

Demand Response Algorithms for Smart-Grid Ready Residential Buildings using Machine Learning Models

Fabiano Pallonetto^{a,b,*}, Mattia De Rosa^{a,c}, Federico Milano^b, Donal P. Finn^{a,c}

^a*UCD Energy Institute, University College Dublin. Ireland*

^b*School of Electric Engineering, University College Dublin. Ireland*

^c*School of Mechanical and Materials Engineering, University College Dublin. Ireland*

Abstract

This paper assesses the performance of control algorithms for the implementation of demand response strategies in the residential sector. A typical house, representing the most common building category in Ireland, was fully instrumented and utilised as a test-bed. A calibrated building simulation model was developed and used to assess the effectiveness of demand response strategies under different time-of-use electricity tariffs in conjunction with zone thermal control. Two demand response algorithms, one based on a rule-based approach, the other based on a predictive-based (machine learning) approach, were deployed for control of an integrated heat pump and thermal storage system. The two algorithms were evaluated using a common demand response price scheme. Compared to a baseline reference scenario, the following reductions were observed: electricity end-use expenditure (20.5% rule-based and 41.8% predictive algorithm), utility generation cost (18.8% rule-based and 39% predictive algorithm), carbon emissions (20.8% rule-based and 37.9% predictive algorithm).

Keywords: Building demand response, Optimisation, Machine learning, Control algorithms, Smart grids, Energy efficiency

*Corresponding author

Email addresses: fabiano.pallonetto@ucd.ie (Fabiano Pallonetto), mattia.derosa@ucd.ie (Mattia De Rosa), federico.milano@ucd.ie (Federico Milano), donal.finn@ucd.ie (Donal P. Finn)

1. Introduction

Buildings account for 39% of the total primary energy consumption [1]. As a result, various EU legislation initiatives on energy efficiency have been passed in the recent years to reduce their carbon emissions [2]. New energy efficiency measures, such as increasing thermal envelope insulation, more efficient heating and cooling systems and the penetration of Renewable Energy Systems (RES), are contributing to the reduction of the overall energy consumption [3] and are key to a low-carbon European economy [4]. However, renewable energy penetration at a building and system level are increasing the supply/balance variability in the power system due to the intermittent nature of their generation and widely dispersed geographical locations. These challenges could cause congestion and atypical power flows which would strain the underlying electricity transmission and distribution network [5].

Proposals for new buildings envision a system, equipped with connected sensors and smart Energy Management System (EMS), which can control heating systems and appliances, responsive to smart grid signals [6]. The controllers can be embedded in a Home Area Network (HAN) and become an integrated part of the dwelling. Such features would enable buildings to adapt their electricity demand, thereby assisting in reducing the grid frequency instability [7], generally caused by the increase of variable generation [8].

In this scenario, a network of connected buildings would be able to actuate balancing strategies triggered by demand response signals. However, the realisation of these advanced smart grid features requires rapid prototyping and validation of EMS control algorithms. Fuller et al. [9] notes that until equipment and algorithms can be analysed, it is difficult for utilities and regulators to install, operate and exploit these new resources.

Building simulation software can be utilised to assess the value and the risks associated with the adoption of new technologies capable of providing electricity demand flexibility. Building simulation models can save considerable resources compared to experimental analysis and, if properly calibrated, are accurate to within 10% and 30% compared to the metered data [10]. Furthermore, building simulation software offers the opportunity to perform energy assessment across different time intervals from annual to sub-hourly assessment, without the need for hardware and communication network prototypes.

In the current literature, there are numerous examples of controllers developed to reduce energy consumption while implementing Demand Response

(DR) measures. DR is one of the Demand Side Management (DSM) measures that has been promoted since the 1970's in the UK and other countries, so as to reduce high winter peaks as well as avoiding associated grid upgrade costs [11]. More recently, there has been renewed interest in DR as a mechanism to increase the percentage of renewable energies in the system [12].

DR has been defined as "*changes in electricity use by demand-side resources from their normal consumption patterns in response to changes in the price of electricity or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardised*" [13]. A recent overview on the development of demand response programs in Europe can be found in De Rosa et al. [14]. DR measures require building automation systems which have not yet widely adopted in residential buildings [15]. Furthermore, quantifying the savings of DR measures and the flexibility provided by a single building is challenging. Several studies aiming to find effective means of shifting loads and developing EMSs with embedded optimisation algorithms are currently ongoing. A recent review on algorithms and modelling techniques was carried out by Vzquez-Canteli and Nagy [16].

For instance, Mohsenian-Rad and Leon-Garcia [17] describes a residential control algorithm capable of reducing peak energy load up to 37% and shifting loads to periods of lower electricity pricing. More recently, Parizy et al. [18] developed an appliance scheduling technique to implement demand response reducing the overall peak average ratio. The optimisation algorithm shows a reduction in energy consumption and a more efficient use of the appliances, however, the benchmark was exclusively numeric, without considering occupancy profiles or specific weather conditions. Additionally, Bahrami et al. [19] developed a distributed controller for optimal appliance scheduling. The controller was able to reduce the average peak consumption up to 17%. Despite a significant average peak ratio reduction in the building consumption, the authors do not evaluate the impact of the controller on the internal building temperature and on the thermal comfort for the occupants.

Other researchers focused on EMS hardware design which incorporates demand response capabilities. Hu and Li [20] propose a controller that can provide DR capabilities without providing a metric for the assessment of the flexibility exploited. The proposed controller, does not have an interface and it cannot be integrated to a smart grid system. On the other hand, Ren et al. [21] focused on a smart grid control infrastructure without a comprehensively assessing the impact of the energy profile on the building.

Kolokotsa et al. [22] developed an integrated indoor management system for buildings using a fuzzy controller. The authors considered the comfort constraints of the occupants by tuning and optimising the system using two demonstration buildings in Greece. The objective function of the controller aimed to reduce the total energy consumption, however, assessing the controller from a grid integration perspective was out of the scope of the study. The research demonstrated how the fuzzy controller satisfies the indoor comfort requirements, while reducing the energy use by 23%.

A relevant work in the area is the one carried out by Alimohammadisagvand et al. [23]. The research compares four rule based algorithms in a residential building modelled with a Building Energy Simulation (BES) software equipped with a ground source heat pump using different control techniques. The results show a decrease in energy cost and an overall increase in the heating system efficiency of up to 15%. Additionally, Arabzadeh et al. [24] describes the integration of a data-driven predictive demand response controller for residential buildings with heat pump and on-site energy generation reducing the heating costs of up to 12%. The authors investigate the effect of the prediction error on energy cost saving for a predictive DR control using a synthetic data set. Liu and Heiselberg [25] evaluates the performance and potential of the energy flexibility for a commercial building with weather prediction control of a convective building energy system with different metrics. The application of the predictive control strategy integrated with the two-pipe heating and cooling system provides better performance of energy flexibility compared to simple rule-based control.

The majority of advanced DR control algorithms examined in the literature to date have been developed and trained on simulated data [26]. This is because the evaluation of the effectiveness of DR control algorithms in real buildings often requires considerable periods of analysis before a consistent validation can be arrived at, and test conditions are not easy to reproduce because of the unpredictable nature of human behaviour or weather conditions. Moreover, when the evaluation involves critical infrastructure such as a power grid or advanced heating and cooling equipment, trial-and-error approaches can compromise the integrity of the test bed and the related systems. Therefore, the ability to develop control algorithms using measured data, and assess their performance within a co-simulation environment which exploits a calibrated BES model is a novelty element and provides a step forward to narrow the gap between research, development and implementation. In comparison, in order to perform a full cycle test on a physical building,

this hybrid methodology can reduce the test cycle time, reduce the hardware infrastructure and also allows for the replication of the experimental conditions [27].

This paper outlines the design, the development and the testing of an EMS which provides DR capabilities for residential buildings using a calibrated energy building model. The novelty of the algorithm resides, not only in the combination of the optimisation technique and the machine learning model used for finding an optimal strategy, but also in the use of metered data to train and test the algorithms. Furthermore, the algorithm benchmark has been carried using an open source co-simulation framework which allows decoupling of the building simulation model and the controllers.

The assessment is divided into two stages: (i) the first EMS prototype is based on a rule-based algorithm, while (ii) the second more advanced algorithm, identified as a smart algorithm, aims to reduce the electricity expenditure by optimising Thermal Energy Storage (TES) charging/discharging cycles. In the following sections, the methodology, the algorithm descriptions and the simulation results are described in more detail.

2. Methodology

The current work describes the development of a rule based and a smart algorithm, where both algorithms were subjected to a dual purpose objective function. The algorithms were developed to find the optimal strategy in a DR scenario where a price tariff signals the DR event. The algorithms were tested using the building co-simulation framework [28]. The open source co-simulation platform was coupled with the calibrated building model and the exposed Application Program Interface (API). The infrastructure represented the test harness for the two developed EMS approaches; the rule-based and the smart algorithm. The assessment includes the perspectives of the main stakeholders involved in the development of the EMS with DR capabilities.

2.1. Rule based

The rule-based control algorithm was developed and tested on the building with the objective to reduce the energy expenditure and be responsive to a DR signal exploiting the thermal storage connected to the heat pump. A heuristic response to a price signal was then embedded in the overall EMS architecture. The initial test of the rule-based algorithm has been outlined

in Pallonetto et al. [29] and is summarised in the current paper in Section 5.

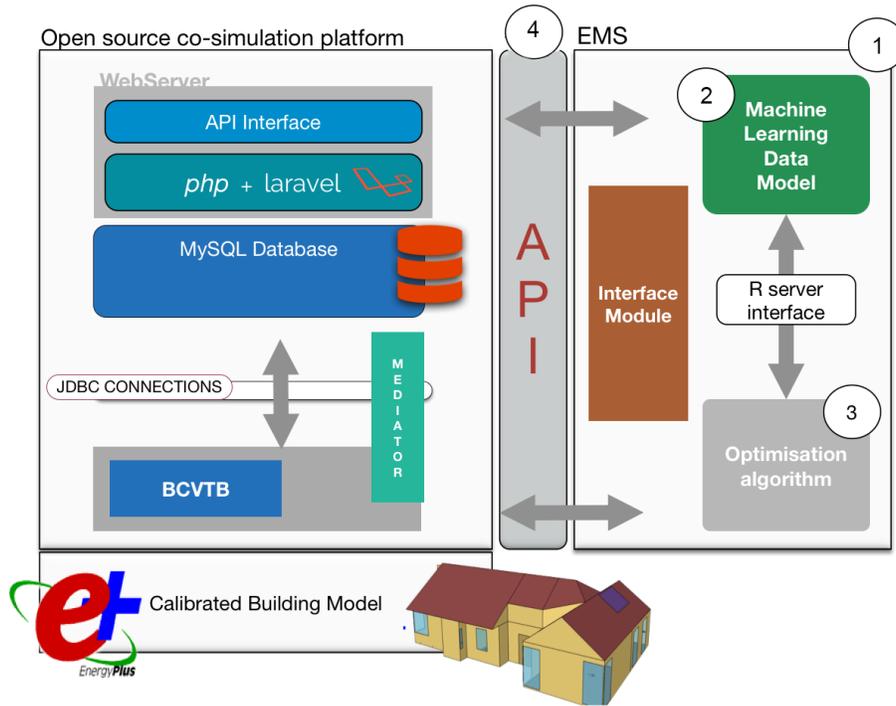


Figure 1: Overview of the software infrastructure linked with references to the development phases

2.2. Smart algorithm

The smart algorithm was developed in different stages which are referenced in Figure 1. The development process was divided into four main phases as follows:

1. EMS development (Labelled 1 in Figure 1). This phase involved the development of an object oriented code skeleton and the data structure for the search tree. The outcome of this phase is a high-level flow diagram of the algorithm.
2. Data modelling (Labelled 2 in Figure 1). Machine learning techniques were used to identify a suitable statistical model that can provide the

Table 1: Time of Use electricity tariffs (€/kWh)

	Weekdays						Weekends					
	A	B	C	D	Flat	SMP (avg)	A	B	C	D	Flat	SMP (avg)
00:00-08:00	0.12	0.11	0.1	0.09	0.135	0.046	0.12	0.11	0.1	0.09	0.135	0.044
08:00-17:00	0.14	0.135	0.13	0.125	0.135	0.065	0.14	0.135	0.13	0.125	0.135	0.062
17:00-19:00	0.2	0.26	0.32	0.38	0.135	0.097	0.14	0.135	0.13	0.125	0.135	0.088
19:00-23:00	0.14	0.135	0.13	0.125	0.135	0.071	0.14	0.135	0.13	0.125	0.135	0.067
23:00-00:00	0.12	0.11	0.1	0.09	0.135	0.053	0.12	0.11	0.1	0.09	0.135	0.053

foundation for the predictor module. The output of this phase was a prediction model used by the optimisation algorithm.

