

Energy demand prediction for the implementation of an energy tariff emulator to trigger demand response in buildings

Sarah Noyé^{1,*}, Unai Saralegui¹, Raphael Rey², Miguel Angel Anton¹ and Ander Romero¹

¹TECNALIA, Building Technologies department, 20009 San Sebastian, Spain

²Mines de Nancy, Energy and Fluids, 54042 Nancy cedex, France

Abstract. Buildings are key actors of the electrical grid. As such they have an important role to play in grid stabilization, especially in a context where renewable energies are mandated to become an increasingly important part of the energy mix. Demand response provides a mechanism to reduce or displace electrical demand to better match electrical production. Buildings can be a pool of flexibility for the grid to operate more efficiently. One of the ways to obtain flexibility from building managers and building users is the introduction of variable energy prices which evolve depending on the expected load and energy generation. In the proposed scenario, the wholesale energy price of electricity, a load prediction, and the elasticity of consumers are used by an energy tariff emulator to predict prices to trigger end user flexibility. In this paper, a cluster analysis to classify users is performed and an aggregated energy prediction is realised using Random Forest machine learning algorithm.

1 Introduction

The increase of renewable energy in the energy mix implies the need to find new ways to manage the energy grid. Indeed, intermittency of energy production makes it difficult to align the production with the energy demand. In the traditional energy production model, centralized energy plants are turned on or off to supply the demand, but with renewable energy, there might be periods when the energy demand is higher than the production, while excess energy is produced at other moments. There is also an interest to reduce energy consumption peaks to avoid, on one hand to provide stability to the grid and avoid power failure, on the other hand to reduce energy prices as turning additional plants on for short amount of time is expensive. Due to the cost of electricity storage, a change of paradigm is required, where the energy demand is adapted to the energy available. This is referred to as demand response (DR). Demand response apply primarily to electricity.

DR is traditionally applied in large commercial buildings and industries with high energy consumption and where the main process is directly related with energy consumption, as a few actors can provide significant flexibility to the grid [1]. To achieve the objectives set by the EU Renewable Energy Directive of 20% of renewables by 2020 [2] and 27% by 2030 [3], more sources need to be available for the demand response mechanism. Because they consume around a third of the final energy consumption, buildings are an important part of the energy grid and a potential reserve of energy flexibility, notably through their HVAC systems that can make use of thermal energy storage [4].

The extension of demand response to the commercial and residential building market require the development of new mechanism to balance production and demand. One of them is the use of tariff signals to encourage users to reduce or displace their consumption when there is an energy demand peak. Several pricing schemes already exist [5]. One of the key aspects of demand response through price signal is a good understanding of the client portfolio and prediction of its aggregated energy consumption.

The paper is structured as follow. First some background on demand response and the stakeholders involved is given. Then the principle of an energy tariff emulator to trigger demand response in residential and small/medium commercial building is presented. The method and preliminary analysis to devise this price emulator will then be presented and discussed. It includes a cluster analysis of the customer and the training of a model to predict aggregated energy consumption.

2 Background

To understand the stakes of demand response in the context of buildings, this section presents the stakeholders involved and the main demand response scheme categories.

2.1 Stakeholders

There are 4 main actors involved in demand response schemes: the end user, the utility, the retailer and the aggregator. End users and utilities are the traditional actors of the sector. New energy services like demand

* Corresponding author: sarah.noye@tecnalia.com

response or energy efficiency see the emergence of other actors such as the aggregator or the retailer.

2.1.1 End user

The end user is the final client that consumes the energy. It is the household in the case of residential buildings and the facility manager in non-residential buildings.

Both actors have similar objectives regarding demand response, which is to pay low energy bills while optimising their comfort.

Demand response can give residential and medium/small tertiary consumers an active role in their own electricity consumption. User engagement is one of the challenges of demand response.

2.1.2 Utility

Utilities produce and distribute the energy. Those two functions can be carried out by different companies or different filial of the same company. Their objective is to minimize the cost of energy production and of grid management while satisfying energy needs. This is easier to achieved for regular electricity production as it is costly to turn energy plants on and off for a short period. In this sense, it is in the interest of the utility to decrease the consumption peaks.

Furthermore, utilities are under the pressure of producing more electricity from renewable sources, which means that they become more reliant on the external conditions (sun, wind, etc). Therefore, they want the energy consumption to adapt itself to the production.

2.1.3 Aggregator

The necessity to optimise the energy grid has led to the apparition of new actors that serve as intermediary between the end user and the utility. One company can play both the roles of aggregator and retailer, depending on the contract that it has with its clients.

