# Exploiting IoT-based Sensed Data in Smart Buildings to Model its Energy Consumption

M. Victoria Moreno* and Antonio F. Skarmeta*
Luc Dufour†, Dominique Genoud† and Antonio J. Jara†
Bruno Ladevie‡ and Jean-Jacques Bezian‡
* Computer Science Faculty, University of Murcia
Espinardo Campus, 30100 Murcia, Spain
{mvmoreno, skarmeta}@um.es
† Institute of Information Systems
University of Applied Sciences Western Switzerland (HES-SO)
Sierre, Switzerland
{Luc.Dufour, Dominique.Genoud, Antonio.Jara}@hevs.ch
‡ Mines Telecom, Albi, France
{Bruno.Ladevie, Bezian}@mines-albi.fr

***Abstract*—Due to the high impact that energy consumption of buildings has at global scale, it has been stated the need of achieving energy-efficient buildings to reduce $CO_2$ emissions and energy consumption at global scale. In this work we propose to model the energy consumption associated with services provided in buildings to help select the best strategies to save energy. To verify the feasibility of the proposed approach using measurements of relevant parameters affecting, we carry out some analysis in a reference building of which we have contextual data. Firstly, we provide a complete characterization of this building in term of its energy consumption and generate accurate building models able to predict its energy consumption given a concrete set of inputs. Finally, considering the generated energy usage profile of the building, we propose some concrete control actions and strategies to save energy.**

***Keywords*—Data intelligence analysis; Microgrid; Energy information management; Advanced Metering Infrastructure;***

## I. 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 [1]. 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 own energy sources to ensure their energy sustainability), but also of improving occupant experience and productivity.

Analysis of the energy efficiency of the built environment has received growing attention in the last decade [2] [3]. 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 [4].

The integration and development of systems based on Information and Communication Technologies (ICT) and, more specifically, the Internet of Things (IoT) [5], 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 analysed deeply in order 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 analyse such datasets [6]. 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

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

used to select the optimal measurements and strategies to save energy. Firstly, 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: Section II presents our approach to generate energy consumption maps of buildings to help us define the best strategies and action to save energy. Section III details the energy usage characterization of the building used as reference, the process of generation of the energy consumption maps of the building. Section IV describes the analysis extracted from the models generated following the approach proposed in this paper. Finally, Section V provides conclusions and future directions of our work.

## II. Methodology

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.

## III. Energy consumption characterization

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 is 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.

It has been stated that the impact of HVAC on the energy consumption of a building represents 76% of the total in European countries [7]). As regards the thermal comfort provided in our reference building, due to its high volume, we focus on modeling this consumption.

Different time periods can be identified in which the number of occupants and their behavior are usually similar in the target building. 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.

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 analyse the impact of changes in environmental parameters in each one of the energy consumption patterns associated to the same production process and occupancy pattern.

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.

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 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]

Bearing all these aspects in mind, we distinguish between when the building is empty and when not. For the first case, we propose some strategies to reduce the energy consumption only considering the impact of environmental conditions (see Section IV). 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.

### *A. 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 Section II, 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.

Table I. 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) |

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:

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

where $N$ is the number of data instances collected during one hour of monitoring, $Z^{(t)} \in R^k$ and $EC^{(t)} \in R^n$ 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:
- 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 I.
- Filtering. During this process a filter is applied that removes features extracted from the training data set that do not vary at all or that vary 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:

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

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[2]. Therefore, the number of features was reduced from the initial proposal of 6 to 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:

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

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 Maximisation (EM)* and the *Simple K Means* [8] in terms of success in classification by different classifiers (that is the next stage). These techniques were evaluated with 10-fold 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 9 subsamples used as training data. The results can be found in Tables II and III. 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 [9]. 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 $\mu_{Ci}$ and $\sigma_{Ci}$ 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. In order to select a suitable classification technique, we analysed the performance of different classifiers. The corresponding results are summarised in Tables II and III. As can be seen from the table, the meta-classifier *LogitBoost* [10] provided the highest classification success rate for the case of constant occupancy, and the tree-classifer *Random Tree* [11] when the occupancy is variable. Both evaluations are performed with 10-fold 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* [12] 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 condition; when a variable number of centroids has been identified previously.

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

### B. 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 can be seen in Figure 1. 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.

