

DISCRETE OCCUPANCY PROFILES FROM TIME-USE DATA FOR USER BEHAVIOUR MODELLING IN HOMES

Dorien Aerts¹, Joeri Minnen², Ignace Glorieux², Ine Wouters¹ and Filip Descamps¹

¹Department of Architectural Engineering, Vrije Universiteit Brussel, Brussels, Belgium

²Department of Sociology, Vrije Universiteit Brussel, Brussels, Belgium

ABSTRACT

User behaviour plays a key role in the energy demand of residential buildings, and its importance will only increase when moving towards nearly-zero-homes. However, little information is available on how users interact with their homes. Due to the lack of information, user behaviour is often included in building performance simulations through one standard user profile. To obtain more accurate building simulations, we need user profiles that capture the wide variations in behaviour without making simulations overly complicated. To this end, we defined realistic occupancy profiles that include three possible states: (1) at home and awake, (2) sleeping or (3) absent. This paper reports on the methodology used to obtain these occupancy profiles based on the 2005 Belgian time-use survey, that contains detailed activity data of 6400 individuals from 3474 households. Using hierarchical clustering, we found seven profiles. These profiles include highly differentiated yet general behaviour that is relevant to building simulations.

INTRODUCTION

To obtain a more energy efficient building stock, we need accurate prediction and modelling methods for energy demand that take into account both building characteristics and user behaviour. The current energy performance calculation method ISO 13790 (ISO, 2007) focuses primarily on building characteristics. The result of this method is a theoretical energy consumption. However, once the dwelling is occupied the actual energy consumption may differ greatly from the predicted theoretical consumption (Bishop and Frey, 1985). Researchers believe that this difference is caused by the diversity of user behaviour, since they have found a wide variation of the energy consumption of dwellings with similar building characteristics (Beerepoot and Beerepoot, 2007; Guerra Santin et al., 2009; Gram-Hanssen, 2010; Andersen, 2012).

User behaviour influences the energy demand of a building both passively and actively. On the one hand, the presence of people in a building will lead to passive effects such as the change of heating or cooling demand, depending on the hygrothermal conditions in the building. On the other hand, active effects include the operation of control devices (e.g. window opening, lighting control, heating thermostat) or the use

of electrical appliances (e.g. computers, washing machines) (Mahdavi, 2011). Both effects are closely related: the presence of people is required for the majority of control actions and the use of most appliances. Understanding of both the passive and the active effects of user behaviour is needed when modelling (nearly-)zero-energy buildings because these buildings are primarily heated by the sun, metabolic heat of the users and heat emitted from electrical home appliances. However, these effects are difficult to predict because they are based on the user's preferences and habits rather than on purely rational choices.

For the modelling of user behaviour we may distinguish between deterministic and probabilistic approaches (Borgeson and Brager, 2008; Fabi et al., 2011). In the deterministic approach human behaviour is treated as a fixed schedule or is based on the assumption of purely rational behaviour. For example, internal gains are typically computed based on a fixed occupancy schedule. Control actions such as window opening or heating set-point adjustment are often linked with building related variables such as the indoor temperature or the CO₂ level. This approach is the simplest way of integrating user behaviour, but it has two important limitations. On the one hand, fixed schedules lead to repeatable and predictable behaviour. Consequently, the variations of behaviour are lost. On the other hand, it is not realistic to assume that users make perfectly rational choices, for example towards the control of the indoor environment. In many cases, personal preferences or habits play an important role in the decision making (Gram-Hanssen, 2009). Probabilistic models typically use statistical data to predict the probability that a certain activity occurs. Many of these profiles focus on one specific activity, for example the opening of windows (Herkel et al., 2008; Yun et al., 2009) or lighting control (Reinhart, 2004; Bourgeois et al., 2006). These probabilistic models take into account correlations between observed behaviour and building related variables such as the indoor temperature. As a result, these models capture more variations in behaviour and include "irrational" behaviour.

We developed a set of occupancy profiles for deterministic user behaviour modelling by performing cluster analysis on a Belgian time-use survey. This survey contains detailed information on the whereabouts

and activities of 6400 respondents from 3474 households with a time resolution of 10 minutes. For each of the respondents we analysed the occupancy data from their survey entries. We used hierarchical clustering analysis to detect similarities in the occupancy data. Using hierarchical clustering, we found seven profiles that apply to weekdays, Saturdays and Sundays. From the average occupancy data of these clusters, we formulated discrete occupancy profiles that are both simple and characteristic for a subgroup of the population. This paper presents the methodology used to develop these user profiles. First, it discusses the time-use survey data. Second, it describes the clustering techniques used to subtract profiles from the survey data. Finally, it illustrates the results with user profiles for weekdays.