3. Optimisation algorithm (Labelled 3 in Figure 1). This phase involved the development of the optimisation algorithm within the EMS infrastructure, the data structures to support the search tree evaluation and the tree pruning techniques.
4. Model integration (Labelled 4 in Figure 1). This involved the assembly of the optimisation algorithm, the data model and the EMS skeleton and the integration with the API. This phase required several iterations of bug fixing and feature implementation. It also entailed the synchronisation of the API and the controller workflows.

Each of the development phases required one or more test cycles and the alignment of the interface in order to facilitate the communication of the modules. At the end of the development phases, the smart algorithm was tested against the baseline and the rule-based algorithm using January 2014 data during the peak heat demand for the year. The assessment was based on several standard metrics as described in Section 7.

2.3. Electricity price

Time of use electricity tariffs are increasingly being utilised in different parts of the world and to a lesser extent in European markets [30]. In 2010, the Irish Commission for Energy Regulation initiated a residential smart meter trial, with associated tariffs (A, B, C, D, Flatrate) as shown in Table 1 [31]. The price scheme is structured on the basis of; *peak*, *off-peak* and *night* tariffs, to reflect the average Irish System Marginal Price (SMP) and consequently of the overall electricity demand.

3. Test bed dwelling

The selected building is a detached bungalow-type house, which represents 40% of the Irish building stock and is the most common single building category [32]. It was constructed in 1973 with increased thickness of insulation materials in its opaque elements compared to the contemporary standards.



Figure 2: Test bed house and EnergyPlus model

The dwelling, a single storey building, was constructed using a two leaf concrete wall with core insulation. Therefore, the inner wall displays significant passive thermal energy storage capacity. The floor area is $205m^2$ and the overall window to wall ratio is 15%, with a 22% and 10% ratio on the south and north facades, respectively. The house has 12 rooms and an unused attic space at roof level. Although its architectural characteristics are those of a typical rural Irish bungalow dwelling of the 1970s, its fabric specifications are very close to the current Irish building regulation values [33] as outlined in Table 2. The building thermal performance is also aligned with the average residential dwelling outlined in the 2020 scenario research published in the Residential Energy Roadmap for Ireland [3].

As part of a retrofit measure, a photovoltaic panel array consisting of 30 panels, of a total nominal power of 6 kWp was installed and solar thermal collectors were installed for the supply of Domestic Hot Water (DHW).

The space heating system is a 12 kW (thermal output) Ground Source Heat Pump (GSHP). The heat pump uses water as a heat source. Mea-

Table 2: U-Value of different building elements

Building Element	U-Value Test building (W/m^2K)	U-Value Irish Building Regulations[33] (W/m^2K)
Walls	0.25	0.21
Roof	0.25	0.21
Windows	1.7	1.6
Floor	0.21	0.21

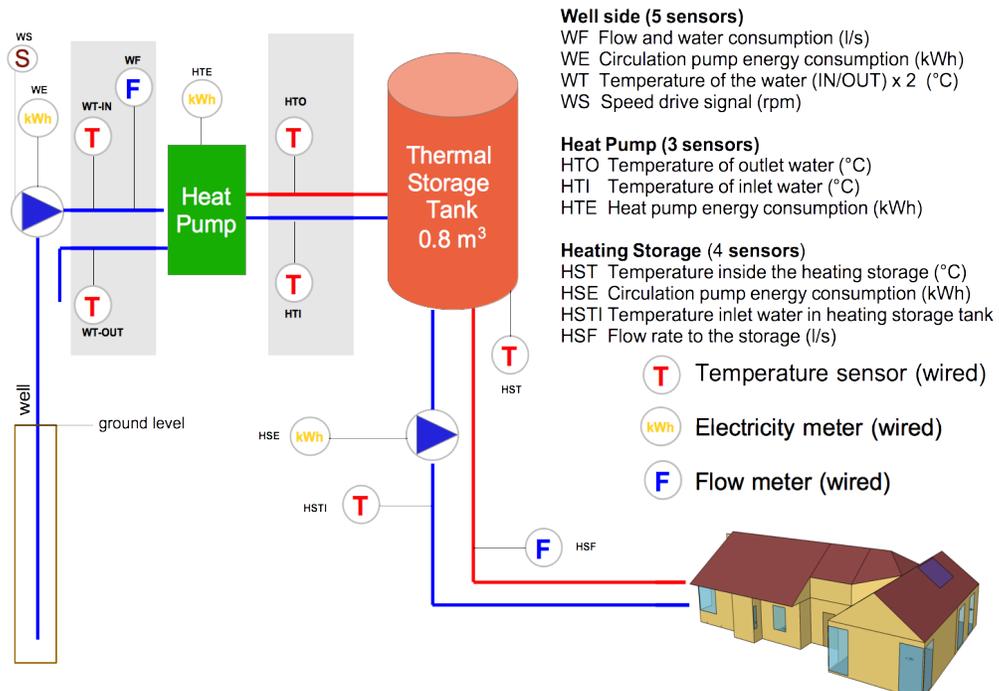


Figure 3: Test bed building, heat system design and sensor metering

measurements indicated that the water temperature of the source varied over the heating season between 8/6/8°C for Oct/Feb/May, respectively. For the provision of thermal energy storage, the heat pump was equipped with a hot water storage tank of 0.8 m³.

The system illustrated in Figure 3, was metered and the EnergyPlus model was calibrated using ASHRAE standards using sensor data from the year 2014. Although there is a lack of standards for sub-hourly calibrations, the thermal envelope, the Heat Pump (HP) and the PhotoVoltaics (PV) were calibrated to a 15-minute resolution, by adopting hourly ASHRAE standards to make the model suitable for the assessment of DR strategies.

4. EMS Design and Implementation

In the following section, the design and implementation of the EMS is described from the point of view of the software architecture, simulation settings and control flow. The core design and implementation described applies to both the rule-based and the smart algorithm.

4.1. Software Architecture

The developed EMS was connected to the building simulation model using the open source software infrastructure described in [28]. The system remotely controls a building simulation through a cloud infrastructure, called SimApi. SimApi is developed using a Model View Control (MVC) framework (Laravel 4.2 [34]) on top of the PHP [35] programming language stack. At each simulation timestep, the routing controller stores all the API requests from the EMS in a relational database [36], while it retrieves sensor readings from the model. The communication between the database and the building simulation model is realised through the Building Controls Virtual Test Bed (BCVTB) [37]. BCVTB and the API are connected to the database via Java Database Connectivity (JDBC) technology [38].

The developed EMS exchanges a JavaScript Object Notation (JSON) data structure, which contains building thermostatic settings and actions for controlling the heating system, with the API [39]. For each simulation of N time-steps, the EMS sends $N - 1$ actions that represent the control flow of the system, and receives $N + 1$ JSON data structures with the sensor readings. A software package developed in Java acts as the simulation manager and in this work, is referred to as a Mediator.

Table 3: Thermostatic set points (°C)

	Weekdays	Weekend
00:00 - 06:30	17 °C	20 °C
06:30 - 09:00	19 °C	20 °C
09:00 - 16:00	16 °C	20 °C
16:00 - 00:00	18 °C	20 °C

When the EMS invokes the API endpoint, the simulation begins and the Mediator transfers the appropriate EnergyPlus files, the weather data, and the settings to the simulation server. A BCVTB instance is then triggered, which initialises the EnergyPlus environment. In the current work, the EMSs developed, both rule-based and predictive, are tested on a calibrated building simulation model of the residential test-bed house.

The thermal comfort settings used by the building owner are given in Table 3 and they represent hard constraints in the algorithm. The settings were mutually agreed upon with the building owner. The occupants are not in the house during weekdays between 0900-1600 hrs. Consequently, the thermal comfort score during the day is expected to align with the lower PMV boundary. However, the temperature sensor is located in the corridor which is one of the colder zones within the building while the thermal comfort score is calculated within the living area, which is affected by internal heat gains of the kitchen and the wood stove. The weather data used in the current work is retrieved from a weather station located at Dublin Airport.

4.2. Algorithms objective function

This research focuses on the implementation and testing of two EMSs, developed using a real-time smart meter data and weather data. The purpose of the objective function is to minimise the electricity expenditure and is described as:

$$\min_{(T_{tk}, C_{set}, PV_e)} (C^T P(T_{tk}, C_{set}, PV_e)) \quad (1)$$

$$T_{set}(t) - T_{bd} \leq T_{int}(t) \leq T_{set}(t) + T_{bd} \quad t = 1..N \quad (2)$$

$$T_{min} \leq T_{tk} \leq T_{max} \quad t = 1..N \quad (3)$$

where N is the number of time-steps, C is an array with the price of electricity at each time-step (€/kWh) and P represents the electricity consumption.

$P(T_{tk}, C_{set}, PV_e)$ is the energy consumption in kWh determined by the building simulation model at each time-step t . The cost optimisation depends on the following three variables: (i) the temperature of the tank (T_{tk}) which is maintained within the range T_{min} and T_{max} , (ii) the thermal energy supplied to the zones by means of the circulation pump which can be either enabled or disabled (C_{set}) and, (iii) the energy produced by the RES, labelled PV_e .

At each time-step, the zone temperature (T_{int}) is constrained at the appropriate set point and an associated 2°C bandwidth (i.e., +/- 1°C around the set point), as shown in Equation 2; while Equation 3 constrains the storage tank temperature between its MIN and MAX temperature settings. The function and the constraints reported in Equations 1 to 3 have the objective of minimising the electricity cost and energy consumption of the building, by controlling the temperature set point of the storage tank and maintaining the internal thermal comfort. The detailed description of the strategy of the smart algorithm is reported in Section 6.4.2, while the rule-based algorithm achieves a sub-optimal result by charging the thermal storage before a price signal and reducing the electricity consumption during the peak hours as described in Section 5.

4.3. Simulation settings

The assessment of the two EMS systems was performed during the month of January 2014, a period during which peaks of Heating Ventilation and Air Conditioning system (HVAC) system usage occur. The total simulation time is 30 days with a time-step resolution of 15 minutes. Three controllers, the baseline, the rule-based and the smart algorithm were compared.

The baseline simulation does not use any controller to charge and discharge the storage tank. Instead the heating system is controlled by the thermostatic set points detailed in Table 3. For the baseline case to meet the dwelling comfort constraints, the heat pump is switched on whenever needed, even during peak times.

4.4. Description of the control flow

The simulation period is one month, equivalent to 2881 time-steps at 15 minute resolution. At each time-step, the EMS retrieves the data from the co-simulation infrastructure via the API. The software infrastructure is synchronised to the BES model simulation through the database. If the sensor information is updated, then, based on the control algorithm, the system controls the heating system by the TES thermostatic setting. The

command sent to the SimApi server embeds the setting point of the thermal storage and the status of the circulation pump.

After each iteration, the EMS changes the status of the existing instance. When the update triggers the API controller, the instance status is updated in the persistent layer and the mediator waits the completion of the building simulation step via BCVTB, which coordinates the EnergyPlus instance. During the simulation, all the sensor results are sent and stored in the SimApi database by BCVTB. Thus, the EMS can retrieve the previous readings during the simulation and it controls the simulation by the API calls.

In the overall control flow mechanism, a scheduler component collects periodic reports from a monitor and a price predictor component and it analyses the data and decides an optimal choice for energy consumption scheduling. The controller manages all household appliances submitting the specified power levels over either wired or wireless HAN.

5. EMS Rule-based

In the rule-based EMS, the heating system is controlled according to the flow chart shown in Figure 4 which has been adapted for a Hypertext Transfer Protocol (HTTP) communication scenario. The objective function of the system, as per Equations 1 to 3, is to minimise energy consumption, maintaining the inside temperature within a comfort set point temperature illustrated in Table 3. The four rules (as per Figure 4) are designed to adapt to the specific tariff peak hours and occupant lifestyle, e.g., when they are at home or at work. The strategy to achieve this goal is to shift the energy consumption of the heating system two hours before the peak time, by charging the thermal storage.