The aggregator role is to constitute a portfolio of users that can offer flexibility to the utility. The aggregator signs a contract with the end user to determine the conditions of the flexibility it can provide. This includes the type of action possible (e.g.: change of set points, turning off some equipment) and the related time constraint (e.g.: a certain number of hours during the year, not more than two hours in a row). The end users are monetarily compensated for their flexibility.

The aggregator then responds to the utility's need for demand response by activating the flexibility in its portfolio.

The objective of the aggregator is to be a mediator between the utility and the end-user. It makes his profit from the difference between the income it receives from the utility to provide flexibility and the compensation it pays its client.

2.1.4 Retailer

A retailer buys energy from the utility and sells it to the end user. Various companies can play the role of retailer. It can go from being a department of a utility company or an independent private company to be a municipality or an association of users.

In the first case the objective would be to maximize profit through energy trading, while in the second case it is more likely to be about providing the best energy prices for the end users.

2.2 Explicit demand response

When the utility triggers a demand response event, there are two ways to obtain flexibility from the end user. The first one, called explicit demand response, is when there is a direct external control on the systems of the end user under conditions agreed in advance. This is usually done via the aggregator.

In this scenario, the utility would ask the aggregator to provide a certain amount of flexibility, either by reducing or displacing the load of its clients. Based on the flexibility provided by its clients, the aggregator determines the best strategy to provide this flexibility. This can include a phase of negotiation both between the utility and the aggregator and between the aggregator and its clients. Using specific protocol and hardware [6], the aggregator sends direct commands to the client's equipment. The utility rewards the aggregator for the flexibility and the aggregator passes down part of the remuneration to the clients that provided a change in their consumption.

This type of DR is primarily applied to non-residential buildings where significant reduction in consumption can be achieved by turning-off or reducing the load of HVAC or lighting system with high power consumption. Explicit DR permits a more precise control on the load, because the aggregator has detailed information about its clients (building characteristics, habit, preferences) and has some degree of direct control over its clients' consumption.

2.3 Implicit demand response

By opposition to explicit DR where the aggregator can control part of the end user load, implicit DR consists in asking for a change in consumption from the consumer but leaving them the choice to do it or not. One way to do this is through energy prices, where prices are high at time the consumption needs to be reduced and prices are low when there is an excess of energy to encourage load displacement. This can go from static peak/off-peak pricing to encourage the end user to use energy at night to dynamic pricing, adjusted to energy production in real time. This scenario involves the retailer, who would adapt prices based on the contract signed with the end-user.

The utility sells energy at a previously fixed wholesale price. This price is published the day before and depends on a prevision of the cost of production and of global energy demand. The retailer buys electricity at its

wholesale price and will later modify such price before selling the energy to the end user, fixing the retail price.

The retail price can be static (e.g.: time of use) or dynamic. In case of dynamic pricing, a mechanism such as a phone app can be used to send the variation of prices to the end user. There is also the possibility to automate some of decisions via connected equipment such as a smart thermostat that would change its set points depending on the energy price according to rules established by the end user.

Implicit demand response through price signal rely on the economical concept of elasticity, which is the likelihood of the consumer to change behaviour because of a change in price under fixed circumstances. Correctly estimating elasticity to obtain the adequate response from the end user in one of the challenges of demand response, especially in the case of the dynamic pricing scheme [7].

Implicit demand response relies strongly on consumers behaviours. The degree of control is thus smaller. However, it opens the door to reach a larger consumer population as the barrier to enter the scheme is lower.

This type of demand response is suitable both for commercial and residential buildings. In the case of dynamic prices for residential buildings, the implementation of some form of automation of the decision is suitable, as involvement is likely to decrease overtime if users must look for prices variation constantly.

3 Method

3.1 HOLISDER project

In the framework of the H2020 program funded by the European Union, the HOLISDER project works on introducing a Holistic Demand Response Optimization Framework to reduce the total energy bill for the consumer and improve the electric network stability [8]. Existing technologies are brought together to create a framework that covers the entire demand response value chain. One of the key aspects of this framework is the definition of a common information model integrating existing standards (e.g: OpenADR) and bridging the gaps to create an interoperable solution for building demand response.

The HOLISDER framework addresses both implicit and explicit demand response. In this paper, we will focus on the implicit demand response and in particular on how to send price signals to obtain flexibility.

