

Big data: the key to energy efficiency in smart buildings

M. Victoria Moreno, Luc Dufour, Antonio F. Skarmeta, Antonio J. Jara, Dominique Genoud, Bruno Ladevie, Jean-Jacques Bézian

▶ **To cite this version:**

M. Victoria Moreno, Luc Dufour, Antonio F. Skarmeta, Antonio J. Jara, Dominique Genoud, et al.. Big data: the key to energy efficiency in smart buildings. Soft Computing, 2016, 20 (5), p. 1749-1762. 10.1007/s00500-015-1679-4 . hal-01620293

HAL Id: hal-01620293
https://hal.science/hal-01620293

Submitted on 7 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Big data: the key to energy efficiency in smart buildings

M. Victoria Moreno[1] · Luc Dufour[2] · Antonio F. Skarmeta[1] · Antonio J. Jara[2] · Dominique Genoud[2] · Bruno Ladevie[3] · Jean-Jacques Bezian[3]

Abstract Due to the high impact that energy consumption by buildings has at global scale, energy-efficient buildings to reduce CO_2 emissions and energy consumption are needed. In this work we present a novel approach to energy saving in buildings through the identification of the relevant parameters and the application of Soft Computing techniques to generate predictive models of energy consumption in buildings. Using such models it is possible to define strategies for optimizing the day-to-day energy consumption of buildings. To verify the feasibility of this proposal, we apply our approach to a reference building for which we have contextual data from a complete year of monitoring. First, we characterize the building in terms of its contextual features and energy consumption, and then select the most appropriate techniques to generate the most accurate model of our reference building charged with estimating the energy consumption, given a concrete set of inputs. Finally, considering the energy usage profile of the building, we propose specific control actions and strategies to save energy.

Keywords Internet of things · Big data · Smart buildings · Energy efficiency

Communicated by A. Jara, M. R. Ogiela, I. You and F.-Y. Leu.

✉ M. Victoria Moreno
mvmoreno@um.es

Luc Dufour
Luc.Dufour@hevs.ch

Antonio F. Skarmeta
skarmeta@um.es

Antonio J. Jara
Antonio.Jara@hevs.ch

Dominique Genoud
Dominique.Genoud@hevs.ch

Bruno Ladevie
Bruno.Ladevie@mines-albi.fr

Jean-Jacques Bezian
Bezian@mines-albi.fr

[1] Department of Information and Communications Engineering, University of Murcia, Murcia, Spain

[2] Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland

[3] Mines Telecom, Albi, France

1 Introduction

By 2020 there will be 7.5 billion people in the world and consumption will increase by 75 % compared to 2000, equally split between developing and developed countries. This means an increase of 37.5 % in energy consumption every 10 years. These factors have driven research on sustainability in energy production, distribution, storage and consumption.

Furthermore, there are the European 20-20-20 objectives[1]: to decrease by 20 % gas emissions with greenhouse effect (GHG); to decrease by 20 % the energy consumption; and to increase by 20 % the production of renewable energy. It is important to highlight that buildings are responsible for 40 % of total EU energy consumption and generate 36 % of GHG (Petersen et al. 2009). This indicates the need to achieve energy-efficient buildings to reduce their CO_2 emissions and their energy consumption. Moreover, the building environment affects the quality of life and work of all citizens. Thus, buildings must be capable of not only providing mechanisms to minimize their energy consumption (even integrating their

[1] http://ec.europa.eu/clima/policies/package/index_en.htm.

own energy sources to ensure their energy sustainability), but also improving occupant experience and productivity.

Analysis of the energy efficiency of the built environment has received growing attention in the last decade (Agarwal et al. 2010; Lindberg et al. 2004). Various approaches have addressed the energy efficiency of buildings using predictive models of energy consumption based on usage profile, climate data and building characteristics. Nevertheless, most of the approaches proposed to date only provide partial solutions to the overall problem of energy efficiency in buildings, where different factors are involved in a holistic way, but until now have been addressed separately or even neglected by previous proposals. This division is frequently due to the uncertainty and lack of data and inputs included in the modeling process, so that analysis of how energy is consumed in buildings is incomplete (Voss et al. 2010).

The integration and development of systems based on Information and communication technologies (ICT), and more specifically, the internet of things (IoT) (Perera et al. 2014), are important enablers of a broad range of applications, both for industries and the general population, helping make *smart buildings* a reality. But most of the approaches to the problem of energy efficiency in smart buildings present partial solutions regarding monitoring, data collection from sensors and control actions. The IoT has provided vast amounts of data that can be analyzed deeply to reveal interesting relationships, which can be used to generate models able to anticipate and respond efficiently to certain events. Methods, techniques and tools from diverse disciplines can be combined to help analyze such datasets (Berthold et al. 2012). In that sense, Big data and IoT are a perfect combination that can be applied to Smart Buildings scenarios for energy efficiency.

The approach of this paper involves applying insights from Big data algorithms to sensed data in smart buildings. We select the most suitable soft computing (SC) techniques to manage these data with the aim of enabling real-time systems anticipation and optimization of the energy consumption in buildings. We propose a solution for data processing to generate energy consumption models of buildings which can be used to select the optimal measurements and strategies to save energy. First, we analyze what the main drivers of the energy consumed in buildings are. For this analysis, we use the data measured by sensors installed in the building and thus generate the predictive model that estimates its daily energy consumption. As a real case where energy saving must be achieved, we present an industrial building with high levels of monthly energy consumption involved in thermal comfort provision. In this building, the first stages of experimentation have been already carried out following the approach proposed in this work. Analysis of the generated models has led to energy saving strategies being applied.

The structure of this paper is as follows: Sect. 2 reviews some related work proposed in the literature. Section 3 describes the key issues involved in energy efficiency in buildings. Among these issues, relevant parameters affecting the energy consumed in buildings are described and proposed as input data of the building energy consumption model. Finally, this section presents our approach to propose optimum strategies and actions to save energy in buildings. Section 4 details the energy usage characterization of our reference building, the process of generation of its energy consumption model and the analysis extracted from such models. Finally, Sect. 5 provides conclusions and future directions of our work.

2 Related work

As regards the analysis of buildings to understand how energy is consumed, initial solutions were mainly focused on using non-deterministic models based on simulations. A number of simulation tools are available with varying capabilities. In Al-Homoud (2001) and Crawley et al. (2008) a comprehensive comparison of existing simulation tools is provided. Among these tools are ESP-r (Clarke et al. 2002) and Energy Plus (Crawley et al. 2001). However, this type of approach relies on very complex predictive models based on static perceptions of the environment. For example, a multi-criteria decision model to evaluate the whole life cycle of a building is presented in Chen et al. (2006). The authors tackle the problem from a multi-objective optimization viewpoint and conclude that finding an optimal solution is unreal, and that only an approximation is feasible.

With the incessant progress made in the field of ICT and sensor networks, new applications based on using extensive number of different sensors to monitor building environments are being proposed to improve energy efficiency of buildings through the integration of huge volume of data. For instance, in office spaces, timers and motion sensors provide a useful tool to detect and respond to occupants, while providing them with feedback information to encourage behavioral changes. The solutions based on these approaches are aimed at providing models based on real sensor data and contextual information.

