

# Occupant behavior in identical residential buildings: A case study for occupancy profiles extraction and application to building performance simulation

Antonio Muroi<sup>1,2</sup>, Isabella Gaetani<sup>1,3</sup> (✉), Pieter-Jan Hoes<sup>1</sup>, Jan L. M. Hensen<sup>1</sup>

1. Building Physics and Services, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

2. Enel S.p.A., Viale Regina Margherita, 137, 00198 Rome, Italy

3. Arup, 13 Fitzroy St, Bloomsbury, W1T 4BQ London, UK

## Abstract

This study employs a simplified Knowledge Discovery in Database (KDD) to extract occupancy, equipment and light use profiles from a database referred to 12 all-electric prefabricated dwellings in the Netherlands. The profiles are then integrated into a building performance simulation (BPS) model using the software TRNSYS v17. The significance of the extracted profiles is verified by comparing the total and end-use yearly electricity consumption of the investigated dwellings as predicted by the simulation tool with on-site measurements. For the considered dwellings, using standard OB modeling results in an underestimation of the energy use intensity (EUI) by 5.9% to 42.5%, depending on the case. The integration of the occupant behavior (OB) profiles improves the total electricity consumption prediction from an initial 22.9% average deviation from measurements to 1.7%. The results corroborate that the 1.6x discrepancy observed in the buildings' energy use intensity could be entirely ascribed to OB. Then, the knowledge extracted from the households' database is used to propose a local electricity market framework to reduce the electricity bill and grid dependency of all households. This study confirms the need for appropriate OB modeling in BPS, it shows the potential of the KDD method for successful OB profiles extraction, and is a first example of data-mined OB profiles integration in BPS, as well as of OB profiles deployment for a practical application other than energy use prediction.

## Keywords

Occupant behavior, identical dwellings, data-mining, occupant behavior profiles

## Article History

Received: 18 March 2019

Revised: 28 July 2019

Accepted: 29 July 2019

© The Author(s) 2019

## 1 Introduction

In recent years, the importance of occupant behavior (OB) for building energy performance has been widely recognized (Zhang et al. 2014; Attia et al. 2013; Hensen 2011; Daniel et al. 2015). The relative impact of OB on building performance is shown to increase as building standards become more stringent and building envelopes and systems more efficient (Hong and Lin 2012; Clevenger and Haymaker 2006). In residential buildings, such influence appears to be even more crucial, due to a higher level of freedom and control over the indoor environment (Urban and Gomez 2013; Andersen 2012; Bahaj and James 2007; Saldanha and Beausoleil-Morrison 2012; Gram-Hansen 2010; Maier et al. 2009; Juodis et al. 2009). As pointed out by Guerra Santin et al.

(2009), different household types and occupancy patterns can lead to a variation in the electricity consumption with a factor of 3 in the Dutch building stock. A similar study by Andersen (2012) confirmed the importance of OB in buildings' energy consumption, finding a factor 20 difference in the heating consumption of 290 identical townhouses in Denmark. Since the buildings considered in the study were standardized, OB appears to be the main reason for the discrepancies in the results. Also in similar low energy houses (Bahaj and James 2007), different OB patterns led to a dramatic difference in energy consumption, which is shown to reach up to 80% during certain periods of the year.

An incorrect evaluation of occupants' influence on buildings may query the reliability of simulation results due to discrepancies between predicted energy performance

E-mail: i.i.gaetani@tue.nl

and actual one (Yu et al. 2011). A number of studies (Raftery et al. 2011; Pan et al. 2007; Monetti et al. 2015; Royapoor and Roskilly 2015) show that wrong assumptions related to occupants are among the main causes for the incorrect building performance estimation by simulation tools. Samuelson et al. (2015) performed a calibration study to show how on-site measured data can reduce this discrepancy, improving the gap between simulation and measured consumption from an initial deviation of 36% to a net 7%.

To address this issue, BPS developers are focusing their attention on defining general and robust methods to predict and model OB in buildings through different approaches (Yan et al. 2015; Gaetani et al. 2016; Yu et al. 2011). Real-time measured data, collected in building databases, have found a direct application in non-probabilistic models. The data mining (DM) process allows determining patterns of behavior and general occupants' profiles (Yu et al. 2013; D'Oca et al. 2015; Yu et al. 2016), which provide a more efficient alternative to real-time data for integration in BPS and have a higher generalizability potential. Yu et al. (2013) and Fan et al. (2015) developed two general frameworks for the application of DM in the building sector. They included a step-by-step analysis process from the problem definition to the knowledge discovery. In a more specific perspective, Basu et al. (2013) developed a decision tree-based model to predict the usage of home appliances for the hour ahead. D'Oca and Hong (2015) proposed a three-step DM learning framework to extrapolate occupancy patterns and user profiles from big data streams.

Currently, DM-techniques have been applied to well-established databases characterized by high quality of data in terms of reliability, completeness, consistency, and resolution. However, data storage systems seem to have found a wide implementation mainly in office buildings, where the interest to reduce consumption and thus energy expenditure is higher than in other sectors. In the residential sector, monitoring systems are still not common and data analysis is usually carried out with highly summarized data (e.g. monthly, annual energy bills), which does not allow to perform a detailed analysis nor to extract information about occupancy state. The few existing studies which concern residential buildings are primarily dedicated to extracting OB profiles. Due to the relatively recent nature of this field of investigation, little research has been devoted to the verification of the extracted profiles, their implementation in BPS software or their use for real-life applications. This study seeks to fill this knowledge gap.

The first objective of this study is to improve building energy performance predictions in residential buildings by identifying occupant behavior (OB) profiles and household types using DM-techniques. OB is analyzed in identical prefabricated dwellings in the Netherlands through an online

database which collects real-time data. OB profiles are derived from the on-site measured data in a simplified knowledge discovery in database (KDD) process and integrated into a BPS model with the software TRNSYS. The improvement in the building performance prediction is estimated by comparing the simulation results, after the integration of the extracted profiles, with the actual measured data.

A more accurate knowledge regarding OB can be useful when modeling a realistic energy demand per household in the design of neighborhoods with connected buildings (e.g. district system, smart grid, etc.). As a further objective of this study, the derived household types and OB profiles are used to investigate the potential of a local electricity market framework for a district. A local electricity market is a cooperative system in which the members of a community can exchange the on-site generated but not consumed electricity within the neighborhood, avoiding to feed it into the grid. This system allows reducing the electricity expenditure and the related carbon dioxide (CO<sub>2</sub>) emissions. At the same time, relying on locally generated electricity increases the self-consumption and energy matching for the whole community. In this study, the benefits of the local market in terms of bill reduction and grid-independency are estimated with the on-site energy matching (OEM) and on-site energy fraction (OEF) indicators.

This paper is structured as follows: the characteristics of the buildings under investigation are presented in Section 2. The steps followed to perform the study are presented in Section 3. The results of the database analysis, the building simulation and the local electricity market analysis are reported in Section 4. These results are discussed and interpreted in Section 5, with a focus on the issues and limits encountered. Finally, the conclusions are drawn in Section 6.

## 2 Dwellings under investigation

The building database refers to 150 all-electric terraced dwellings built in the Netherlands since 2014 (Fig. 1). The houses are characterized by low-energy demand with high insulation of the building envelope and highly efficient equipment installed. The embedded system includes a 4-kW air-to-water heat pump to cover the space heating and domestic hot water demand, and a 3-kW balanced ventilation unit with heat recovery for the fresh air supply. The electricity demand is partly covered on-site by a 5.5-kW rooftop PV system, reducing the import of electricity from the national grid. The monitoring system, installed in each dwelling, registers and collects online a large amount of data, including the CO<sub>2</sub> concentration in the indoor environment, the imported and exported electricity, the usage of installed appliances and the water temperature in the system. The

dwellings are all prefabricated and heavily standardized, with the same number of floors, layout, installed equipment and thermo-physical properties. Table 1 summarizes the main characteristics of the dwellings under investigation.

The dwellings are designed according to the concept of Nota Null (Dutch: “Zero Energy Bill”), which should guarantee a net zero electricity bill on annual basis. This result is achievable through net metering and tax compensation for the installation of renewable technologies (Boekhoud and Behrendt 2013). However, the goal is not always reached, with several households finding themselves to pay a certain amount for the imported electricity at the end of the year.