3.2 Energy Tariff emulator

One of the elements of the HOLISDER framework is an energy tariff emulator, aimed at the retailer. The tariff emulator calculates the prices that will trigger the corresponding amount of elasticity to approach the ideal consumption curve.

The energy tariff emulator uses historical and context (e.g. weather) data to predict the consumption of the retailer's portfolio. Renewable generation is also considered. Based on elasticity profiles, the emulator

calculates real time energy prices that will result in the consumption desired by the retailer. In a first estimation, the required consumption profiles are calculated based on the wholesale price, however the objective is for the retailer to have the freedom to input the profile that suits him, as other parameters can come in play.

The tariff emulator will support three types of dynamic pricing rates:

- Time of Use (also called dual tariffs): two time-zones are defined – peak and off-peak, and the tariff for each of the zone is recalculated daily.

- Critical Peak Pricing: the electricity price can be punctually increased to obtain energy reduction when a critical peak of consumption threatens the stability of the grid.

- Real-Time Pricing: the customers are charged hourly (or half-hourly) with the price fluctuating based on the wholesale energy prices.

In this paper, the focus is on the prediction of the total energy consumption of retailer's portfolio. To simulate a retailer portfolio, the data set described in the following section is used.

3.3 London case study

Between November 2011 and February 2014, the Low Carbon London project recorded energy consumption from 5,567 London Households recruited as a representative sample of the Greater London population [9].

This openly available dataset contains half-hourly energy consumption (kW/hh), associated with a unique household identifier, the date and time and the CACI Acorn group of the household, a UK consumer classification.

In 2013, a subset of 1100 end users have been applied a Dynamic Time of Use (dTou) energy tariff where the periods for three fixed tariffs (low, normal and high) were given a day ahead to analysis the benefit of flexible tariff for the grid [10]. Most time periods were set to normal, so the data to evaluate elasticity of the consumer from this data set is limited, but it is useful for energy prediction. All the data analysis was carried out by using the Python programming language [11].

3.4 End user consumption pattern analysis

As a first step, the consumption pattern of the group of users was analysed. The objective is to identify different profiles of costumers that can be used at the time of predicting the consumption and elasticity. After an exploratory analysis to better understand the data, a series of variables were calculated to use for the classification. A PCA was carried out to sort out which variables are most relevant to model the consumption patterns. Based on these variables, a cluster analysis was carried out to determine several groups of consumers.

3.4.1 Exploratory data analysis of historic energy data

During the exploratory analysis, it was observed that some users had incomplete data. Since the users with incomplete data was 108 out of the total of 5,567. They were eliminated of the data set, leaving the data of 5,459 consumers.

Time series of a sample of those users were analysed to get a preliminary understanding of the consumption patterns and the following observations were made:

- First, there are often peaks in consumption during the 5am-10am and 17am-21am intervals. Some consumers also have a peak of consumption around noon.
- Moreover, there is a certain periodicity in the consumption patterns, which will help to predict the demand in Section 4.3.
- However, the energy consumption cycle is not stationary over the year and a correction factor will need to be applied to take this into account, in particular the winter consumption is higher than the summer consumption.

3.4.2 Principal component analysis

The first step to perform the classification is to identify the variables that are the best to explain the users' consumption. A first set of variables were calculated from the energy consumption to help identify different consumption behaviour. Those variables consist in calculations from the energy consumption on various time intervals. For each day, four intervals are considered: 0am-8am, 8am-8pm, 8pm-12pm, and 12am-12pm. For each of these intervals, calculations are carried out on the entire period, the weekday and the weekend days. The average on the entire period and on the weekend is also considered.

The average energy consumption for each of these intervals is evaluated for each user over the 2-year period of the dataset. The standard deviation from each interval compared to the average over the entire period is also calculated, as the variations in the consumption is as interesting as the value itself. This results in 24 variables to explain the consumption behaviour.

This set of variables was not able to satisfactorily explain the data. To improve this, a new set of variables was calculated to better capture the variation over each interval. The standard deviation of each interval compared to the interval average of each day was calculated. These standard deviations were then averaged and the standard deviation over the period was also calculated, created two new variables for each interval to replace the standard deviation calculated before.

The correlation between those variables was then analyzed. Weekly means and standard deviation were removed as they provided nearly the same information as the mean values split between weekend and weekdays. The PCA was carried on the remaining variables.

3.4.3 CLUSTER analysis

A clustering analysis was then carried out to identify groups of users with similar consumptions. The analysis used the variables identified by the PCA. A K-means method using Lloyd's algorithm which aims to minimize the within-cluster sum of squared criterion is used [12].