Nevertheless, some current monitoring systems, such as automated lighting systems, have limitations such as those identified in Garg and Bansal (2000). In these proposals, the time delay between the response of the automated system and the actions performed can reduce any energy saving, while an excessively fast response of the system can produce inefficient actions. These monitoring systems, while contributing toward energy efficiency, require significant investment in an intelligent infrastructure that combines sensors and actuators to control and modify the overall energy

consumption. The cost and difficulty involved in deploying such networks often constrain their viability. In this sense, rather than solving the above drawbacks, the approach of this paper involves predictive models based on a combination of real data and predictive patterns that represent the evolution of the parameters affecting energy consumption of buildings.

We propose to generate accurate predictive models of energy building consumption to be used to select the best strategies for saving energy in light of the specific characteristics of each building. Closely reflecting our work there are approaches that have aimed to improve control beyond what off-the-shelf products currently offer. For example, Mozer et al. describe a "Neurothermostat" which utilizes a hybrid occupancy predictor, making use of an available daily schedule and a neural network which was trained using 5 consecutive months of occupancy data (Scott et al. 2011). Lu et al. (2010) formulate a hidden Markov model (HMM) to predict occupancy and to control heating, ventilation and air conditioning (HVAC) systems. They collected occupancy data in eight US households for 1–2 weeks. Using leave-one-out cross-validation to train and test the HMM, they evaluate their solution and the energy savings for each day in a week using the EnergyPlus simulator.

In Hagras et al. (2003) the authors propose an intelligent system able to manage the main comfort services provided in the context of a smart building, i.e. HVAC and lighting, while user preferences concerning comfort conditions are established according to the occupants' locations. Nevertheless, the authors only propose the inputs of temperature and lighting to make decisions, while many more factors are really involved in energy consumption and should be included to provide an optimal and more complete solution to the problem of energy efficiency in buildings. Furthermore, no automation platform is proposed as part of the solution. In Severini et al. (2013) end-user comfort constraints are addressed for the home thermal characterization. This is accomplished by a nonlinear model relating the energy demand with the required temperature profile. A genetic algorithm based on such a model is then used to optimally allocate the energy request to match the user thermal constraints, and therefore to allow the mixed-integer deterministic optimization algorithm to determine the remaining energy management actions. And, as a last example of related work, in Costa and La Neve (2015) the authors present the application of a neuro-fuzzy system (ANFIS) for air conditioning systems to reduce electricity consumption.

As can be noted, most of the approaches proposed to date only provide partial solutions to the overall problem of energy efficiency in buildings. No solutions have been proposed tackling the full integration of information related with all the relevant aspects directly involved in the energy consumption of buildings (which are described in Sect. 3). Therefore, a more integral vision is required to provide accurate models of the energy consumed in buildings.

3 Approach to design optimum strategies for energy efficiency in buildings

Optimizing energy efficiency in buildings is an integrated task that covers the whole life cycle of the building. For buildings to have an impact at city level in terms of energy efficiency, different challenges have been identified in the building value chain[2] (from design to end-of-life). During these phases it is necessary to continuously re-engineer the indexes that measure energy efficiency to adapt the energy consumption profile to the building's conditions. Hereinafter, we refer only to electrical energy consumption since other kinds of energy such as fuel oil, gas or water are beyond the scope of this work.

The energy performance model for buildings proposed by the CEN Standard EN15251 (British Standards Institution 2007) proposes criteria for dimensioning the energy management of buildings, while indoor environmental requirements are maintained. According to this standard, there are static and dynamic conditions that affect the energy consumption of buildings. Therefore, it is first necessary to identify the main drivers of energy use in buildings. After monitoring these parameters and analyzing the associated energy consumed, we can model their impact on energy consumption, and then propose control strategies to save energy. The main idea of this approach is to provide anticipated responses to ensure energy efficiency in buildings.

During the monitoring phase, information from heterogeneous sources is collected and analyzed before specific actions are proposed to minimize energy consumption. Since buildings with different functionalities have different energy use profiles, it is necessary to carry out an initial characterization of the main contributors to their energy use. For instance, in residential buildings, the energy consumed is mainly due to the indoor services provided to their occupants (associated to comfort), whereas in industrial buildings energy consumption is associated mostly with the operation of industrial machinery and infrastructures dedicated to production processes. Considering this, and taking into account the models for predicting the comfort response of building's occupants given by the *ASHRAE* (Berglund 1977), we describe below the main parameters that must be monitored and analyzed before selecting optimum strategies to save energy. From this set of parameters affecting energy consumption in buildings, we can extract the input data to be included in the mechanism to generate predictive models of the energy consumption of buildings.

[2] http://www.ectp.org/.

1. *Electrical devices always connected to the electrical network* In buildings, it is necessary to characterize the minimum value of energy consumption due to electrical devices that are always connected to the electrical network.
2. *Electrical devices occasionally connected to the electrical network* Depending on the kind of building under analysis, different electrical devices may be used with different purposes. Whatever the case, recognition of the operation pattern of devices must be included in the final system responsible for estimating the daily electrical consumption of the building. To obtain these patterns, it is necessary to monitor previously the associated energy consumption of every device or appliance. To monitor each component separately in the total power consumption in a household or an industrial site over time, cost-effective and readily available solutions include non-intrusive load monitoring (NILM) techniques Zoha et al. (2012).
3. *Occupants' behavior* Energy consumption of buildings due to the behavior of their occupants is one of the most critical points in every building energy management system. This is mainly because occupant behavior is difficult to characterize and control due to its uncertain dynamic.
4. *Environmental conditions* Parameters such as temperature, humidity, pressure, and natural lighting have a direct impact on the energy consumption of buildings.
5. *Information about the energy generated in the building* Sometimes, alternative energy sources can be used to balance the energy consumption of the building. Information about the amount of daily energy generated and its associated contextual features can be used to estimate the total energy generated in the future. This information allows us to design optimal energy distribution or/and strategies of consumption to ensure the energy-efficient performance of the building.
6. *Information about energy consumed in the building* Knowing the real value of the energy consumed hourly or even daily permits the performance and accuracy of the building energy management program to be evaluated, and makes it possible to identify and adjust the system in case of any deviation between the consumption predicted and the real value. In addition, providing occupants with this information is crucial to make them aware of the energy that they are using at any time, and encourages them to make their behavior more responsible.

Our approach to design optimum strategies to save energy in buildings proposes to monitor the contextual conditions of buildings to identify what parameters (among those presented in the previous section) are involved in energy consumption. In this way, from this set of parameters affecting energy consumption, we can extract the input data to be included in the estimation of the target building energy consumption model.

Bearing in mind all these parameters, it is possible to design optimum strategies to save energy taking into account both the evolution of the affecting parameters and the consequence of such evolution in the energy consumption of the target building. Therefore, the approach proposed to design optimum strategies of energy saving in buildings is the following:

1. Analyze the energy consumption profile associated to each service provided in the building. In this way, it is possible to identify variables affecting the energy consumption of each service.
2. Analyze the relation among the evolution of such variables and the energy consumed. Thus, it is possible to select variables with the most relevant impact in the energy consumption.
3. Provide behavior patterns of the variables identified as relevant, including their uncertainty. We propose to include as inputs of the model such behavior patterns together with the associated sensed data.
4. Implement a predictive building model able to estimate the evolution of the energy consumption given such a set of inputs.
5. Design optimum strategies of control to save energy in the building based on the estimated evolution of the energy consumption.

Regarding the 4th step of our approach, which proposes to implement predictive models of the energy consumption of buildings, in this paper we propose a procedure based on applying different SC techniques according to the specific goal to be achieved. This general procedure will be instanced later into the specific case of our reference building used for experimentation. After carrying out these steps, an estimator is able to predict the energy consumption of the analyzed building in an on-line way using the building model generated.