**Table 1** Characteristics of the dwellings under investigation

	Building characteristics
Dimensions [m <sup>3</sup> ]	9.00 × 5.70 × 9.53 ( <i>h</i> )
Location	the Netherlands
Year of construction	2014-on going
<i>U</i> -value [W/(m <sup>2</sup> ·K)]	Walls: 0.20; windows: 0.93
Number of floors	3 (ground, first floor and attic)
Observed floors	Ground floor (day activity) and first floor (night activity)

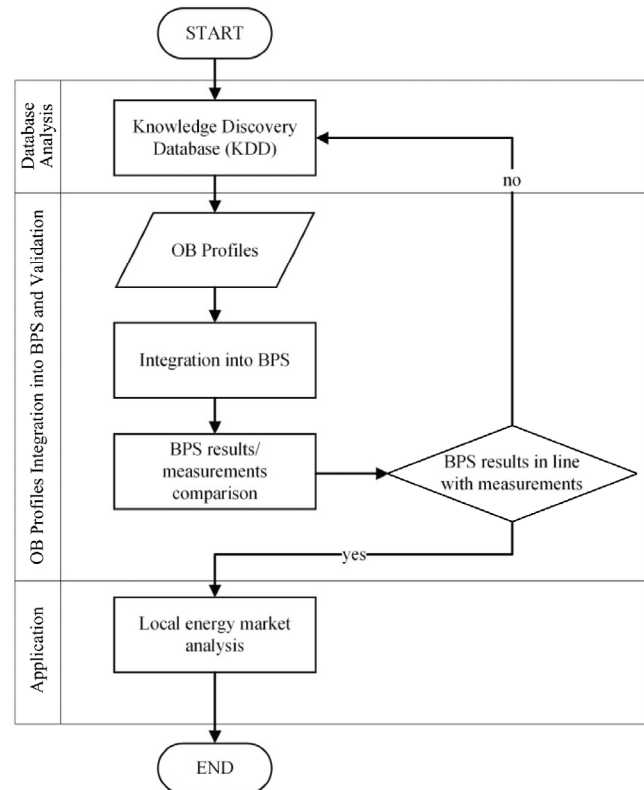


**Fig. 1** External view of the dwellings under investigation

### 3 Methodology

The steps followed throughout this study belong to three main sections, as displayed in Fig. 2. In the first section (Section 3.1), the OB profiles were extracted for the selected households through a simplified Knowledge Discovery in Database (KDD) process.

In the second section (Section 3.2), the extracted OB profiles were integrated into a building simulation model developed in TRNSYS v17. They were verified by comparing the electricity consumption predicted in the simulation model with the actual one measured in the database, both



**Fig. 2** Methodology overview

on a yearly and monthly basis. The comparative analysis was performed for the total and end-use consumption.

In the third section (Section 3.3), a direct application of the study’s results was proposed. The knowledge extracted from the database regarding OB was used to propose a local electricity market framework. The benefits of such system were evaluated through OEM/OEF indicators.

#### 3.1 Database analysis

In this study, a simplified KDD process is used to analyze the database. The KDD is gaining popularity as a partially automated process of identifying valid, useful, and ultimately understandable patterns in data (D’Oca et al. 2015; D’Oca and Hong 2014). It involves the application of six steps (Hong et al. 2015):

- *Data selection*: the creation of a target data set, namely the identification of a subset of variables or data samples, on which the analysis should be performed.
- *Data cleaning and pre-processing*: removals of outliers and missing data fields to reduce the errors in the KDD results.
- *Data transformation*: finding useful features to represent the data depending on the main goal of the discovery process.
- *Data mining*: matching a particular data mining method

(e.g. summarization, classification, regression, clustering, etc.) to meet the goal of the KDD process.

- *Data interpretation and evaluation*: once a data mining method has been applied, obtained patterns should be interpreted and evaluated.
- *Knowledge extraction*: consolidating discovered knowledge that can be used for further analysis.

The general framework proposed in (Khan et al. 2014; D'Oca et al. 2015; Yu et al. 2013) aimed to extract occupancy patterns from a building database. In the data mining process, a decision tree model and a rule induction algorithm were used to extract the occupancy state (label attribute) based on several predictor attributes. The methodology required the prior knowledge of the label attribute to be performed. In this study, the information about occupancy state as label attribute was unknown and not included in the database. Therefore, the KDD process had to be modified using the general framework as baseline method.

### 3.1.1 Data selection

A total number of 220 variables were measured for each dwelling. Among them, 12 variables were directly or indirectly related to occupant behavior. Therefore, they represented the predictor attributes used to extrapolate the occupancy state in the analysis process. The variables are listed in Table 2, according to the respective database category. An increase in the CO<sub>2</sub> concentration normally present in air is used as an indicator of human presence in the ground floor (day activities) and in the first floor (night activities). Considering that the houses are all-electric dwellings, emissions related to cooking with gas stove are equal to zero. Therefore, it is possible to assume that all variations in

**Table 2** Predictor attributes

Database category	Variable	Indication
Balanced ventilation	CO <sub>2</sub> concentration first floor [ppm]	Occupancy presence
	CO <sub>2</sub> concentration ground floor [ppm]	
Smart meter	Power from the grid [W]	Electricity usage
	Power generation [W]	
	Power from the grid [W]	
	Power generation [W]	
	Power to the grid [W]	
	Power use heat pump [W]	
	Power use dishwasher [W]	
	Power use washing machine [W]	
Heat pump	Vessel temperature set point [°C]	DHW usage
	Domestic hot water temperature [°C]	

CO<sub>2</sub> concentration in the indoor environment are related to human presence. Appendix A describes in further detail the use of CO<sub>2</sub> as a proxy for occupancy.

### 3.1.2 Data cleaning and pre-processing

Data cleaning and pre-processing were employed to exclude buildings with a high number of missing fields and evident issues in the monitoring systems. The location of Holten, the Netherlands was selected among various as it included the highest number of buildings with suitable data. The choice of selecting only one site allowed to exclude uncertainties related to weather. At the end of the process, 4 weeks of data (one per season) of 12 dwellings in Holten were considered for the analysis. This decision resulted from the strong seasonal behavior of the data discovered during the pre-processing stage.

### 3.1.3 Data transformation to knowledge extraction

In this study, a two-step learning framework to extract OB profiles without prior knowledge of the occupancy state was proposed (Fig. 3). The data mining process is used to define the hourly occupancy and average electricity consumption schedule. For each day of the week the occupancy schedule was paired with the corresponding electricity consumption profiles. The electricity profiles only accounted for the consumption by equipment (installed appliances and unknown apparatus) and lighting, due to their dependence on OB. The consumptions by other end-uses (space heating, domestic hot water, etc.) were also considered in the simulation model. However, their dependency on external factors, such as weather conditions, required to consider them separately from the OB inputs.

In the first step, the predictor attributes in Table 2 were analyzed individually within their database category. From each category, an indication regarding the occupancy state was extrapolated observing the instantaneous value of each attribute and the evolutionary trend during the day. An assumption-based rule was then generated combining the information extracted in each category (occupancy presence, electricity consumption, and DHW usage) to determine the hourly occupancy state as a binary value (0 = absent, 1 = occupied) and to extract the hourly electricity consumption (see Appendix A). At the end of the process, each household was characterized by daily occupancy schedules and the respective daily electricity consumption profiles.

In the second step, a clustering analysis with the *k*-means algorithm is performed with the open-source software Rapid Miner (RapidMiner 2017). The choice of this method over other clustering methods was determined by the ease of use, as well as by the fact that the methodology was successfully implemented in D'Oca and Hong (2014) for a similar scope.



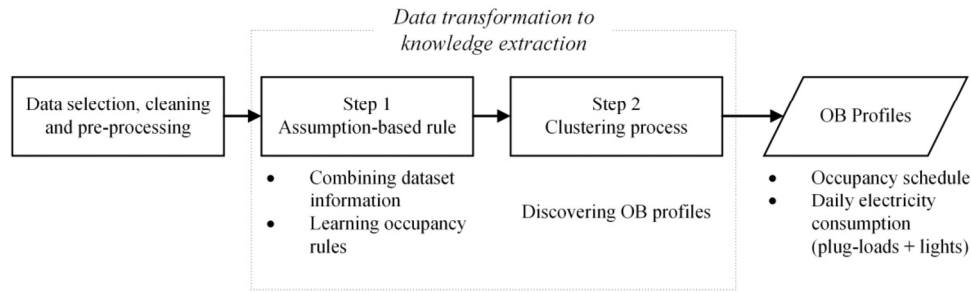


Fig. 3 Two-step data mining framework for the extraction of OB profiles

In the research presented here, the methodology introduced by D’Oca and Hong (2014) is adapted to the case-study of Holten and expanded to provide integration within a BPS software. The clustering process allowed to generalize the results of the first step of the data mining framework, combining the different OB profiles and defining a finite number of mean daily occupancy schedules and mean daily electricity consumption for the analyzed households. The clustering process was performed studying separately weekdays and weekend due to a previously identified different behavior of the occupants during the course of the week. Because similar dwellings were studied, no normalization or transformation of the obtained data was needed.

