAC) equipment whose consumption in most of the cases is far higher than the specific appliances related with each sector (Pérez-Lombard et al., 2008).

2.1.2. Types of days and working hours

In addition to the definition of nine economic sectors, two additional situations were considered. On the one hand, those interviewees that completed the survey for a weekday (Monday to Friday), have been distinguished from the ones that did it for a weekend (Saturday or Sunday). On the other hand, the configuration of the working hours during the day was also taken into account differentiating between continuous working hours, meaning that the person is actively working more than 6 h with breaks no longer than 30 min or split working hours, where the working time is divided into two main shifts no longer than 6 h with a break of at least 1 h between them.

2.2. Model fundamentals

The goal of the model is to determine for each instant of the day the number of active workers that can be found in each one of the previously defined sectors and under certain conditions of working hours and type of day. Hence, the first element that must be defined is the concept of an active worker.

An active worker is defined as a person who is in the workplace at a given instant, independently of the activity that he or she is developing (e.g., a worker who is having lunch at the workplace is not developing a work during that period, however, he or she is using energetic resources that are found there). This fact should be pointed out since the mere occupancy of the workplace implies an energy consumption due to the human activity and subsequently represents the cornerstone of many of research fields such as energy simulation, building efficiency assessing or energy planning.

Regarding all the previously exposed, the proposed methodology to model this occupancy was based on the Markov chains theory, whose principle of operation as well as some practical applications can be found in Gamerman (2006) and Gilks et al. (1996). This theory defines a stochastic process that consists of a discrete number of states where the probability of moving from one state to the next one is only determined by the current state. In our case, the space of possible states is defined as follow.

$$S = \{w, n\} \tag{1}$$

where w represents the state of a worker being active, whereas n means a person who is not in the workplace. The possibility to move from one state to other is determined by the so-called transition probabilities p_{ij}. Therefore, if a person is in the workplace, there is a probability p_{ww} that in the next instant of time he or she remains in the same state, as well as a probability p_{wn} meaning that he or she becomes an inactive worker. Likewise, for a person who is not currently in the workplace, he or she will either stay inactive with a probability p_{nn} or will move to the active state, according to p_{nw}. All this process can be clearly seen in Fig. 1 where the different states and transition probabilities for the proposed model have been represented.

These transition probabilities are not usually expressed

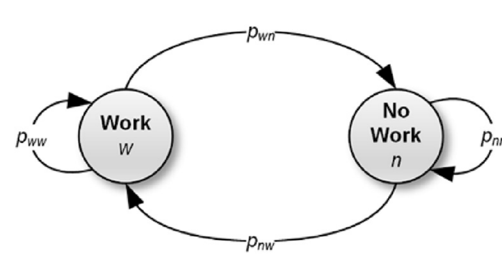

Fig. 1. Markov chain states and possible transitions between them.

independently, but the combination of every transition probability between the defined states is joined into the so-called matrix of transition probabilities or transition matrix **P**, as it can be seen in (2), where each row represents the current state, whilst the columns are the next state.

$$\mathbf{P} = \begin{pmatrix} p_{ww} & p_{wn} \\ p_{nw} & p_{nn} \end{pmatrix} \tag{2}$$

In the case of using only one transition matrix **P** for the whole system, the Markov process is called homogeneous. However, the TUS offers information recorded with 10-min resolution or 144 samples for a day. Therefore, for each instant t a different transition matrix $\mathbf{P}(\mathrm{t})$ was calculated forming a non-homogeneous transition process. In addition, the model not only distinguishes between the nine above-mentioned economic activities, but, as previous exposed, it also takes into account the type of day (weekday or weekend) and the type of working hours (continuous or split). Thus, 144 transition matrices were calculated using the TUS data for each combination of conditions (economic activity, type of day and type of working hours).

The process to obtain this number of transitions between states and calculate the transition probabilities was fully automatized using an SQL database where the TUS data were stored following a relational philosophy and the environment and high-level programming language MATLAB was used to perform the necessary queries. The steps for these queries were as follow:

1. Select those interviewees whose economic activity corresponds with the studied group.
2. Select those interviewees with a specific type of working hours.
3. Select those interviewees who fill out the diary or activity notebook for a specific type of day.
4. From 1 to 144 select the time step s and the next time step $s+1$ and count the number of occurrences for each transition.
5. Calculate the transition probabilities according to (3) where n_{ij} is the number of transitions between the initial state i and the next state j and between the current instant t and the following instant $t+1$, whereas n_i is the total count of people being at state i in the instant t.

$$p_{ij}(t) = \frac{n_{ij}(t|t+1)}{n_i(t)} \tag{3}$$

An example of this calculus is illustrated in Table 1 where the number of transitions between 00:00 h and 00:10 h were obtained from the TUS in the manufacturing sector (C), in a weekday and working with continuous shifts. As it can be seen, if the current state of a worker is to be at the workplace (w), the probability of staying in this place at 00:10 h is 0.9615 against a 0.0385 probability of becoming inactive. Opposite, if a person is inactive this state will remain at 00:10 h, since the probability of not being in the workplace at that time is 1. All the values calculated in this example corresponds to the previously mentioned probabilities $p_{ij}(t)$ for a specific sector, type of day, and type of working hours, consequently, they can be also expressed using matrix format as in (2).

2.3. Simulation procedure

The previous methodology allows calculating the transition matrices for the Markov process. Nevertheless, due to the stochastic nature of the model, a simulation algorithm had to be implemented for obtaining the results. The simulation procedure was based on Monte Carlo methods using repeated random samplings in order to generate the initial state of the system, as well as the successive transitions. The following subsections address this process.

2.3.1. Input parameters

As it was exposed, various economic activities have been modeled and among them, different conditions regarding the type of day and working hours have been established. Therefore, three inputs parameters were provided to indicate each of these characteristics. In addition, the Markov process models the daily profile of only one worker. Thus, a variable that indicated the number of workers to be simulated was added to the system. Subsequently, the set of input parameters which must be provided before running the process contains the below input data.

- Economic Activity or sector (s_n): It establishes the sector for which the simulation will be carried out. The possibilities are the nine sectors discussed in Section 2.1.1.
- Type of day (*Day*): The parameter distinguishes whether the simulation is performed for a weekday (WD) or a weekend (WE).
- Type of working hours (*Hours*): This parameter differentiates between those employees that work continuously (Cont.) and the ones that have split shifts (Split).
- Number of workers (N): It represents the number of people to be simulated. For instance, this number can correspond to the number of employees in an office or workers in a manufacturing factory.

The three first parameters (economic activity, type of day and type of working hours) allow the selection of the transition matrices suitable for the process, whereas the number of workers determines the number of iterations.

2.3.2. Initial state

One of the main problems of a Markov process is defining the initial state of the system. This state will determine the following transitions and it is necessary for the simulation algorithm to start running. Moreover, the proposed Markov model is non-homogeneous so not only the method for determining the initial state had to be defined but also the instant for which this initial state will be generated.

For the study, the 00:00 h was selected as the initial instant for the simulation process. Subsequently, for each conditions of sector, type of day and type of working hours, and using the TUS data, the probability mass function (PMF) $p_o(X)$ of a person being in the workplace for this time was calculated. In this function, X can take two possible discrete values: 0 when the worker is inactive or 1 for a worker in the workplace. This PMF in the form of discrete

Table 1
Example of the number of transitions in the manufacturing sector (C) for a weekday with continuous working hours.

Current state 00:00 h (i)	Next State 00:10 h (j)				Total
	Work (w)		No Work (n)		
Work (w)	25	25/26 = 0.9601	1	1/26 = 0.0385	26
No Work (n)	0	0/288 = 0	288	288/288 = 1	288

cumulative distribution function (CDF) $P_o(X)$ will be used in the first step of the algorithm to generate the initial state of each worker to be simulated.

2.3.3. Algorithm

The stochastic process to generate the occupancy profile is similar regardless of the input parameters provided to the system. It consists of two different parts, as it can be observed in Fig. 2. In the first part, the initialization of the system is performed. Hence, after providing the system with the required input parameters, the initial state probability function and the transition matrices for the different time steps are loaded into the simulation system.

After this previous initialization, the remaining process is equivalent for each of the worker (n) and it is repeated as many times as the number of selected workers N. Therefore, as it was exposed in the previous section, the first state is calculated before starting the Markov process using the CDF at 00:00 h. For this aim, first a random number R_0 uniformly distributed between 0 and 1 is generated.

$$R_0 \sim U([0-1]) \tag{4}$$

As it is indicated in (5) and using this number R_0, a random inverse transform sampling is performed in the initial state CDF, $P_0(X)$, obtaining the activity state $O_n(t)$ of this person n for the first step (00:00 h, $t = 0$). The state of the worker $O_n(t)$ is a binary variable that can take the same values as the CDF for the initial state, i.e., 0 if the worker is inactive or 1 if the worker is in the workplace.

$$O_n(t=0) = P_0^{-1}(R_0) \tag{5}$$

Once this initial state has been established, the iteration for each interval of time t starts. Since the TUS sample was taken every 10 min, 144 intervals are possible, but the first state is already determined so only 143 transitions and, therefore, different transition matrices, are required to simulate a day. Nevertheless, the transition matrix between 23:50 h and 00:00 h was also calculated so various consecutive days can be simulated.