4 Application and evaluation

To verify the feasibility of the approach presented in the previous section to propose optimum strategies to save energy in buildings, we apply it to a reference building used for experimentation. There are available contextual data associated to the energy consumption for thermal comfort of this building.

First, we provide a complete characterization of the building in terms of both its context and its energy consumption, and then generate the building model able to predict its energy consumption in the future given a concrete set of inputs (considering the steps described in Sect. 3), and based on these

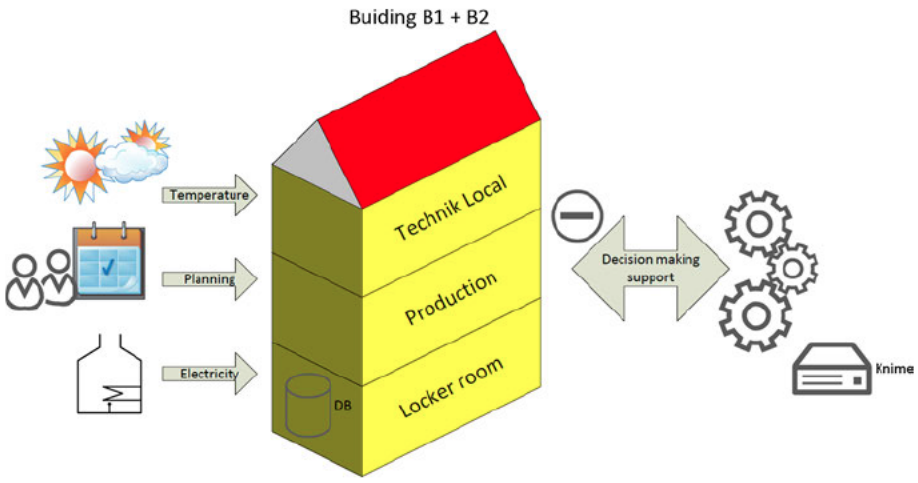


Fig. 1 Information system deployed

models, we propose some relevant strategies to save energy in the target building (whose Information system is showed in Fig. 1).

4.1 Test bed description

Our test bed is located in the buildings of the Debiopharm Group. Debiopharm is a Swiss-based global biopharmaceutical group of companies with a focus on the development of innovative prescription drugs that target unmet medical needs. It is located in Martigny, Switzerland. In this activity sector, the security and the quality of the productive process of the building are crucial. The information system deployed enables the visualization in real time of different stages of the production process. Each parameter monitored in the building is gathered every second in an SQL database. The automation system is flexible enough to integrate new functions and facilitates the maintenance of the installation.

Inside and outside the building, the temperature, the humidity and the pressure are monitored for each room of each floor. Global energy consumption for each building is also monitored and registered for each device (heat pump, boiler, water production). The building automation system uses these parameters to monitor and control the status of the production process. To access these collected data, in this work we use the software analysis Knime for the connection between the information system from Debiopharm and our system. The KNIME platform (the Konstanz Information Miner) is an open source data analytics, reporting visualization and integration platform, which integrates various components for machine learning and data mining through its modular data pipelining concept. Thus, the sensed data are sent toward the automaton component of Knime to our databases.

It has been stated that the impact of HVAC on the energy consumption of a building represents 76 % of the total in European countries (Perez-Lombard et al. 2008). As regards the thermal comfort provided in our reference building, due to its high volume, we focus on modeling this consumption. We have information about the energy consumption of the HVAC systems belonged to buildings B1 and B2. Figure 2 represents the HVAC systems distribution. Building B1 consumes 41.6 %, and B2 consumes 22.2 % of the total energy consumption. The input data of the automation system of these buildings are the outdoor and indoor temperature, humidity and pressure, as well as the electricity consumption of buildings B1 and B2 for each second, which correspond to the consumption from the global HVAC system. We focus our experiments on buildings B1 and B2, where the production line is. Hereafter, we make reference to B1 + B2 as if it were a single building.

The context of the selected building for experimentation is industrial, where people work 5 days per week. It should be noted that the productive process, the number of people and their working behavior are similar every working day. Due to the requirements of the production process carried out in this building, specific thermal comfort parameters are provided in different zones of the building. These thermal comfort values depend on the people working there, the activities carried out, the equipment installed and the products manipulated in each zone. In this way, the thermal comfort characterization of the building can be considered as the same every working day, and so the energy consumption associated to this service can be modeled, analyzing how the dynamic contextual conditions of the building affect the energy consumption for thermal comfort.

The global system B1 + B2 has an interval of confidence for the configuration of indoor temperature which is function to the activities carried out. This interval of confidence depends on the quality and the security of products, and the comfort of users (see Fig. 3). If the building is in a production phase, the indoor temperature must be between 5 and 26 °C. If the building is not in the production period, the interval of confidence for indoor temperature is between 17 and 25 °C. Therefore, it would be possible to change the indoor temper-