METHODOLOGY

A wide range of parameters may influence household behaviour that ultimately determines the energy consumption. In literature, building related parameters such as the dwelling type, surface area of the dwelling and appliance holdings are frequently proposed to have strong explanatory power. Household related variables that are often put forward are income, number of household members, household composition and age. However, it is suggested by McLoughlin et al. (2012) that not these variables but their underlying features are the source of their explanatory power. For example, the age of the head of the home (HoH) appears to have an inverse effect on energy consumption. This may be due to the fact that middle-aged HoHs generally have more children living in the home, resulting in a higher energy consumption. Retired HoHs are likely to spend more time at home, explaining higher energy consumption. Ultimately, as suggested by Yao and Steemers (2005), it all boils down to the number of people and the amount of time they spend at home.

Time-Use Survey

We derived realistic occupancy data from the combined Belgian Time-Use Survey (TUS) and Household Budget Survey (HBS) that were collected in 2005 (Glorieux and Minnen, 2008). The combined surveys TUS and HBS include 6400 respondents from 3474 households. In the TUS all household members over twelve years old were asked to complete the diary for the same two days; one weekday and one weekend day. In these diaries the respondents described their activities and movements from 4:00 AM until 3:50 AM the next day. The Belgian diaries use a continuous registration system with fixed time slots of 10 minutes. The continuous registration system forces respondents to provide an activity for each time slot, preventing gaps in the time-use pattern. For each of these time slots, the respondent could enter up to two activities, a primary activity and a secondary activity, so that multitasking may be captured. The activities

are described by the respondent in his/her own words, and afterwards recoded into 272 activity codes. Furthermore, for each activity the respondents mentioned if they were at home and if they were accompanied by someone. In addition to the TUS, a HBS was collected that comprises individual and household questionnaires. The former include information about the age, position within the household, education, income and employment, whilst the latter contain details about the family home, the ownership of (electrical) appliances and vehicles, as well as their expenditures on goods and services.

The large TUS dataset allows us to analyse general occupancy patterns and, more importantly, study correlations between behaviour and socio-economic characteristics. To this end, we generated three-state occupancy patterns for all 6400 respondents by extracting the activity and location data from their diaries.

For each time step we included three possible states: (1) at home and awake, (2) at home and sleeping or (3) absent. Each of these state have an impact on the likelihood of energy consumption in the dwelling. Energy consumption is most likely to occur during state 1, whilst users are at home and awake, since they are able to control appliances, lighting and heating. Whilst asleep, no control actions occur but the users' presence in the home is taken into account. When users are absent, their actions are of no relevance to the energy balance of the home.

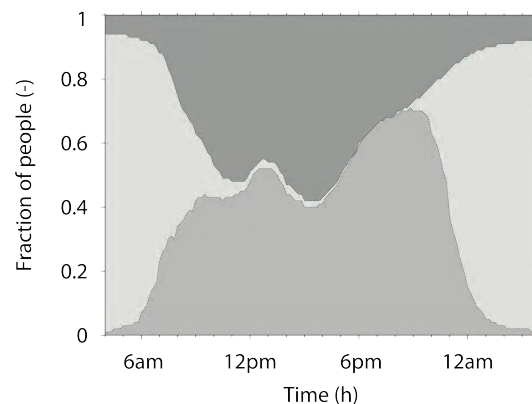


Figure 1: Average Occupancy profile for the population

Average occupancy data for the whole population can easily be deduced from the TUS database, but more specific occupancy profiles are needed to understand the diversity of user behaviour. The average occupancy profile for the Belgian population (fig. 1) shows three coloured zones that represent the three occupancy states (at home, sleeping, absent). The horizontal axis represents the time of day, whilst the vertical axis illustrates the fraction of people within the population that are engaged in a certain state. Clearly, the vast majority of people is asleep between midnight and 5 AM. Between 8 AM and 6 PM, however, large variations in behaviour are found, which indicates that one standard user is insufficient. Therefore, it is worth-

while to investigate whether the population can be divided in subgroups that each show distinct, characteristic behaviour.