To determine the optimal number of clusters which better define the OB in the analyzed households, a square Euclidean distance performance operator was considered. In this study, the Davies-Bouldin Index (DBI) was used for the evaluation (Davies and Bouldin 1979). The DBI is the ratio of the sum of average distance inside clusters to distance between clusters (Davies and Bouldin 1979) and it can be defined as Eq. (1):

$$DBI = \frac{1}{n} \sum_{i=1}^n \max_{i \neq j} \left[ \frac{R_i + R_j}{M_{i,j}} \right] \quad (1)$$

where  $n$  is the number of clusters,  $R_i$  is the average distance inside cluster  $i$ ,  $R_j$  is the average distance inside cluster  $j$ ,  $M_{i,j}$  is the distance between the cluster centers.

The number of clusters  $n$  was varied until achieving the smallest value  $k$  of DBI, which indicates a better performance of the clustering algorithm. The  $k = n$  algorithm that produced clusters with low intra-cluster distances (high similarity between the cluster elements) and high inter-cluster distance (low similarity between the elements of different clusters) was considered the  $k = n_{opt}$  cluster algorithm.

### 3.2 OB profiles integration into BPS and validation

A model with standard OB inputs (Hoes 2014; Yang and Tysoe 2016; Aerts et al. 2013), referred to as base-case

model, was developed for the analyzed dwellings with the software TRNSYS v17 (TRNSYS 2017). Table 3 summarizes the variations in the inputs related to OB applied in the final simulation model. In TRNSYS, occupant behavior is primarily modeled in terms of heat gains from occupants, equipment use, light use, space heating and DHW use. Occupants were modeled in the base-case model by multiplying the average heat gain per person by the number of occupants and the occupancy state (absent state = 0, occupied state = 1) according to the profiles proposed by Aerts et al. (2013) and presented in Fig. 4.

Conversely, the profiles were extracted from the database in the OB-integrated models. A similar approach was used to model the heat gains resulting from equipment use and light use. Standard inputs from Hoes (2014), expressed in  $W/m^2$ , were used in the base-case model (Fig. 5).

Table 3 Occupants behavior modeling in the base-case model and OB-integrated model

Occupancy inputs	Base-case model	OB-integrated model
Presence	(Aerts et al. 2013)	Extracted occupancy schedule
Equipment gains	(Hoes 2014)	Extracted consumption profiles
Lighting gains	(Hoes 2014)	Extracted consumption profiles
Occupants gains	(Hoes 2014)	(Hoes 2014)
DHW daily usage	(Yang and Tysoe 2016)	Database analysis (mean value)
$T$ set point	Constant at 20 °C	Database analysis (mean value)

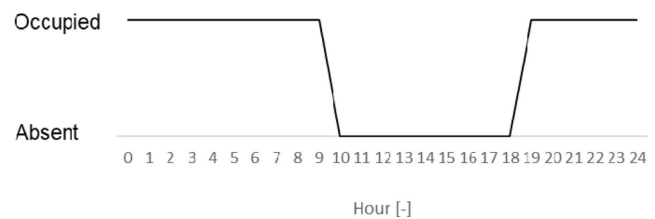


Fig. 4 Occupancy schedule used in the base-case model

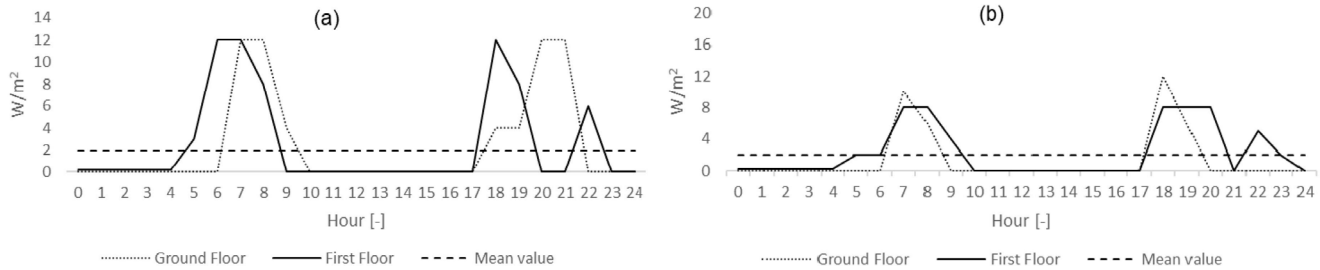


Fig. 5 Light use power density (a) and equipment use power density (b) used in the base-case model

Such standard inputs were modified in the OB-integrated model according to the values extracted in the dataset. Finally, a fixed setpoint of 20 °C was used for heating in the base-case model, while it was adjusted according to the mean values extrapolated in the database analysis of each dwelling. Similarly, the DHW usage profiles based on Yang and Tysoe (2016) were refined according to the data for each dwelling in the OB-integrated model.

The final simulation model, defined as OB-integrated model, was then simulated. To verify the OB profiles and the database analysis process, a comparative analysis of the measured data was carried out aiming to investigate the accuracy of the simulation results in predicting the annual and monthly electricity consumption and the improvement from the base-case model prediction.

### 3.3 Application to the local electricity market

As a further step of the study, the consumption levels obtained by means of the extracted OB profiles were used to propose a local electricity market as a possible solution to achieve a net zero electricity bill. A lower electricity bill can be achieved reducing the imported electricity from the grid and thus increasing the energy independency through self-consumption.

In this study, 12 households located in Holten were considered as members of a local community. The households were divided into groups in which the exchange of the surplus electricity was made possible on hourly basis. The grouping process was done considering the net annual consumption of each household, namely the total consumption reduced of the on-site generation.

The members were paired in groups aiming to guarantee that each household took the same advantage as the others in participating to the local market, regardless of in which group it was inserted. Therefore, the cumulative net consumption, namely the sum of the net consumption of each group member, was calculated for each possible group combination. At the end of the pairing process, groups characterized by a similar cumulative net consumption between one and another were obtained.

The improvements in the self-consumption, energy matching and reduction in carbon emissions were analyzed through OEM and OEF indicators, defined as follows (Cao et al. 2014):

$$\text{OEM} = \frac{\int_{t_1}^{t_2} \min[G(t); L(t)] dt}{\int_{t_1}^{t_2} G(t) dt} \quad (2)$$

$$\text{OEF} = \frac{\int_{t_1}^{t_2} \min[G(t); L(t)] dt}{\int_{t_1}^{t_2} L(t) dt} \quad (3)$$

The equations are defined according to the two standard power curves in Fig. 6.  $G(t)$  and  $L(t)$  represent the on-site generated and the load power curve. The variables  $t_1$  and  $t_2$  represent the starting and final point of the time span, respectively. In this study, the OEM and OEF were calculated on an hourly basis for the whole year. Therefore, the minimum between the instantaneous value of the generation power curve  $G(t)$  and the load power curve  $L(t)$  was calculated hourly.

The OEM and OEF indicators were compared between an initial case, where each household was considered individually according to its consumption profile, and the local market case, in which households are paired in groups. The results concerning the local energy market are obtained only by means of the profiles extracted from the database.

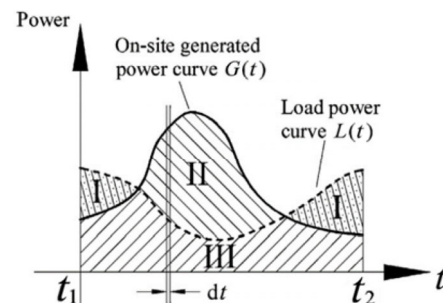


Fig. 6 Representative power curves for the definition of OEM and OEF indicators (Cao et al. 2013; reproduced with permission © Elsevier)

## 4 Results

The analyzed households, despite being pre-fabricated, standardized, in the same location and with the same orientation, are characterized by fairly different annual electricity consumption levels (Fig. 7), with the highest consumer consuming about 1.6 times as much as the lowest consumer. It is reasonable to ascribe these differences entirely to OB. The imported electricity is on average equal to  $4000 \pm 850$  kWh/y. The self-consumed electricity, or the electricity generated on-site by PV panels and directly used by the household, is equal to  $2600 \pm 570$  kWh/y. As the average generation is 5650 kWh/y, the self-consumed electricity represents on average 38% of the total. The remaining part of the generated electricity is exported to the grid.