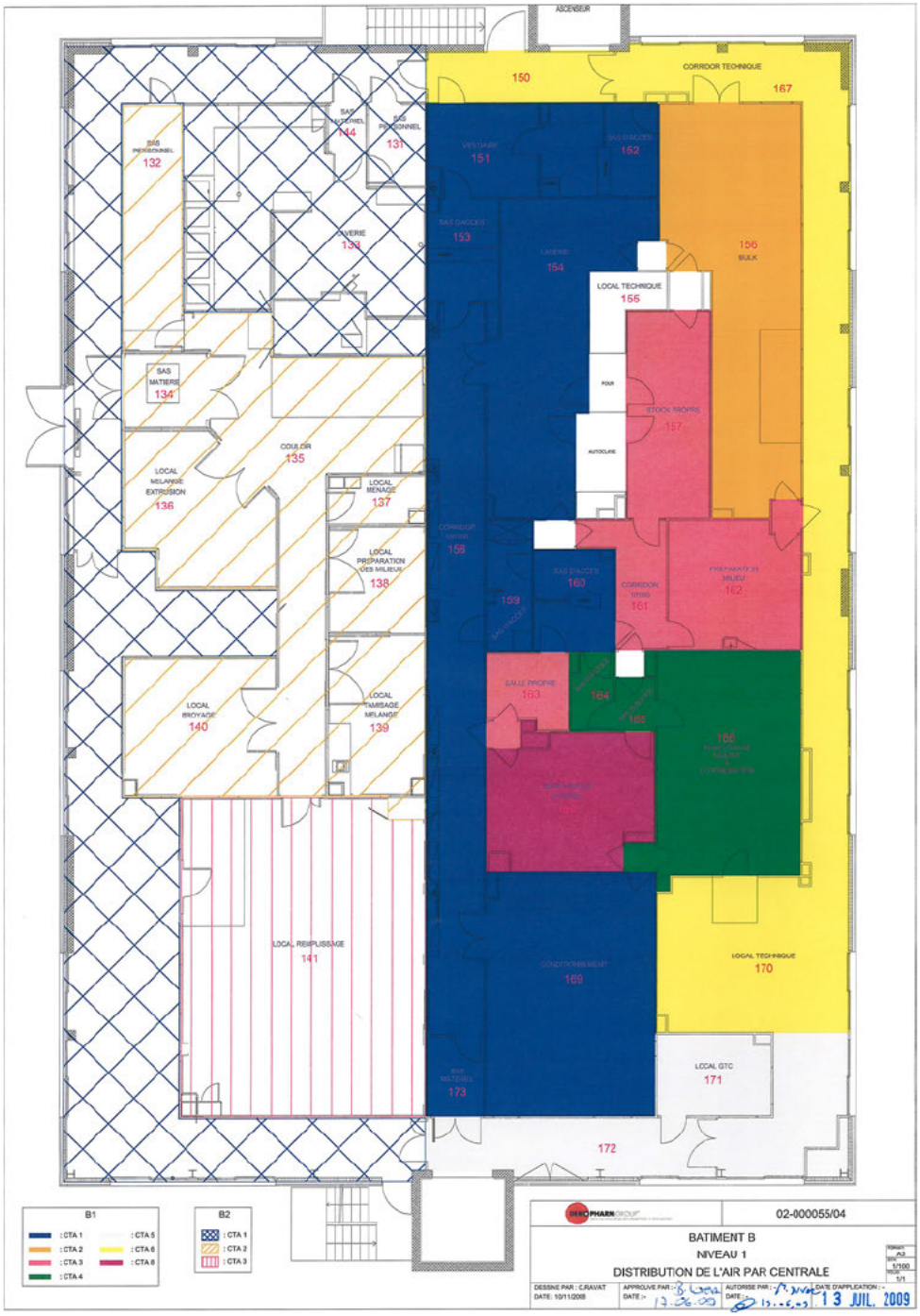


Fig. 2 HVAC Systems distribution in the buildings B1 + B2

ature to between 10 and 25 °C during a day without stopping the production process of the company.

On the other hand, different time periods can be identified in which the number of people and their behavior are usually similar. These cases are when the building is empty (between 00:00 and 07:00, and between 17:00 PM and 23:59), when workers come into the building to start their tasks (between 07:00 and 09:00), and when people leave the building to have lunch (between 11:00 and 14:00) or finish their work day (between 15:30 and 17:00). Furthermore, in this company it is obligatory to be working during the time periods 09:00–11:00 and 14:00–15:30. Bearing this in mind, we can use these periods of time to model the patterns of occupancy and people's behavior in this building, and then generate the model of the building energy consumption based on these time ranges.

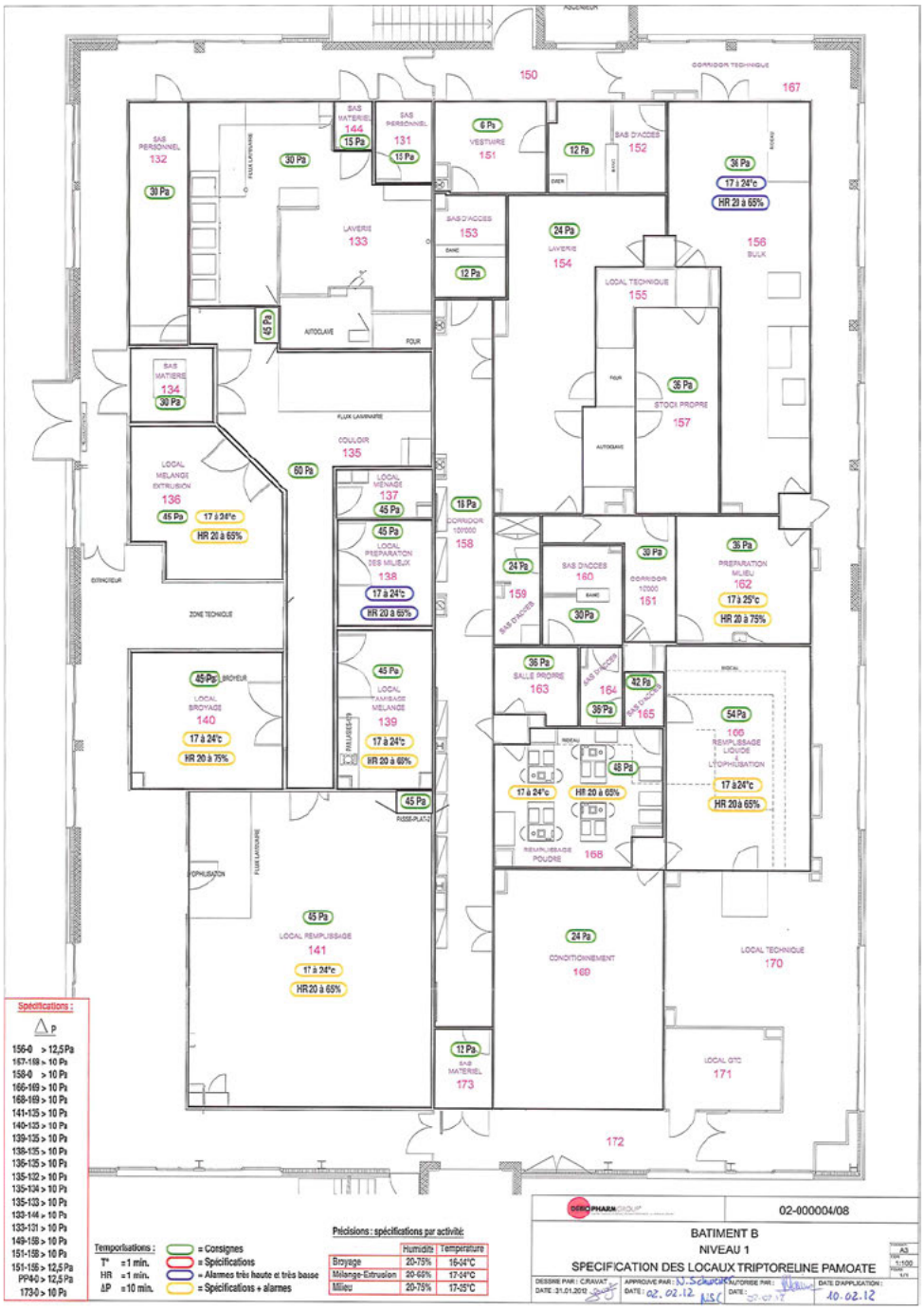


Fig. 3 Specifications of indoor pressure, temperature and humidity requirements of the buildings B1 + B2

4.2 Energy consumption characterization

The energy consumption patterns of this building for thermal comfort, taking only into account the productive process and its occupancy patterns, are similar every working day. Thus, we decided to analyze the impact of changes in environmental parameters in each one of the energy consumption patterns associated to the same production process and occupancy pattern.