Hierarchical Clustering

We performed hierarchical cluster analysis on the TUS dataset to discover groups of the population that show different behaviour. Hierarchical clustering is a method of cluster analysis that orders differences between elements in a dataset hierarchically, enabling the definition of clusters on the needed degree of precision (Jain and Dubes, 1988; Jain et al., 1999; Mirkin, 1996). The differences, or distances, between the elements depend on the chosen metric. In our case, each element represents one occupancy profile. Each profile in turn is constructed from 144 digits, one for each 10-minute time step, with a value that corresponds to one of the three possible occupancy states. Each element was handled as a string.

Two well-known metrics for strings are the Hamming distance and the Levenshtein distance (Jacobs, 2004). The Hamming distance was designed to compare two strings of the same length. The i -th character of the first string is compared with the i -th character of the second string. Every mismatch between two corresponding characters has a cost of one. The total cost, or distance, is the amount of changes needed to convert one string into the other string and may vary between zero and the length of the strings. The Levenshtein distance, sometimes referred to as the edit distance, is defined as the minimum number of edits needed to transform one string into the other, allowing insertions, deletions and substitutions (Levenshtein, 1966). Each edit operation may have a different cost. However, the Levenshtein distance typically uses a cost of one for each edit operation.

We used the Levenshtein distance since the distance definition is more nuanced. Whilst the Hamming distance only compares corresponding characters, the Levenshtein distance allows insertions and deletions. As a result, strings that are almost identical but shifted by one character will be attributed a much higher difference using the Hamming distance than to the Levenshtein distance. For example, the strings 'abcdefg' and 'bcdefga' will have a distance of 7 using the Hamming distance, but only a distance of 2 using the Levenshtein distance (one deletion at the beginning and one insertion at the end of the string). For our time series, a shift of the series by one character (= one 10-minute time step) should only have a small effect on the distance.

The results of the hierarchical clustering are presented in a dendrogram or tree diagram (fig. 2). The higher the branches split, the higher the distance between elements within two branches.

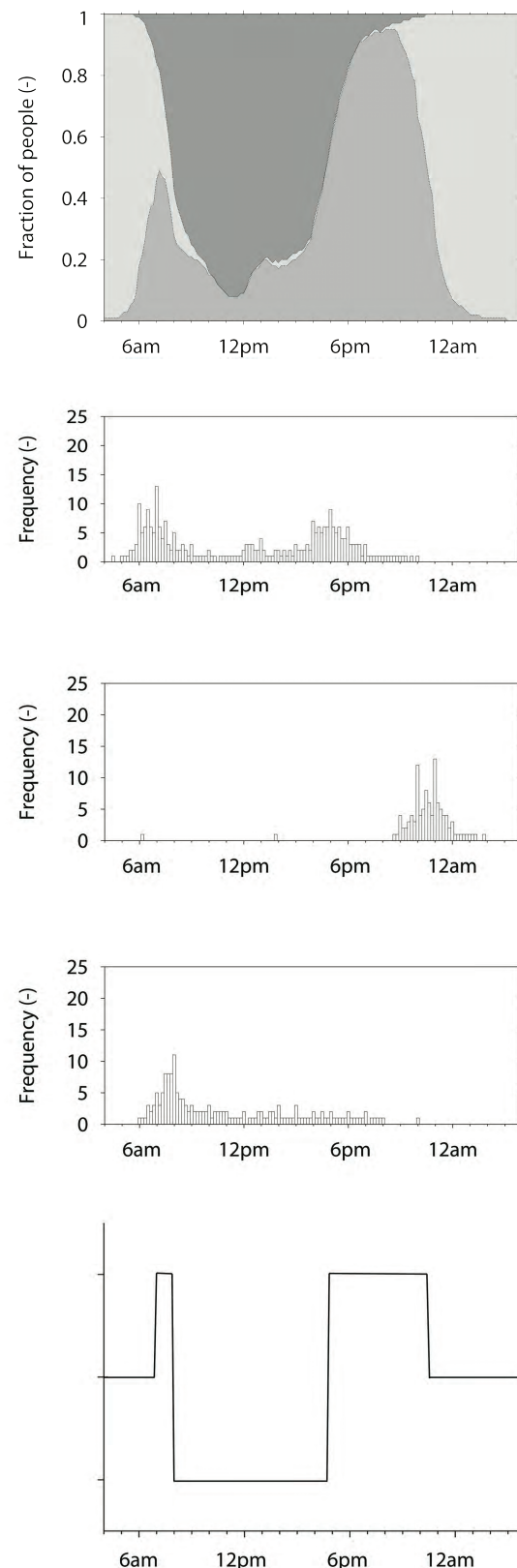


Figure 3: Discrete profile (e) based on the average profile (a) and the start distributions (b,c,d) - Example methodology for profile 2