An analysis of the annual electricity consumption by end-use (Fig. 8) reveals that the equipment (installed appliances and unknown devices) and lighting represent the

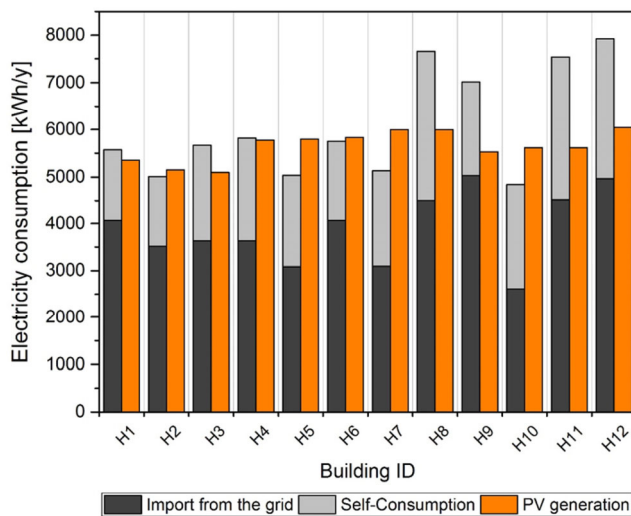


Fig. 7 Holten: annual PV generation and total consumption (self-consumption and import from the grid) in kWh/year

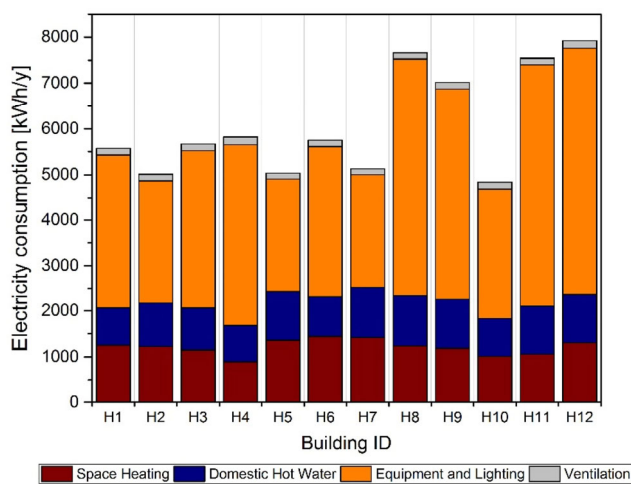


Fig. 8 Holten: annual consumption by end-use in kWh/year

highest share of electricity use in each household, accounting for 61.7% of the mean annual consumption with a standard deviation of 1109.2 kWh/y between households.

The heat pump consumption, equal to 2179.9 kWh/y on average (35% of the total consumption) is characterized by fewer discrepancies, with a standard deviation of 248.5 kWh/y. The space heating accounts for 58.5% of the mean heat pump electricity consumption. The remaining part is consumed to cover the DHW demand.

### 4.1 Database analysis

The clustering process allowed to combine the observed predictor attributes into average OB profiles. The clustering process was used to define daily occupancy schedule and average daily electricity consumption during weekdays (Mon–Fri) and weekend (Sat–Sun). The same four weeks of the year (second week of January, April, July and November), one per season, were analyzed for 12 dwellings. This period was chosen as characterized by a complete data coverage for each predictor attribute and each building under investigation. Moreover, a preliminary analysis confirmed low intra-seasonal discrepancies in OB.

During the process, a different behavior in terms of time spent at home between the cold (Autumn/Winter) and warm (Spring/Summer) season was noticed. A differentiation between these two periods rather than each season was found to be representative of the occupancy schedule patterns. Instead, a similar profile in the electricity consumption by equipment was registered during the year. The electricity use for lighting showed a reduction during spring and summer, presumably due to increased daylight. However, for the purpose of integrating the profiles in the simulation model, the dependency of light use on solar radiation was not considered essential as it is already accounted for in the model.

Therefore, the clustering process was performed considering separately two periods of the year (cold and warm season) for the occupancy schedule and without any distinction for the electricity consumption profiles.

In order to achieve the optimal number of clusters that better defined the OB, the DBI was used as cluster distance performance operator. This operation resulted in the following number of clusters:

- 4 daily occupancy schedule for the weekdays (autumn/winter and spring/summer)
- 2 daily occupancy schedule for the weekend (autumn/winter and spring/summer)
- 3 daily electricity use for equipment and lighting for the weekdays
- 3 daily electricity use for equipment and lighting for the weekend

The daily occupancy schedule during weekdays and weekends are displayed in Fig. 9 and Fig. 10, respectively. During the weekdays of autumn and winter, the occupancy with the highest (Schedule A) and the lowest (Schedule D) number of occupancy hours represent the two most common schedules with a weekday mean percentage of 30% each. Schedule D, with the lowest number of occupancy hours, is instead the most common schedule in summer (35% of the cases). During the weekend, similar schedules were found in both seasons. In the first case, Schedule E (at home all day) occurs in 52.9% of the time. Its frequency decreases during summer to 38.5%, due to less hour of occupancy state observed. Instead, Schedule F occurs during most days (61.5%).

Figure 11 displays the electricity consumption profiles for equipment and light use during the weekdays and weekend. The consumption profiles are labeled according to

their value as low, middle and high consumption respectively. Considering the weekdays, the profiles have similar average distribution with the high consumption profile occurring 38.2% of the time. A similar behavior occurs during the weekend, with the high consumption profile covering 40.9% of the time.

At the end of the process, each household was defined according to daily occupancy schedules and daily consumption profiles.

#### 4.2 OB profiles integration into BPS and validation

The OB profiles were then integrated in a simulation model as OB inputs, differentiated for each household in agreement with the results of the clustering process. The occupancy schedules are defined separately for the cold and warm season, as different behavior in terms of occupancy schedules during

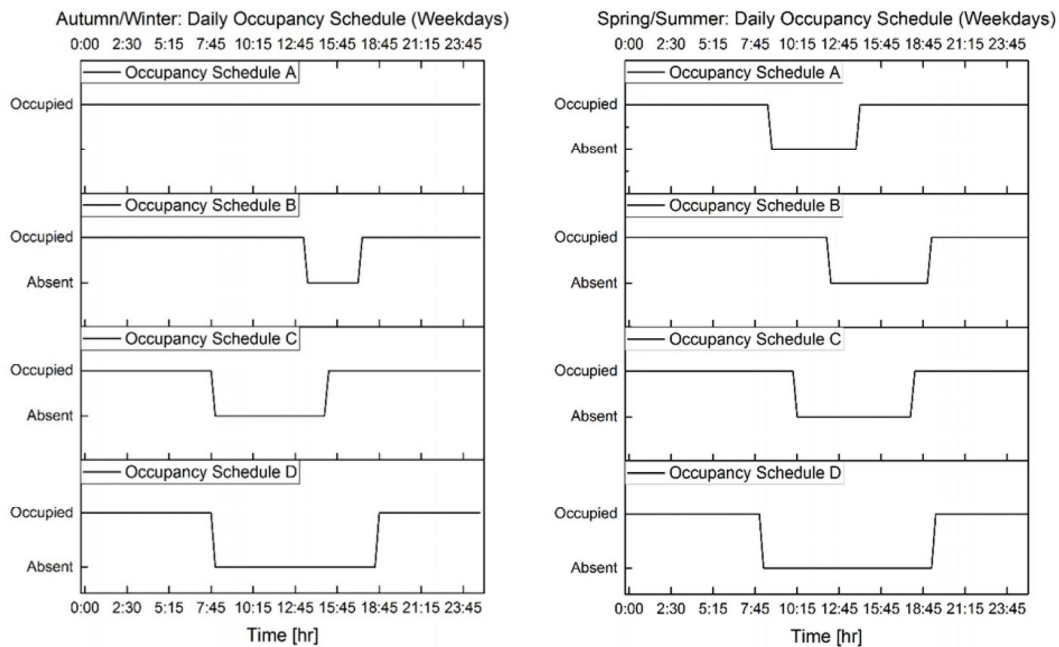


Fig. 9 Daily occupancy schedule during weekdays: autumn/winter (left), and spring/summer (right)

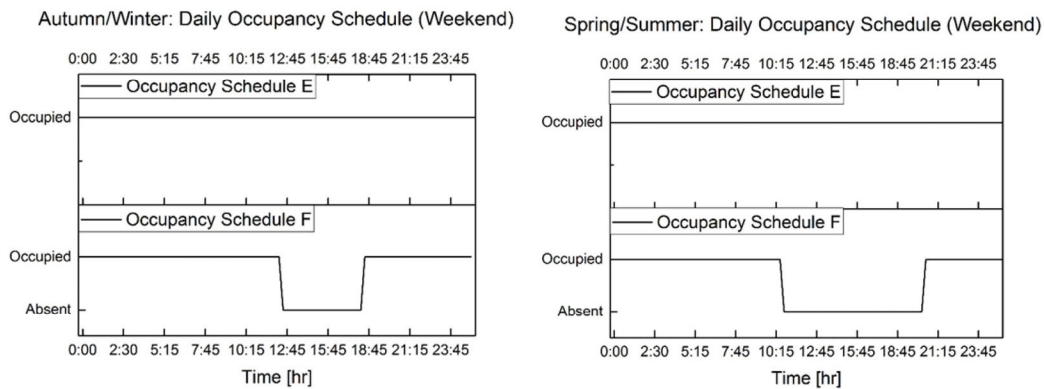


Fig. 10 Daily occupancy schedule during weekends: autumn/winter (left), and spring/summer (right)



these seasons was noticed in the database analysis. The mean temperature set-point and DHW usage, as they were extracted from the database analysis of each household, were also integrated OB inputs into the simulation model.