## IV. 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 term of the success in classification and the estimation error in the energy consumption. Table IV summarizes the results. The first column refers to the number of data (ND) considered for each cluster automatically generated after clustering (NC),

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

Table II. 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 III. 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% |

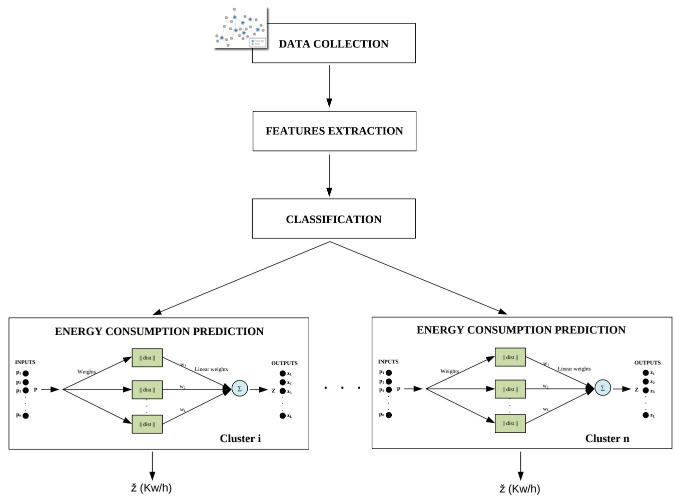


Figure 1. Computational model for estimating the energy consumption of buildings

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 success in classification (SC) and the estimation error (EE), all of them evaluated with 10-fold 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,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°C, 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.
- **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°C, 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.

## V. CONCLUSION AND FUTURE WORK

The ICT and especially IoT provide a quantity of information and our role is to provide the maximum of knowledge of different uses cases identified. In this work, we want decrease the energy consumption by the prediction with the outdoor temperature and the work planning of people.

Such an analysis permits us to propose an optimum prediction concerning the daily energy consumed in buildings by

Table IV. NUMBER OF TRAINING DATA (ND), MAXIMAL ENERGY CONSUMPTION (MAXEC), NUMBER OF CLUSTERS (NC), AVERAGE SUCCESS IN CLASSIFICATION ($\mu_{SC}$), DEVIATION OF SUCCESS IN CLASSIFICATION ($\delta_{SC}$), AVERAGE ESTIMATION ERROR ($\mu_{EE}$) AND DEVIATION OF ESTIMATION ERROR ($\delta_{EE}$) FOR EACH BUILDING MODEL

| **Models** | **ND** | **maxEC (kW)** | **NC** | $\mu_{SC}$ **(%)** | $\delta_{SC}$ **(%)** | $\mu_{EE}$ **(kW)** | $\delta_{EE}$ **(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 |

integrating the most relevant input data involved in energy consumption. Our approach is based on using the sensed data measured by sensors installed in the building to generate the predictive model that estimates the energy consumption for thermal comfort, considering the behavior patterns of the parameters identified as main contributors.

This models are validated by debiopharm and a new project created enable the implementation. The first step is the control of temperature when we don't have a people in the building. Furthermore, we add the humidity and pressure outdoor in the input data. 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. In order to provide the good strategies, it's essential to provide the electricity prices information to adapt its consumption.

## ACKNOWLEDGMENT

This research was undertaken as part of the Adaptive IES Project and has been funded by The Ark Energy foundation (Project No. 712-07). We would like to thank the contributions from the Debiopharm partners: Vincent Griffoul and Lionel LeRoux.

## REFERENCES

[1] D. Petersen, J. Steele, and J. Wilkerson, "Wattbot: a residential electricity monitoring and feedback system," in *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*. ACM, 2009, pp. 2847–2852.

[2] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng, "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, 2010, pp. 1–6.

[3] R. Lindberg, A. Binamu, and M. Teikari, "Five-year data of measured weather, energy consumption, and time-dependent temperature variations within different exterior wall structures," *Energy and Buildings*, vol. 36, no. 6, pp. 495–501, 2004.

[4] K. Voss, I. Sartori, A. Napolitano, S. Geier, H. Gonçalves, M. Hall, P. Heiselberg, J. Widén, J. A. Candanedo, E. Musall *et al.*, "Load matching and grid interaction of net zero energy buildings," 2010.

[5] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Sensing as a service model for smart cities supported by internet of things," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 1, pp. 81–93, 2014.

[6] M. R. Berthold, C. Borgelt, F. Höppner, and F. Klawonn, *Guide to intelligent data analysis*. Springer, 2012.

[7] L. Perez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and Buildings*, vol. 40, no. 3, pp. 394–398, 2008.

[8] A. McGregor, M. Hall, P. Lorier, and J. Brunskill, "Flow clustering using machine learning techniques," in *Passive and Active Network Measurement*. Springer, 2004, pp. 205–214.

[9] F. Luna, C. Estébanez, C. León, J. M. Chaves-González, A. J. Nebro, R. Aler, C. Segura, M. A. Vega-Rodríguez, E. Alba, J. M. Valls *et al.*, "Optimization algorithms for large-scale real-world instances of the frequency assignment problem," *Soft Computing*, vol. 15, no. 5, pp. 975–990, 2011.

[10] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors)," *The annals of statistics*, vol. 28, no. 2, pp. 337–407, 2000.

[11] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[12] H. Simon, *Neural networks: a comprehensive foundation*. Prentice Hall, 1999.