In a previous work (Moreno et al. 2014), we demonstrated the relation between the energy consumption associated to the thermal comfort, environmental conditions and the occupancy patterns of the building. In that work we designed and carried out some associated control actions to save energy in

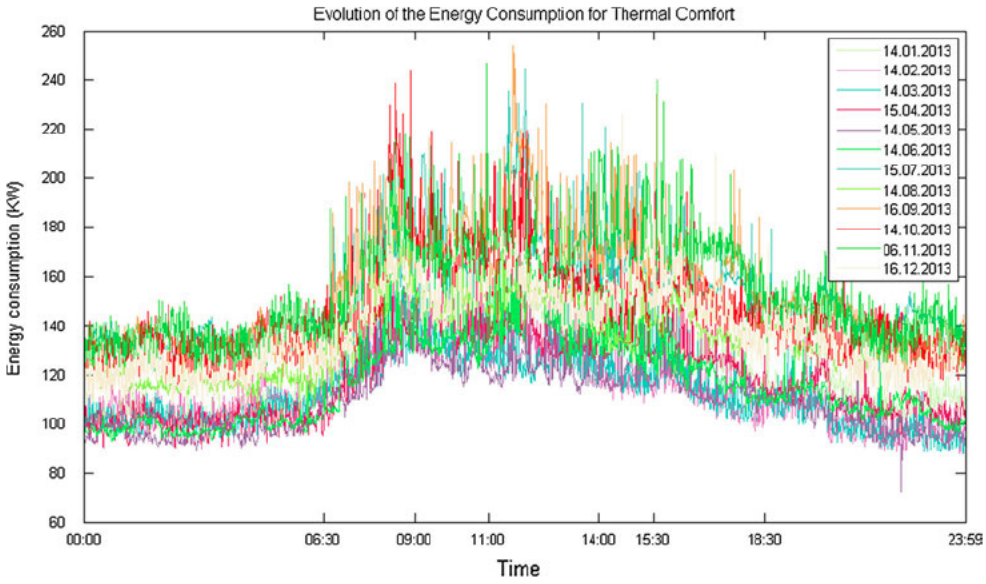


Fig. 4 Evolution of energy consumption for thermal comfort during 2013

buildings in different contexts. Now, our interest centers on obtaining the predictive models of the energy consumption of buildings for thermal comfort, taking as inputs the environmental conditions, the occupancy patterns and the productive process patterns.

For the energy consumption characterization of the reference building, we only consider the work days of the year 2013. Note that there is energy consumption due to thermal comfort in the target building during bank holidays and weekends due to the maintenance of products and the associated thermal comfort restrictions, but during these days there is no controllable parameter affecting energy consumption, i.e. there are no people, and so the comfort service provided these days is already adjusted to the minimal comfort requirements for the safe maintenance of products.

The energy consumed in this building for thermal comfort depends on the work planning of the people working there. To demonstrate this issue, in Fig. 4 we show the evolution of the energy consumed in the building for thermal comfort for each hour of different work days in the previous year (2013). We can observe how the evolution of the energy consumption follows a similar distribution for different days/months. Furthermore, different time periods can be distinguished in which energy consumption behaves in a similar way (constant, increasing or decreasing). As we shall see later, such time periods are useful for the data classification which will be carried out to generate the energy consumption model of the building.

Once the influence of time planning in the energy consumption of the building has been identified, we split the productive time planning of this building into different time periods with the aim of providing a different model of the energy consumption associated to each one of them. In this way, for the model associated to the time period when the building is empty, the energy consumption is due only to the thermal comfort requirements associated to the quality and safety of the products' maintenance. For the model associated to the time in which the building presents a constant number of people in it, the energy consumption for thermal comfort is due to both the quality and safety of the maintenance of products, and the comfort requirements of people. Furthermore, taking into account the flexibility in the working time of people, we distinguish the following three time periods which will be used to provide a different predictive energy consumption model of the building associated to them:

- No occupancy: [00:00, 07:00] and [17:00–23:59].
- Constant occupancy: [09:00, 11:00] and [14:00, 15:30].
- Variable occupancy: [07:00, 09:00] and [11:00, 14:00] and [15:30, 17:00].

Analyzing the energy consumption for thermal comfort provided in this building in each one of the mentioned time periods (see Fig. 5), we can see that when the building is empty, the energy consumption is the lowest, and when the number of occupants inside the building is constant, the energy consumption is the highest. This is because when the number of occupants in the building is constant, it presents its highest occupancy level, and so its energy consumption level is also the highest. In Fig. 5 such consumption levels are depicted.

Bearing all these aspects in mind, we distinguish between when the building is empty and when not. For the first case,

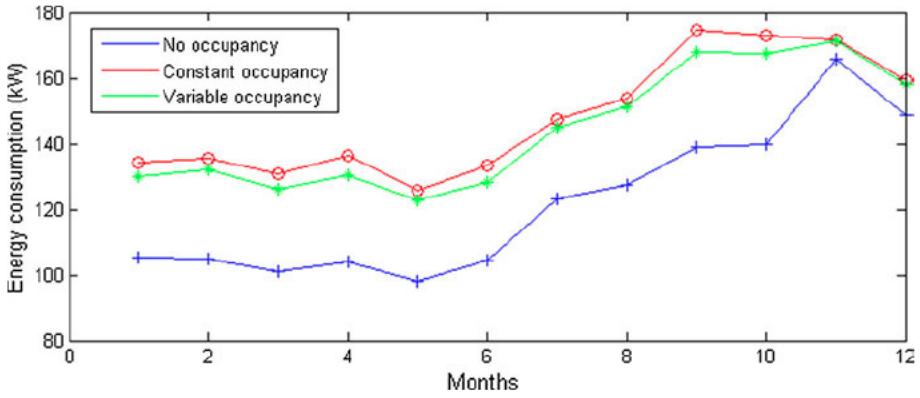


Fig. 5 Energy consumption levels of the building for different time periods

we propose some strategies to reduce the energy consumption only considering the impact of environmental conditions (see Sect. 4.5). But for the case when the building is not empty, we decide to generate a building model able to estimate the energy consumption for thermal comfort, taking into account both the environmental conditions and the occupancy patterns; in this way, it is possible to design more specific strategies to save energy in the target building since people can be involved in such strategies. The implementation of such a model is described below.

4.3 Generating the energy consumption models of the reference building

As already mentioned, due to the features of our reference building, we focus on modeling its energy consumption associated to the time periods in which the building is occupied. Taking into account the approach to generate building models presented in Sect. 3, we describe now the computational techniques selected for generating the energy consumption model of the target building. These techniques were selected because they presented the best results.

1. *Data collection* During this first stage, we collect data about outdoor/indoor temperature, humidity and pressure. Considering each one of the sensed parameters, "snapshots" of the energy consumption (EC) are collected over short periods of time (each minute of every day during a year). Such measurements are associated to specific vectors of environmental parameters measured outside and inside the building ($Z^{(t)}$). Several data collection processes are carried out, considering different context conditions. Thus, the building models generated will be sufficiently representative to cover different contextual conditions (different seasons for instance). So, the set of data pairs for the training of our building model is:

$$(\mathrm{EC}^{(t)}, Z^{t}), \quad t = 1, 2, \ldots, N \tag{1}$$

Table 1 Summary of the extracted features

	Feature
1	Mean outdoor temperature (°C/h)
2	Mean outdoor humidity (%/h)
3	Mean outdoor pressure (Pa/h)
4	Mean indoor temperature (°C/h)
5	Mean indoor humidity (%/h)
6	Mean indoor pressure (Pa/h)

where N is the number of data instances collected during 1 h of monitoring, $Z^{(t)} \in R^k$ and $\mathrm{EC}^{(t)} \in R^n$ refer to the environmental parameters vector associated to the energy consumption measured at the instant t.