The OB profiles were verified by comparing the measured electricity consumption and the BPS prediction in terms of annual energy use intensity (EUI) [kWh/(m<sup>2</sup>·y)] (Fig. 12). The base-case model estimates a EUI of 44.6 kWh/(m<sup>2</sup>·y). Since no database analysis was performed in this initial case, OB was modeled with the same inputs for all the analyzed

buildings, and the results do not differentiate between one household and the other. This approach reflects the common-practice representation of OB. For the considered dwellings, using standard OB modeling results in an underestimation of the EUI of 5.9% to 42.5%, depending on the case.

Once the OB profiles were integrated into the BPS software, each household is defined with more accurate, specific OB inputs based on the database analysis. As a result, a better prediction of the EUI is achieved. The variations between measured and estimated EUI is equal to 1.7% on average. The reliability of the simulation prediction is evaluated through a number of aggregate statics as proposed in Samuelson et al. (2015). In particular, the monthly normalized mean bias error (nMBE), defined as

$$nMBE = \frac{\sum_{i=1}^{12} (pred_i - meas_i)}{meas \times (12 - p)} \quad (4)$$

and the coefficient of variation of the root mean square error CV(RMSE)

$$CV(RMSE) = \frac{1}{meas} \times \sqrt{\frac{\sum_{i=1}^{12} (pred_i - meas_i)^2}{(12 - p - 1)}} \quad (5)$$

are evaluated for each building. Moreover, the weighted mean (WM) of all buildings are calculated for both indicators.

$$WM = \frac{\sum_{i=1}^{12} |Z_i \times A_i|}{\sum_{i=1}^{12} (A_i)} \quad (6)$$

where *pred* = predicted monthly EUI, *meas* = measured monthly EUI,  $\overline{meas}$  = mean of measured monthly EUI, *p* = number of predictor variables (1 in this case), *Z* = either nMBE or CV(RMSE) for each building, and *A* = annual measured energy consumption for each building.

Table 4 summarizes the results for nMBE and CV(RMSE) for each building as well as their weighted mean. The maximum nMBE (H12) varied from an initial -46.4% in the base-case model to a final 2.1%. Similar values are obtained for the CV(RMSE), varying from 51.0% to 2.3%. Considering the combined statistics for the 12 households under analysis, the nMBE improves from -26.8% to 2.0% and the CV(RMSE) from 29.5% to 2.2%.

Figure 13 is an illustrative example of the variation in the EUI prediction after each step of the integration of the OB inputs (dwelling H7). The household has a measured EUI equal to 49.4 kWh/(m<sup>2</sup>·y), differing from the EUI predicted in the base-case model of 9.7%. In the first step of the process, the lighting gains were corrected according to the extracted consumption profiles. Since in the base-case model the considered profile, based on Hoes (2014),

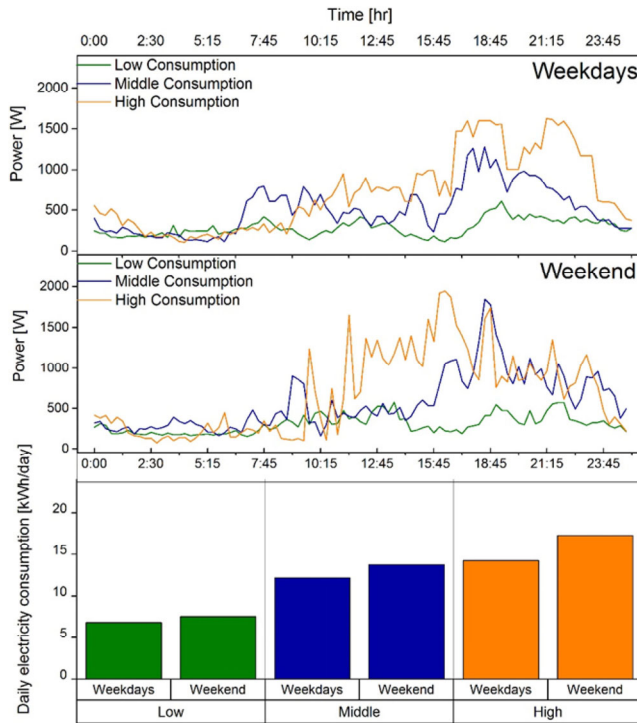


Fig. 11 Electricity consumption profiles by equipment and lighting for the weekdays and the weekend

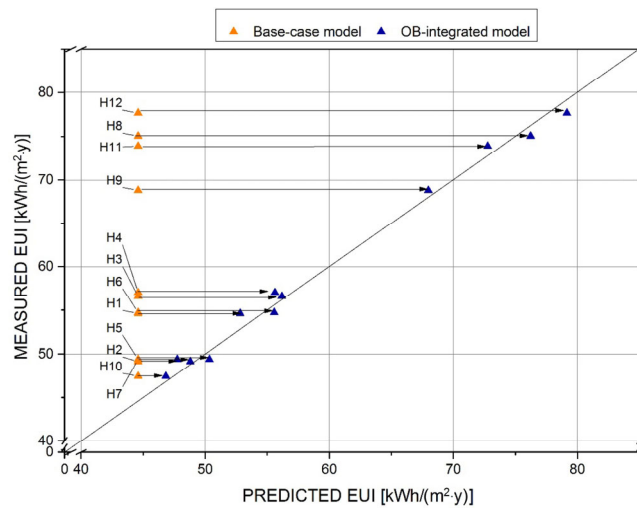
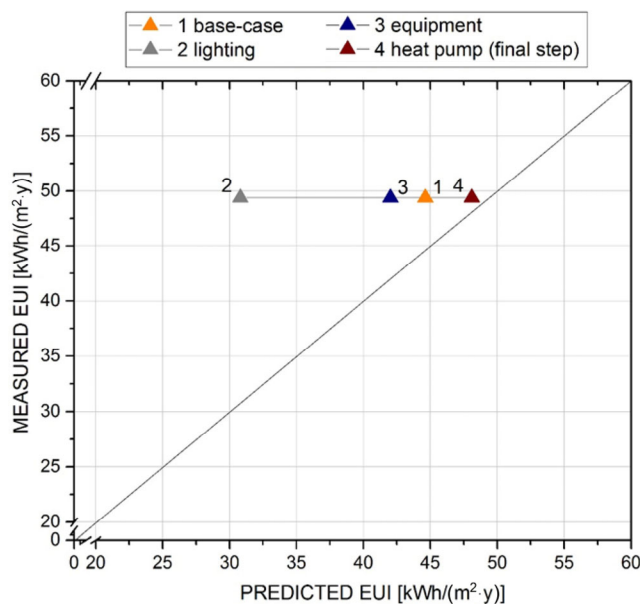


Fig. 12 Predicted EUI vs. measured

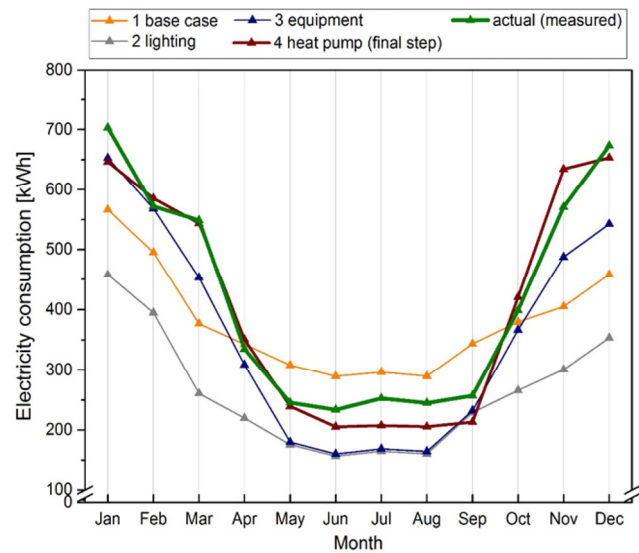
**Table 4** Normalized mean bias error (nMBE) and coefficient of variation of the root mean square error CV(RMSE) in the base-case and OB-integrated model

#House	Base-case model		OB-integrated model	
	nMBE	CV(RMSE)	nMBE	CV(RMSE)
H1	-19.9%	21.9%	-3.6%	3.9%
H2	-10.0%	11.0%	-0.8%	0.8%
H3	-23.0%	25.3%	-0.8%	0.8%
H4	-23.7%	26.1%	-2.7%	2.9%
H5	-10.5%	11.6%	2.1%	2.3%
H6	-20.1%	22.1%	2.1%	2.3%
H7	-10.5%	11.6%	-3.6%	4.0%
H8	-44.2%	48.6%	1.7%	1.9%
H9	-38.3%	42.1%	-1.3%	1.4%
H10	-6.5%	7.1%	-1.4%	1.6%
H11	-43.2%	47.5%	-1.7%	1.9%
H12	-46.4%	51.0%	2.1%	2.3%
<b>Weighted mean</b>	<b>-26.8%</b>	<b>29.5%</b>	<b>2.0%</b>	<b>2.2%</b>



**Fig. 13** H7: EUI variations after the integration of OB inputs

overestimates the actual lighting consumption, the correction leads to an initial reduction of the predicted EUI. Integrating the equipment consumption, the EUI increases due to an underestimation of the consumption in the base-case model, equal to 58.6% in the analyzed case. The modification of the temperature set-point and DHW usage (step 4), according to the mean values observed in the database for the household, improved the final estimation of the EUI to 47.7 kWh/(m<sup>2</sup>·y) with a difference of 3.3% from the measured EUI. Figure 14 shows the monthly comparison in the consumption after each step of the integration of the OB inputs. Compared to



**Fig. 14** H7: Monthly electricity comparison between base-case model, OB-integrated model and measured data

the base-case model, the final prediction shows a similar pattern with the measured monthly consumption.