In this way, for each instant t, a random sampled number R_t uniformly distributed between 0 and 1 is generated as in (4). After this, the random number is employed to calculate the following state of the worker by means of the transition matrix for the step t and the previous activity state of the worker $O_n(t-1)$.

$$O_n(t) = P_{ij}^{-1}\left(t|O_n(t-1), R_t\right) \tag{6}$$

This second part is repeated as many times as the number of workers selected (N), calculating for all of them the period of occupancy of the workplace for each instant. Finally, the total daily occupancy profile $O_T(t)$ for the selected input parameters is obtained as the aggregated of each individual employee $O_n(t)$, using (7).

$$O_T(t) = \sum_{n=1}^{N} O_n(t) \tag{7}$$

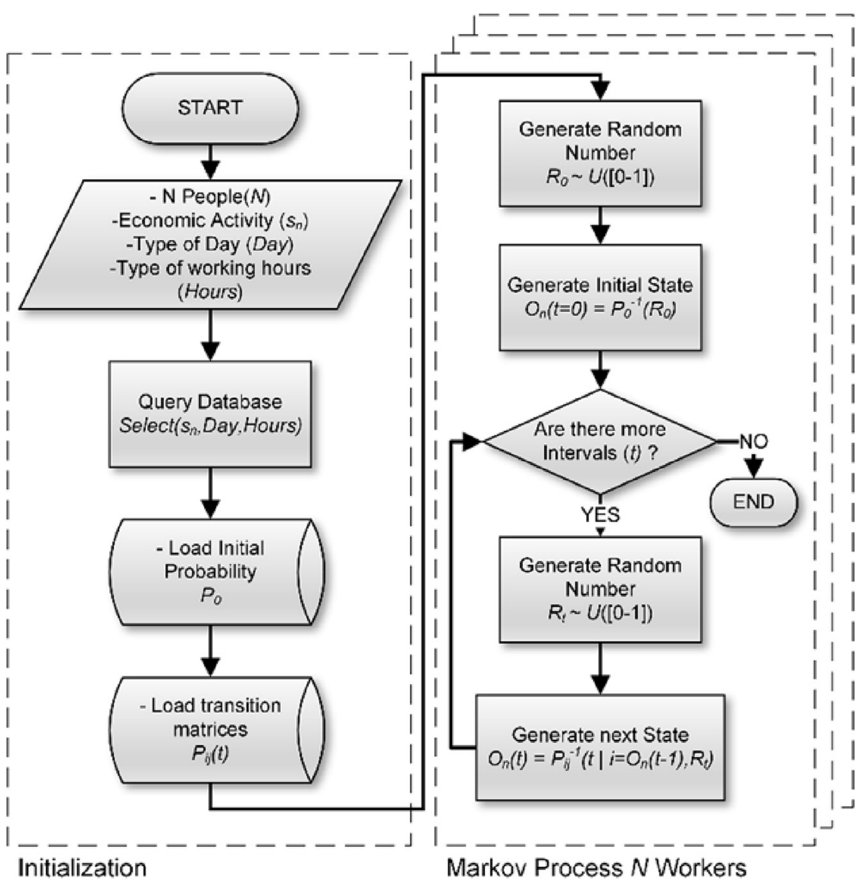

Fig. 2. Flowchart of the simulation process.

2.4. Output data

Using the above-described input data and calculation algorithm, the implemented model can generate the workers' daily activity profile named as $O_T(t)$ for nine different economic activities, with a temporal resolution of 10 min and being able to distinguish between the type of day and working hours. Such a profile will have a deterministic nature since each individual employee has associated exact periods when he or she is in the workplace, but as an aggregate, it will represent the average trend observed in the sector.

2.5. Model implementation

The calculation of the matrix for the model and simulation algorithm were implemented using the MATLAB software. The transition matrices for each sector, type of day, and type of working hours were programmatically calculated and stored in the MAT-File format, so they can be easily loaded into the development environment. Likewise, the simulation algorithm was programmed as a MATLAB function using the high-level programming language provided by this tool. This function can be called from the MATLAB command line, nevertheless, to ease the simulation and analysis process a general user interface (GUI) was also implemented.

2.6. Model update mechanisms and application to other regions

One of the main features of the system that should be pointed out is the functional independence of the simulation algorithm and the transition matrices. The simulation system, described in Section 2.3 encapsulates the functionalities for loading the transition matrices and performing the Markov Chain procedure. Nevertheless, the transition matrices are not part of the simulation but are loaded from the MAT files indicated in Section 2.5 based on the given input parameters.

Therefore, if new data are available or new locations are to be included, the system can be updated in a seamless way, only by including the new MAT files with the transition matrices and the initial probabilities. The calculation of these transition matrices for the MAT files is the one described in Section 2.2, which can be applied to any other TUS since they are standardized all around Europe. Therefore, the system is not only flexible but also easy to maintain and update.

3. Results

Once the methodology of the model has been explained, the obtained results achieved running the implemented model are shown. These results are presented for the different sectors considered to be modeled, as well as the different additional conditions of type of day, and working hours. Moreover, the implemented GUI is also described in the following subsections.

3.1. User application

The GUI provides a friendly interface for inserting all the necessary input parameters required by the model, visualizing the daily profiles or saving the numeric results of each simulation. As it is shown in Fig. 3, it is composed of three main panels. The one on the top left named as *Input Parameters* allows the user to indicate the simulation parameters, as well as the initialization of the simulation procedure by clicking the button label as *Simulate Results*.

Regarding the panel on the right side, it is devoted to the representation of the results once the simulation is finished. Finally, the panel named as *Measures* shows several statistical indicators of the aggregated results such as the mean active workers along the day, as well as the maximum and the minimum level of occupancy. Likewise, an estimation of the main occupancy periods is performed. The system can detect up to two periods calculating the starting and end time of the day, and the number of hours of high occupancy. Furthermore, the individual worker details can be saved to a CSV file using the provided button in the same panel and the field *FName* to assign the name of this file.

3.2. Distribution of interviewees among economic activities

First, and before the calculation of the transition matrices or the initial state of the Markov process, the distribution of interviewees among sectors were studied, taking into account their subdivision

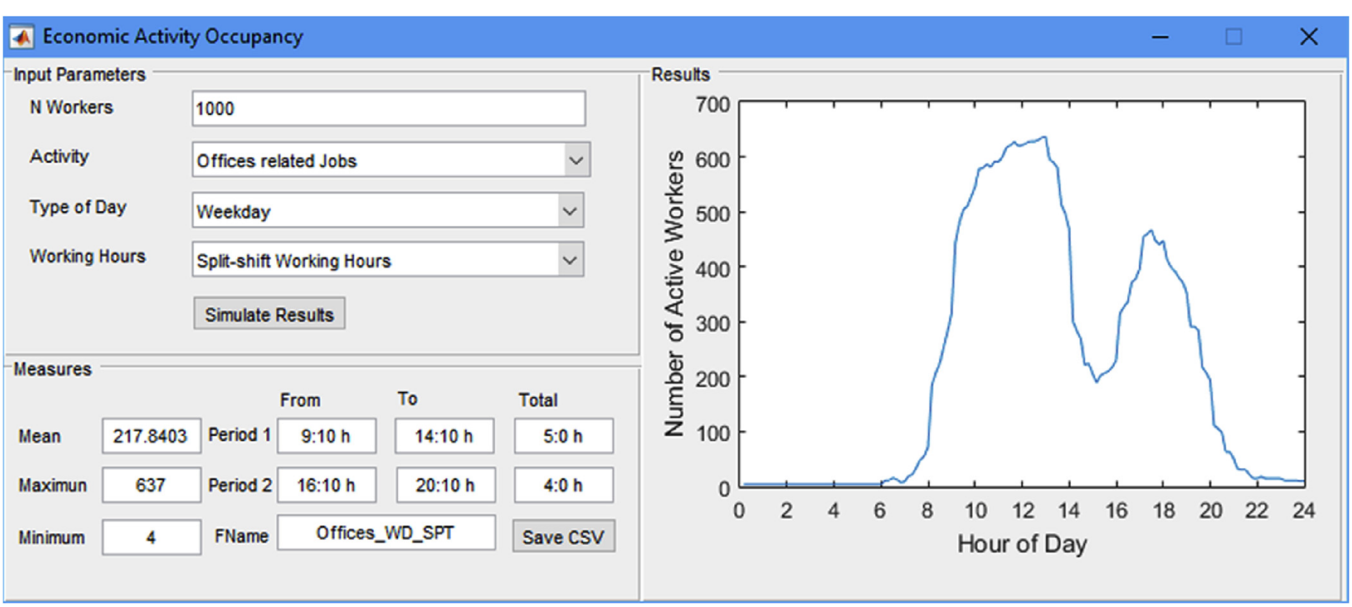

Fig. 3. Snapshot of the General User Interface (GUI) built for performing the simulations.

Table 2
Distribution of interviewees among sectors, type of day and type of working hours.

Economic Activity	Continuous		Split		Total
	WD	WE	WD	WE	
A	71	32	148	70	321
C	314	164	338	170	986
F	90	50	276	136	552
G	216	120	448	238	1022
H	137	73	104	39	353
I	170	84	111	60	425
J-O	707	368	391	231	1697
P	175	107	103	69	454
Q	284	146	50	29	509

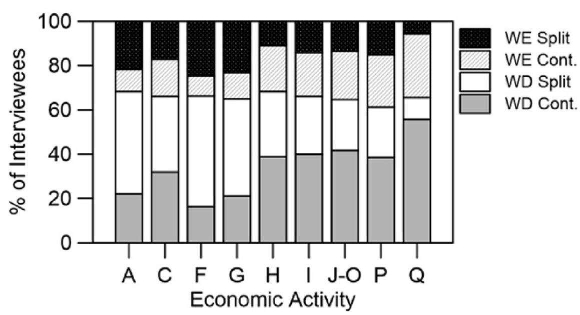

Fig. 4. Relative distribution of type of day and working hours among sectors.

regarding the type of workings hours and type of day for which they filled out the diary. As exposed in Section 2.1 the number of interviewees in the TUS was 19,285, of this figure, only 7810 developed a pay economic activity at the time the Survey was performed. Within this fraction, the distribution among the study conditions is collated in Table 2 where each sector was expressed using its group letter and each column represents the number of records in the survey for those conditions. However, since not all the sector where included in the model, the sum of the column total in Table 2 is lower than 7810.

As it can be observed in Table 2 the activities with the larger number of interviewees are the jobs related to offices (J-O), the wholesale and retail trade sector (G) and the manufacturing industry (C). Opposite, the sectors of agriculture, forestry and fishing (A) and transportation and storage (H) present the lower number of records in the survey. In addition, as it could be expected the number of interviewees developing a paid activity during the weekend is lower.

This fact can also be better observed in Fig. 4 where the distribution of type of days and type of working hours was represented as a percentage of the total number of records for each sector. The results depict that for all the sectors the percentage of interviewees working during weekdays is higher than 60%. Furthermore, regarding the type of working hours, the continuous shifts prevail in the sectors of human health (Q), education (P), offices related jobs (J-O), accommodation (I), and transportation (H), whereas in agriculture (A), manufacturing (C), construction (F) and retails (G) is more common to work in split shifts.