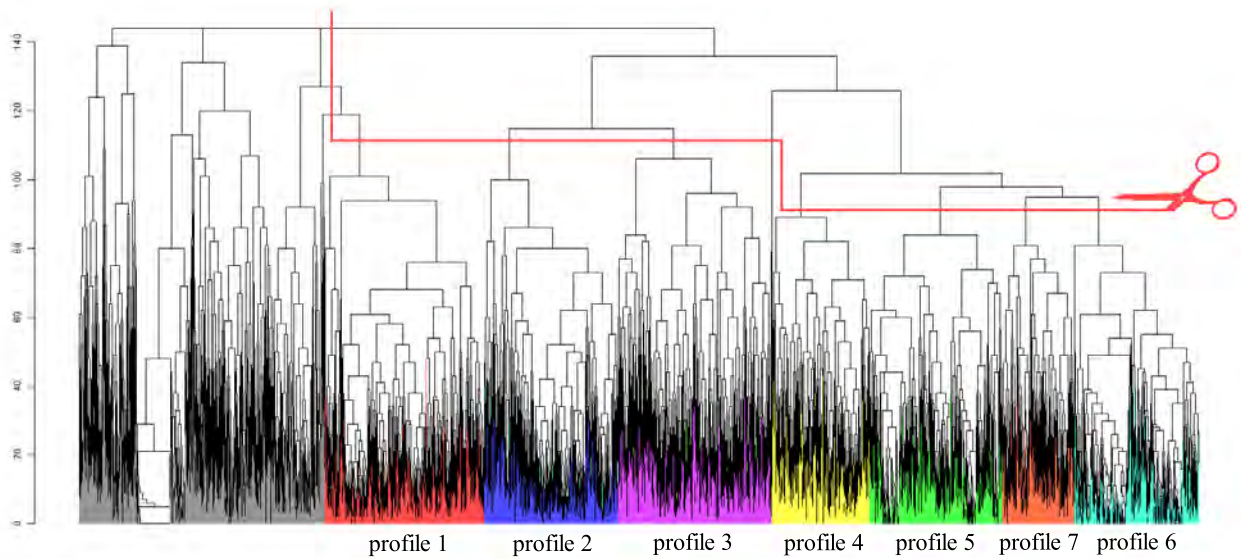


Figure 2: Results of the hierarchical clustering

Occupancy Profiles

Using hierarchical clustering, we found seven profiles that apply to weekdays, Saturdays and Sundays. However, some profiles are more prevalent during weekdays, whilst others might apply primarily to Saturdays and/or Sundays (table 1). For reasons of brevity, this paper only discusses the results for weekdays.

Occupancy profiles were constructed from the results of the clustering analysis by applying hierarchical clustering on the dataset (fig. 2). We found seven weekday profiles with relatively strong average behaviour that showed branch nodes at approximately the same height in the dendrogram. In total, 81.4% of weekday entries in the database were assigned to one of the user profiles. The remaining users could not be clustered into a meaningful, strong average profile.

Table 1: Profile numbers according to increasing occupancy level

Profile	Weekdays	Saturdays	Sundays
1	21.3%	8.2%	6.2%
2	17.2%	5.2%	7.8%
3	10.0%	16.1%	18.3%
4	6.9%	9.1%	11.6%
5	6.8%	7.2%	4.9%
6	8.2%	12.3%	18.5%
7	11.0%	9.6%	12.7%
Coverage	81.4%	67.7%	80.0%

Although average occupancy profiles contain valuable information, we need discrete occupancy profiles that can be implemented in building simulation software. These discrete profiles were obtained from the average occupancy profiles and from the distribution of the start of each state. The start state distribution (fig. 3b-d) shows a histogram for each of the occupancy states.

Peaks in the histogram indicate that a high number of users switched to this particular occupancy state at a particular time of day. These peaks were used to discretize the user profiles. The average profiles (fig. 3a) were used to ensure that the total amount of time spent on each state fits with the average data. The discrete occupancy profiles that are obtained (fig 3e) may be considered to be representative for their user profiles. A full set of discrete profiles is provided in fig. 4.