Irrespective of the status of the circulation pump, the storage tank minimum and maximum temperatures are set at $40^{\circ}C$ and $55^{\circ}C$, respectively. The maximum temperature outlet of the heat pump is $60^{\circ}C$, therefore taking into account the thermal losses from the storage tank, the effective maximum temperature set point achievable is slightly less than $55^{\circ}C$. The minimum temperature set point is the result of a parametric analysis performed on the building model and validated on the physical building as described in [40]. These settings allow the system to also meet the energy demand during the winter period.

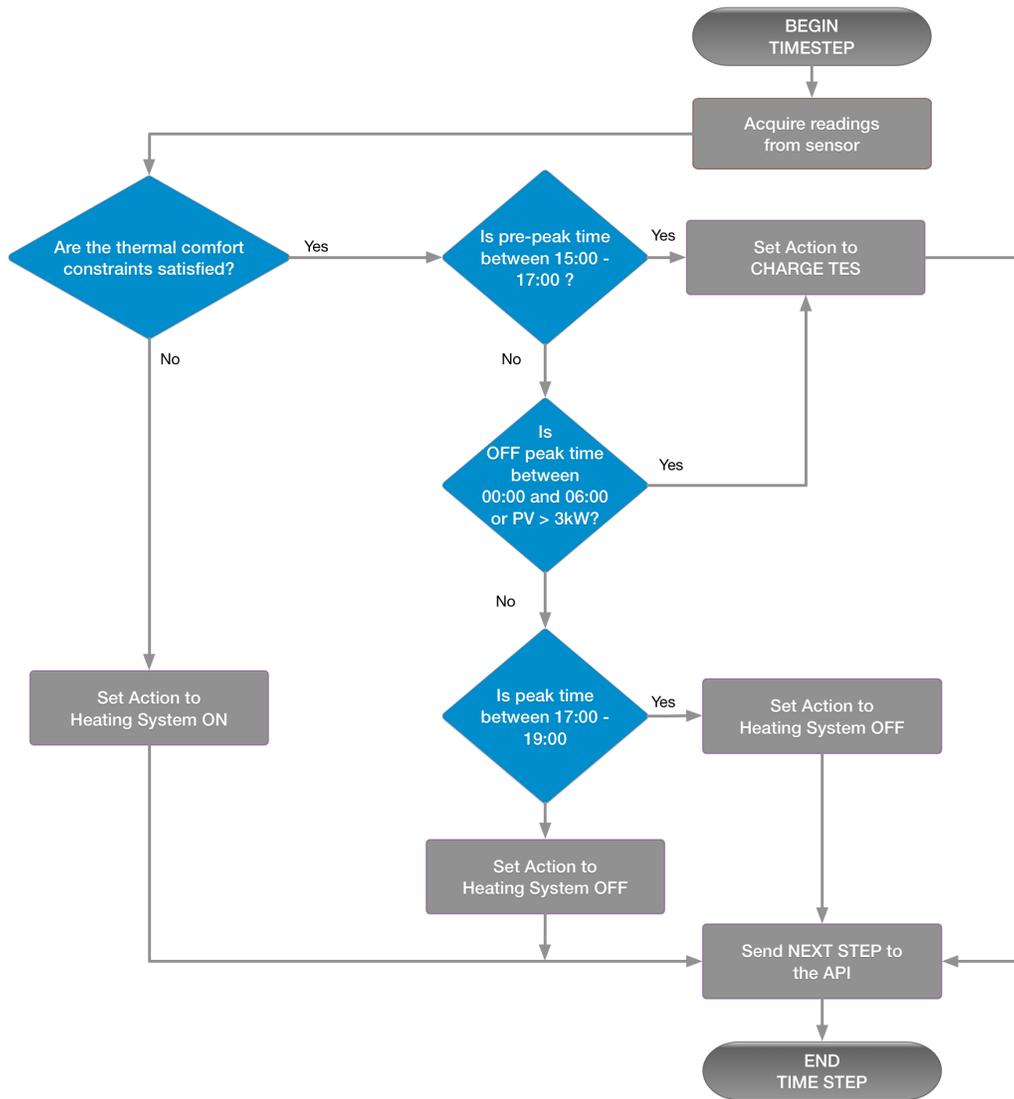


Figure 4: Control flow rule-based algorithm

5.1. Algorithm description

The rule-based algorithm creates and uses a single instance of a HTTP client class during its execution for all communication to the API. In the code repository, a utility class that outputs all the initial settings and the static variables such as temperature, server name and others was developed.

The program has two conditions based on the sensor reading at each timestep. The first condition checks whether the data point for the timestep is produced or not. The timestep counter is incremented only if a new data point is generated, whereas the second condition checks if the control needs to disable the heating system. This is implemented by controlling the status of the circulation pump and lowering the storage tank set point temperature at its minimum temperature.

The second condition is verified only between 0900 hrs and 1500 hrs, which represents the period when there is no one at home. Between 1500 hrs and 1700 hrs, the controller enables the heating system, keeping the circulation pump switched off to charge the thermal storage. The control instructions and sensor readings are stored in the SimApi database during the whole simulation.

From the perspective of control flow, the algorithm embedded in the EMS is equivalent to the rule-based which was tested in Pallonetto et al. [29]. However, instead that controlling the BES directly, the EMS sends the commands to the API.

6. EMS - Smart algorithm

This section describes the design and implementation of the smart algorithm for the described residential building. The algorithm is defined by four main modules developed as classes. Each class exposes a set of methods which are invoked by a special class called the main class. The four modules are described in the following sections. The predictors have been developed as machine learning algorithms using a black box modelling approach. As described by Alpaydin [41], the advantages of a machine learning algorithm is the possibility to capture the behaviour of the building by exclusively using a data driven approach. A measurement dataset consisting of internal temperatures, weather data and equipment state such as PV output or storage tank temperature, can be analysed to develop an approximate model of the inside temperature of the building or of the total electricity consumption. The main advantage of such a black box model is the possibility to automatically apply

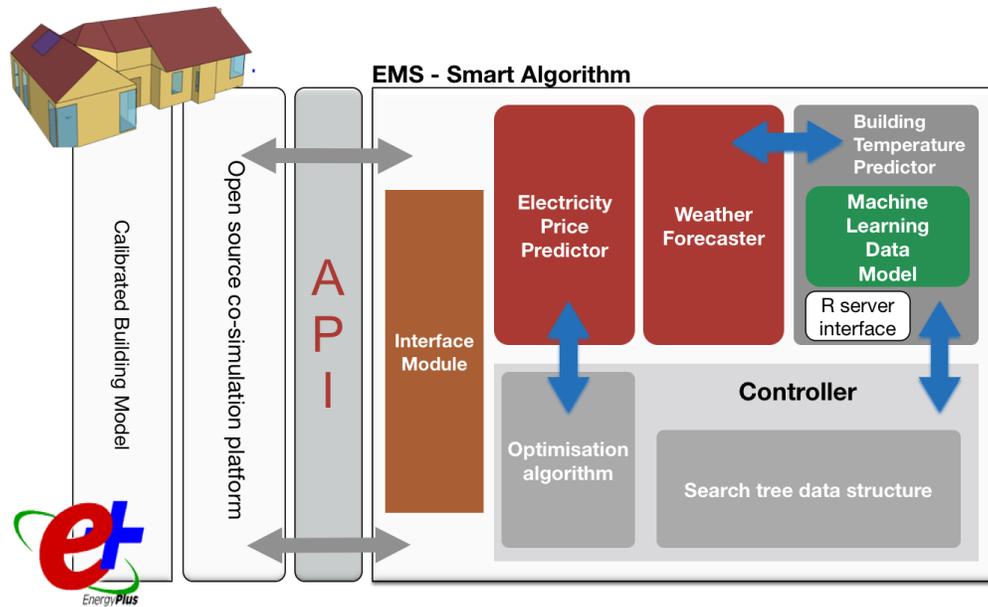


Figure 5: Smart algorithm components layout

the model on different buildings and so replicate the methodology and the algorithms. The interactions of these four modules are illustrated in Figure 5 and each component is described in detail the next sections.

6.1. Weather forecaster

This module forecasts the weather conditions for subsequent hours. The module was designed to retrieve forecast data from an API which provides a forecast of the outside temperature and the solar radiation [42]. In the assessment of the algorithm described in Section 7, however, stored weather data forecast for the month of January 2014 was used in order to reduce network overload.

6.2. Electricity price predictor

This module forecasts the electricity price of a future point in time. In the current research, the electricity price is dependent only on the time of the day and day of the week. Consequently, in the implementation tested, given these two parameters, the predictor outputs the correspondent Time

of Use (TOU) tariff for that time and day. The current implementation does not limit the time horizon window, however, the algorithm uses the module with a maximum time horizon of two hours. The resolution of the prediction is 15 minutes.

6.3. Building temperature predictor

The inside temperature dataset for a 12 month period (2014) was used to develop the building temperature predictor. The dataset has 35040 data points, equivalent to a measurement every 15 minutes. The average inside temperature in the building is 19 °C with a standard deviation of 1.6 °C. The maximum recorded temperature was 27.5 °C (July). A minimum temperature of 14.8 °C was recorded in November, when the building was unoccupied for prolonged period and all systems were non-operational.

The objective of the building temperature predictor module is to provide a 15 minute short term forecast of the inside temperature of the building. Given a set of selected features, the output is the internal temperature variation for the next timestep. The predictor developed was tuned to forecast with a 15 minute time resolution to match the EnergyPlus simulation.

The predictor was developed in three phases: data collection and analysis, feature selection, model description and training. The three phases are described in the next three sub-sections.

6.3.1. Data Collection and Analysis

The data repository used to develop the building temperature predictor was collected from the sensors in the HAN of the test bed building and a weather station located at Dublin Airport. The time span for the collected data was the whole heating season of 2013/2014, from 1st of October to the 30th of April.

The objective of the algorithm was to control the heating system and the associated TES, maintaining the thermostatic setpoints and reducing the energy expenditure and consumption. The designed building temperature predictor has a number of proxy variables as input and an inside delta temperature as output. Consequently, the first step was the selection of the proxy variables which, after a second statistical selection process, would become the features of the model. Using the domain knowledge of the system and the available weather and sensor data, a dataset with the most relevant variables was combined as follows:

1. *Outside temperature* (°C) (**outT**). The indoor temperature is considered correlated to the outdoor temperature variations. Source: air temperature sensor installed outside the house.
2. *Wind speed* (m/s)(**ws**). In the presence of high levels of infiltration the wind speed can affect the inside temperature. Source: weather conversion software using Dublin Airport weather station data.
3. *Inside temperature* (°C). (**inT**) This variable has been selected because it is directly correlated to the target of the model. Source: air temperature sensor installed in the corridor, zone Z3:Corridor.
4. *PV production* (kWh) (**PV**). This variable can be a proxy for internal heat gains. Source: electricity meter beside the inverter in the utility room.
5. *Storage tank temperature* (°C) (**tesT**). The temperature of the TES modulates the effectiveness of the heat transfer to the fluid circulating in the heat emitters. Source: water temperature sensors inside the TES.
6. *Heat pump electricity consumption* (kWh) (**hpC**). The heat pump electricity consumption can be a proxy to the TES temperature. Source: electricity meter.
7. *Circulation pump electricity consumption* (kWh) (**cpC**). This can be correlated to the inside temperature. Source: electricity meter.

The data was normalised (feature scaling [43]) and prepared for the feature selection according to best practice [44]. The normalisation process used the min-max algorithm which, according to Al Shalabi et al. [45], displays the highest accuracy for machine learning and decision tree algorithms.

After a data cleaning process and deletion of outliers and inconsistent sensor readings, the working data set size was composed of 19,643 data points at 15 minute time resolution. The data cleaning process affected less than the 5% of data and it is not relevant for the accuracy of the model.

The dataset was divided into two subsets based on the circulation pump operation schedule. Splitting the dataset facilitated the development of two different models, one for each dataset. The dataset *HeatOFFDB* was used to develop a model for capturing the temperature decay of the building when the heating system was off. The second model was based on *HeatONDB* subset, and approximated the inside temperature variations while the heating system was on.

6.3.2. Feature selection

The feature selection process was based on the calculated linear correlation between input variables and target variable for the two different datasets (*HeatOFFDB* and *HeatONDB*). The target variable calculated was the inside temperature difference at 15-minute resolution called DT and expressed in degrees Celsius ($^{\circ}\text{C}$).