3.3. Start states

After the analysis of the survey population, the discrete probability function for the initial states was calculated following the previously exposed methodology. Those initial results indicate significant differences between the sectors as it can be seen in Fig. 5 where the initial probability for a worker to be in the workplace

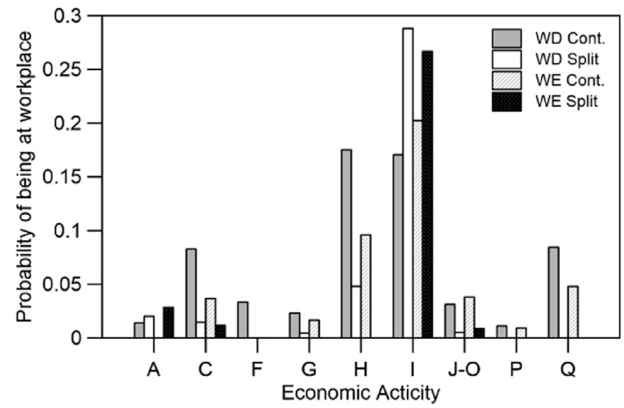

Fig. 5. Probabilities of a worker to be in the workplace at the initial state 00:00 h.

was represented for the different types of days and working hours.

Fig. 5 shows that, as it might be anticipated, most of the sectors have a low probability of having active workers at midnight. Nevertheless, four sectors showed a relatively high level of occupancy. The highest values were found for the accommodation and food service sector (I) where the average percentage of active workers at midnight for any type of day or working hours is around 20%. This may be justified due to the necessity of having workers that serve the guest during this period.

With lower values, the transportation sector (H) can be also pointed out with an average occupancy higher than 10%, although in the case of split working hours the level is significantly lower. Finally, the human health (Q) and manufacturing (C) sectors have also a level of activity higher than the rest. This can be related to the 24/7 schedules that many of these works imply.

3.4. Simulations results

Using the previous distribution functions for the initial states, together with the transition matrices and the implemented simulation procedure several simulations were performed selecting different values of the input parameters. The following subsections will discuss these results.

3.4.1. Individual results

As it was exposed in the methodology, each employee or worker is individually simulated using the calculated transition matrices for the sector, the type of day and the type of working hours. Therefore, the daily profile of an individual simulation presents a deterministic pattern where the state can be either 0, inactive worker, or 1 employee in the workplace. This fact is depicted by Fig. 6 whose top chart shows the occupancy periods (black rectangles) of 20 different persons working in the wholesale and retail sector (G).

It can be observed that there are no two equal profiles, therefore, confirming the stochastic nature of the process and the high degree of determinism in the individual results. It can also be seen that one employee is continuously active despite working in split shifts. Nevertheless, as it was previously commented the model only accounts for the occupancy of the workplace. Thus, a worker that is having lunch in the workplace is not developing any work but it is considered as an active worker in the model. However, as shown in the bottom graph of Fig. 6 if the individual results are aggregated following (7) a general trend can be observed, which depicts two differentiate occupancy peaks.

This is one of the most useful features of this model. Whereas other simulation methods only provide aggregate results, using the

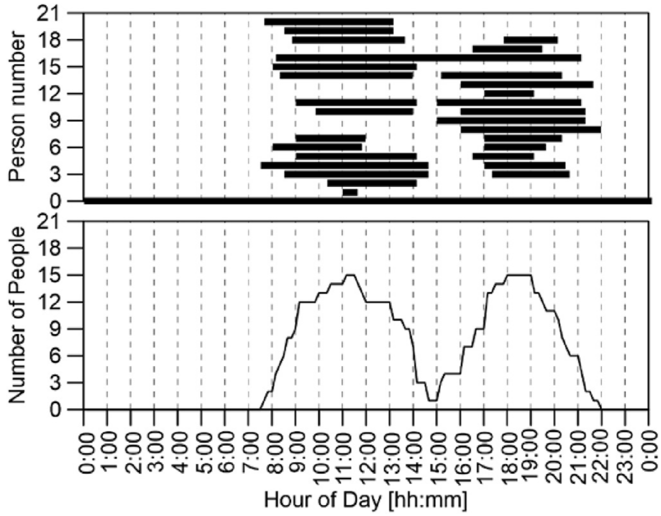

Fig. 6. Individual simulations and aggregate results of active occupancy for economic activity G with split-shift working hours for a weekday.

proposed model the total profiles are obtained as a consequence of the individual human behavior. Therefore, if the system is to be employed for predicting or assessing the energy demand, some consumptions such as those related to PCs, tools, individual HVAC system or other appliances can directly be attributed to each individual worker, whilst also having a global knowledge of the aggregate pattern and the common energetic needs. For instance, in this case, the simulation was performed for split working hours, so two different peaks are observed that correspond to the two daily shifts, being the periods where the higher occupancy is concentrated and, consequently, the higher energetic demand for this sector.

3.5. Aggregate results

The previous section showed the benefits of this methodology where having both individual and aggregate results provide a broader perspective of the human behavior and, subsequently, a better estimation. However, when the aggregate results are considered the different profiles can also be studied aiming to recognize the periods where the highest occupancy is located and the relationship between them in order to improve the energy planning policies.

In this way, the aggregated daily profiles for 1000 workers were obtained for each sector and for a weekday distinguishing between continuous and split working hours. The results can be observed in Fig. 7 where each sector is presented using a different graph. The X-Axis indicates the hour of the day, whereas the Y-Axis shows the aggregate percentage of active workers. In addition, the profiles for continuous working hours are illustrated with a solid black line, whilst a solid gray line was used for the split working hours.

First, it can be observed that some characteristics are common to every sector, specifically those related to the type of working hours. In the case of continuous working hours, all the sectors present a period of high occupancy that starts in the first hours of the morning and continues until the afternoon. Likewise, the occupancy after this first period keeps normally values that are relatively low, although in some sectors such as the agriculture, forestry, and fishing (A), the wholesale and retails (G) and the accommodation and food service (I) a second occupancy peak can be found. Nevertheless, the level of this peak is significantly lower than the first one.

This scenario contrasts with the case of split working hours where two occupancy peaks are observed during mornings and afternoons. The second peak is also lower than the first one, but its value is higher compared with the case of the continuous working hours. In this way, just the mere knowledge of the type of working hours for a certain sector provides valuable information about the major occupancy periods and, therefore, when the highest demand is likely to occur. However, each of the selected sectors presents individual characteristics that should be pointed out. These differences can mainly be observed in three aspects which are the starting and end hours of the occupancy peaks, the width of the periods and the base occupancy that remains out of the periods of high activity. All these characteristics can not only be depicted from Fig. 7, but Table 3 also gathers together this information in a numeric form.

As far as the starting hours are concerned, they are all comprised between 06:00 h and 09:00 h. Within this interval, the sectors that present a high level of occupancy during the early hours are the agriculture, forestry, and fishing sector (A), the construction sector (F) and the transportation and storage sector (H). Nevertheless, in the case of split working hours, the starting time is delayed an hour except for the construction sector (F) where it remains unaltered. For the rest of the sectors, the most common starting hour is around 08:00 h for both continuous and split working hours. However, a sudden gradient can be seen in sectors such as offices related jobs (J-O), education (P) or human health (Q), whereas in others this transition is more gradual as in the case of the wholesale and retail sector (G), as well as in the accommodation and food service sector (I).

Regarding the duration of the first occupancy peak, in the case of continuous working hours, it comprises in most of the cases a larger time interval since there is no activity during the afternoon. Opposite, the width of the first peak is smaller for the split working hours, distinguishing in all the cases an occupancy valley between mornings and afternoons. The only exception is the accommodation and food service sector (I) where not only is the morning occupancy peak wider, but the transition valley is not as pronounced as in the rest of the sectors. This is justified due to the activity of the sector itself, where the highest activity is found during meals hours. In addition to this, another unique characteristic is observed in the profiles of offices related jobs (J-O) for continuous working hours. This feature is a short but clearly visible period of low occupancy around 11:00 h which coincides with the breakfast break in Spain.

On the other hand, for the end time and the base occupation, two different trends are observed. In the case of continuous working hours, a negative gradient can be seen between 14:00 h and 15:00 h except for the accommodation and food service sector (I) where this decrease does not take place until 16:00 h. Moreover, whilst in all the sector the level of occupancy has low and decreasing values until 22:00 h, the accommodation, and food service sector (I) has an opposite trend, reaching a second peak at that time. Finally, regarding the base occupancy during no peak hours, three sectors stand out with relatively high values. These sectors are the manufacturing (C), the transportation, and storage (H), and the human health (Q). In addition, the accommodation and food service sector (I) presents again substantial differences which an occupancy trend that decreases from 22:00 h to 04:00 h in a profile that resembles the one observed in the residential sector.

In the case of split working hours, the situation is significantly different and after the second occupancy peak, a pronounced negative gradient takes place until reaching low values of occupancy. Furthermore, the base occupancy is also lower than for continuous working hours, except in the accommodation and food service sector (I) that differs again from the other with a higher occupancy.