To get an insight of the number of groups to use for the cluster analysis, a hierarchical clustering algorithm was used, where the data points are grouped based on the distance between them and have been represented on a dendrogram (Figure 1). 20 is identified as a discriminating distance suggesting classifying the consumers in three groups.

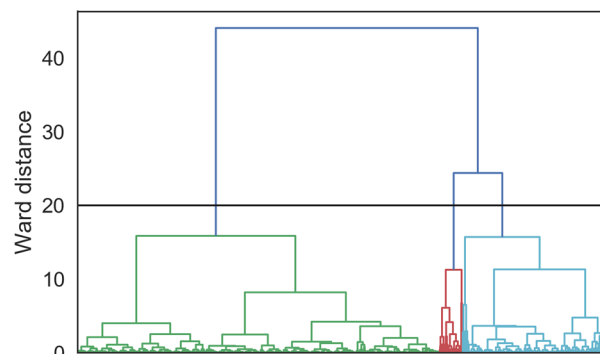


Fig. 1. Hierarchical clustering of users

Based on the dendrogram analysis the hierarchical clustering algorithm was run again with a maximum number of clusters between 2 and 5 and using five different methods and four different distance metrics to determine which one gives the best results. During the tests, an outlier consumer, whose consumption is very different from the rest, was identified and removed to improve the clustering results. The best result is shown on Figure 2 and was obtained with the complete method (at each step, the two clusters which are separated by the shortest distance are combined) and with the Manhattan distance (sum of the absolute differences of the coordinates of the points, as seen in equation (1)).

$$d_{Man}(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\| = \sum_{i=1}^n |u_i - v_i| \quad (1)$$

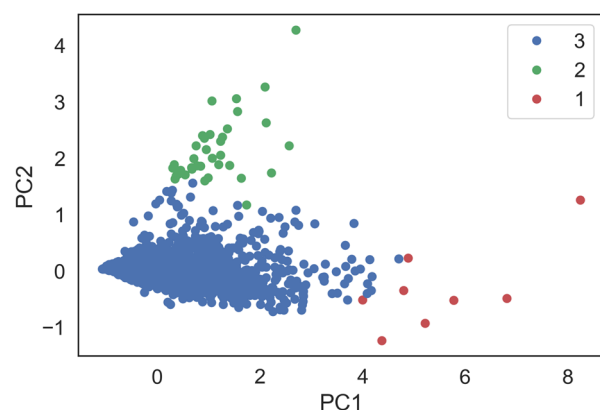


Fig. 2. Final CLUSTER with 3 different groups of consumers

3.5 Energy demand behaviour prediction

The next step is to develop a model to predict the energy consumption of the users. Because this prediction is aimed at the retailer, we are interested in the aggregated consumption. A Random Forest (RF) method is used. Random Forest can be used for classification and regression. The algorithm is based on decision trees. It creates random subsets from the data. With each subset, it creates a decision tree. Different variations of the main classifier are created. In the case of classification, the most repeated output in the individual trees is the result. In the case of regression, which is the case in this paper, the result is the mean of the outputs of each individual tree [13].

To train the RF model, two types of data are used:

- Historical energy consumption
- Variables linked to the date and time (seasons, day of the week, if it is a weekend or a weekday, hour of the day).

The model was trained using 85% of the 40,403 data entries, while the rest was reserved for testing purposes. A 10-fold cross validation, where the model is trained on 9 parts of the dataset and tested on the 10th part, repeating this on the 10 possible combinations, was used to test the quality of the model.

For the actual prediction, the train model was used with the consumption at the time to predict during the 7 preceding days.

4 Results

In this section, we describe the results of the analysis described in Section 3.

4.1 Classification of consumer profiles

4.1.1 Principle component analysis

Figure 3 shows the result of the PCA after adjusting the variable as described in Section 3.4.2.

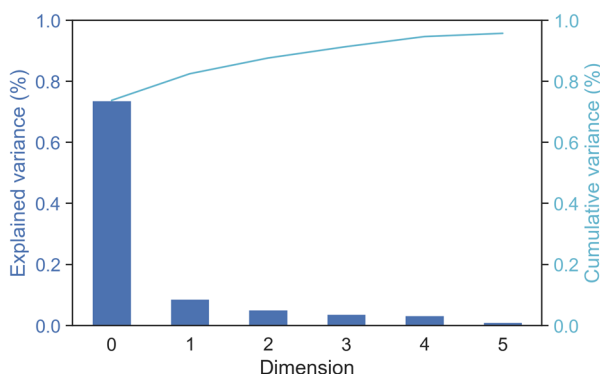


Fig. 3. PCA analysis results for the 6 first components

The model is considered sufficiently determined when the variables explain more than 95% of the variance. This is achieved with 6 variables. The two first variables will help to visualise the clusters.