2. *Pre-processing* The pre-processing unit is responsible for transforming the measured data. Besides, feature vectors are extracted from the data for use in energy consumption estimation. The different processing techniques applied in this stage are given as follows:
   - Transformation based on the raw dataset collected. During the transformation, compact representations of the input data, namely features, are extracted, which will be used later for energy consumption estimation. The values within the dataset are grouped into windows of 60 samples (one sample per minute), and each window is processed by several feature extraction methods, producing a feature vector that can be used to generate the clusters and train the classifier. The features adopted are summarized in Table 1.
   - Filtering. During this process a filter is applied that removes features extracted from the training data set that does not vary at all or that varies too much.
   - Normalization. All values in the given dataset are normalized during this phase. The resulting values are in the [0,1] interval for every feature extracted from the initial dataset.
   - Feature selection. We apply *principal components analysis (PCA)* in conjunction with a ranker search mechanism. PCA is a widely used technique for

reducing dimensionality in high-dimensional data, identifying the directions in which the observations most vary. If we consider EC(i) as multi-dimensional observations and u as an arbitrary direction in this multi-dimensional space, the principal components are calculated by optimizing the following equation:

$$\frac{1}{m} \cdot \sum_{i=1}^{m} (\mathrm{EC}(i)^T \cdot u)^2 \tag{2}$$

Dimensionality reduction is accomplished by choosing a sufficient number of vectors to account for a given percentage of the variance in the original data (by default 0.95). With the aim of reducing the final computational load of the estimation mechanism, we searched the optimum number of attributes to represent the energy consumption profile of our reference building. After this analysis, we found that outdoor temperature, humidity and pressure were the features selected by the ranked feature combination technique used by the PCA mechanism implemented in the WEKA toolkit.[3] Therefore, the number of features was reduced from the initial proposal of 6–3. Which will be denoted as $f1$, $f2$, $f3$. Note that indoor environmental conditions are directly the consequence of outdoor conditions, which is why not all of them are selected as principal features by the PCA. So, the energy consumption associated to thermal comfort of our reference building is due to outdoor environmental changes.
Considering this vector of features, Eq. (1) can be rewritten as:

$$\{[f1^{(t)}, f2^{(t)}, f3^{(t)}], Z^t\}, \quad t = 1, 2, \ldots, N \tag{3}$$

At this point, we generate the maps of the building based on the selected features. The stages described below refer to the mechanism based on such building maps.

3. *Clustering* During this stage, the input data division according to the distribution of the values of these features is carried out, the data being grouped according to the identified clusters, whose centroids are associated to landmarks.
We compared two techniques commonly used for clustering, the *simple expectation maximization (EM)* and the *simple K means* McGregor et al. (2004) in terms of success in classification by different classifiers (that is the next stage). These techniques were evaluated with tenfold cross-validation over our dataset, in which the original sample was randomly partitioned into 10 equal-size subsamples, a single subsample being retained as the validation data for testing the model, and the remaining nine subsamples used as training data. The results are found in Tables 2 and 3. These results show that the use of EM clustering yields a better classification. We then selected EM for our implementation.
EM assigns a probability distribution to each instance, indicating the probability of belonging to one of the identified clusters. EM can decide how many clusters to create by cross-validation, although the number of clusters to be generated can also be specified a priori. We propose an automatic search for the number of clusters that optimizes both classification success and accuracy in the energy consumption estimation (carried out later). For this, we follow a similar approach to that presented in Luna et al. (2011). Each one of the generated clusters is a vector of mean values of the outdoor environmental conditions forming the centroid of the cluster, and a vector of deviation values associated to the clusters. These vectors can be represented mathematically as: $\mu_{Ci} = [\mu_{f1}, \mu_{f2}, \mu_{f3}, \mu_Z]$, and $\sigma_{Ci} = [\sigma_{f1}, \sigma_{f2}, \sigma_{f3}, \sigma_Z]$; where μ_{Ci} and σ_{Ci} denote the mean and deviation of the centroid of the landmark i, respectively.
4. *Landmark classifier* The landmark classifier assigns each new vector of features to a specific landmark previously determined by the clustering algorithm. To select a suitable classification technique, we analyzed the performance of different classifiers. The corresponding results are summarized in Tables 2 and 3. As can be seen from

Table 2 Classification success rate of different clustering and classification techniques when occupancy is constant

Classification	EM (%)	Simple K means (%)
Decorate	94.7	92.6
LogitBoost	95.7	91.7
Bagging	92.8	89.2
J48	94.7	91.3
Random forest	94.5	91.1
Random tree	94.9	91.6

Table 3 Classification success rate of different clustering and classification techniques when occupancy is variable

Classification	EM (%)	Simple K means (%)
Decorate	96.9	93.4
LogitBoost	97.1	94.2
Bagging	96.7	93.2
J48	97.5	95.2
Random forest	97.9	93.1
Random tree	98.5	95.4

[3] http://www.cs.waikato.ac.nz/ml/weka/.

the table, the meta-classifier *LogitBoost* (Friedman et al. 2000) provided the highest classification success rate for the case of constant occupancy, and the tree-classifier *random tree* (Breiman 2001) when the occupancy is variable. Both evaluations are performed with tenfold cross-validation over the input dataset.

After classifying the energy consumption landmark for each new measurement, we can focus on the outdoor temperature characterization of such landmark and ignore the rest of the sensed values to carry out the energy consumption estimation.

5. *Energy consumption estimator* The next step consists of carrying out an energy consumption estimation using the knowledge available for the associated landmark. For this, a *Radial basis functions network* (Haykin 1999) for each landmark is computed as regression technique, which uses all training data associated to every landmark to estimate the energy consumption according to its associated outdoor temperature vector.

RBF networks find approximate solutions in the form of weighted sums of basis functions based on reference data. The main advantages of using RBF to solve our estimation problem are its scalability and easy deployment under different contextual conditions; when a variable number of centroids have been identified previously.

In our case, for each energy building consumption division associated to one landmark, an RBF network is implemented.

The input space P of our RBF is the vector of the mean values of the outdoor environmental parameters. These data can be denoted as

$$P \in R, P = \{p_i\}, \quad \forall p_i = \left[p_1, p_2, \ldots, p_n\right] \tag{4}$$

where n is the number of measurements gathered and classified within the chosen subset associated to a landmark. The target class Z represents the energy consumption. This is denoted as

$$Z \in R^k, Z = \{z_i^k\}, \forall z_i^k = \left[z_1^k, z_2^k, \ldots, z_n^k\right] \tag{5}$$

where k is the dimension. In our case: $k = 1$. Then, given the training values $\{(p_i, z_i^k), \ldots, (p_n, z_n^k)\}$, our goal is to find a function that will allow us to classify the monitored energy consumption (z_i), giving its vector of features $(p_i = [f1_i, f2_i, f3_i])$.

The vector p_j is provided as input to all functions of our RBF network, and the output $f(p_j)$ is given by

$$f(p_j) = \sum_{i=1}^{c} w_i \cdot \varphi(\parallel p_j - c_i \parallel) \tag{6}$$

where $\parallel p_j - c_i \parallel$ is the Euclidean distance between p_j and the RBF function with center c_i.

The number of RBFs is C, and w_i are the weights of the network. The value of β specifies the width of the basis functions and allows their sensitivity to be adjusted. As β decreases, the basis functions become wider and overlapping may increase. An appropriate value of β is usually selected experimentally based on the reference data, and this can be further adjusted when testing data are available. Another common practice is to use a heuristic method to set the width β according to Eq. (7), where $d_{\max} = \parallel p_j - c_i \parallel$ for $i = 1, \ldots, L$.

$$\beta = \frac{1}{2 \cdot d_{\max}} \tag{7}$$

From this equation it is deducted that when the distance between centers in input features increases, the value of β is reduced to ensure that the basis functions still overlap sufficiently to produce accurate energy consumption estimates. In our implementation, and following this scheme, the value of β is easily adjusted to provide a high level of accuracy, even when a variable number of inputs are used.