The candidate features were the aforementioned seven variables of the dataset. The selection was performed using a threshold and a correlation function between the single candidate feature and the target variable. The correlation result between feature and target variable was called cR .

The correlation function used was the absolute value of the Pearson correlation linear coefficient, which results span from -1 to +1 where -1 indicates a negative correlation, +1 a positive correlation and 0 is equal to no correlation [46]. In the current research, the following guide was used to score the correlation among the target variable DT and the single feature by using the absolute correlation value cR :

- 0.00 - 0.19 *very weak*
- 0.20 - 0.39 *weak*
- 0.40 - 0.59 *moderate*
- 0.60 - 0.79 *strong*
- 0.80 - 1.00 *very strong*

The selected threshold value to consider a variable relevant for the predictive model is 0.5, as suggested by Kapetanakis et al. [47].

Figure 6 illustrates the results of the feature selection phase for each of the dataset. For the subset *HeatONDB*, the selected features were the outside temperature, the inside temperature, storage tank temperature and the consumption of the circulation pump that is correlated with the flow rate of the water in the heating system. For the subset *HeatOFFDB*, the features selected for the model fitting phase were the outside temperature, the inside temperature, and PV production.

It is noted that the wind speed correlation with the inside temperature of the building was not considered sufficient and therefore ignored in both models. Moreover, the PV system is installed outside of the house and electricity output is directly correlated to the irradiation at ground level. Consequently,

Data set	Outside Temperature	Wind speed	Inside temperature	Storage tank temperature	PV production	GSHP consumption	Circulation pump consumption
HeatONDB (Target DT)	Strong	Very Weak	Strong	Very Strong	Very Weak	Weak	Moderate
HeatOFFDB (Target DT)	Strong	Very Weak	Strong	X	Moderate	X	X

Very Strong

Strong

Moderate

Weak

Very Weak

X

 Not relevant

Figure 6: Results of the feature selection phase

the solar gains of the site were estimated using the PV electricity production as a proxy variable. As illustrated in Figure 6, the correlation with the subset *HeatOFFDB* was found higher than with the subset *HeatONDB* because of the way the heating system was configured. In fact, more than 70% of the data points in that subset do not have a PV production output because, in the test bed model, the GSHP operates mostly in the evening and night hours. Therefore the PV feature was ignored for the model fitting for the *HeatONDB*, while it was included in the *HeatOFFDB*.

As expected, the GSHP consumption is not strongly correlated to the target variable. The reason for the non-correlation is because the TES functions as a thermal buffer and decouples the heating demand represented from the energy supply, represented by the GSHP.

6.3.3. Model description and training

During the feature selection phase, a linear regression model was used to assess a linear correlation among features. In this phase, a linear regression model and a more sophisticated model were compared. Three indexes were used to evaluate the model accuracy and appear as follows:

- The Pearson correlation coefficient is a statistical measure that illustrates the degree of linear dependence of actual and predicted outputs and it was described in the previous section.

- Mean Absolute Error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The mean absolute error is the sum over all the instances absolute error divided by the number of instances in the test set and can be defined as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y - \hat{y}_i| \quad (4)$$

where n is the size of the dataset, y is the measured data and \hat{y} is the predicted value output of the model.

- Root Mean Square Error (RMSE) is a quadratic scoring index that measures the average magnitude of the error. It represents the sample standard deviation of the differences between predicted values and observed values, and is defined as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y - \hat{y}_i)^2} \quad (5)$$

where n is the size of the dataset, y is the measured data and \hat{y} is the predicted value output of the model.

Two linear regression models were calculated from the datasets. The objective of the model was to predict the temperature differential for the next timestep (15 minute forecast) using the selected features according to Table 6. The linear regression model relative to *HeatONDB* was calculated as follows:

$$DT = 0.0092 * outT + 0.0443 * tesT + 22.34 * cpC - 0.0292 * inT - 2.3705 \quad (6)$$

while the linear model for the *HeatOFFDB* was calculated as follows:

$$DT = 0.0081 * outT + 0.030 * inT + 0.4032 * PV - 2.59 \quad (7)$$

where DT is the internal temperature difference, $outT$ is the outside temperature, $intT$ is the inside temperature, $tesT$ is the TES temperature and cpC is the electricity consumption of the circulation pump.

The results of the linear regression fitting based on these indexes are reported in Table 4. The Pearson correlation score for the linear regression

Table 4: Linear regression fitting results

	HeatONDB	HeatOFFDB
Pearson Correlation	0.339	0.6831
MAE	0.1196	0.1781
RMSE	0.2923	0.2484

was found to be *weak* for *HeatONDB*. It was therefore necessary to find a more accurate model for the predictor.

It is noted that for both datasets, the correlation between *DT* and the selected features is not linear. Therefore, in order to increase the accuracy of the model, it was necessary to select an algorithm capable of capturing the non-linear correlation. For this particular problem, a tree-based technique such as M5P algorithm resulted more accurate in comparison to the linear model [48]. This technique, where the P stands for "prime", is a rational reconstruction of Quinlan M5 with some algorithm optimisation. M5P is a reconstruction of the Quinlan M5 algorithm for inducing trees of regression models and it combines a conventional decision tree with the possibility of linear regression functions at the nodes. The MP5 model creates a decision tree and uses a linear model at each node to make a prediction, rather than using an average value. In this case, the leaves of the model tree produced by M5P algorithm have linear functions [49].

In the first step of the model creation, a decision-tree induction algorithm is used in order to construct a tree, but instead of maximising the information gain at each inner node, a splitting criterion was used that minimises the intra-subset variation in the class values down in each branch. The splitting procedure in M5P stops if the class values of all instances that reach a node have a minimum variance, or if only a few instances remain.

In the second step, the tree was pruned back from each leaf. When pruning an inner node, a regression plane substitutes the leaf element. The objective of the third step was to avoid sharp discontinuities between subtrees. For this purpose, a smoothing procedure was applied that combined the leaf model prediction with each node along the path back to the root, smoothing it at each of these nodes by combining it with the value predicted by the linear model for that node [50].

Thus, M5P was able to generate predictive models which are both compact and computationally stable. With regards to the M5P algorithm, a pre-compiled package that contains this tree-based algorithm has been used

Table 5: MP5 tree fitting results

	HeatONDB	HeatOFFDB
Pearson Correlation	0.9235	0.7067
MAE	0.0382	0.0724
RMSE	0.122	0.144

in the implementation. Based on the feature selection, the M5P algorithm generates a modular tree which is composed of 109 linear models. As illustrated in Table 5, this technique has significantly improved the accuracy of the model when compared to the linear regression model, minimising the errors for the two datasets.

The outcome was two independent models: a model for the first dataset *MP5OFF* based on *HeatOFFDB*, and *MP5ON* based on *HeatONDB*. The prediction of each model is a temperature variation inside the building for the following timestep given the current status.

From an implementation perspective, the building temperature predictor was modelled using the R language and was interfaced to the Java implementation through JRI [51]. During the initialization, the R engine loads two data files that contain the two models, *MP5ON* and *MP5OFF*. The JRI interface runs on the server when the EMS is launched and it creates a two-way communication channel between the Java environment and the R instance.

The described models were used in the predictor class which implements an interface with a single method. The method decides which model to call based on the input parameters. It is noted that the predictor class was implemented to be instantiated as a singleton to avoid heap space issues.

6.4. Controller

The controller embeds the control logic of the predictive algorithms (*MP5ON* and *MP5OFF*). It uses three helper classes: the weather forecaster, the electricity price predictor and the building temperature predictor. In the next section, the predictive algorithm and its data structures are described. The control flow of the smart algorithm is described in Figure 7. After acquiring the reading from the sensors, it creates an instance of the class state with the current status of the system.

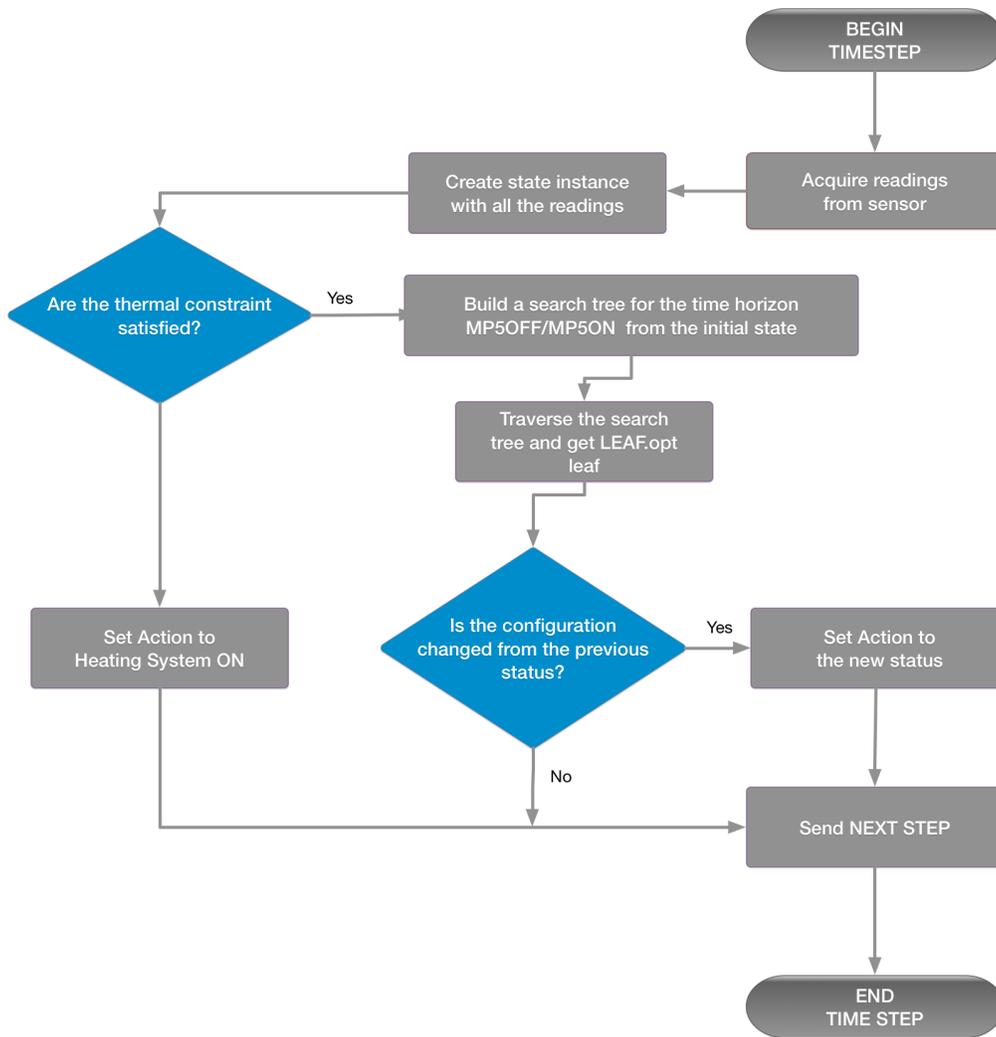


Figure 7: Control flow of the smart algorithm

6.4.1. Algorithm flow description

The controller checks if the temperature is within the thermostatic set point temperature range. If not, it will skip the prediction and enable the heating system. In the other case, the controller builds the search tree and performs a status evaluation. If a generated state cannot guarantee the minimum comfort temperature, then it will be pruned from the searching tree and will not be used during the tree traversal. In that case, a tree will not be built because the only valid child will be a child state with the GSHP and circulation pump active. During the evaluation of all the possible states, the algorithm reaches the end child states that correspond to the time horizon window.

For each node, the controller sums up the energy consumption of each state and the associated energy cost so the parent state will retrieve the least energy child state consumption and update the instance data structure. Furthermore, in order to reduce the size of the search tree, an heuristic function for the selection of the storage tank temperature set point (charging or discharging) has been constructed based on the electricity price for the forecast horizon. If a price peak signal is detected within the forecast horizon of two hours, then the storage tank temperature is set to the charging mode, so it will be ready to be exploited at a later stage of the simulation. In an instance where an off-peak price is detected within the forecast horizon, then the storage tank temperature is set to discharging mode so the system can start exploiting the stored thermal energy.

At the end, the algorithm returns the results of the cascade tree of states with updated energy consumption details. Thus, a root child state can provide an the optimal strategy to minimise the energy expenditure and consumption for the next timestep.