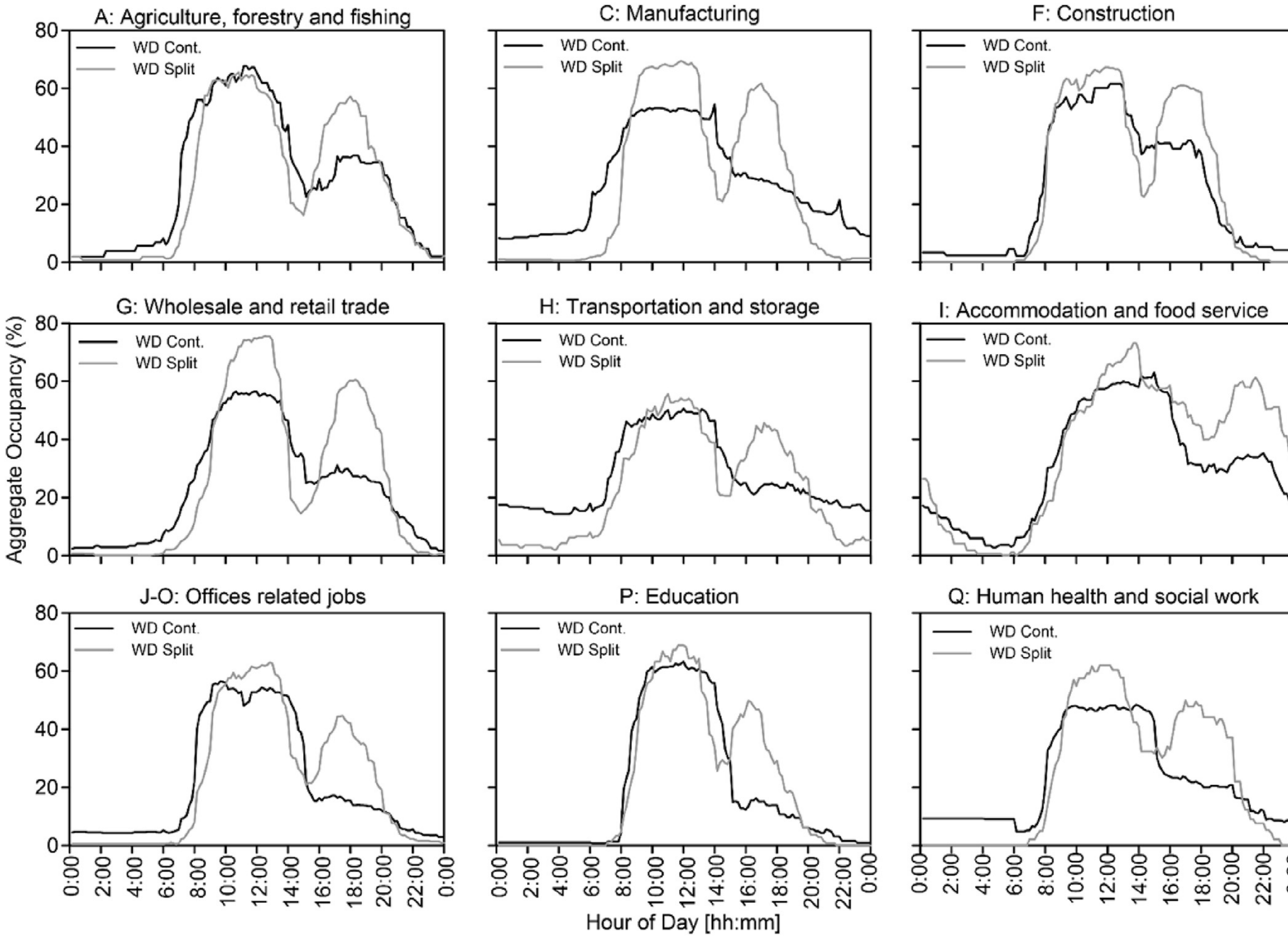

Fig. 7. Aggregated results for 1000 workers in different sectors for a weekday distinguishing between continuous and split working hours.

Table 3
Occupancy periods and statistics during weekdays for each sector.

Sector	Working Hours	Peak 1			Peak 2			Mean	Max	Min
		From	To	Total	From	To	Total			
A	Continuous	07:00 h	14:00 h	07:00 h	17:00 h	20:30 h	03:30 h	28%	70%	1%
	Split	08:00 h	14:00 h	06:00 h	16:00 h	19:00 h	03:00 h	27%	66%	1%
C	Continuous	06:00 h	14:00 h	08:00 h	–	–	–	27%	54%	8%
	Split	08:00 h	13:00 h	05:00 h	15:00 h	18:00 h	03:00 h	27%	72%	1%
F	Continuous	08:00 h	14:00 h	06:00 h	–	–	–	25%	62%	2%
	Split	08:00 h	13:00 h	05:00 h	15:00 h	19:00 h	04:00 h	26%	68%	0%
G	Continuous	08:00 h	14:00 h	06:00 h	17:00 h	20:00 h	03:00 h	24%	58%	2%
	Split	09:00 h	14:00 h	05:00 h	17:00 h	20:00 h	03:00 h	26%	74%	0%
H	Continuous	07:00 h	14:00 h	07:00 h	–	–	–	28%	51%	14%
	Split	08:00 h	14:00 h	06:00 h	15:00 h	20:00 h	05:00 h	24%	56%	2%
I	Continuous	08:00 h	16:00 h	08:00 h	20:00 h	23:00 h	03:00 h	32%	62%	3%
	Split	09:00 h	14:00 h	05:00 h	20:00 h	23:00 h	03:00 h	37%	75%	0%
J-O	Continuous	08:00 h	15:00 h	07:00 h	–	–	–	20%	54%	3%
	Split	09:00 h	14:00 h	05:00 h	16:00 h	20:00 h	04:00 h	22%	63%	1%
P	Continuous	08:00 h	15:00 h	07:00 h	–	–	–	19%	65%	1%
	Split	08:00 h	14:00 h	06:00 h	15:00 h	18:00 h	02:00 h	21%	68%	0%
Q	Continuous	08:00 h	15:00 h	07:00 h	–	–	–	24%	50%	6%
	Split	08:00 h	14:00 h	06:00 h	16:00 h	20:00 h	04:00 h	21%	68%	0%

Thus, summarizing the above-discussed trends, four types of occupancy patterns can be distinguished, especially in the case of continuous working hours. First, the manufacturing (C), transportation (H) and human health (Q) sectors present a high base occupation, as well as a long period of activity. The second type is characterized by two periods of high occupancy during mornings and afternoons that despite the continuous working hours are significant. This profile is found in the agriculture, forestry and

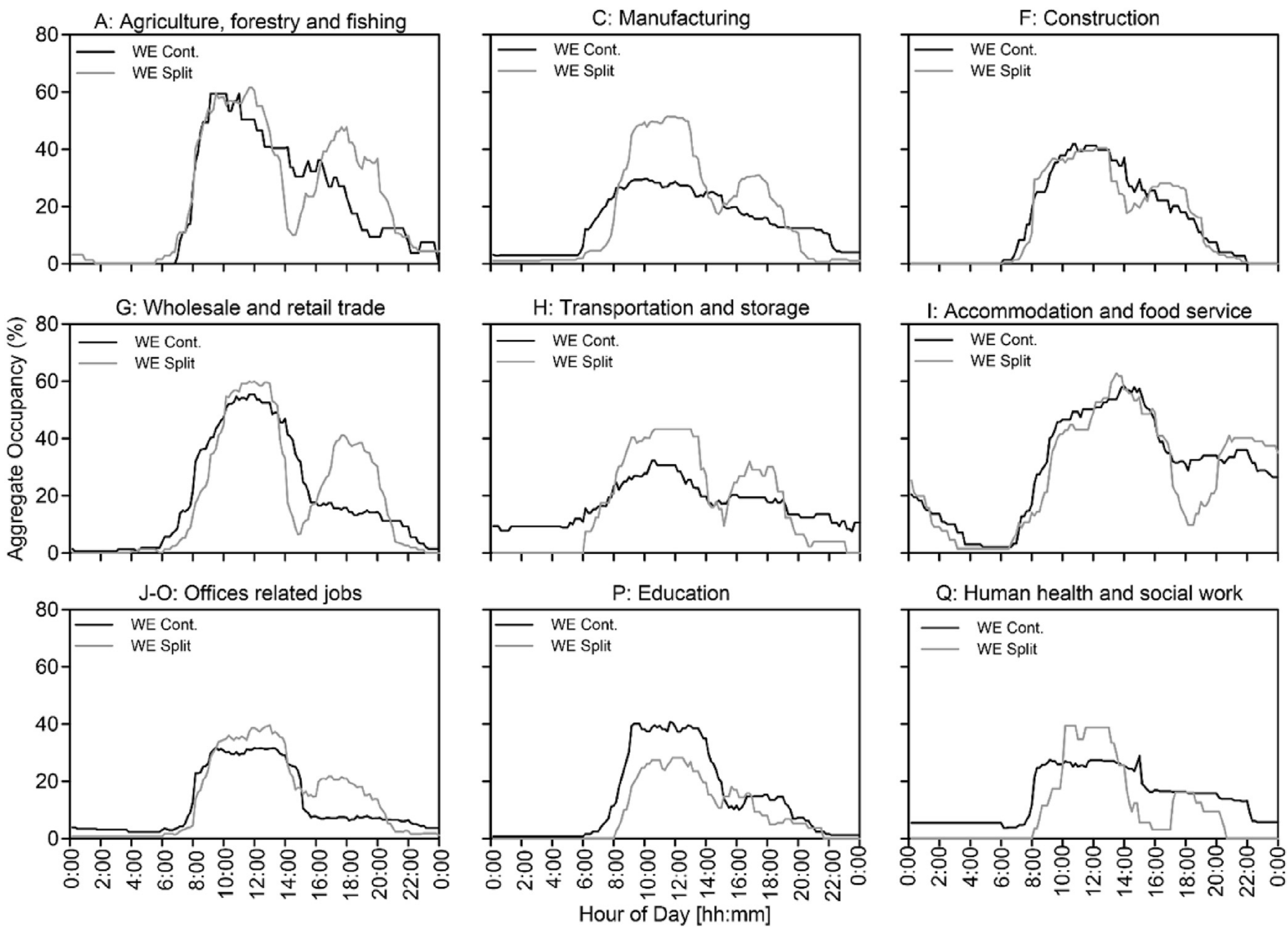

Fig. 8. Aggregated results for 1000 workers in different sectors for a weekend distinguishing between continuous and split working hours.

fishing sector (A), the wholesale and retail sector (G) and the construction sector (F). The jobs related to offices (J-O) together with the education sector (P) constitute a third group where the main occupancy period is depicted with very low values for the rest of the day. Finally, and with some features that differ from the rest of sectors, the accommodation and food service sector (I) stands out with occupancy peaks that take place where the occupancy valleys of the other sectors are found.

Further conclusions can be detected when the profiles for weekends are observed. Those patterns are illustrated in Fig. 8 where the aggregate results for 1000 workers distinguishing between continuous and Split working hours can be seen. Likewise, the numeric results for the occupancy periods and some statistic are summarized in Table 4.

The first difference between weekdays and weekends is the lower levels of occupancy in general terms. Nevertheless, some exceptions are found such as the agriculture, forestry, and fishing (A) or the accommodation and food service sectors (I) whose levels are similar for both types of days. The duration of the occupancy peaks is also lower, generally at least an hour shorter than for the weekdays. In the same way, the two peaks of the split working hours are less pronounced and the afternoon occupancy period presents a lower activity, especially in the sectors of construction (F), offices related jobs (J-O), education (P) and human health (Q).

In conclusion, therefore, the results show the diversity of existing profiles, not only due to the economic activity but also because of the type of day and type of working hours within the same sector. Thus, a good estimation of the energetic consumption and the effective energy planning of the systems must take into account all these particular features that this model allows us to recognize.

4. Validation

Once the results have been presented and discussed the validation process is addressed. This process aims to estimate the ability of the proposed model to simulate occupancy profiles which match the ones observed in the TUS and the uncertain in the modeled data. Therefore, the validation was carried out using the original data recorded in the TUS and generating daily occupancy profiles with the model per sector. The selected number of workers matched the population of the TUS which was indicated in Table 2, so the samples can be compared. Two tests were developed to test the global error per sector and the individual uncertain for each input parameter.