Considering the consumption by end-use (Fig. 15), the mean variation between the base-case model predictions and actual consumption ranges between 1.8% for the ventilation consumption and 371.1% for the lighting consumption. As for the other end-uses, the average discrepancy with the actual consumption is equal to 14.9%, 31.5% and 71.8% for the space heating, DHW and equipment consumption, respectively. In the OB-integrated model, the lighting consumption shows the highest discrepancy with the measured data, with an underestimation of 3.2%. The prediction of the space heating, DHW and equipment consumption, after the integration of the extracted OB profiles, improved to 6.1%, 4.3% and 2.8%, respectively.

Figure 16 illustrates the monthly electricity consumption divided by end-use for H7. Comparing the base-case with the final model, the annual nMBE improves from -10.5% to -3.6%. The maximum variation between the monthly total electricity consumption predicted in the base-case model and the measured consumption occurs in September. It is equal to 34.6% in the base-case model, improving to 16.5% in the final model. Considering the end-use consumption, the major discrepancy between base-case model prediction and measured data appears in the lighting consumption of May, with a mean discrepancy of 343.9%. The prediction of the monthly electricity consumption for lighting in May improves in the OB-integrated model with an overestimation of 35.8%.

### 4.3 Application to the local electricity market analysis

Figure 17 shows the annual total electricity consumption

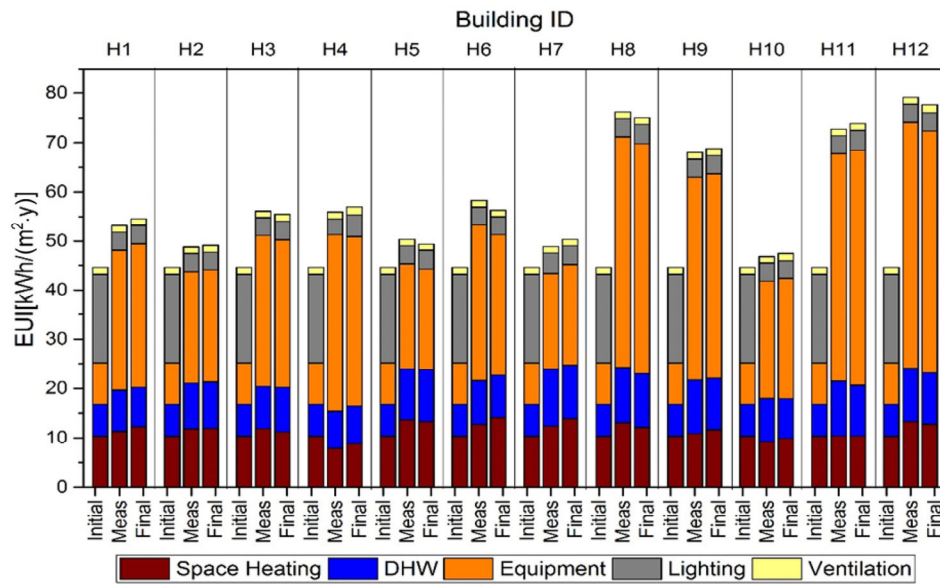


Fig. 15 EUJ by end-use: comparison between base-case model (initial) measured data (meas) and OB-integrated model (final)

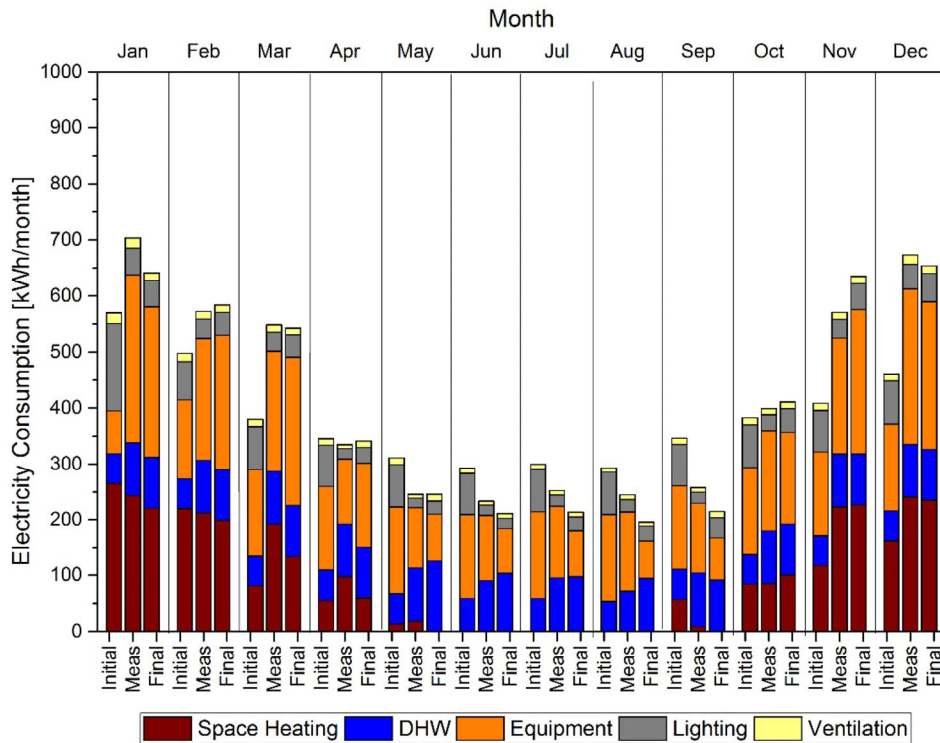


Fig. 16 H7: Monthly electricity consumption by end-use: comparison between measured data, base-case model, and OB-integrated model

and PV generation in the analyzed households which represent the 12 members of the local electricity market. In order to study different configurations, households were divided into groups, each composed of the same number of members.

Table 5 displays the variations in the on-site energy matching (OEM) and on-site energy fraction (OEF) indicators

between the initial case and the local market case. The indicators increase in all the analyzed cases except in two households for OEM and in three households for OEF. The mean increase in the OEM and OEF are equal to 29% and 28%, respectively.

Figure 18 shows the variations in the imported electricity in H7 between the initial and the local market case in two

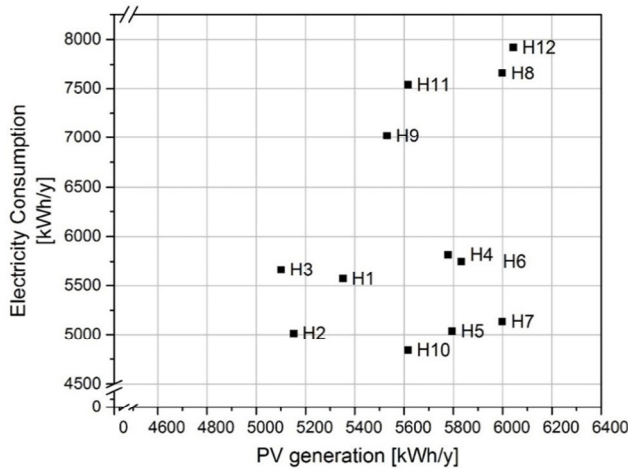


Fig. 17 Electricity total consumption vs PV generation in 12 households in Holten

Table 5 Variations in OEM/OEF indicators between the initial case and the local market case

		OEM			OEF		
		Initial	Local market	%	Initial	Local market	%
Group 1	H2	0.25	0.48	95.8%	0.22	0.41	84.7%
	H5	0.37	0.48	29.9%	0.42	0.41	-2.9%
	H9	0.40	0.48	20.4%	0.24	0.41	71.3%
	H4	0.37	0.48	29.4%	0.37	0.41	9.8%
	H8	0.59	0.48	-17.6%	0.47	0.41	-12.3%
	H12	0.52	0.48	-6.8%	0.43	0.41	-5.8%
Group 2	H10	0.32	0.47	45.8%	0.34	0.41	18.4%
	H1	0.34	0.47	37.2%	0.31	0.41	29.9%
	H6	0.33	0.47	42.8%	0.35	0.41	17.3%
	H7	0.38	0.47	21.9%	0.33	0.41	24.3%
	H3	0.36	0.47	29.5%	0.26	0.41	57.1%
	H11	0.41	0.47	12.7%	0.30	0.41	37.5%

typical days, January 15th, and July 15th. In the initial case, the PV generation of the household is not enough to cover the total demand and it is necessary to rely partly on the import from the grid. In the local electricity market case, the availability of surplus electricity generated and not consumed by the neighbors allows reducing the import from the grid. On January 15th, the number of hours of self-consumption increases from 0 to 6. The major part of the import does not vary between the two cases due to weather conditions (winter season, low number of daylight hours, etc.). During summer, on July 15th, the more favorable weather conditions lead to a high number of self-consumption in the initial case without local electricity market, equal to 10 hours. The local market framework helps to increase the value to a final 13 hours, covering almost all the daylight hours through the electricity generated by the neighbors.