6.4.2. Optimal strategy search

A tree is instantiated to build the search space data structure in memory where each node of the tree invokes the helper classes and store the electricity expenditure (**Cost**) and consumption (**Elect**). The helper class invokes also the *MP5ON* and *MP5OFF* to predict the internal temperature for the time horizon. In order limit the number of nodes in the search tree, if a node violates the thermal comfort constraint, it is directly pruned and not added to the data structure. As illustrated in Figure 8, the maximum depth of the tree is equal to the time horizon of the controller prediction, which is set to eight timesteps, equivalent to two hours. For each node, the helpers provide

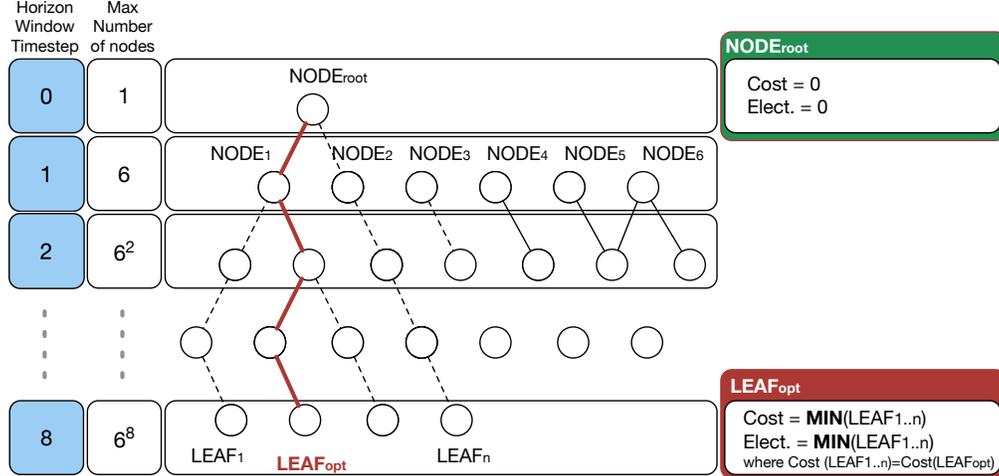


Figure 8: Optimal strategy search on solution tree

forecasts with 15 minute time horizons (1 timestep) to each child of the node, which corresponds to a control configuration.

The control configuration is based on three control variables: circulation pump (active or inactive), TES (charging or discharging) and GSHP (active or inactive). These three variables can output eight possible scenarios. However, the two scenarios wherein the TES is charging but the GSHP is inactive, are pruned.

After the construction of the tree, the child state class calls a method of the model class that in constant time finds the leaf with minimum energy expenditure and, if more than one leaf has the same energy expenditure, the algorithm chooses the leaf which minimises energy consumption. As shown in Figure 8, the selected leaf is called $LEAF_{OPT}$.

At this stage, the control configuration for the next timestep to reach the $LEAF_{OPT}$ status is stored in the $NODE_{root}$ child node that is the parent of $LEAF_{OPT}$. The configuration is retrieved by the controller and sent to the BES via the API. The instantiation and tree search are repeated for each timestep of the simulation to produce an optimal control strategy.

6.4.3. Selection of the forecast time horizon

The two hour prediction horizon is calculated based on computational time required to explore all the possibilities within the prediction horizon, which is time limited to 15 minutes, equivalent to the length of the time step. If a single time step computation longer than 15 minutes was utilised, the methodology would not have been implementable in the actual building, as the size of the possible scenarios would grow exponentially. Each computation has eight possible solutions, therefore, considering a time step of 15 minutes, for a two hour horizon, it needs to calculate 8^8 possibilities (reduced to 6^8 with a pruning technique), resulting in $1.678 \cdot 10^7$ combinations. Therefore, using the developed optimal algorithm, multiple runs on a testing machine, showed that the maximum horizon reachable within the 15 minutes time constraint was two hours, which was equivalent to eight time steps.

7. Simulation Results

Although, the overall objective of the developed optimisation system is to minimise the energy expenditure and as a secondary goal, the energy consumption, other metrics were considered so as to include the additional stakeholder perspectives. This resulted in the following metric list:

- Consumer electricity profile (kWh)
- Consumer electricity consumption (kWh)
- Consumer electricity cost (€)
- Utility electricity cost (€)
- Environmental impact (kg of CO_2)

The consumer electricity profile was evaluated based on five selected representative days of the heating season from the month of January 2014. This period was selected because the weather data showed the presence five representative days for the assessment of the electricity profile using a percentile rank approach.

Figure 9 shows the hourly based cumulative simulated electricity consumption for 2014 versus the metered data. Additionally, on the right vertical axis is reported the average Mean Bias Error (MBE) index for the calibration for each hour. For the remaining metrics, the simulation was performed

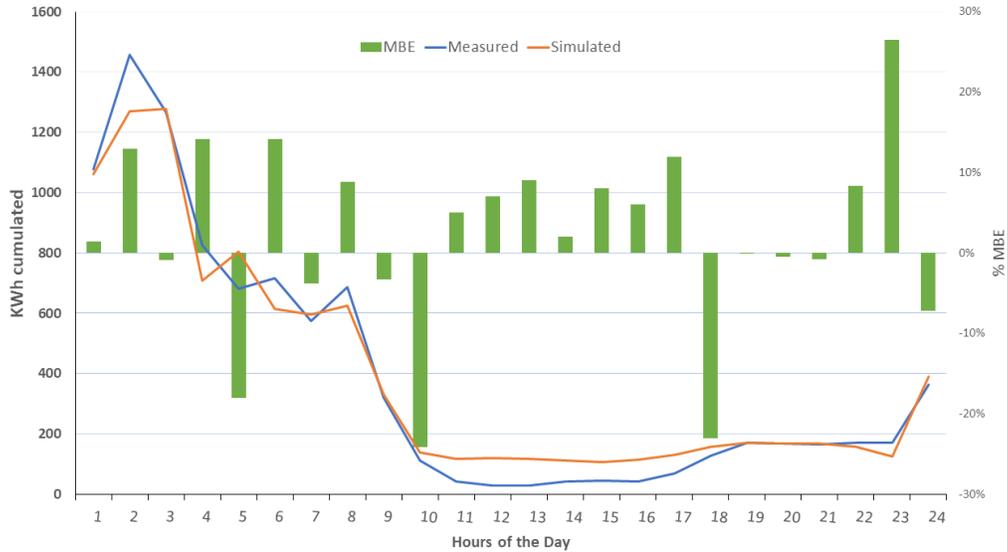


Figure 9: Cumulative annual electricity consumption on an hourly basis: simulated, metered and MBE (2014)

for the whole month of January 2014, noting that January was the second coldest month of the heating season 2013/2014 after December 2013. Each of these metrics are considered in the following sections and are assessed against the rule based algorithm and baseline scenario.

7.1. Consumer electricity profile

Figure 10 shows the aggregated energy consumption for the baseline and for the two controllers for the month of January 2014. A 15 minute time resolution is used, where each data point is determined by summing each time associated electricity consumption value (kWh) for the entire month. The PV contribution is determined in a similar way, but is subtracted from the consumed electricity to allow a net value to be determined. However, the limited power generation resulting from January solar irradiation levels reduces the PV contribution.

7.1.1. Baseline

As shown, the baseline consumption follows the thermostatic set points established by the building owner, not exploiting the thermal mass of the building and the TES. The outcome of the thermostatic controller is an

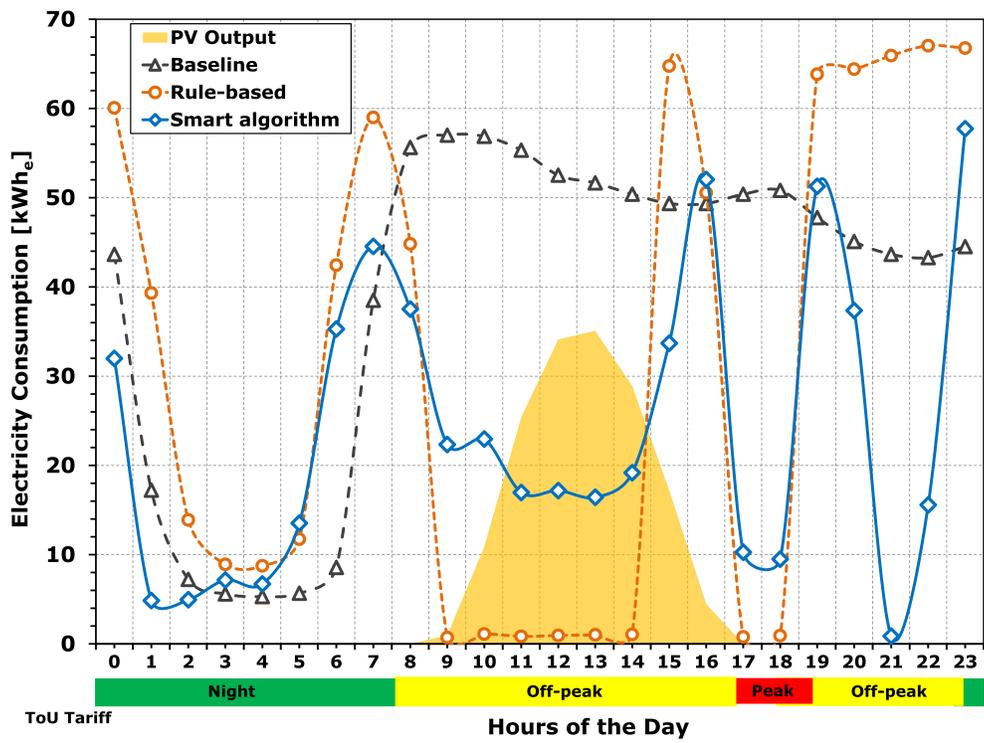


Figure 10: Electricity consumption profiles for the month of January 2014

increased consumption and electricity cost compared to the rule-based and the smart algorithm.

7.1.2. Rule-based

The rule-based algorithm is characterised by steep consumption spikes (0700 hrs and 1500 hrs) and extended valleys (0900 to 1400 hrs and 1700 to 1800 hrs) caused by the binary operation control. The operation of the heating system depends exclusively on hourly rules defined to reduce consumption in the TOU tariff scenario and evaluated with design day data. Hence, the hourly based rules requires the holding of a thermal energy reserve for the coldest days. Consequently, in the case of a different heat demand compared to the designed days, the energy stored in the TES always needs to be balanced to maintain the thermal comfort which causes consumption spikes close to peak periods. However, the rule-based algorithm reduces the energy consumption of the building by 32% and improves the efficiency of the heating system when compared with the baseline. During the day, the rule-based algorithm consumes less electricity because it switches off the heating system. Until 1500 hrs, the system control flow logic disables the heating system, allowing the inside temperature and the storage tank temperature to decrease to the lower boundary.

The strategy implemented in the rule-based algorithm is to reduce the peak time consumption between 1700 hrs to 1900 hrs, by charging the storage tank before the peak. During the peak hours, as can be seen in Figure 10, the system extracts heat from the storage tank, which it needs to be restored to the set point condition after one hour. Therefore, after 1900 hrs, the heating system is switched ON, which causes a rebound effect because the storage tank temperature breached the lower set point.

7.1.3. Smart algorithm

The smart algorithm cumulated profile shows a more balanced energy consumption pattern. It is noted that during the DR peak price period between 1700 hrs and 1900 hrs, the profile shows a consumption valley because of the shifting event triggered by the smart algorithm. The valley is preceded by a charging period at 1600 hrs. A rebound effect is also present after the peak time at 1900 hrs, followed by a consumption spike during the first hour of night tariff.

The consistency of the optimal strategy on the TOU tariff is supported by the increased efficiency operation and the reduction in electricity costs as

Table 6: Selected days based on 2014-2016 HDD data

Percentile	Quantile HDD	Actual HDD	Deviation	Heat demand	kWh
2nd	1.71	1.73	1.1%	Low	12.4
25th	6.00	6.17	2.8%	Medium-Low	15.8
50th (median)	8.30	8.89	6.7%	Median	17.7
75th	10.60	10.61	0.0%	Medium-High	18.4
98th	14.68	14.61	-0.4%	High	25.3

discussed in the following sections.