First, the global error of the model for each sector was evaluated by mean of the root mean squared error (RMSE). This indicator accounts for the absolute instant discrepancies between the TUS data and the synthetic profiles generated by the model. The expression used for the RMSE is given in (8)

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=0}^{n}\left(O_M(t) - O_{TUS}(t)\right)^2} \tag{8}$$

where n is the number of instants or time steps in the day (144 due

Table 4
Occupancy periods and statistics during weekends for each sector.

Sector	Working Hours	Peak 1			Peak 2			Mean	Max	Min
		From	To	Total	From	To	Total			
A	Continuous	08:00 h	14:00 h	06:00 h	–	–	–	204	581	0
	Split	08:00 h	14:00 h	06:00 h	16:00 h	20:00 h	04:00 h	239	628	0
C	Continuous	06:00 h	14:00 h	08:00 h	–	–	–	149	206	32
	Split	09:00 h	13:00 h	04:00 h	16:00 h	20:00 h	04:00 h	178	494	6
F	Continuous	08:00 h	14:00 h	06:00 h	–	–	–	157	445	0
	Split	08:00 h	13:00 h	05:00 h	16:00 h	19:00 h	03:00 h	143	407	0
G	Continuous	08:00 h	15:00 h	07:00 h	–	–	–	203	558	8
	Split	10:00 h	14:00 h	04:00 h	17:00 h	20:00 h	03:00 h	194	610	0
H	Continuous	07:00 h	14:00 h	07:00 h	–	–	–	177	333	69
	Split	08:00 h	14:00 h	06:00 h	16:00 h	19:00 h	03:00 h	170	430	0
I	Continuous	09:00 h	16:00 h	07:00 h	–	–	–	303	601	11
	Split	10:00 h	17:00 h	07:00 h	20:00 h	23:00 h	03:00 h	275	639	19
J-O	Continuous	08:00 h	15:00 h	07:00 h	–	–	–	123	319	28
	Split	08:00 h	14:00 h	06:00 h	16:00 h	19:00 h	03:00 h	141	404	9
P	Continuous	08:00 h	05:00 h	07:00 h	17:00	19:00 h	02:00 h	137	419	8
	Split	09:00 h	14:00 h	05:00 h	15:00 h	17:00 h	02:00 h	80	272	0
Q	Continuous	08:00 h	15:00 h	07:00 h	–	–	–	146	294	32
	Split	10:00 h	14:00 h	04:00 h	17:00 h	20:00 h	03:00 h	98	377	0

to the 10-min resolution), O_M is the aggregated occupancy profile obtained using the model and with the same population of the TUS, and O_{TUS} is the daily pattern registered in the TUS for a given sector, type of day and type of working hours. To increase the number of reference samples from the TUS, each sector was validated as an aggregate. This means that the profile used in the validation process for a given sector is the sum of the four combinations of inputs parameters of the model (i.e. weekday continuous working hours, weekday split working hours, weekend continuous working hours, and weekend split working hours).

The results of this process are illustrated in Fig. 9 where the patterns generated by the model for each sector (solid black line) are presented together with the original TUS data (cross symbol). As it can be seen, the nine sectors match almost perfectly the original data with very low RMSE values, which proves the validity of this methodology for simulating the human behavior in certain economic activities by means of a stochastic process and using individual workers' profiles as the basis for the calculus.

The same conclusions can be drawn if the relative error ($Err\%$) is observed. This value was calculated according to (9) and it is the RMSE referenced to the number of samples per sector n_s used for the validation, which was taken from Table 2.

$$Err\% = \frac{RMSE}{n_s} \tag{9}$$

As it can be seen, the relative error is always below 1%, even when the reference sample is small, which numerically confirms the great ability of the model for reproducing the TUS profiles with a high accuracy.

The RMSE showed a low error in the aggregate profiles, nevertheless, the stochastic nature of the simulation procedure means that the model results have an associated uncertain component. Subsequently, after testing the RMSE of the model the dispersion and variability of the results referred to the data collected by the TUS were analyzed. For this aim, the Normalized Variation Factor (NVF) was employed. This indicator is a pseudo-variance in which the mean squared error (MSE) existing between the generated profiles and the original data are normalized using the squared mean of the reference population as it is indicated in (10).

$$\mathrm{NVF} = \frac{\mathrm{MSE}}{\overline{\mathrm{O}_{\mathrm{TUS}}}^2} = \frac{\sum_{t=0}^{n}\left(O_M(t) - O_{TUS}(t)\right)^2}{n\left(\frac{1}{n}\sum_{t=0}^{n} O_{TUS}(t)\right)^2} \tag{10}$$

The variables included in this equation have the same meaning as the ones in (8). Due to the stochastic nature of the model, in order to evaluate the maximum and minimum limits of uncertain, the NVF was calculated for 100 different simulation selecting among all of them the observed extreme values for this indicator.

The obtained figures can be seen in Table 5 for each different input condition of the type of day and working hours, as well as for the aggregated case. All of them took values lower than one, this means that the variability of the results is below the 10% of the original results' mean. An exception is observed in the human health sector (Q) for weekends and split working hours where the maximum NVF resulted in 1.293. This high value is justified due to the small number of records found in the TUS, which are not statistically significant. The same problem, although with a lower impact, can be depicted from the results of the agriculture, forestry, and fishing (A), or the transport and storage sectors where both high NVF values and a small sample in the TUS exist too.

This relationship between the uncertain accounted by the NVF and the number of samples can be better understood with the scatter plot in Fig. 10. As previously stated, the number of employees simulated for each case was equal to the number of samples recorded in the TUS. This data is represented in the X-Axis, whereas the Y-Axis illustrates in a logarithmic way the obtained scattered mass of NVF values. The Figure depicts how the uncertain of the simulation results decreases with the number of employees being simulated, which is totally consistent with the random nature of the simulation process.

Hence, after the validation process, it can be seen that the capacity of the model to reproduce the original information is extremely good, as indicated the low RMSE values. Likewise, the low NVF figures confirm the quality of the model, although it was detected that small simulation populations led to higher variability and uncertain in the individual generated profiles.

5. Comparison with other works

In addition to the validation process, the model was compared with previous works in order to confirm the consistency of the generated results. Due to the lack of research papers addressing this problem the Commercial Buildings Energy Consumption Survey (CBECS) was selected as the reference source for the comparison (U. S Energy Information Administration, 2012). This survey carried out periodically by the U. S. Energy Information Administration (EIA)

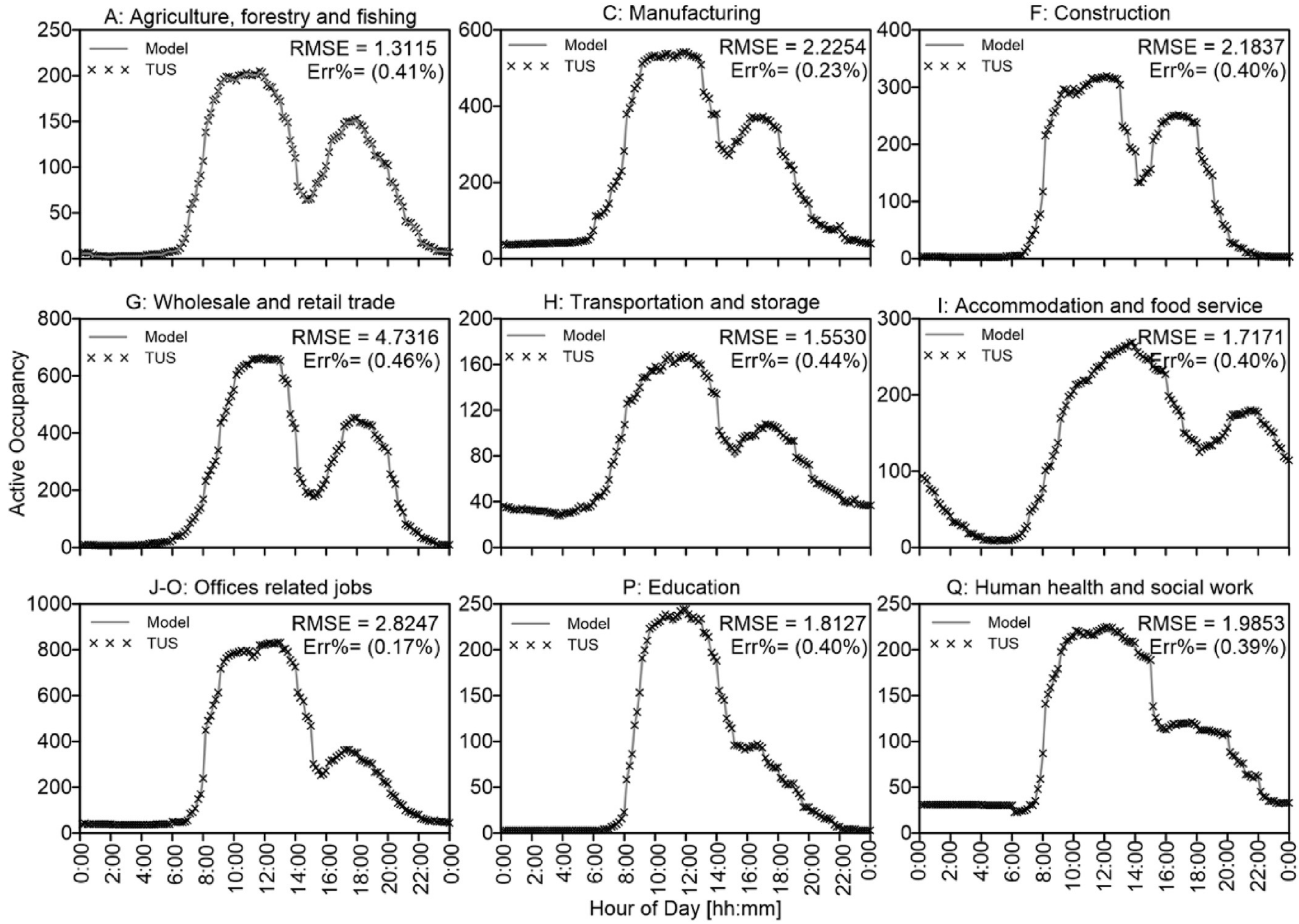

Fig. 9. Model Validation. Comparison of TUS data with the model. Aggregated results of weekdays and weekends, as well as continuous and split working hours.