The proper values for C and the centers c_i are not unimportant since they affect the performance of the RBF network. A common practice is to use each reference feature value to define the centers, so if there are L vectors of input features, there will be L basis functions. However, this architecture has high memory requirements when there are many reference fingerprints. In such cases, the computational complexity is high, both for the calculation of w_i and energy consumption estimation. Nevertheless, in our system the number of reference inputs per cluster is low, meaning that there are no problems related to computational complexity, and it is possible to use the centroids obtained previously during clustering phase to provide the center of our basis functions. For these reasons, our RBF system has one solution and its design guarantees the exact fitting of all reference data.

The reference input values and their corresponding energy consumption estimations (EC_i) are used to train the network and adjust the weights each time new information is registered. Thus, given a target vector of features p_j associated to the energy consumption z_j, the output of the RBF network may be expressed as a weighted sum of normalized basis functions:

$$z(p_j) = \sum_{i=1}^{c} w_i \cdot \frac{\varphi(\parallel p_j - c_i \parallel)}{\sum_{k=1}^{c} \varphi(\parallel p_j - c_k \parallel)} \tag{8}$$

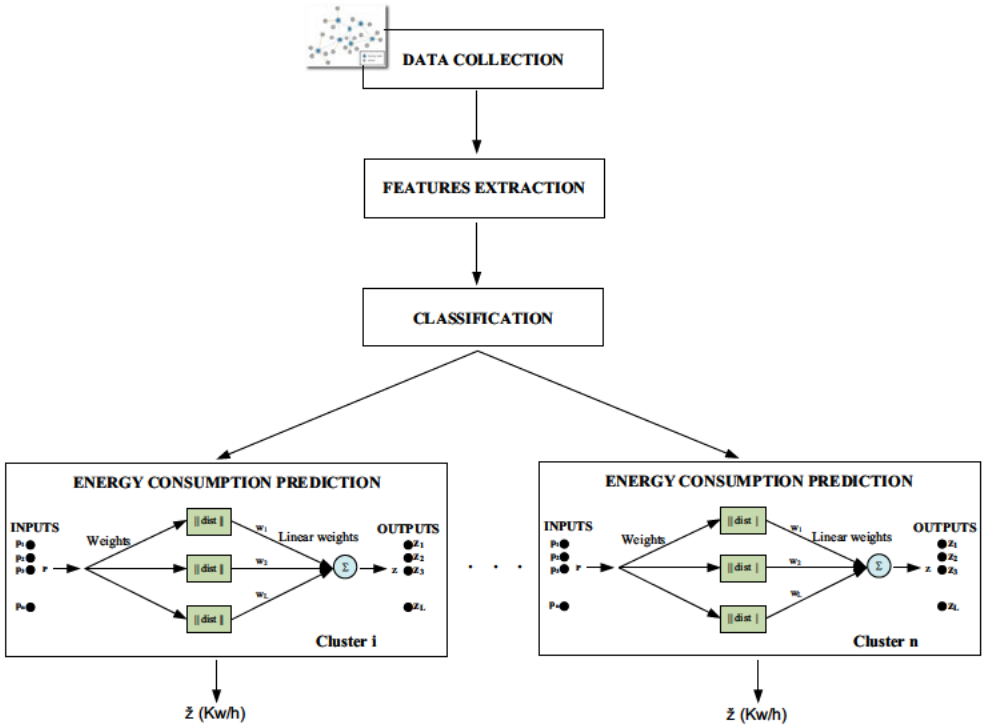


Fig. 6 Computational model for estimating the energy consumption of buildings

where w_i are one-dimensional weights. The parameter w_i may be determined to obtain a good approximation by optimizing the fit represented by the difference between the input values of the reference data and the test targets [Eq. (8)]. Thus, we form the following set of equations:

$$z(p_k) = \sum_{i=1}^{c} w_i \cdot u(\| p_k - c_i \|), \quad k = 1, \ldots, L \tag{9}$$

we calculate w_i by solving the system of linear equations based on Eq. (9) and using the reference values of the database and their corresponding energy consumption estimations. Therefore, our resulting RBF avoids over-fitting. Subsequently, given a new vector of features p'_j, the weights w_i are used during the estimation process to obtain an energy consumption estimate $\hat{z}$ according to:

$$\hat{z}(p'_j) = \sum_{i=1}^{c} w_i \cdot u(\| p'_j - c_i \|) \tag{10}$$

4.4 On-line energy consumption estimation

After the off-line phase, energy consumption can be estimated using the building maps generated during the off-line stage, and the RBF-based estimator implemented. A schema of the steps carried out during the on-line phase of our mechanism is shown in Fig. 6. The input data are the outdoor environmental measurements around the buildings, from which the features are extracted. This feature vector is classified as belonging to a particular landmark cluster. Finally, the building energy consumption is estimated using the corresponding RBF that has been implemented for this landmark.

4.5 Evaluation and analysis

Using all data collected about the outdoor environmental parameters and the energy consumed for thermal comfort in the reference building, we analyze the results obtained in terms of the success in classification and the estimation error in the energy consumption. Table 4 summarizes the results. The first column refers to the number of data (ND) considered for each cluster automatically generated after clustering (NC), and the column with maximal energy consumption (maxEC) refers to the maximum value of energy consumption for such case. Finally, the four last columns refer to the average and deviation of the percentage of success in classification (PSC) and the estimation error (EE), all of them evaluated with tenfold cross-validation over the input dataset.

We can see that for the first model generated (when the number of people in the building is constant), the error in the estimation is lower than for the case with a variable number of people. This is because the 2nd model presents greater variability related with the number of people in the building,

Table 4 Number of training data (ND), maximal energy consumption (maxEC), number of clusters (NC), mean percentage of success in classification (μ_{PSC}), deviation of success in classification (δ_{PSC}), average estimation error (μ_{EE}) and deviation of estimation error (δ_{EE}) for each building model

Models	ND	maxEC (kW)	NC	μ_{PSC} (%)	δ_{PSC} (%)	μ_{EE} (kW)	δ_{EE} (kW)
Constant occupancy	490	195.4	8	95.7	2.3	8.8	3.0
Variable occupancy	735	186.8	5	98.5	1.6	20.5	4.3

so that the model loses accuracy in the estimation, which is translated into a greater error in the estimation. The results obtained for these models demonstrate the suitability of the techniques finally selected for implementing the mechanism proposed to generate energy building consumption models.

After proposing these models, and according to the results obtained from them, the optimal strategies to save energy in the target building can be selected. Some examples of strategies to save energy, and which are being performed currently, are enumerated below for the two cases proposed in this problem.

- *When the building is empty*, the indoor temperature necessary inside the building to ensure the good state of the different products can have a value in the confidence interval: [5, 25 °C]. Then, depending on the outdoor temperature measured, the indoor temperature can be configured with the most similar value to this, always being within the mentioned comfort interval, and considering the expected time when the comfort conditions must change, ensuring the energy conservation in the building for the minimum associated energy consumption. For this, the energy consumption building management is able to reason over a set of rules based on the current contextual data processing carried out. Therefore, when it is known that the building is empty according an operational time scheduling, some rules are activated, and the corresponding commands are configured in the HVAC system of this building.
- *When the building is occupied*, depending on the energy consumption estimated by the implemented model of the building for the next hour, different strategies to save energy can be carried out, such as:
  1. Adjusting the work of people to shorten the periods of time with a variable occupancy in the building.
  2. Providing occupants with comfort conditions that save energy while maintaining suitable levels of comfort. For this, it must be considered that when the building is not empty, the indoor temperature that must be present in this building should have a value within to the confidence interval: [16, 26 °C]. Then, depending on the estimated energy consumption, we select the value most similar to the outdoor temperature for establishing the indoor temperature while taking into account comfort conservation at the same time.
  3. If there is a source of renewable energy, its optimum distribution in the building should be designed in accordance with any expected abrupt changes in the outside environmental conditions, and consequently, in the energy consumption.