4.1.2 Cluster analysis

The CLUSTER analysis identifies 3 groups of consumers. The consumers the closest to the cluster centre of each group were extracted to identify the main consumption's characteristics of each group.

Members classified in Group 1 have an energy consumption average much higher than the other groups for each period. The average over the total period is 1.15 kWh/hh for group 1 versus 0.42 kWh/hh for group 2 and 0.21 kWh/hh for group 3. Group 1 also has the highest standard deviation values.

Group 2's consumption is higher at night for weekdays. It has a higher consumption during the 8pm-0pm and much higher during the 0am-8am period than for the group 3. During the day group 2 and 3 show similar behaviours. During the weekend, the group 2 shows a higher energy consumption only for the 0am-8am and not for the 8pm-0pm period anymore.

Group 3 has, in average, a much lower standard deviation than group 2. Their energy consumption is more constant over each period. Group 3 is representative of a bigger number of users.

4.1.3 Classifier

To be able to use the results above for predicting individual energy consumption, a classifier is built to sort users between the three identified groups. A tree classifier is used as represented on Figure 4.

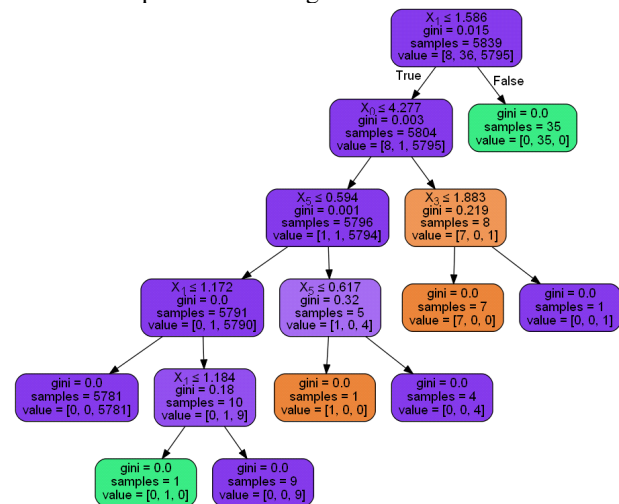


Fig. 4. Decision tree to classify the users according to their consumption

4.2 Prediction of the consumption

This section presents the results linked with energy consumption prediction. Figure 5 shows the error of the RF model for each of the data point of the 2 year-period. The coefficient of determination (R^2) of the model is 0.66.

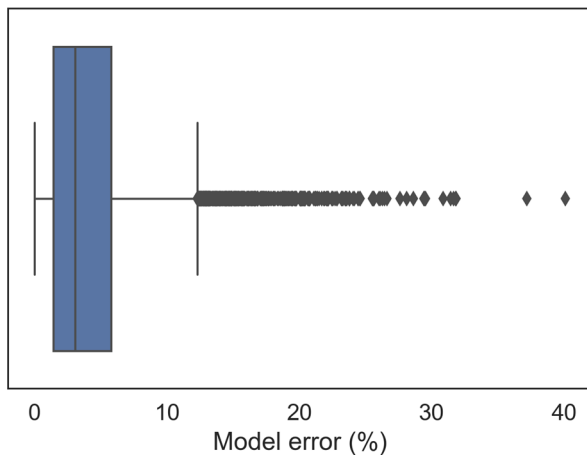


Fig. 5. Error of Random Forest model

An R^2 of 0.66 might seem low, but as it can be seen in Figure 6, this is sufficient to detect the peaks and the valleys of the aggregated energy consumption of the portfolio of end users. This information is crucial for the retailer at the moment of planning tariffs based on his desired consumption curve.