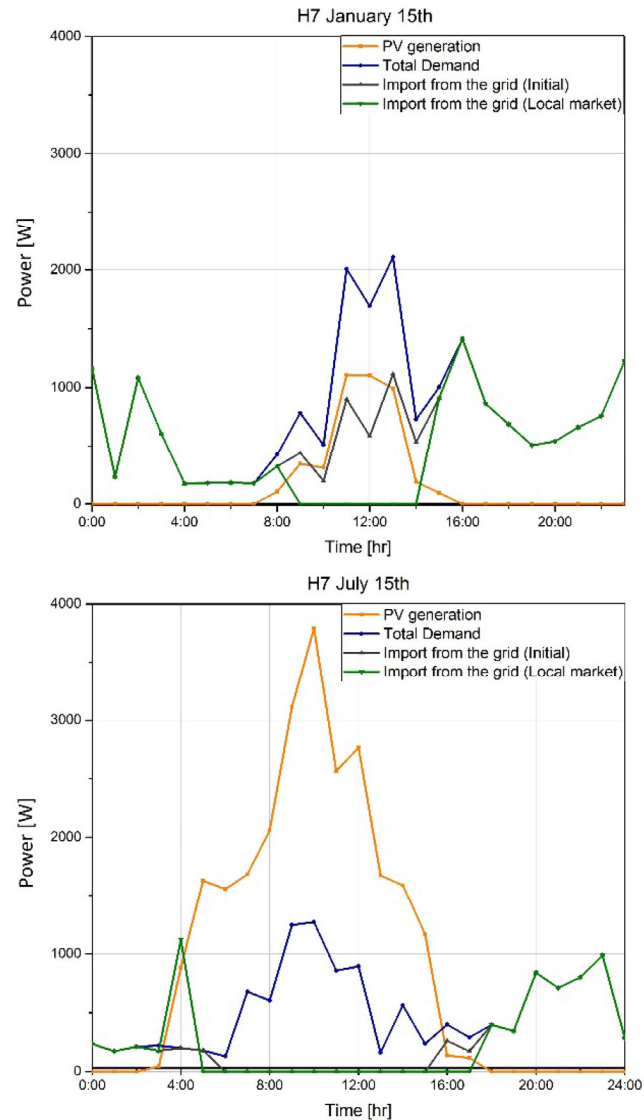


Fig. 18 H7: Comparison between the electricity imported from the grid in the initial case and the local electricity market case in two days (January 15th and July 15th)

### 5 Discussion

The results presented in Section 4 confirm the importance of appropriately modeling OB for energy performance predictions (O'Brien et al. 2017a,b). The KDD is deemed to be a suitable technique for extracting OB profiles from a stream of raw data for implementation in BPS. The implementation of OB profiles increased the accuracy of the model energy use predictions for all 12 dwellings, both on a yearly and monthly timescale. However, the results are dependent on the case study and generalizations concerning the typologies of inhabitants cannot be made. Similarly, there is no guarantee that applying the same profiles to different households would result in an improved accuracy of the model. As the dataset becomes richer, the presented technique will be



applied to a higher number of dwellings to possibly derive standardized Dutch families profiles that can be easily implemented in BPS.

This study makes a number of simplifications that need to be highlighted. Firstly, although the complete database referred to over 150 all-electric dwellings, only 12 were deemed suitable for this analysis. This choice significantly reduces the sample size and may not be statistically significant. However, the decision was dictated by two constraints: i) the usability of the collected data; and ii) the necessity of removing any factor that could have an influence on electricity use rather than OB. The 12 analyzed dwellings are identical in each aspect other than OB and hence represent an ideal case study for researching the influence of OB on energy use. Secondly, only four weeks of the calendar year were used to extract OB profiles. The assumption that the selected four weeks were representative of the whole year has been verified by means of a preliminary analysis. Were the dataset of the 12 dwellings complete for the whole year, it is nonetheless recommendable to use the full data.

## 6 Conclusions

A dataset concerning 12 identical, all-electric dwellings in Holten, the Netherlands was analyzed with the purpose of investigating the effect of OB on residential buildings' energy use. The dwellings show a factor 1.6x discrepancy in the electricity use due to OB, with the lowest consumer using 4839.9 kWh/y and the highest consumer using 7920.8 kWh/y. As for the different end-uses, the highest differences in energy use are related to the use of equipment (standard deviation of 1007 kWh/y between households). The KDD DM-technique was employed to derive occupancy and electricity use profiles for lighting and equipment. The extracted OB profiles were integrated in a base-case model developed in TRNSYS v17. The more accurate representation of OB by means of the profiles improved the predictive ability of the model from an initial 22.9% average deviation from measurements to 1.7%. This result confirms the validity of the extracted profiles. Moreover, the profiles allowed to evaluate the potential for a local energy market in the considered neighborhood. Sharing the surplus energy with neighbors resulted in a mean improvement in OEM and OEF of 29% and 28%, respectively. This study confirms the importance of appropriate OB modeling for analyzing the energy use of dwellings, as highlighted by IEA EBC Annex 66 (Yan et al. 2017) and Annex 79. To the best knowledge of the authors, the presented research is a first example of DM-derived OB profiles integration in BPS, as well as implementation to a practical application such as evaluating the potential for a local electricity market.

## Acknowledgements

The support of PIT/VABI, the SPARK consortium and TKI EnerGO TRECO-office is greatly appreciated. This research greatly benefitted from discussions with members of IEA-EBC Annex 66.

**Open Access:** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A: Assumption-based rule to determine occupancy schedule

In order to determine the occupancy schedule without any prior knowledge of the occupancy state, an assumption-based rule was generated analyzing 12 database parameters related to OB. CO<sub>2</sub> concentration, electricity consumption and appliances usage and DHW usage were considered as an indication of the occupied state in the analyzed dwellings. Each category was studied individually with the related parameters and the information extracted combined to obtain the final occupancy state as a binary value (0 = unoccupied, 1 = occupied state).

### CO<sub>2</sub> concentration

In this study, human activities are assumed as the only source of CO<sub>2</sub> in the buildings. Since the houses are all-electric, the installation of gas stoves, fireplaces, furnaces, boiler and other similar CO<sub>2</sub> sources can be excluded. Other possible causes of emissions are lit candles and pets. For safety reasons, candles are assumed to be not used when the house is unoccupied while pets, if present, emit a smaller amount of CO<sub>2</sub> compared to the household members and for this reason their contribution can be neglected.

The following assumptions were considered to extract the occupancy state:

- (1) Outdoor CO<sub>2</sub> is usually between 350 and 450 ppm (The Engineering Toolbox 2016). In case the CO<sub>2</sub> is lower than 450 ppm the dwelling is considered unoccupied.

- (2) A steady value of CO<sub>2</sub> suggests that the occupancy state is not changing between two consecutive intervals. Therefore, the state in the interval  $n$  is assumed the same of the previous  $n-1$  interval.
- (3) Since the houses are all-electric, an increase in the CO<sub>2</sub> is assumed to be caused mostly by the presence of occupants. Therefore, if a continuous increase in the CO<sub>2</sub> is reported, the building is considered occupied.
- (4) A decrease of CO<sub>2</sub> does not necessarily mean that the occupants left the house but it can be caused by a different activity, variations of the number of occupants, etc. For this reason, an unoccupied state is considered only if the reduction in the CO<sub>2</sub> is continuous not only between  $n$  and  $n+1$  intervals with 15 minutes difference but for at least one hour.

Analyzing the last assumption, a decrease in CO<sub>2</sub> can be also caused by the ventilation of the indoor environment. In the analyzed dwellings, a balanced ventilation unit for fresh air supply is installed. The ventilation unit turns on automatically when the CO<sub>2</sub> concentration exceeds 1000 ppm. Since the occupants are assumed as the main source of CO<sub>2</sub> emissions and other possible sources were excluded, this concentration can be reached only due to human activities in the house. The manual opening of windows cannot be excluded as a cause of reduction in the CO<sub>2</sub> and the database does not contain this information. However, considering the installation of the ventilation system, the operation with windows are assumed to be less frequent in the analyzed dwellings.