Table 5
Normalized Variation Factor (NVF) for each sector, type of day and type of working hours. Maximum and minimum values for 100 simulations using the TUS sample size.

Sector	WD Continuous		WD Split		WE Continuous		WE Split		Total	
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
A	0.006	0.079	0.002	0.031	0.019	0.313	0.010	0.091	0.002	0.025
C	0.002	0.021	0.002	0.020	0.005	0.120	0.006	0.076	0.001	0.010
F	0.005	0.116	0.001	0.021	0.014	0.485	0.005	0.211	0.001	0.017
G	0.003	0.032	0.001	0.012	0.005	0.095	0.003	0.036	0.000	0.006
H	0.005	0.063	0.009	0.074	0.010	0.151	0.018	0.316	0.002	0.025
I	0.002	0.033	0.003	0.031	0.003	0.070	0.008	0.184	0.001	0.015
J-O	0.001	0.018	0.001	0.015	0.003	0.086	0.006	0.079	0.000	0.010
P	0.003	0.055	0.006	0.094	0.010	0.209	0.018	0.485	0.002	0.038
Q	0.001	0.028	0.007	0.112	0.006	0.157	0.038	1.293	0.001	0.017

contains information regarding the characteristic of different buildings according to the economic activity that is being developed in them.

Among the data collected in the CBECS, the number of hours for which the buildings were weekly occupied was recorded. Since the study aims to account for the energy consumption and expenditures exclusively in buildings not all the studied economic activities have their equivalence in the CBECS. However, most of them can be associated with a certain type of building as it is shown in Table 6 where the equivalences adopted between our model and the CBECS are stated.

In this way, by using the codes normalized in the survey for the field primary building activity (PBA) the probability mass function for the hours of activity in each sector was calculated. Those distributions were subsequently compared with the ones obtained with the proposed model. The results of this comparison are shown in Fig. 11 where the black bars represent the probability distributions of the model, whereas the white bars correspond with the CBECS data.

In all the graphs, the X-Axis is the number of hours a day that the building is occupied, whilst in the Y-Axis the probability is represented. The main observed difference is that the model predicts with a relatively high probability short periods of occupancy which are not visible in the CBECS. Nevertheless, whereas our model

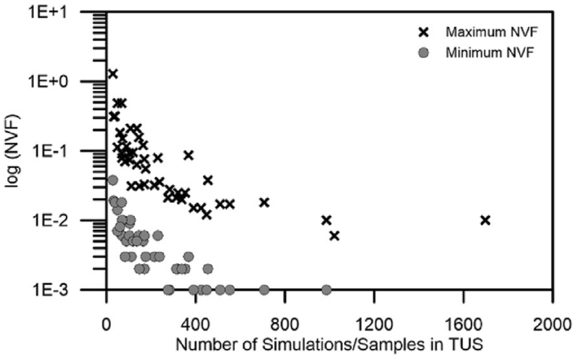

Fig. 10. NVF vs number of samples in the TUS/Number of employees simulated for the validation.

simulated the occupancy using the individual human behavior, the CBECS only studied the occupancy of the building taken as a whole. Thus, the existence of high probabilities for short periods is unlikely. Moreover, the results were obtained for different countries. The CBECS data belongs to the USA, whereas the model was developed with the data collected in the Spanish version of the TUS. Therefore, the significant differences in the schedules that exist directly affect the mean number of occupancy hours.

However, both studies present similarities that prove the capability of the proposed model for predicting the daily activity profiles. As it can be observed in Fig. 10 the jobs related to offices (J-O), the transportation (H), the education (P) and the human health (Q) sectors reach the maximum probability for similar intervals of occupancy. In contrast, in the wholesale and retail sector (G), as well as in the accommodation and food service sector (I) the CBECS projects occupancy periods larger than the proposed model. Nevertheless, as previously commented, this can be justified by the differences in schedules among countries. In this specific case, one of the behaviors that might provoke these differences is the period devoted to having lunch, since in Spain people usually go back home during that time, which results in lunch breaks longer than 2 h in most of the cases.

Finally, another important result that should be highlighted is the characteristic shape of both the CBECS and the model probability mass functions, which resemble a Weibull distribution. This indicates that in other fields of study where a high-resolution model is not needed, a probabilistic model can be an alternative to the Markov Chain processes whose accuracy is higher but at the expense of a more complex methodology and computational intensive calculations.

Table 6
Equivalences between the studied economic activities and the CBECS codes to filter the data.

Economic Activity	CNAE Group/s (codes)	CBECS Code for field PBA
Wholesale and retail trade	Group G (45–47)	25 – Retail other than mall
Transportation and storage	Group H (44–53)	05 – Nonrefrigerated warehouse
		11 – Refrigerated warehouse
Accommodation and food service	Group I (55–56)	15 – Food Service
		18 – Lodging
Offices related jobs	Groups J-O (58–84)	02 – Office
		04 – Laboratory
Education	Group P (85)	14 – Education
Human health and social work	Group Q (86–88)	08 – Outpatient health care
		16 – Inpatient health care
		17 – Nursing

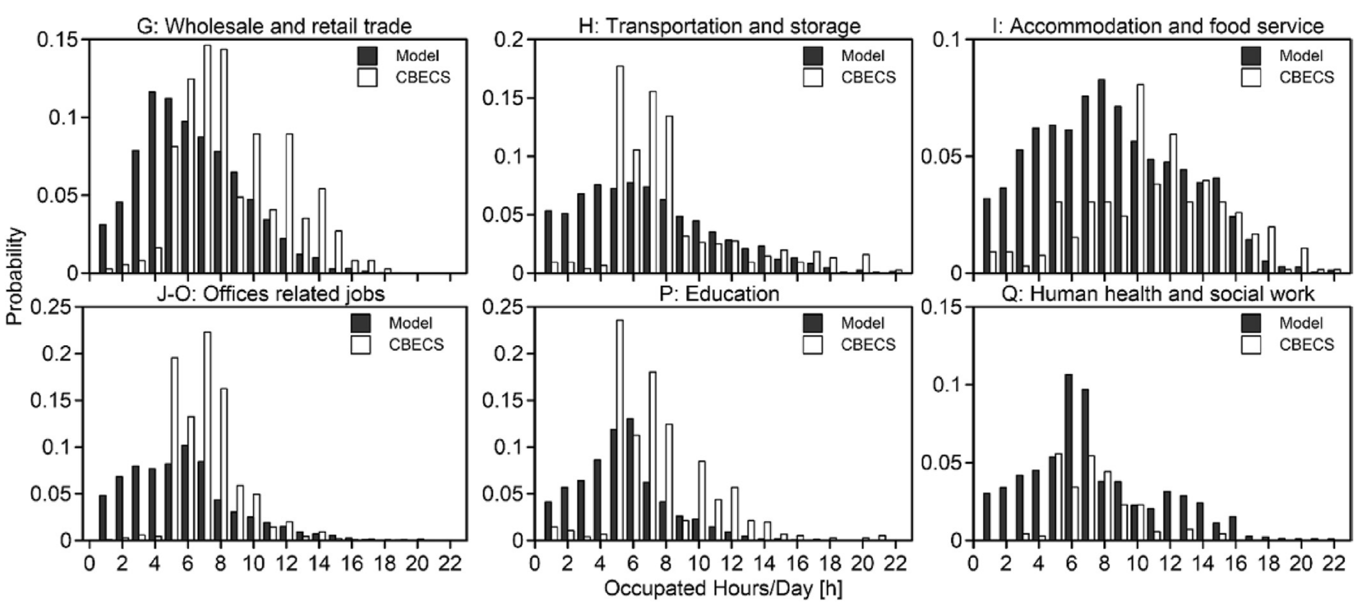

Fig. 11. Comparison between the mean daily hours of occupancy. Probability mass functions for the model and the CBECS.

6. The occupancy-energy link

The occupancy profiles provided by the model mean a valuable source of information by themselves. The mere occupancy data allowed identifying the main daily peaks where the consumption will be concentrated, as it was indicated in Tables 3 and 4, as well as the hours of activity in building-related sectors, as exposed in Fig. 11. Furthermore, its high flexibility also makes possible to use it as a base system for generating coherent profiles that can be used in other advanced buildings or energy simulations tools as the ones analyzed in Al-Homoud, 2001, Mohd-Nor and Grant, 2014 such as EnergyPlus, TRNSYS or BLAST.

Although nowadays smart meters and smart metering infrastructures provide detailed information regarding energy consumption in buildings, as well as other indicators such as voltage events or basic power quality indicators (Palacios-Garcia et al., 2017), there are still a wide variety of scenarios where metered data cannot be used. For instance, in a new construction building, since it has not been operating yet, no metered data are available. Nevertheless, the estimation of the building energy performance in the design stage is essential and is usually done by the aforementioned tools which require detailed input occupancy profiles such as the ones provided by this model. Likewise, smart meters data might provide an accurate estimation of the current consumption situation, but they cannot be used to predict future scenarios and the long-term impact that new energy policies or technological changes might have on the system.

Therefore, the previous knowledge of occupancy profiles is of special interest for new buildings planning, when detailed metering data are not available, or for assessing energy-saving actions or potential future scenarios. In order to illustrate the possibilities of the system, two modeling scenarios are proposed in the following subsections, both aiming to use the occupancy profiles as an input parameter in the energy quantification problem for some of the modeled sectors.

6.1. Top-down estimations

The top-down estimations or modeling processes are those in which the input or base parameters for the model are at a conceptual level above the consumption profiles. These data are usually socio-economic indicators, aggregated energy figures, employment rates, appliances usage rates, etc. In our case, two main indicators were used which are the annual electricity consumption $E_{year/emp}$ and the total working hours in a year $h_{year/emp}$, both figures referred to an employee in a given sector.

The results for the energy figures were obtained from the Odyssee-Mure project database (ADEME, 2017). This project, coordinated by ADAME, contains detailed energy consumption figures and indicators of all EU countries, as well as Norway, Switzerland and Serbia. Among the information, the main indicators regarding agriculture and services sectors were studied, which allowed us to obtain the figures included in Table 7. Regarding the working hours in a year for each selected sector, the statistic report on working hours of the National Statistics Institute of Spain was used (National Statistics Institute of Spain. Ministry of Economy and Competitiveness., 2000).