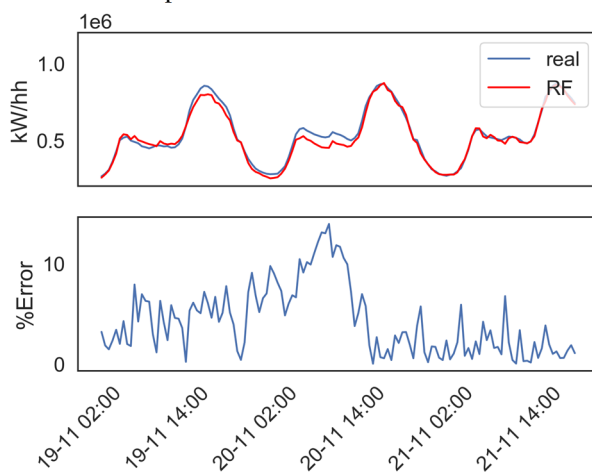


Fig 6. Prediction with RF model compared to real data and error of the model.

5 Discussion

Demand response can lead the way to change the energy grid paradigm and is necessary for a higher integration of renewable energy in the energy mix. Specific DR schemes need to be implemented to activate the DR potential of buildings. One option is to send price signal to encourage users to reduce or displace their consumption. In this paper, the concept of an energy tariff emulator to assist the Retailer in providing the right price signals is introduced. The energy tariff emulator rests on three elements: the prediction of the consumption from their client portfolio, the prediction of the elasticity of the portfolio, and the ideal consumption curve.

In this paper, a machine learning algorithm based on the Random Forest method was developed based on a 2 years dataset from the Low Carbon London project. The median error of prediction was 3% and peak and valley in

consumption are correctly predicted. Further work to improve the precision of the model will include using weather as an additional parameter of the model and testing the replicability of the method on another dataset.

A cluster analysis was also carried out to understand the consumption pattern of the user. Three groups were identified with group 3 being majority (5795 users). The next step will be to define the elasticity of the users in this clusters and possibly divide further group based on their response to price signals.

Elasticity of the users is difficult to determine with precision, because there is a lack of available data. This can be partly explained by the limitation of the current regulations, that mean price signals are often simulation (i.e.: consumers are sent artificial price signals that don't correspond to their real energy bills). This creates a bias in the data especially regarding the consumer engagement. There is a need for a large scale and long-term dataset to evaluate users' elasticity.

6 Conclusion

In this paper, a tariff emulator to activate demand response potential of the building sector was presented. In particular, a model for classifying the individual consumers and one for predicting the aggregated consumption of a portfolio of consumers were developed. The clustering model identified three groups of consumers, with one of them being applicable to the majority of users. The predictive model using a Random Forest machine learning algorithm showed satisfactory prediction of peaks and valleys with a R^2 of 0.66.

Acknowledgment

This paper is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 768614. This paper reflects only the author's views and neither the Agency nor the Commission are responsible for any use that may be made of the information contained therein.

References

1. J. Aghaei, M.-I. Alizadeh, *Renew Sust Energ Rev*, **18**, pp. 64–72 (2013)
2. EU Directive 2009/28/EC, *Directive of the European Parliament and of the Council on the Production of the use of Energy from Renewable Sources* (2009)
3. COM/2016/0767, *Proposal for a directive of the European Parliament and of the Council on the promotion of the use of energy from renewable sources (recast)* (2016)
4. P. Kohlhepp, H. Harb, H. Wolisz, S. Waczowicz, D. Müller, V. Hagenmeyer, *Renew Sust Energ Rev*, **101**, pp. 527–547 (2019)
5. X. Yan, Y. Ozturk, Z. Hu, Y. Song, *Renew Sust Energ Rev*, **96**, pp. 411–419 (2018)

6. O. Macías, S. Noyé, N. Tellado, D. Tsagkrasoulis, T. Tsitsanis, M. Bucur, J. Southernwood, S. Oxisidis, *NOVICE D2.1 Lifecycle performance of building energy upgrade technologies* (2018)
7. D. S. Kirschen, G. Strbac, P. Cumperayot, D. de P. Mendes, *IEEE Trans. Power Syst.*, **2**, pp. 612–617 (2000)
8. A. Romero, P. DE Agustín, T. Tsitsanis, *EEEIC/I&CPS Europe 2018*, pp.1-6 (2018)
9. UK Power Networks, *SmartMeter Energy Consumption Data in London Households* (2014)
Retrieved from <https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households>
10. UK Power Networks, *LCL Learning Report - AI - Residential Demand Side Response for outage management and as an alternative to network reinforcement* (2014)
11. Python Software Foundation. *Python Language Reference, version 3.6*. Available at <http://www.python.org>
12. S. Lloyd, *IEEE Trans. Inf. Theory*, **28**, pp 129-137 (1982)
13. T. K. Ho, *IEEE Trans. Pattern Anal. Mach. Intell.*, **20**, pp 832-844 (1998)