RESULTS

To acquire better insight in user behaviour, we analysed the composition of the clusters behind each occupancy profile based on a number of socio-economic variables fig. 4. These variables are employment, individual income, household size and age. For reasons of brevity, this paper only discusses these variables for weekdays.

For weekdays, the strongest correlation between is found for the employment type. Most children and respondents who are employed full time are captured within occupancy profiles 1, 2 and 3, three profiles with a large period of absence during daytime. Conversely, inactive and retired respondents are found most frequently in profile 4, 5, 6 and 7, four profiles with low absence during the daytime. People who work part time are found in all profiles, which is caused by the fact that only one weekday was included in the diary. This day may either be a day they worked all day, part of the day or not at all.

As expected, age is closely related to employment. The majority of respondents aged between 12 and 55 are found in profiles with a high-absence profiles, whilst respondents older than 55 are typically charac-

terised by low-absence profiles.

Respondents with a higher income are increasingly likely to be found in the profiles with high absence during the day. Respondents with very low incomes, possibly unemployment allowances, are typically represented by the low-absence profiles. An exception to this finding are children under 18, who clearly do not receive an income but nevertheless show absence during daytime.

The profiles mainly differ during daytime, sleeping patterns only show small variations. During daytime, users are either absent or at home. Key variables to distinguish between profiles are the (1) the time at which users leave the house and (2) the duration of their absence. We found little variation in the time of day people wake up or go to sleep.

DISCUSSION

We presented a set of profiles that describe the occupancy behaviour of users and linked these profiles to the employment type and age of the users. The profiles may be implemented in building simulations. Also, they may serve as a foundation for the improvement of user behaviour calculations in energy performance regulations.

Implementation in Building Simulations

The set of occupancy profiles we presented in this paper is based on the deterministic approach. These profiles are suitable for building energy simulations, keeping in mind the limitations regarding the predictability and the lack of interactions between user and building. In some cases these limitations may not be important, for example when the impact of different scenarios is tested or for benchmarking purposes. The main contributions of these profiles are the identification of characteristic behaviour for subgroups of the population and connections between these subgroups and a number of socio-economic variables.

In future work, we will develop a probabilistic model based on the current deterministic profiles. This way, the assets of the profiles are combined with the advantages of probabilistic models. The combination of the current profiles with a probabilistic model is valuable for energy modelling, including the modelling systems of distributed generation on neighbourhood level or for smart-grid applications. In these cases, where the equilibrium between energy demand and production is important, the implementation of realistic variations of user behaviour is crucial.

Implementation in Building Standards

We will derive an updated method for the calculation of internal gains in energy performance standards. The current energy performance calculation method ISO 13790 (ISO, 2007) states that internal gains from occupants and appliances should be determined on a national basis. In the calculation method for Belgium, internal gains only depend on the building vol-

ume. The combination of our user profiles and socio-demographic data enables the development of a 'standard user' that is representative for the Belgian population. This should improve the accuracy of the calculation method.

Other future work

This methodology for occupancy profiles will be repeated for activities that might lead to energy consumption. The use of electrical appliances (e.g. computer, television, dishwasher) and hot water (e.g. shower, cooking) will be analysed based on the TUS diary entries. To estimate household energy consumption, the activity data will be linked to the appliance ownership data from the HBS. Similar to the occupancy profiles, a probabilistic model will be constructed to include variations in behaviour. Whilst we analysed occupancy on an individual level, activities will be analysed on household level.

CONCLUSION

We developed a set of user profiles that reflects realistic user behaviour in homes. We used an existing Belgian time-use survey (TUS) to analyse actual user behaviour. On the TUS data we applied an hierarchical clustering algorithm to detect significant differences in the Belgian user behaviour. We used the Levenshtein distance to quantify the differences between the occupancy patterns. The clustering resulted in seven profiles that apply to weekdays, Saturdays and Sundays.

For weekdays, we found relationships between employment and the seven corresponding profiles. Users that are employed full-time are mainly situated in two profiles with low occupancy levels during daytime. Conversely, those who are either unemployed or retired are largely represented by two profiles with high occupancy levels during daytime.