7.2. Consumer electricity profile

The effectiveness of the smart algorithm in reducing the consumer electricity consumption and expenditure was assessed utilising a whole month simulation period. Within the whole month several representative days have been extracted on the basis of a statistical analysis. The analysis to extract the most representative day of the heating season, was performed on the daily data for a three-year heating period from 2014-2016, using the associated daily Heating Degree Days (HDD) available from a weather data provider [52]. A percentile rank approach on the meteorological daily data ordered by HDD has been used to extract five days which represents the five corresponding quantiles (2nd, 25th, 50th, 75th and 98th).

From the simulation data from January 2014, the closest days to the three-year quantiles have been extracted and utilised in the assessment. These days are called Actual HDD and are close to the Quantile HDD extracted from the three years historical data.

In Table 6, the sample days in the three years assessment are called quantile HDD, while the actual days used for the assessment are called actual HDD. The table also shows the deviation in percentage between the quantiles and the selected actual days used for the assessment. The deviation is below 3% for all days except the median day, that shows a 6.7% heat demand increase.

The selected days aim to represent different heat demand conditions: low, medium-low, average, medium-high, high. After a first analysis of the distribution, the 100th and 1st quantiles have been substituted for the 98th and the 2nd to provide more representative extreme weather days. The selected extreme percentiles (98th and 2nd) identify a wider range of days for low and high heat demand profiles compared to the 100th and 1st percentiles.

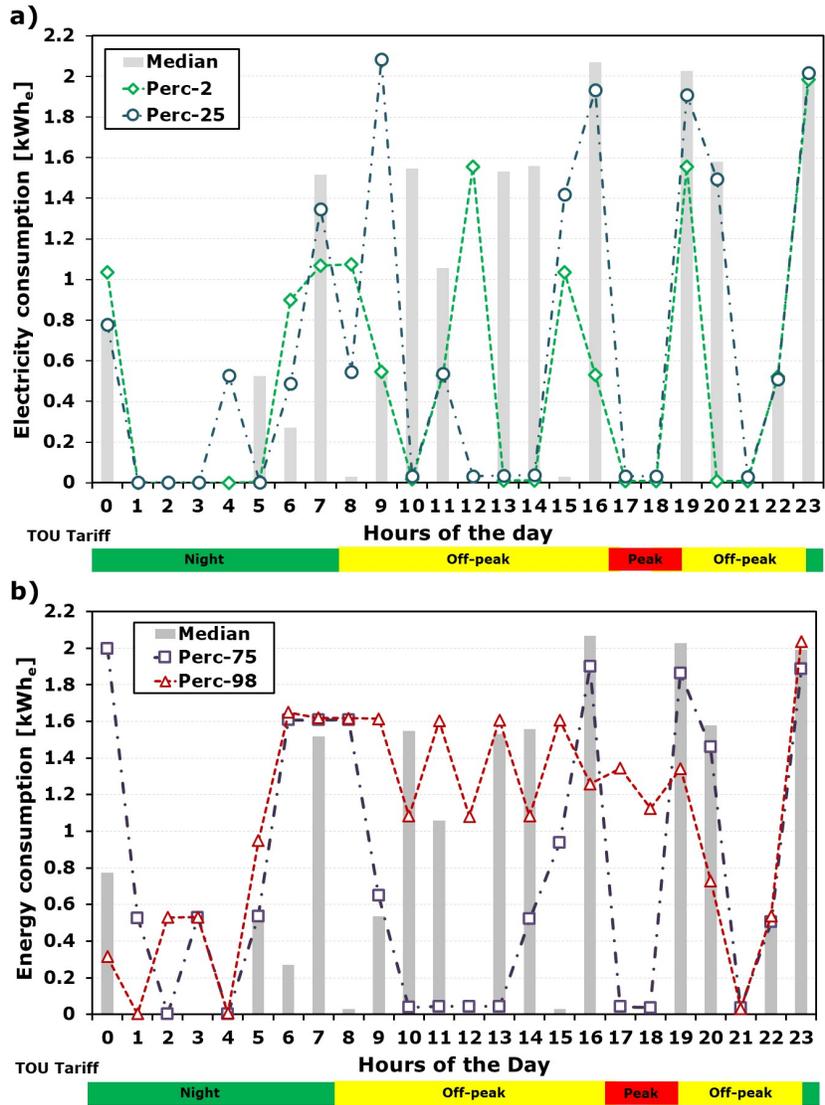


Figure 11: Energy consumption for three heat demand scenarios. (a) Low, Low-medium, Median; (b) Median, Medium-high, High

7.2.1. Electricity profile: Low, Medium-Low, Median

Figure 11a illustrates the electricity consumption pattern in the case of low, medium-low and median heat demand. Each data point represents the electricity consumption in kWh for an hour, i.e., the first data point is the electricity consumption of the building between 0000 hrs and 0059 hrs. The curves are identified with the related percentile described in the Table 6. Moreover all the curves are compared with the median profile that is the 50th percentile and represented by a grey histogram in the background. The median histogram shows a peak consumption at 1600 hrs, before two hours shifting load at peak time followed by a rebound effect at 1900 hrs. A second peak consumption is noted the first hour of the night tariff, 2300 hrs.

The low heat demand curve (2-Perc) exhibits a power reduction during peak hours, and two spikes during the off-peak hours at 1200 hrs and 1500 hrs to charge the TES before the peak, and after the peak at 1900 hrs as a rebound effect from DR peak price event.

Steeper peaks characterise the medium-low heat demand profile (Perc-25) before and after peak hours. The profile shows a consumption peak at 0900 hrs to restore the TES temperature within the boundaries after the morning heat demand peak. The same peak in the median profile is deferred between 1000 and 1100 hrs. The profile also exhibits two hours shifting load at peak time followed by a rebound effect at 1900 hrs. It also highlights an electricity demand peak during the first hour of night tariff, 2300 hrs. A peak in consumption is also evident during the hours with maximum solar irradiation at 1200 hrs because the algorithm maximises the utilisation of electricity produced by on-site RES by activation of the GSHP. It is also noted that the medium-low profile is similar to the median, revealing a consistent optimal strategy of the controller.

7.2.2. Electricity profile: High, Medium-High, Median

Figure 11b illustrates the medium-high and high heat demand profiles and the comparison with the median profile. The medium-high (Perc-75) shows a similar curve as the average case with a valley consumption during peak hours and a rebound effect after. In this case, the first two night tariff hours are characterised by greater electricity consumption due to the medium-high heat demand. The curve exhibits a peak during late night hours between 0200 and 0400 hrs because of the temperature drop, and although the algorithm reduces the consumption during the peak hours, to keep the temperature at the established comfort level, the valley during peak tariff time (1700 - 1800

hrs) is not as deep as in the previous instances. In this case, the system is switched on during the day, regardless of the price oscillations in order to meet thermal comfort.

The smart algorithm shows a consistent operational schedule during the day. It produces an increase in electricity consumption before and after peak in order to charge the TES. The algorithm prediction horizon limits the TES charging/discharging time to two hours before the price change which does not reduce the efficiency of the heating system because the GSHP is able to fully charge the TES within the time window horizon.

As a general indication for a predictive algorithm, the capacity of the storage in kWh divided by the primary heating source power output should be less than, or equal to the window time horizon. In case of larger storage, the algorithm does not exploit the full capacity of the system, limiting the shifting and forcing capabilities. In this research the ratio between the storage capacity and the GSHP is equal to 1.56 hours.

7.3. Consumer electricity consumption

Although the objective function of the smart algorithm is to minimise expenditure, the secondary goal is the minimisation of overall consumption. Figure 12 illustrates the total electricity consumption of the heating system during the test period (January 2014).

In comparison to the baseline, the rule-based algorithm shows an electricity reduction of 20.9%, while the smart algorithm exhibits a 39% reduction. When compared with the rule-based algorithm, the smart algorithm shows an electricity reduction of 22.9%. Additionally, an estimate of the 2014 total heating season consumption using the three algorithms was calculated in proportion to the monthly HDD. The consumption, net of the PV production for the heating season, was 2236 kWh. Additionally, the annual heating energy consumption of the building was normalised for its floor area (208 m^2). The data was then compared and the results are reported in Table 7. It should be noted that the rule-based algorithm compared to the baseline contributes to reduce the consumption of the building and consequently it raises the building energy efficiency. Additionally, the smart algorithm reduces the heating energy consumption, thereby increasing the efficiency of the system. This result further evidences the potential contribution of an EMS equipped with a smart algorithm to a dwelling energy rating.

Table 7: Building energy rating for the three algorithms

Algorithm	Annual heating energy demand (<i>kWh</i>)	Annual heating per floor area (<i>kWh/m²/yr</i>)
Baseline	13228	63.6
Rule-based	8841	42.5
Smart algorithm	5036	24.2

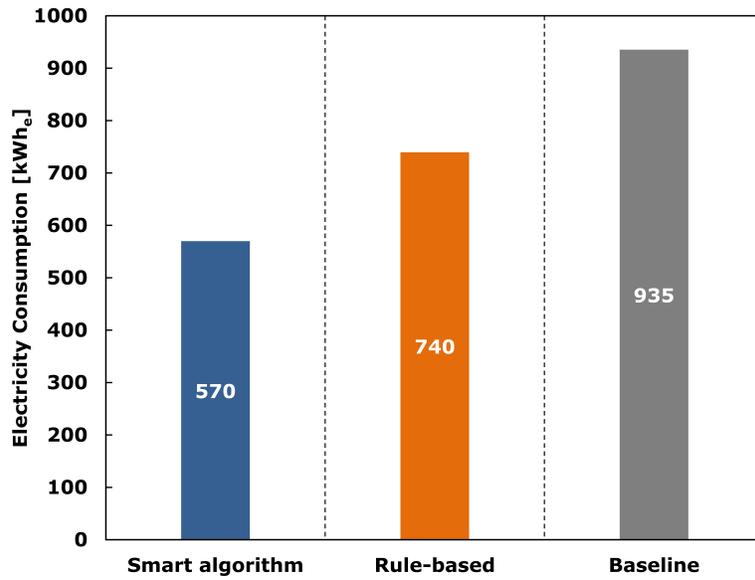


Figure 12: Total heating electricity consumption for (January 2014)

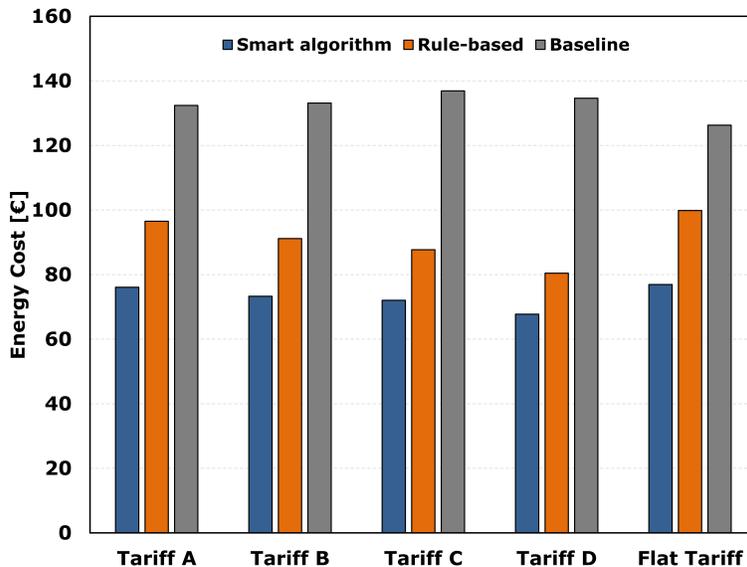


Figure 13: Energy expenditure for the three algorithms for different tariff structures (January 2014)

7.4. Consumer electricity cost

In the current work, TOU tariffs used for the Irish smart meter trial [31] were used to assess the consumer electricity expenditure and to compare the savings with the rule-based algorithm and the baseline. Figure 13 shows the electricity cost for the baseline, the rule-based and the smart algorithm using the TOU tariffs outlined in Table 1 for the month of January 2014. While the baseline exhibits the highest cost, the rule-based is the second highest cost, which can be attributed to the use of a fixed set of operational rules to control the heating system, resulting in additional charging cycles of the TES. With reference to the rule based, in comparison to the baseline, savings of between 27% for tariff A and 40% for tariff D (which most closely follows the SMP price) are evident.