#### *Electricity consumption and appliances usage*

In the analyzed database, the electricity consumption of 4 installed appliances (oven, dryer, washing machine, dishwasher) and the heat pump are collected. The consumption by lighting and other devices are unknown. However, they are extrapolated from the total electricity consumption reduced of the exported electricity and the consumption of the mentioned appliances and heat pump. Therefore, the electricity consumption can be used as an indication of occupied state in the analyzed dwellings. The following assumptions were developed:

- (1) *Oven consumption*: for safety reason the house is considered occupied during all the period in which the oven consumption is higher than zero.
- (2) *Dryer consumption*: an occupied state is considered at the beginning of the power cycle when the consumption is higher than zero.
- (3) *Base load*: a base load of 400 W is considered through observation of the collected data during the analyzed period. The base load is registered during the day in which the level of activity is assumed to be lower. Therefore,

value higher than 400W suggests the usage of equipment and lighting and an occupied state in the house.

The consumption of the washing machine and dishwasher are excluded from the analysis because the appliances might have an internal time switch installed. Since they might turn on in periods in which occupants are not at home they cannot be used to estimate the occupancy state.

#### *DHW usage*

The water vessel contains DHW at the constant temperature of 55 °C. A drop in the water temperature implies the usage of DHW in the house. In order to exclude possible fluctuation in the water temperature related to external causes, the usage of DHW is considered only with a standard deviation higher than 4 between two consecutive intervals, which corresponds to a mean difference of 10 °C.

## References

- Aerts D, Minnen J, Glorieux I, Wouters I, Descamps F (2013). Discrete occupancy profiles from time-use data for user behaviour modelling in homes. In: Proceedings of the 13th International IBPSA Building Simulation Conference, Chambéry, France.
- Andersen RK (2012). The influence of occupants' behaviour on energy consumption investigated in 290 identical dwellings and in 35 apartments. In: Proceedings of the 10th International Conference on Healthy Buildings, Brisbane, Australia.
- Attia S, de Herde A, Gratia E, Hensen JLM (2013). Achieving informed decision-making for net zero energy buildings design using building performance simulation tools. *Building Simulation*, 6: 3–21.
- Bahaj AS, James PAB (2007). Urban energy generation: The added value of photovoltaics in social housing. *Renewable and Sustainable Energy Reviews*, 11: 2121–2136.
- Basu K, Hawarah L, Arghira N, Joumaa H, Ploix S (2013). A prediction system for home appliance usage. *Energy and Buildings*, 67: 668–679.
- Boekhoud A, Behrendt L (2013). Taxes and Incentives for Renewable Energy. KPMG International.
- Cao S, Hasan A, Sirén K (2013). On-site energy matching indices for buildings with energy conversion, storage and hybrid grid connections. *Energy and Buildings*, 64: 423–438.
- Cao S, Hasan A, Sirén K (2014). Matching analysis for on-site hybrid renewable energy systems of office buildings with extended indices. *Applied Energy*, 113: 230–247.
- Clevenger CM, Haymaker J (2006). The Impact of the building occupant on energy modeling simulation. In: Proceedings of Joint International Conference on Computing and Decision Making in Civil and Building Engineering, Montreal, Canada.
- D'Oca S, Hong T (2014). A data-mining approach to discover patterns of window opening and closing behavior in offices. *Building and Environment*, 82: 726–739.
- D'Oca S, Hong T (2015). Occupancy schedules learning process through a data mining framework. *Energy and Buildings*, 88: 395–408.
- D'Oca S, Corngati S, Hong T (2015). Data mining of occupant behavior in office buildings. *Energy Procedia*, 78: 585–590.

- Daniel L, Soebarto V, Williamson T (2015). House energy rating schemes and low energy dwellings: The impact of occupant behaviours in Australia. *Energy and Buildings*, 88: 34–44.
- Davies DL, Bouldin DW (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2): 224–227.
- Fan C, Xiao F, Yan C (2015). A framework for knowledge discovery in massive building automation data and its application in building diagnostics. *Automation in Construction*, 50: 81–90.
- Gaetani I, Hoes P-J, Hensen JLM (2016). Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy. *Energy and Buildings*, 121: 188–204.
- Gram-Hansen K (2010). Residential heat comfort practices: Understanding users. *Building Research & Information*, 38: 175–186.
- Guerra Santin O, Itard L, 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: 1223–1232.
- Hensen JLM (2011). Building performance simulation for sustainable building design and operation. In: Proceedings of the 60th Anniversary Environmental Engineering Department, Prague, Czech Technical University.
- Hoes P (2014). Computational performance prediction of the potential of hybrid adaptable thermal storage concepts for lightweight low-energy houses. PhD Thesis, Universiteit Eindhoven, the Netherlands.
- Hong T, Lin H-w (2012). Occupant behavior: Impact on energy use of private offices. In: Proceedings of Asim IBSPA Asia Conference.
- Hong T, D'Oca S, Taylor-Lange SC, Turner WJN, Chen Y, Corgnati SP (2015). An ontology to represent energy-related occupant behavior in buildings. Part II: Implementation of the DNAS framework using an XML schema. *Building and Environment*, 94: 196–205.
- Juodis E, Jaraminiene E, Dudkiewicz E (2009). Inherent variability of heat consumption in residential buildings. *Energy and Buildings*, 41: 1188–1194.
- Khan A, Nicholson J, Mellor S, Jackson D, Ladha K, Ladha C, Hand J, Clarke J, Olivier P, Plötz T (2014). Occupancy monitoring using environmental & context sensors and a hierarchical analysis framework. In: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings – BuildSys'14.
- Maier T, Krzaczek M, Tejchman J (2009). Comparison of physical performances of the ventilation systems in low-energy residential houses. *Energy and Buildings*, 41: 337–353.
- Monetti V, Davin E, Fabrizio E, André P, Filippi M (2015). Calibration of building energy simulation models based on optimization: A case study. *Energy Procedia*, 78: 2971–2976.
- Saldanha N, Beausoleil-Morrison I (2012). Measured end-use electric load profiles for 12 Canadian houses at high temporal resolution. *Energy and Buildings*, 49: 519–530.
- O'Brien W, Gaetani I, Carlucci S, Hoes P-J, Hensen JLM (2017a). On occupant-centric building performance metrics. *Building and Environment*, 122: 373–385.
- O'Brien W, Gaetani I, Gilani S, Carlucci S, Hoes P-J, Hensen J (2017b). International survey on current occupant modelling approaches in building performance simulation. *Journal of Building Performance Simulation*, 10: 653–671.
- Pan Y, Huang Z, Wu G (2007). Calibrated building energy simulation and its application in a high-rise commercial building in Shanghai. *Energy and Buildings*, 39: 651–657.
- Raftery P, Keane M, Costa A (2011). Calibrating whole building energy models: Detailed case study using hourly measured data. *Energy and Buildings*, 43: 3666–3679.
- RapidMiner (2017). "Rapid Miner Studio."
- Royapoor M, Roskilly T (2015). Building model calibration using energy and environmental data. *Energy and Buildings*, 94: 109–120.
- Samuelson HW, Ghorayshi A, Reinhart CF (2016). Analysis of a simplified calibration procedure for 18 design-phase building energy models. *Journal of Building Performance Simulation*, 9: 17–29.
- TRNSYS (2017). TRNSYS V17.
- The Engineering Toolbox (2016). Carbon Dioxide Concentration - Comfort Levels.
- Urban B, Gomez C (2013). A case for thermostat user models. In: Proceedings of the 13th International IBPSA Building Simulation Conference, Chambéry, France.
- Yan D, O'Brien W, Hong T, Feng X, Gunay HB, Tahmasebi F, Mahdavi A (2015). Occupant behavior modeling for building performance simulation: Current state and future challenges. *Energy and Buildings*, 107: 264–278.
- Yan D, Hong T, Dong B, Mahdavi A, D'Oca S, Gaetani I, Feng X (2017). IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings. *Energy and Buildings*, 156: 258–270.
- Yang X, Tysoe B (2016). Comparing standard domestic hot water modeling methods for multi-residential buildings. In: Proceedings of ASHRAE and IBPSA-USA SimBuild 2016 Building Performance Modeling Conference, Salt Lake City, UT, USA.
- Yu Z, Fung BCM, Haghghat F, Yoshino H, Morofsky E (2011). A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy and Buildings*, 43: 1409–1417.
- Yu Z, Fung BCM, Haghghat F (2013). Extracting knowledge from building-related data — A data mining framework. *Building Simulation*, 6: 207–222.
- Yu Z, Haghghat F, Fung BCM (2016). Advances and challenges in building engineering and data mining applications for energy-efficient communities. *Sustainable Cities and Society*, 25: 33–38.
- Zhang W, Tan S, Lei Y, Wang S (2014). Life cycle assessment of a single-family residential building in Canada: A case study. *Building Simulation*, 7: 429–438.