5 Conclusion and future work

Energy building management is recognized as a fundamental piece for ensuring energy sustainability in modern cities and the planet as a whole. The solutions put forward to date concerning the great amount of energy consumed by buildings have numerous limitations and many are extremely complicated to undertake.

The many advances made in ICT, and especially IoT, represent a great potential as regards the quantity of real-world information that can be generated and, at the same time, permit interaction with the environment to change behavior and provide more efficient services.

In this work, we analyze the main parameters affecting energy consumption of buildings. Such an analysis permits us to propose an optimum prediction concerning the hourly energy consumed in buildings by integrating the most relevant input data involved in energy consumption.

For the extraction of relevant knowledge from all the sensed data, we apply sophisticated SC techniques to model the energy consumption profile of buildings. Thus, our approach to address energy efficiency in buildings is based on using the data measured by sensors installed in the building to generate the predictive model that estimates the energy consumption of the same.

Once energy usage profiles have been extracted, we can design and implement actions to save energy; for instance, proposing strategies to adjust the operation time and configuration of the involved appliances or devices, selecting the optimal distribution of energy to maximize the use of alternative energies, etc.

As a real use case where energy saving must be achieved, we present a scenario where the first stages of experimentation have been carried out following the approach proposed

in this work. From this, accurate predictive models of the energy usage of a reference building have been obtained. For the generation of such predictive models, we carry out an exhaustive selection of the data mining techniques that provide the optimum results, taking into account the data collected during a monitored year. Then, considering such models, we propose some measurements of control to save energy in the building taken as reference.

At present we are carrying out experiments to analyze the impact of implementing the strategies proposed to save energy in the building under experimentation. Furthermore, we are testing the performance of our predictor when weather forecast is included as input of the model for making prediction of the energy consumed by the building during next hours or days, and in this way, it is possible to anticipate and design the optimum strategies to save energy. As future work we will apply the approach proposed in this paper to other buildings, with the aim of demonstrating its applicability in different contexts.

Acknowledgments This work has been sponsored by European Commission through the FP7-SMARTIE-609062 project.

References

Agarwal Y, Balaji B, Gupta R, Lyles J, Wei M, Weng T (2010) Occupancy-driven energy management for smart building automation. In: Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building. ACM, pp 1–6

Al-Homoud MS (2001) Computer-aided building energy analysis techniques. Build Environ 36(4):421–433

Berglund L (1977) Mathematical models for predicting thermal comfort response of building occupants. In: Ashrae Journal—American Society of Heating Refrigerating and Air-Conditioning Engineers, vol. 19(12). Amer Soc Heat Refrig Air-Conditioning Eng Inc 1791 Tullie Circle Ne, Atlanta, 30329:38-38

Berthold MR, Borgelt C, Höppner F, Klawonn F (2012) Guide to intelligent data analysis. Springer, London, Dordrecht, Heidelberg, New York

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Chen Z, Clements-Croome D, Hong J, Li H, Xu Q (2006) A multicriteria lifespan energy efficiency approach to intelligent building assessment. Energy Build 38(5):393–409

Clarke J, Cockroft J, Conner S, Hand J, Kelly N, Moore R, O'Brien T, Strachan P (2002) Simulation-assisted control in building energy management systems. Energy Build 34(9):933–940

Costa HRDN, La Neve A (2015) Study on application of a neuro-fuzzy models in air conditioning systems. Soft Comput 19:929–937. doi:10.1007/s00500-014-1431-5

Crawley DB, Lawrie LK, Winkelmann FC, Buhl WF, Huang YJ, Pedersen CO, Strand RK, Liesen RJ, Fisher DE, Witte MJ et al (2001) Energyplus: creating a new-generation building energy simulation program. Energy Build 33(4):319–331

Crawley DB, Hand JW, Kummert M, Griffith BT (2008) Contrasting the capabilities of building energy performance simulation programs. Build Environ 43(4):661–673

British Standards Institution (2007) Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics. European committee for Standardization. Standardization, 2007

Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). Ann Stat 28(2):337–407

Garg V, Bansal N (2000) Smart occupancy sensors to reduce energy consumption. Energy Build 32(1):81–87

Hagras H, Callaghan V, Colley M, Clarke G (2003) A hierarchical fuzzy-genetic multi-agent architecture for intelligent buildings online learning, adaptation and control. Inf Sci 150(1):33–57

Haykin SS (1999) Neural networks: a comprehensive foundation. Prentice Hall, Upper Saddle River, NJ

Lindberg R, Binamu A, Teikari M (2004) Five-year data of measured weather, energy consumption, and time-dependent temperature variations within different exterior wall structures. Energy Build 36(6):495–501

Luna F, Estébanez C, León C, Chaves-González JM, Nebro AJ, Aler R, Segura C, Vega-Rodríguez MA, Alba E, Valls JM et al (2011) Optimization algorithms for large-scale real-world instances of the frequency assignment problem. Soft Comput 15(5):975–990

Lu J, Sookoor T, Srinivasan V, Gao G, Holben B, Stankovic J, Field E, Whitehouse K (2010) The smart thermostat: using occupancy sensors to save energy in homes. In: Proceedings of the 8th ACM conference on embedded networked sensor systems. ACM, pp 211–224

McGregor A, Hall M, Lorier P, Brunskill J (2004) Flow clustering using machine learning techniques. In: Passive and active network measurement. Lecture notes in computer science, vol 3015. Springer, Berlin, Heidelberg, pp 205–214

Moreno M, Úbeda B, Skarmeta AF, Zamora MA (2014) How can we tackle energy efficiency in IoT based smart buildings? Sensors 14(6):9582–9614

Perera C, Zaslavsky A, Christen P, Georgakopoulos D (2014) Sensing as a service model for smart cities supported by internet of things. Trans Emerg Telecommun Technol 25(1):81–93

Perez-Lombard L, Ortiz J, Pout C (2008) A review on buildings energy consumption information. Energy Build 40(3):394–398

Petersen D, Steele J, Wilkerson J (2009) Wattbot: a residential electricity monitoring and feedback system. In: Proceedings of the 27th international conference extended abstracts on human factors in computing systems. ACM, pp 2847–2852

Scott J, Bernheim Brush A, Krumm J, Meyers B, Hazas M, Hodges S, Villar N (2011) Preheat: controlling home heating using occupancy prediction. In: Proceedings of the 13th international conference on ubiquitous computing. ACM, pp 281–290

Severini M, Squartini S, Piazza F (2013) Hybrid soft computing algorithmic framework for smart home energy management. Soft Comput 17(11):1983–2005

Voss K, Sartori I, Napolitano A, Geier S, Gonzalves H, Hall M, Heiselberg P, Widén J, Candanedo JA, Musall E, Karlsson B, Torcellini P (2010) Load matching and grid interaction of net zero energy buildings. In: Proceedings of EuroSun 2010, Graz, Austria, September 28–October 1, 2010

Zoha A, Gluhak A, Imran MA, Rajasegarar S (2012) Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey. Sensors 12(12):16838–16866