Using the data collected from both sources (ADEME, 2017) and (National Statistics Institute of Spain. Ministry of Economy and Competitiveness., 2000), it is possible to establish for each sector an average energy intensity for a worker performing the economic activity for an hour E_{hour} as it is indicated in (11).

$$E_{hour} = \frac{E_{year/emp}}{h_{year/emp}} \tag{11}$$

In the same way, if this hourly consumption wants to be directly related to one occupant activity, the 10-min interval associated energy that should be demanded E_{10min} can be obtained by dividing (11) by the corresponding temporal factor, six in this case as indicated in (12).

$$E_{10min} = \frac{E_{hour}}{6} \tag{12}$$

Using (11) and (12) the two last columns of Table 7 were calculated. In particular, the figures in the last column allowed directly associating the active number of employees predicted by the model $O_T(t|s)$ for given sector s to the 10-min energy intensity $E(t|s)$ by means of (13).

$$E(t|s) = O_T(t|s) \cdot E_{10min}(s) \tag{13}$$

By using this expression, the daily energy intensity profiles for the five selected sectors were calculated for an aggregate of 1000 employees in a weekday. As it can be observed in Fig. 12, although the total number of employees included in each economic sector is the same, the different consumption per occupant produced different energy intensities.

The sector with the lowest consumption was the education sector (P), followed by the human health and social work (Q) and the offices' sector (J-O). The wholesale and retail (G) and the accommodation and food service (I) sectors stood out as the largest consumers.

Furthermore, the 10-min resolution energy figures estimated by the occupancy model can be now used to predict back the aggregate figures of annual consumption based on the active hours. This will provide an indicator of the accuracy of the model. For that aim, the instantaneous figures of energy for weekdays $E_{10min}(t|s, \mathrm{WD})$ and weekend days $E_{10min}(t|s, \mathrm{WD})$ can be first aggregated to calculate the daily consumption of each type of day and type of working hours, and, subsequently, the annual energy intensity. Since in the studied region the percentage of working days in a year is 68%, a weighting factor was included for considering both types of days. The expression used in the annual energy estimation can be seen in (14).

Table 7
Average annual energy consumption per employee and total working hours in a year.

Economic Activity	$E_{year/emp}$ [kWh]	$h_{year/emp}$ [h]	E_{hour} [kWh]	P_{10min} [kWh]
Wholesale and retail trade	9881.44	1723.8	5.73	0.9554
Accommodation and food service	5409.18	1796.2	3.01	0.5019
Offices related jobs	3781.09	1660.8	2.28	0.3794
Education	2051.21	1449	1.42	0.2359
Human health and social work	3777.57	1663.9	2.27	0.3784

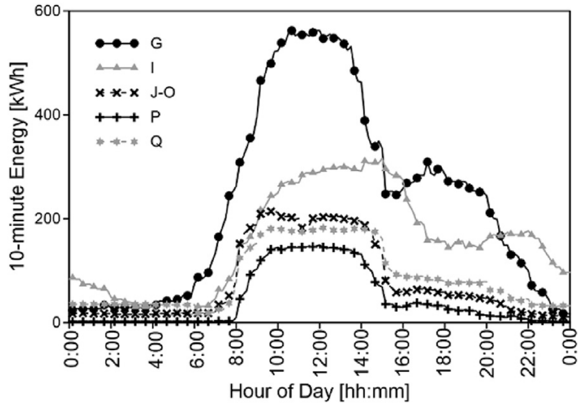

Fig. 12. Daily power profile for 1000 employees with continuous working hours in a weekday.

Table 8
Daily and annual energy figures estimated by the model and relative error to the reference values.

Sector	Continuous Working Hours				Split Working Hours			
	Daily [kWh]		Annual [kWh]		Daily [kWh]		Annual [kWh]	
	WD	WE	E_{year}	Err	WD	WE	E_{year}	Err
G	33.6	27.6	11570.0	17.1	37.3	28.2	12565.4	27.2
I	22.6	21.5	8124.2	50.2	26.5	19.9	8900.8	64.6
J-O	11.4	6.6	3590.3	5.0	11.9	7.5	3836.3	1.5
P	6.4	4.6	2119.5	3.3	7.1	2.9	2093.6	2.1
Q	12.4	7.9	4001.2	5.9	13.1	5.3	3858.6	2.1

$$E_{year}(s) = \left[\sum_{t=0}^{144} E_{10min}(t|s, \mathrm{WD})\cdot 0.68 + \sum_{t=0}^{144} E_{10min}(t|s, \mathrm{WE})\cdot 0.32\right]\cdot 365 \tag{14}$$

In this equation, each sum of the brackets represents the total daily energy for weekdays and weekends respectively, weighted by the scaling factor of working days and multiplied by the total number of days in a year. The results of these daily and annual figures per employee are listed in Table 8, where the relative error between the annual estimation and the reference annual energy figures in Table 7 was also indicated.

As it can be observed, the predictions are accurate for the office (J-O), education (P) and human health and social work (Q) sectors with relative errors lower than 6%. Nevertheless, in the case of the wholesale and retail (G) and the accommodation and food service (I) sectors the relative errors came to represent more than 25% and 50% respectively. The main reason of this mismatch may be attributed to the existence of a large number of devices such as cold storage room, machinery, shop window lighting, etc, which are not directly related to the occupancy but that were accounted in Table 7, therefore resulting in an overestimation of the final aggregate consumption. Hence, in these particular sectors, a previous knowledge of the baseload consumption is needed for assuring accurate estimations as the results depicted.

6.2. Bottom-up estimations

Together with the top-down estimation, an advantageous and novel use of the proposed model is the bottom-up simulation of plug loads consumption. In contrast to the previously presented methodology, the bottom-up approach uses indicators that are below the consumption profiles such as the human behavior, appliances operation cycles, daily activity schedules, etc.

Table 9
Individual appliances characteristic for the high temporal resolution power simulation.

Appliance	P_{ON} [W]	P_{OFF} [W]	Usage Probability
Desktop PC	100	5	1.0
LCD Monitor	16	3	1.0

As stated in the introduction, occupancy patterns are one of the main influence factors in the consumption of plug loads in offices. In contrast to HVAC or lighting systems that are usually centralized, individual loads are activated by the employees and its high-resolution energy consumption is difficult to predict. Therefore, an application case of the proposed model for a seamless energy estimation is shown in this section for the office sector (J-O) one of the larger consumers of these type of loads.

Two simple loads were simulated to show the capabilities of the model, desktop PCs and their associated LCD monitors. These office appliances were assigned an active power P_{ON} and an inactive power P_{OFF} extracted from a report in the USA evaluating the IT equipment in the office sector (Roth et al., 2002). Nevertheless, in a real scenario, the actual power rate of the devices can be used. In addition, each appliance was weighted with a usage probability of 1.0 meaning that every time an employee is active at the workplace the appliance will be active, the rest of the day, if no energy curtailment measures are implemented in the office, the appliance will consume the standby or P_{OFF}. The selected values can be seen in Table 9.

Using the proposed values and the usage probability per employee included in Table 9 for each appliance, the aggregate consumption of a 1000-employee office for a weekday with continuous working hours was simulated distinguishing the active and the inactive consumptions of the selected equipment. The results can be observed in Fig. 13, where the PC consumption is represented with black lines and the monitor energy with gray lines (solid for the active power and dashed for the inactive).

As it can be seen, due to the usage probability of 1.0, the appliances' consumption matches the occupancy pattern. Nevertheless, if additional appliances such as printers, fax or copy machines were included a probability lower than one and a decision mechanism for the switch on/off events should be provided. Furthermore, the system allowed observing the relatively high impact on

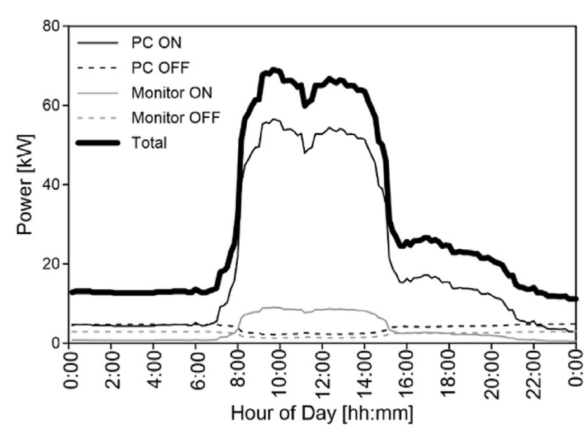

Fig. 13. Aggregate daily consumption profile for 1000-employee office in a working day for desktop PCs and LCD monitors.

Table 10
Daily and annual energy consumptions per appliance and employee. Estimated standby energy saving potential.

Appliance	Daily consumption per Employee		Yearly Consumption per Employee			Standby
	E_{ON} [Wh]	E_{OFF} [Wh]	E_{ON} [kWh]	E_{OFF} [kWh]	E_{Total} [kWh]	
Desktop PC	499.7	95.0	124.0	23.6	147.6	16%
LCD Monitor	79.9	57.0	19.8	14.1	34.0	41%

the load profile of the inactive power consumption of the selected equipment. The energy figures associated with this profile are listed in Table 10, where an estimation of the yearly consumption was included, considering the 68% of working days factor and only a weekday operation of the office.

As the figures depicted, the inactive power consumption of these appliances means a relatively high percentage of the yearly energy demand, being especially high for the LCD Monitors (41%). Thus, the model not only allowed us to estimate the yearly consumption of selected appliances based on their instantaneous demand but also helped highlight the problematic of standby devices in an office. Moreover, as well as the inactive demands, the assessment of appliances replacement campaigns with more efficiency devices could be carried out only by the knowledge of the occupancy profile, being a simple but effective tool in the estimation of future scenarios, something that even the most detailed monitoring or meter collection data cannot directly perform.

Therefore, the occupancy model proposed in the paper is able to establish a seamless link with the energy consumption in both the top-down and the bottom-up perspective. The first approach allowed easy global estimations with widely available data, but with difficulties to accurately represent the high temporal resolution results. The second, the bottom-up approach, showed very good results in the instant power profiles and the capability of assessing future scenarios, but with the limitation of the detailed knowledge of the appliances and devices that is required.

7. Conclusion

The proposed paper has addressed the development, implementation, simulation and validation of a stochastic model for the generation of daily occupancy patterns in different economic sectors with a high impact in the total energy demand of a country.

The TUS was highlighted as a useful source of information for this aim that together with the Markov Chains methodology and a stochastic simulation allowed obtaining the necessaries transition matrices for the process in the context of Spain. Nevertheless, the proposed methodology can easily be adapted to any other location due to the effort to harmonize the TUS at least within Europe.