The presented user profiles may be used for user behaviour modelling in building energy simulations. These profiles were obtained deterministically, which leads to a easy implementation. Limitations of the deterministic approach are the predictability and the lack of interaction between user and building. These limitations will be addressed in future work by developing a probabilistic model based on the presented profiles. The main contributions of these current profiles are the identification of characteristic behaviour for subgroups of the population and connections between these subgroups and a number of socio-economic variables.

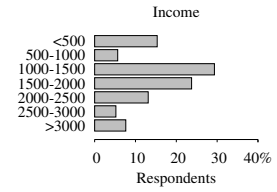
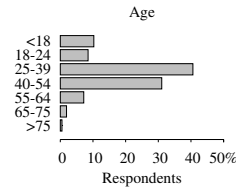
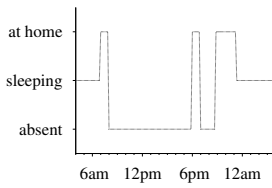
ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support received for this work from the Brussels Institute for Research and Innovation (INNOViris).

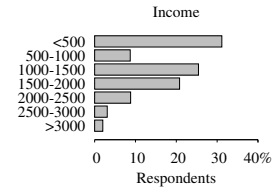
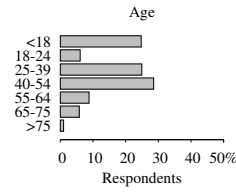
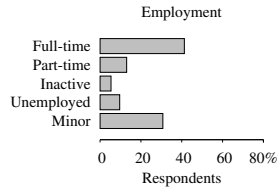
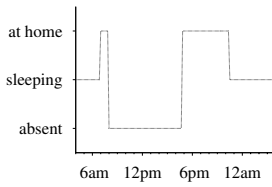
REFERENCES

Andersen, R. K. 2012. The influence of occupants behaviour on energy consumption investigated in 290

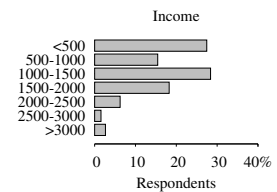
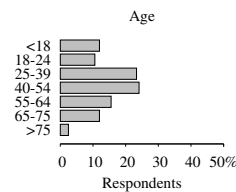
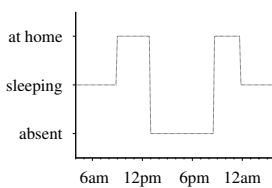
Profile 1 (21.3 %)



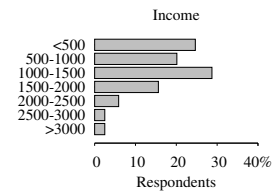
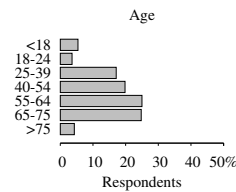
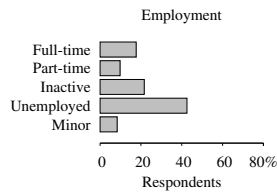
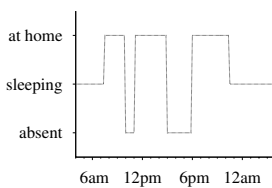
Profile 2 (17.2 %)



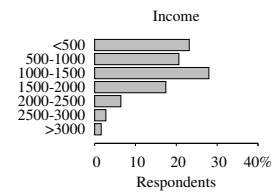
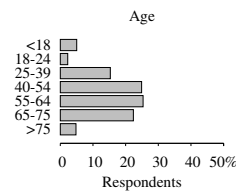
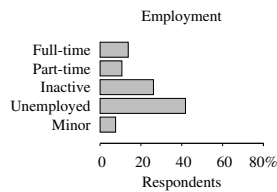
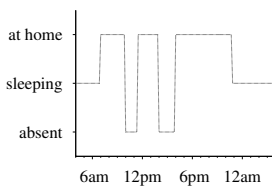
Profile 3 (10.0 %)



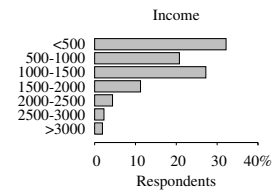
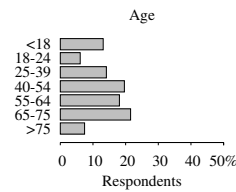
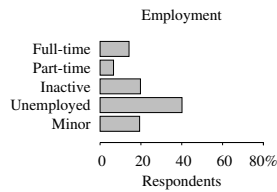
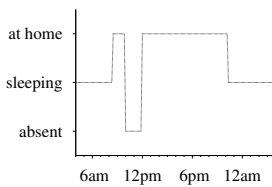
Profile 4 (6.9 %)