Considering the smart algorithm, relative to the baseline, savings of between 42% for tariff A and 49% for tariff D (which most closely follows the SMP price) are evident. As illustrated in Figure 14, the smart algorithm reduces the electricity cost during peak hours, exploiting the storage and the building thermal mass, charging the TES before peak. The rule-based algorithm is characterised by steep cost spikes between peak hours (1500-1600

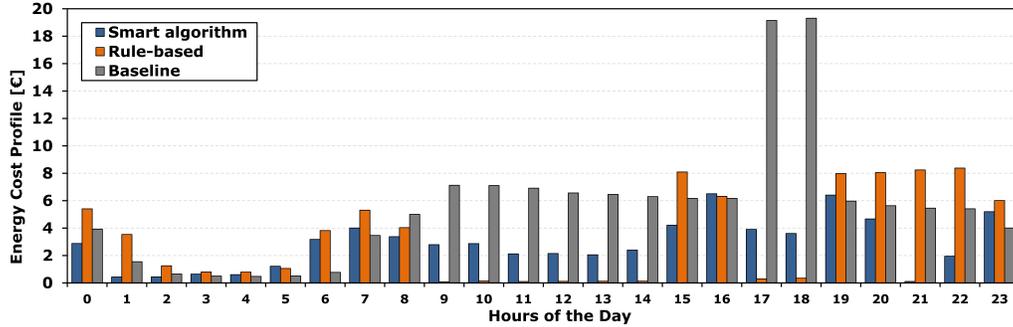


Figure 14: Cumulative electricity expenditure in the three cases for the whole month of January 2014 (Tariff D)

hrs and 1900-2200 hrs) caused by the operational rules. The rule-based algorithm also displays a cost reduction during off-peak hours because the system is completely switched off (1700-1800 hrs). The baseline system is controlled only by the thermostatic set point, so it exhibits a greater overall electricity cost during the hours of the day when the set points are increased. Additionally, as illustrated in Figure 14, the baseline operation of the GSHP during price peaks (1700 to 1800 hrs) produces high expenditure peaks due to the structure of the price tariff, which penalises peak consumption. These peaks are not present for the other controllers demonstrating the effectiveness of the algorithms in reducing the peak expenditure.

7.5. Utility electricity cost

Figure 15a shows the generation cost using Irish electricity SMP prices [53] for the month of January 2014. Electricity consumption is multiplied by the relevant SMP price for each timestep in order to evaluate the test period electricity production cost from a utility perspective. With reference to the rule-based algorithm, where the control strategy aims to shift electricity consumption from peak (1700-1900 hrs) to off peak (1500-1700 hrs) times, a generation cost reduction from €43.08 to €34.02 is evident (21% reduction). With reference to the smart algorithm, a 43% (€ 16.87) reduction is evident.

The difference between the SMP peak price and the off-peak can reach a ratio of 1 to 6. The peaks in SMP price are aligned with the TOU peak periods. Consequently, the algorithms showed a significant reduction of generation cost. As evidence, Figure 15a shows the 2014 SMP hourly price divided into three tertiles and, for each algorithm, the associated percentage

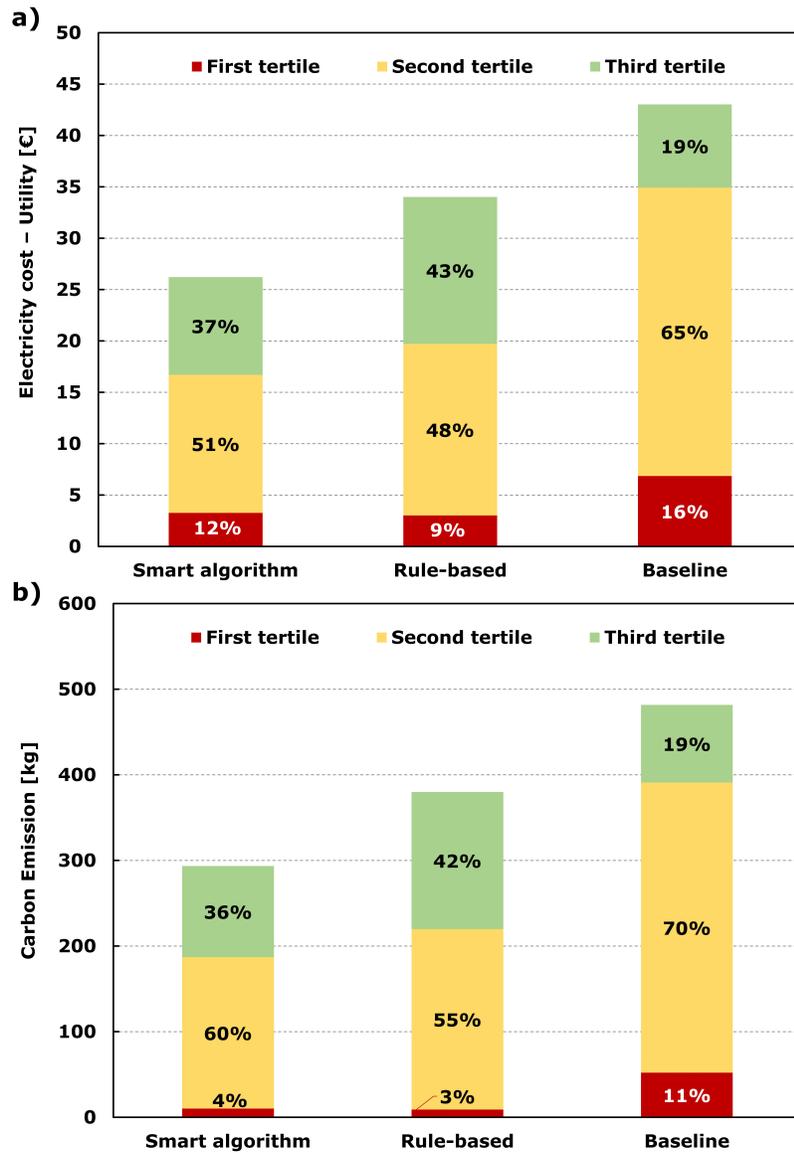


Figure 15: (a) Electricity generation cost and (b) Cumulative carbon emissions for January 2014

of electricity in the tertiles is calculated. The tertiles are based on the SMP price variation during January 2014. The SMP reached its maximum peak at 97.5 €/MWh and a minimum value of 11.85 €/MWh. The first tertile price spans from the maximum to 63.32 €/MWh, while the second tertile spans from 63.32 €/MWh to 48.24 €/MWh.

Although the rule-based algorithm uses a higher percentage of electricity in the third tertile when compared with the smart algorithm, the efficiency of the smart algorithm strategy in reducing the total amount of energy consumed by the system leads to relevant utility cost savings. Additionally, as expected, the baseline thermostatic controller resulted in higher consumption on the first and second tertiles which led to a higher generation cost overall.

7.6. Environmental impact

Eirgrid, the Irish transmission system operator, provides 30 minutes averaged carbon emissions (gCO_2/kWh), based on technical data from all generation units, including renewable energy, for the overall production of electricity [54]. For January 2014, the footprint varied from 251 gCO_2/kWh to 643 gCO_2/kWh . The maximum carbon emission is verified during the peak time hours (1700 - 1900 hrs). Therefore, a significant reduction in CO_2 emission is expected as a result of the control algorithm.

Figure 15b illustrates the time period CO_2 footprint as determined for the baseline, the smart and the rule-based algorithms. The smart-algorithm exhibits the greatest reduction, which is 39% relative to the baseline case, thereby reducing the overall electricity drawn from the grid and consequently the associated carbon emissions.

The assessment divided the hourly emissions in three tertiles to evaluate the impact of the TOU tariffs on the building carbon emissions. The first tertile spans between the maximum, 643 gCO_2/kWh , and 502 gCO_2/kWh . The second tertile spans from 502 gCO_2/kWh to 429 gCO_2/kWh while the third tertile from 429 gCO_2/kWh to the minimum, 243 gCO_2/kWh .

The graph shows how the rule-based, despite having a higher percentage of consumption during lower emissions time, has an overall 20% increased carbon emissions due to the higher energy consumption. The baseline has the majority of consumption during the second tertile, however, the higher carbon footprint is also defined by the increased electricity consumption. It is evident that in order to exploit the penetration of RES at power system level to lower the carbon emissions from buildings, the carbon intensity must be included either within the objective function or in the price scheme utilised by

the optimisation algorithms. Adopting the fixed TOU price scheme results in a reduction of carbon footprint which is exclusively related to DR measures that shift the electricity produced by peak generators. The variability of the carbon intensity at system level is not embedded in the TOU tariffs. A more dynamic price such as Real Time Price (RTP) would be more suitable for the buildings integration of the RES at system level.

In closing, the objective of the paper was to develop a novel optimisation algorithm that provides energy load control and DR capabilities embedded in a EMS and connected to sensors in an HAN. The research has also demonstrated a means of designing, implementing and testing a predictive smart controller which is identified as the most efficient in responding to DR events while maintaining the thermostatic set points. It is noted that the smart algorithm provides a contribution to the research field of optimisation algorithms to enable DR.

The smart controller uses a statistical model to predict the building internal temperature, utilising weather forecasts and a TES model. The computational bottleneck of the algorithm is the construction of the decision tree that affects the memory heap exponentially in relation to the time horizon. Although several pruning techniques were implemented to reduce the complexity of data structure visits, a better optimisation is required to extend the forecast horizon to four hours or a day ahead.

The algorithm relies extensively on weather and price forecasts. It would also be essential to extend the research towards an accurate short-term prediction of electricity price and weather. Short-term forecasts are of vital importance for allocation and control of the loading flexibility provided by buildings, in the context of demand response.

8. Conclusions

The present paper described the design, development and testing of an EMS to provide DR capabilities for residential buildings. A calibrated building energy model based on a real residential building was used to test the capabilities of different methodologies in assessing DR actions. A fully open-source co-simulation tool, based on the combination of optimisation techniques with machine learning models was used to find optimal strategies, while metered data was used to train and test the algorithms. The results obtained can be summarised as follows:

- *Energy consumption*: a significant impact of the smart algorithm on the electricity consumption per floor area has been demonstrated in Section 7.3. The optimal strategy of the algorithm was able to increase the efficiency of the heating system from $63 \text{ kWh/m}^2/\text{yr}$ to $24 \text{ kWh/m}^2/\text{yr}$.
- *Costs*: a reduction of the electricity cost compared to the baseline for January 2014 spans from 42% to 49%, which when compared to the rule-based algorithm, the reduction ranges from 27% to 40%. Although these savings relative to the baseline may appear significant, they are of the order of a few hundred euro per year. Consequently, they need to be evaluated from the perspective of the capital cost and the Return Of Investment (ROI) for the installed technical solution, which is still in the range of 5-10 years.
- *Environmental*: in terms of carbon emissions, the smart algorithm emissions reduction is caused primarily by a combination of peak shifting, overall energy reduction and exploitation of the PV system. It should be noted that the algorithm objective function does not explicitly capture power system carbon emissions intensity and consequently does not maximise carbon footprint reduction, especially during periods with no price tariff variations. Nevertheless, an overall carbon emissions reduction of 38% was achievable by minimising the overall electricity expenditure and the overall consumption. To increase the environmental performance while utilising the smart algorithm, it would be necessary to embed the carbon intensity of the power system into the price tariff, triggering a DR price signal event in periods of carbon emissions peak intensity.
- *Utility*: from the point of view of the utility, it is evident that if the DR price signal is aligned to the market price utility, the smart algorithm can reduce the generation costs for the residential sector. However, high penetration of EMS units equipped with the smart algorithm could lead to shift peak periods to other times of the day, reducing the effect of the DR event to the SMP price. A randomisation of the control mechanism could redistribute the peak electricity consumption throughout the day.
- *Computational*: although the computational time associated with each time step can appear reasonable, a simulation of a building cluster con-

trolled by the smart algorithm requires further optimisation of the statistical model performance. A decoupled infrastructure could facilitate the optimisation process, by separating the controller from the model and from the building simulation. As discussed in Section 7.2, the developed advanced controller reduces the energy consumption of the heating system by 39% and improves the efficiency of the operations, optimising the electricity consumption. Moreover, the smart algorithm optimises the energy consumption throughout the day, reducing the peaks and keeping the temperature of the storage tank between the selected range. This strategy leads to an overall reduction of 22.9% when compared with the rule based approach.

- *Prediction model:* The results demonstrated that the prediction horizon of two hours reduced the energy consumption while maintaining thermal comfort. It is also noted that the MP5 model is sufficiently accurate for forecasting the thermal behaviour of the building and it can be effectively embedded in a predictive controller. Moving beyond the two hour time horizon, requires a longer evaluation time for each time step, because the computational bottle neck of the algorithm is represented by the in-memory search tree. Every time step expansion of the time horizon is equal to an exponential increase of the nodes, leading to oversize errors of the memory heap. Advanced optimisation techniques could have been applied to the tree construction but they were out of the scope of the current research.