Subsequently and using the proposed simulation method both individual and aggregated occupancy results for the different sector were obtained distinguishing between the type of day and the type of working hours. Those results showed the broad variety of daily patterns and pointed out the importance of having the type of day and the type of working hours as additional input parameters.

The results were validated using the original TUS information. Both the RMSE and the NVF were studied due to the stochastic nature of the process. Furthermore, the comparison with other works demonstrated the consistency of the proposed model with other analysis carried out in real buildings. These studies presented some differences in the probability distribution of the mean number hours of occupancy. However, a strong similarity between the probability function and the Weibull distribution was observed, being a simplified alternative for future works.

Finally, the occupancy profiles were used for estimating various energy indicators using the top-down and the bottom-up modeling philosophies. The top-down estimation results showed the accuracy of the model to estimate low-level consumption intensity curves based on global indicators when the main influence factor is the occupancy. However, it presented relatively significant errors in sectors with a large baseload consumption. On the other hand, the bottom-up approach indicated the possibilities of the model to be used as a prediction and assessment tool for future scenarios by means of simulating the high temporal resolution power profiles based on the working cycles of the devices.

Therefore, and considering all the above exposed, the proposed model stands out as an essential first step for the knowledge of the occupancy patterns and linked energy intensity in the selected sectors, offering a high temporal resolution in the results that also makes them an invaluable source of information for building simulation and assessment tools. Future works will be therefore focused on the usage of this occupancy model as the input dataset for the estimation of energy consumption in the commercial sector and the assessment of energy policies aiming to reduce the consumption and the integration of renewable energy sources.

Acknowledgements

This work was supported by the Spanish Ministry of Economy and Competitiveness under Research Project TEC2013-47316-C3-1-P; and the FEDER-INNTERCONECTA project Total Integrated GRid Intelligent System (TIGRIS) ITC-20131002, under contract no. 12013095.

References

ADEME, 2017. Odyssee-mure Project [WWW Document]. URL. http://www.odyssee-mure.eu/ (Accessed 11 January 2017).
Al-Homoud, M.S., 2001. Computer-aided building energy analysis techniques. Build. Environ. 36, 421–433. https://doi.org/10.1016/S0360-1323(00)00026-3.
Arghira, N., Hawarah, L., Ploix, S., Jacomino, M., 2012. Prediction of appliances energy use in smart homes. Energy 48, 128–134. https://doi.org/10.1016/j.energy.2012.04.010.
Official Bulletin of Spanish Government (BOE), 2009. No.102, 2009. R.D. 475/2007 April 13th, in Which Is Approved the National Classification of the Economical Activities for the Year 2009 (CNAE-. Ministry of Economy and Taxes, Spain.
Cardenas, J.A., Gemoets, L., Ablanedo Rosas, J.H., Sarfi, R., 2014. A literature survey on Smart Grid distribution: an analytical approach. J. Clean. Prod. 65, 202–216. https://doi.org/10.1016/j.jclepro.2013.09.019.
de Almeida, A., Fonseca, P., Schlomann, B., Feilberg, N., 2011. Characterization of the household electricity consumption in the EU, potential energy savings and specific policy recommendations. Energy Build. 43, 1884–1894. https://doi.org/10.1016/j.enbuild.2011.03.027.
Ding, Y., Hong, S., Li, X., 2014. A demand response energy management scheme for industrial facilities in smart grid. IEEE Trans. Ind. Inf. https://doi.org/10.1109/TII.2014.2330995, 1–1.
Eurostat, 2008. Statistical Classification of Economic Activities in the European Community. KS-RA-07-015-EN-N.
Gamerman, D., 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, Texts in Statistical Science.
Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov chain Monte Carlo in practice. Technometrics. https://doi.org/10.2307/1271145.
Grandjean, A., Adnot, J., Binet, G., 2012. A review and an analysis of the residential electric load curve models. Renew. Sustain. Energy Rev. 16, 6539–6565. https://doi.org/10.1016/j.rser.2012.08.013.
IDAE, 2015. Spanish Ministry of Industry Energy and Tourism. End energy figures for 2015 [WWW Document]. Annu. Rep. end Consum. energy Balanc. per Sect. energy type. URL. http://sieeweb.idae.es/consumofinal/bal.asp?txt=2015&tipbal=t.
Jia, M., Srinivasan, R.S., Raheem, A.A., 2017. From occupancy to occupant behavior:

an analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency. Renew. Sustain. Energy Rev. 68, 525–540. https://doi.org/10.1016/j.rser.2016.10.011.

Kabalci, Y., 2016. A survey on smart metering and smart grid communication. Renew. Sustain. Energy Rev. 57, 302–318. https://doi.org/10.1016/j.rser.2015.12.114.

Kim, Y.-S., Srebric, J., 2017. Impact of occupancy rates on the building electricity consumption in commercial buildings. Energy Build. 138, 591–600. https://doi.org/10.1016/j.enbuild.2016.12.056.

Lopez, M.A., Santiago, I., Trillo-Montero, D., Torriti, J., Moreno-Munoz, A., 2013. Analysis and modeling of active occupancy of the residential sector in Spain: an indicator of residential electricity consumption. Energy Pol. 62, 742–751. https://doi.org/10.1016/j.enpol.2013.07.095.

McKenna, E., Thomson, M., 2016. High-resolution stochastic integrated thermal-electrical domestic demand model. Appl. Energy 165, 445–461. https://doi.org/10.1016/j.apenergy.2015.12.089.

Mohd-Nor, M.F.I., Grant, M.P., 2014. Building information modelling (BIM) in the malaysian architecture industry. WSEAS Trans. Environ. Dev. 10, 264–273. https://doi.org/10.1016/j.buildenv.2006.10.027.

National Statistics Institute of Spain, 2000. Ministry of Economy and Competitiveness. Working Time Survey, Spain.

National Statistics Institute of Spain, 2010. Ministry of Economy and Competitiveness. Time Use Survey, Spain.

Palacios-Garcia, E.J., Chen, A., Santiago, I., Bellido-Outeiriño, F.J., Flores-Arias, J.M., Moreno-Munoz, A., 2015. Stochastic model for lighting's electricity consumption in the residential sector. Impact of energy saving actions. Energy Build. 89, 245–259. https://doi.org/10.1016/j.enbuild.2014.12.028.

Palacios-Garcia, E.J., Rodriguez-Diaz, E., Anvari-Moghaddam, A., Savaghebi, M., Vasquez, J.C., Guerrero, J.M., Moreno-Munoz, A., 2017. Using smart meters data for energy management operations and power quality monitoring in a microgrid. In: 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE). IEEE, pp. 1725–1731. https://doi.org/10.1109/ISIE.2017.8001508.

Palensky, P., Dietrich, D., 2011. Demand side management: demand response, intelligent energy systems, and smart loads. IEEE Trans. Ind. Inf. 7, 381–388. https://doi.org/10.1109/TII.2011.2158841.

Pérez-Lombard, L., Ortiz, J., Pout, C., 2008. A review on buildings energy consumption information. Energy Build. 40, 394–398. https://doi.org/10.1016/j.enbuild.2007.03.007.

Pournaras, E., Vasirani, M., Kooij, R.E., Aberer, K., 2014. Decentralized planning of energy demand for the management of robustness and discomfort. IEEE Trans. Ind. Inf. 10, 2280–2289. https://doi.org/10.1109/TII.2014.2332114.

Richardson, I., Thomson, M., Infield, D., 2008. A high-resolution domestic building occupancy model for energy demand simulations. Energy Build. 40, 1560–1566. https://doi.org/10.1016/j.enbuild.2008.02.006.

Richardson, I., Thomson, M., Infield, D., Delahunty, A., 2009. Domestic lighting: a high-resolution energy demand model. Energy Build. 41, 781–789. https://doi.org/10.1016/j.enbuild.2009.02.010.

Richardson, I., Thomson, M., Infield, D., Clifford, C., 2010. Domestic electricity use: a high-resolution energy demand model. Energy Build. 42, 1878–1887. https://doi.org/10.1016/j.enbuild.2010.05.023.

Roth, K.W., Goldstein, F., Kleinman, J., 2002. Energy Consumption by Office and Telecommunications Equipment in Commercial Buildings Volume I: Energy Consumption Baseline. Cambridge, MA.

Santiago, I., Lopez, M.A., Gil-de-Castro, A., Moreno-Munoz, A., Luna-rodriguez, J., 2013. Energy consumption of audiovisual devices in the residential sector: economic impact of harmonic losses. Energy 60, 292–301. https://doi.org/10.1016/j.energy.2013.08.018.

Strasser, T., Andren, F., Kathan, J., Cecati, C., Buccella, C., Siano, P., Leitao, P., Zhabelova, G., Vyatkin, V., Vrba, P., Marik, V., 2015. A review of architectures and concepts for intelligence in future electric energy systems. IEEE Trans. Ind. Electron. 62, 2424–2438. https://doi.org/10.1109/TIE.2014.2361486.

Swan, L., Ugursal, V., 2009. Modeling of end-use energy consumption in the residential sector: a review of modeling techniques. Renew. Sustain. Energy Rev. 13, 1819–1835. https://doi.org/10.1016/j.rser.2008.09.033.

Torriti, J., 2012. Demand side management for the european supergrid: occupancy variances of european single-person households. Energy Pol. 44, 199–206. https://doi.org/10.1016/j.enpol.2012.01.039.

U. S Energy Information Administration, 2012. Commercial Buildings Energy Consumption Survey (CBECS) [WWW Document]. URL. http://www.eia.gov/consumption/commercial/data/2012/.

United Nations, 2008. International standard industrial classification of all economic activities (ISIC). Stat. Pap. Series M 306.

Widén, J., Wäckelgård, E., 2010. A high-resolution stochastic model of domestic activity patterns and electricity demand. Appl. Energy 87, 1880–1892. https://doi.org/10.1016/j.apenergy.2009.11.006.

Zhou, S., Brown, M.A., 2016. Smart meter deployment in Europe: a comparative case study on the impacts of national policy schemes. J. Clean. Prod. 144, 22–32. https://doi.org/10.1016/j.jclepro.2016.12.031.

Zhou, B., Li, W., Chan, K.W., Cao, Y., Kuang, Y., Liu, X., Wang, X., 2016. Smart home energy management systems: concept, configurations, and scheduling strategies. Renew. Sustain. Energy Rev. 61, 30–40. https://doi.org/10.1016/j.rser.2016.03.047.