Profile 5 (6.8%)



Profile 6 (8.2 %)



Profile 7 (11.0%)

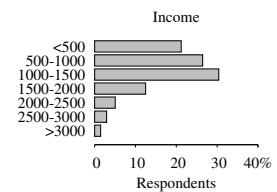
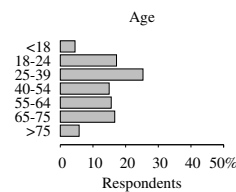
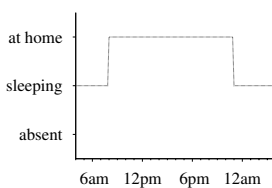


Figure 4: Full set of occupancy profiles for weekdays

- identical dwellings and in 35 apartments.
- Beerepoot, M. and Beerepoot, N. 2007. Government regulation as an impetus for innovation: Evidence from energy performance regulation in the Dutch residential building sector. *Energy Policy*, 35(10):4812–4825.
- Bishop, R. and Frey, D. 1985. Occupant effects on energy performance of monitored houses. In *Proceedings of: The 10th National Passive Solar Conference*, pages 395–400.
- Borgeson, S. and Brager, G. 2008. Occupant Control of Windows: Accounting for Human Behavior in Building Simulation. Technical report, Center for the Built Environment (CBE), University of California, Berkeley.
- Bourgeois, D., Reinhart, C., and Macdonald, I. 2006. Adding advanced behavioural models in whole building energy simulation: A study on the total energy impact of manual and automated lighting control. *Energy and Buildings*, 38(7):814–823.
- Fabi, V., Andersen, R. V., Corgnati, S. P., Olesen, B. W., and Filippi, M. 2011. Description of occupant behaviour in building energy simulation: state-of-art and concepts for improvements. In *Proceedings of Building Simulation 2011*, pages 2882–2889.
- Glorieux, I. and Minnen, J. 2008. website Belgisch tijdsbudgetonderzoek www.time-use.be.
- Gram-Hanssen, K. 2009. Standby Consumption in Households Analyzed With a Practice Theory Approach. *Journal of Industrial Ecology*, 14(1):150–165.
- Gram-Hanssen, K. 2010. Residential heat comfort practices: understanding users. *Building Research & Information*, 38(2):175–186.
- Guerra Santin, O., Itard, L., and Visscher, H. 2009. The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. *Energy and Buildings*, 41(11):1223–1232.
- Herkel, S., Knapp, U., and Pfafferott, J. 2008. Towards a model of user behaviour regarding the manual control of windows in office buildings. *Building and Environment*, 43(4):588–600.
- ISO 2007. *Thermal performance of buildings - calculation of energy use for space heating and cooling ISO/FDIS 13790*, volume 2007.
- Jacobs, N. 2004. *Relational sequence learning and user modelling*. PhD thesis, Katholieke Universiteit Leuven.
- Jain, A., Murty, M., and Flynn, P. 1999. Data Clustering : A Review. *ACM Computing Surveys*, 31(3):264–323.
- Jain, A. K. and Dubes, R. C. 1988. *Algorithms for clustering data*. Prentice Hall advanced reference series. Prentice Hall.
- Levenshtein, V. 1966. Binary codes capable of correcting deletions insertions and reversals.pdf. *Soviet Physics Doklady*, 10(8):707–710.
- Mahdavi, A. 2011. Chapter 4: People in building performance simulation. In *Building Performance Simulation for Design and Operation*, pages 56–83. Routledge.
- McLoughlin, F., Duffy, A., and Conlon, M. 2012. Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. *Energy and Buildings*, 48:240–248.
- Mirkin, B. G. 1996. *Mathematical Classification and Clustering*. Nonconvex Optimization and Its Applications Series. Kluwer Academic Pub.
- Reinhart, C. F. 2004. Lightswitch-2002: a model for manual and automated control of electric lighting and blinds. *Solar Energy*, 77(1):15–28.
- Yao, R. and Steemers, K. 2005. A method of formulating energy load profile for domestic buildings in the UK. *Energy and Buildings*, 37(6):663–671.
- Yun, G. Y., Tuohy, P., and Steemers, K. 2009. Thermal performance of a naturally ventilated building using a combined algorithm of probabilistic occupant behaviour and deterministic heat and mass balance models. *Energy and Buildings*, 41(5):489–499.