In conclusion, advanced demand response control techniques in all-electric residential buildings together with the increase adoption of time of use tariffs in many European countries, can positively contribute to the development of flexible power system frameworks aimed at reducing the carbon footprint of building stocks and supporting the shift towards more sustainable power generation mix.

Acknowledgements

This work was conducted in the Electricity Research Centre, University College Dublin, Ireland, which is supported by the Commission for Energy Regulation, Bord Gás Éireann, Bord na Móna Energy, Cylon Controls, EirGrid, Electric Ireland, ESIPP, Energia, EPRI, ESB International,

ESB Networks, Gaelectric, Intel, SSE Renewables, and UTRC. This publication has emanated from research conducted with the financial support of PRLTI[R12681]. The authors would like to thank the building owners for their essential support.

Nomenclature

C_e	Electricity price (€\kWh)
C_{set}	Circulation pump status
COP_{hp}	Heat pump average COP
cpC	Circulation pump electricity consumption (kWh)
Cpw	Specific heat capacity J\(\text{kg} * \text{K})
DT	Temperature difference ($^{\circ}\text{C}$)
inT	Inside temperature ($^{\circ}\text{C}$)
M	TES water mass
$outT$	Outside temperature ($^{\circ}\text{C}$)
P_e	Building electricity consumption (kWh)
P_{hp}	Heat pump energy consumption (kWh)
PV	PV electricity production (kWh)
S	Shifting flexibility (kWh)
T_{bd}	TES setpoint bandwidth ($^{\circ}\text{C}$)
T_{in}	Internal zone temperature ($^{\circ}\text{C}$)
T_{max}	TES maximum set point temperature ($^{\circ}\text{C}$)
T_{min}	TES minimum set point temperature ($^{\circ}\text{C}$)
T_{out}	Outside temperature ($^{\circ}\text{C}$)
T_{set}	Internal set zone temperature ($^{\circ}\text{C}$)

T_{set}	Temperature set point (°C)
T_{tk}	TES temperature (°C)
$tesT$	TES temperature (°C)

Glossary

- API** Application Program Interface. 5, 7, 10–13, 15, 16, 26
- ASHRAE** American Society of Heating, Refrigerating and Air-Conditioning Engineers. 10
- BCVTB** Building Controls Virtual Test Bed. 10, 11, 13
- BES** Building Energy Simulation. 4, 12, 15, 26
- DHW** Domestic Hot Water. 8
- DR** Demand Response. 2–5, 10, 30, 33, 40, 41
- DSM** Demand Side Management. 3
- EMS** Energy Management System. 2, 3, 5–7, 10–13, 15, 23, 34, 40, 41
- GSHP** Ground Source Heat Pump. 8, 20, 25, 26, 33, 34, 37
- HAN** Home Area Network. 2, 13, 17, 40
- HDD** Heating Degree Days. 31, 34
- HP** Heat Pump. 10
- HTTP** Hypertext Transfer Protocol. 13, 15
- HVAC** Heating Ventilation and Air Conditioning system. 12
- JDBC** Java Database Connectivity. 10
- JSON** JavaScript Object Notation. 10

MAE Mean Absolute Error. 21

MBE Mean Bias Error. 27

MVC Model View Control. 10

PV PhotoVoltaics. 10, 15, 19, 20, 28, 34, 41

RES Renewable Energy Systems. 2, 12, 33, 39, 40

RMSE Root Mean Square Error. 21

ROI Return Of Investment. 41

RTP Real Time Price. 40

SMP System Marginal Price. 7, 37, 39, 41

TES Thermal Energy Storage. 5, 12, 17, 18, 20, 21, 26, 28, 30, 33, 34, 36, 40, 44

TOU Time of Use. 16, 30, 36, 37, 39, 40

References

- [1] IEA, Transition to Sustainable Buildings, Technical Report, International Energy Agency, 2013.
- [2] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, *Energy and Buildings* 40 (2008) 394 – 398.
- [3] SEAI, Residential energy road map, Online, 2011.
- [4] European Commission, Energy roadmap 2050, 2012.
- [5] S. Humeau, T. K. Wijaya, M. Vasirani, K. Aberer, Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households, in: *Sustainable Internet and ICT for Sustainability (SustainIT)*, 2013, IEEE, pp. 1–6.
- [6] M. Balakrishnan, Smart energy solutions for home area networks and grid-end applications, 2012.

- [7] T. Yang, Y. Zhao, H. Pen, Z. Wang, Data center holistic demand response algorithm to smooth microgrid tie-line power fluctuation, *Applied Energy* 231 (2018) 277 – 287.
- [8] H. Farhangi, The path of the smart grid, *IEEE Power and Energy Magazine* 8 (2010) 18–28.
- [9] J. C. Fuller, K. P. Schneider, D. Chassin, Analysis of residential demand response and double-auction markets, in: *Power and Energy Society General Meeting, 2011 IEEE, IEEE*, pp. 1–7.
- [10] ASHRAE, Measurement of energy and demand savings, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta (2002) 17. ASHRAE Guideline 14-2002.
- [11] V. Hamidi, F. Li, F. Robinson, Demand response in the uk’s domestic sector, *Electric Power Systems Research* 79 (2009) 1722–1726.
- [12] P. Palensky, D. Dietrich, Demand side management: Demand response, intelligent energy systems, and smart loads, *IEEE transactions on industrial informatics* 7 (2011) 381–388.
- [13] H. C. Gils, Assessment of the theoretical demand response potential in europe, *Energy* 67 (2014) 1–18.
- [14] M. De Rosa, M. Carragher, D. P. Finn, Flexibility assessment of a combined heat-power system (chp) with energy storage under real-time energy price market framework, *Thermal Science and Engineering Progress* 8 (2018) 426 – 438.
- [15] A. Schumann, J. Ploennigs, B. Gorman, Towards automating the deployment of energy saving approaches in buildings, in: *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, BuildSys ’14, ACM, New York, NY, USA, 2014*, pp. 164–167.
- [16] J. R. Vazquez-Canteli, Z. Nagy, Reinforcement learning for demand response: A review of algorithms and modeling techniques, *Applied Energy* 235 (2019) 1072 – 1089.
- [17] A.-H. Mohsenian-Rad, A. Leon-Garcia, Optimal residential load control with price prediction in real-time electricity pricing environments, *Smart Grid, IEEE Transactions on* 1 (2010) 120–133. 165.

- [18] E. S. Parizy, H. R. Bahrami, S. Choi, A low complexity and secure demand response technique for peak load reduction, *IEEE Transactions on Smart Grid* (2018) 1–1.
- [19] S. Bahrami, V. W. S. Wong, J. Huang, An online learning algorithm for demand response in smart grid, *IEEE Transactions on Smart Grid* 9 (2018) 4712–4725.
- [20] Q. Hu, F. Li, Hardware design of smart home energy management system with dynamic price response, *Smart Grid, IEEE Transactions on* 4 (2013) 1878–1887.
- [21] D. Ren, H. Li, Y. Ji, Home energy management system for the residential load control based on the price prediction, in: *Online Conference on Green Communications (GreenCom)*, 2011 IEEE, pp. 1–6.
- [22] D. Kolokotsa, K. Niachou, V. Geros, K. Kalaitzakis, G. Stavrakakis, M. Santamouris, Implementation of an integrated indoor environment and energy management system, *Energy and Buildings* 37 (2005) 93–99.
- [23] B. Alimohammadisagvand, J. Jokisalo, K. Sirén, Comparison of four rule-based demand response control algorithms in an electrically and heat pump-heated residential building, *Applied Energy* 209 (2018) 167–179.
- [24] V. Arabzadeh, B. Alimohammadisagvand, J. Jokisalo, K. Siren, A novel cost-optimizing demand response control for a heat pump heated residential building, in: *Building Simulation*, volume 11, Springer, pp. 533–547.
- [25] M. Liu, P. Heiselberg, Energy flexibility of a nearly zero-energy building with weather predictive control on a convective building energy system and evaluated with different metrics, *Applied Energy* 233-234 (2019) 764 – 775.
- [26] R. Lu, S. H. Hong, Incentive-based demand response for smart grid with reinforcement learning and deep neural network, *Applied Energy* 236 (2019) 937 – 949.

- [27] F. Schloegl, S. Rohjans, S. Lehnhoff, J. Velasquez, C. Steinbrink, P. Palensky, Towards a classification scheme for co-simulation approaches in energy systems, in: 2015 International Symposium on Smart Electric Distribution Systems and Technologies (EDST), pp. 516–521.
- [28] F. Pallonetto, E. Mangina, D. Finn, F. Wang, A. Wang, A restful api to control a energy plus smart grid-ready residential building: Demo abstract, in: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, BuildSys '14, ACM, New York, NY, USA, 2014, pp. 180–181.
- [29] F. Pallonetto, S. Oxizidis, F. Milano, D. Finn, The effect of time-of-use tariffs on the demand response flexibility of an all-electric smart-grid-ready dwelling, *Energy and Buildings* 128 (2016) 56 – 67.
- [30] CER, Demand side vision for 2020, 2011.
- [31] CER, Smart metering cost-benefit analysis and trials findings reports., 2012.
- [32] Ireland Central Statistics Office, The roof over our heads, 2012. Pg 60.
- [33] Irish building regulations, Building regulations 2011, conservation of fuel and energy - dwellings, 2011. Section L.
- [34] R. Saunier, Getting started with Laravel 4, Packt Publishing Ltd, 2014.
- [35] J. Greenspan, B. Bulger, MySQL/PHP database applications, John Wiley & Sons, Inc., 2001.
- [36] G. M. Nijssen, T. A. Halpin, Conceptual Schema and Relational Database Design: a fact oriented approach, Prentice-Hall, Inc., 1989.
- [37] M. Wetter, Co-simulation of building energy and control systems with the building controls virtual test bed, *Journal of Building Performance Simulation* 4 (2011) 185–203.
- [38] G. Hamilton, R. Cattell, M. Fisher, JDBC Database Access with Java, volume 7, Addison Wesley, 1997.
- [39] D. Crockford, The application/json media type for javascript object notation (json), RFC 4627 - IETF (2006).

- [40] F. Pallonetto, S. Oxizidis, D. Finn, Exploring the demand response potential of a smart-grid ready house using building simulation software, in: IBPSA Building Simulation Conference.
- [41] E. Alpaydin, Introduction to machine learning, MIT press, 2014.
- [42] W. Underground, Historical weather, URL: <http://www.wunderground.com/weather/api> [accessed 2016-01-28][WebCite Cache ID 6erY0WjcK] (2015).
- [43] A. L. Blum, P. Langley, Selection of relevant features and examples in machine learning, *Artificial Intelligence* 97 (1997) 245 – 271. Relevance.
- [44] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, 2016.
- [45] L. Al Shalabi, Z. Shaaban, B. Kasasbeh, Data mining: A preprocessing engine, *Journal of Computer Science* 2 (2006) 735–739.
- [46] D. Kremelberg, *Practical statistics: A quick and easy guide to IBM® SPSS® Statistics, STATA, and other statistical software*, SAGE publications, 2010.
- [47] D.-S. Kapetanakis, E. Mangina, D. P. Finn, Input variable selection for thermal load predictive models of commercial buildings, *Energy and Buildings* 137 (2017) 13 – 26.
- [48] J. R. Quinlan, Learning with continuous classes, in: 5th Australian joint conference on artificial intelligence, volume 92, Singapore, pp. 343–348.
- [49] S. C. Kumar, E. D. Chowdary, S. Venkatramaphanikumar, K. V. K. Kishore, M5p model tree in predicting student performance: A case study, in: 2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT), pp. 1103–1107.
- [50] Y. Wang, I. H. Witten, Induction of model trees for predicting continuous classes, Department of Computer Science, University of Waikato, 1996.
- [51] E. J. Harner, D. Luo, J. Tan, Javastat: A java/r-based statistical computing environment, *Comput. Stat.* 24 (2009) 295–302.

- [52] W. Underground, Degree days, Online, 2017.
- [53] Semo, Single electricity market operator, 2015.
- [54] J. Torriti, M. G. Hassan, M. Leach, Demand response experience in europe: Policies, programmes and implementation, *Energy* 35 (2010) 1575–1583.