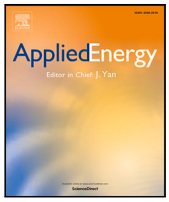
Applied Energy 315 (2022) 118947

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



An ensemble learning-based framework for assessing the energy flexibility of residential buildings with multicomponent energy systems

Adamantios Bampoulas [a,b,*], Fabiano Pallonetto [a], Eleni Mangina [a,c], Donal P. Finn [a,b]

[a] UCD Energy Institute, University College Dublin, Dublin, Ireland
[b] UCD School of Mechanical and Materials Engineering, University College Dublin, Dublin, Ireland
[c] UCD School of Computer Science, University College Dublin, Dublin, Ireland

ARTICLE INFO

Keywords:
Energy flexibility
Flexibility indicators
Residential sector
Ensemble learning

ABSTRACT

A key issue in energy flexibility assessment is the lack of a scalable practicable approach to quantify and characterise the flexibility of individual residential buildings from an integrated energy system perspective without the need to use complex simulation models. In this study, this problem is addressed by explicitly quantifying the flexibility of multicomponent thermal and electrical systems commonly found in residential buildings based on an ensemble learning framework that consists of four algorithms, namely, random forests, multilayer perceptron neural network, support vector machine, and extreme gradient boosting. The day-ahead and hour-ahead prediction models developed are periodically updated considering dynamic feature selection based on residential occupancy patterns. The proposed methodology utilises synthetic data obtained from a calibrated white-box model of an all-electric residential building for two indicative occupancy profiles. The energy systems evaluated include a heat pump, a photovoltaic system, and a battery unit. The daily flexibility mappings are acquired by applying hourly independent, and consecutive demand response actions for each energy system considered, using suitable energy flexibility indicators. The results show that the ensemble models developed for each target variable outperform each of the constituent machine learning algorithms. Moreover, the storage capacity resulting from harnessing the heat pump downward flexibility demonstrates accurate accuracy with a coefficient of determination equal to 0.979 and 0.968 for day-ahead predictions and 0.998 and 0.978 for day ahead predictions for the two occupancy profiles considered, respectively. This framework can be used by electricity aggregators to evaluate a building portfolio in an end-user-tailored manner or optimally exploit its energy flexibility considering multi-step predictions to shift electricity usage to off-peak times or times of excess onsite renewable energy generation.

1. Introduction

The transition of the energy sector to a low carbon energy system is accompanied by several challenges related to the wider adoption of renewable energy sources (RES) – often intermittent – as well as the expected increase in the electrification of several aspects of the end-user sector. Traditionally, power system operators have focused on supply-side flexibility to cope with potential supply and demand mismatches that are likely to compromise electricity system robustness [1]. The advent of new communication technologies (e.g., smart metres, sensors, the Internet of Things, etc.) has paved the way for increased demand-side flexibility or when electricity is the energy vector then demand response (DR) [1–4]. DR can be defined as the ability to change the use of electricity by end-users from their normal consumption patterns by responding to control signals from grid operators and/or financial incentives from electricity generators/aggregators. The scope of these signals is to modulate and optimise electricity usage and to balance electricity production and consumption [2].

Buildings consume approximately 36% of all energy produced globally, thus emerging as promising components of the future smart electricity grid [5]. Buildings are likely to become important sources of energy flexibility for the electricity grid due to their increasing electrification by the wider adoption of heat pumps, on-site electricity generation, storage technologies, and electric vehicles [6]. IEA Annex 67 defines energy flexibility as the ability to manage demand and production according to local climatic conditions, user needs, and grid requirements. The energy flexibility of buildings will thus enable demand-side management (DSM) based on the requirements of the surrounding networks and the availability of RES to minimise CO_2 emissions [7]. The main stakeholders interested in the energy flexibility

* Correspondence to: E0.86, UCD Energy Institute, O'Brien Science Building, University College Dublin, Belfield, Dublin 4, Ireland.
E-mail address: adamantios.bampoulas@ucdconnect.ie (A. Bampoulas).

https://doi.org/10.1016/j.apenergy.2022.118947
Received 10 December 2021; Received in revised form 12 February 2022; Accepted 12 March 2022
Available online 4 April 2022

potential of a particular building or building portfolios are aggregators and end-users (i.e., managers and building owners/occupiers). Aggregators may be more interested in the wider business potential of DR as well as technical issues (capability of building power shifting, response time, etc.) to harness the flexibility potential of their portfolios; end-users are mainly interested in the energy consumption and the potential cost savings or CO_2 emissions reductions from activating the flexibility of their devices [8].

The structural thermal mass of buildings can be used as an energy storage medium that can be readily activated and utilised as a flexibility source by responding to an external signal. Most building model development approaches for DR applications assume that the system model is either perfectly known or can be readily found in the literature; nevertheless, this task could become more complicated than the controller design per se [9]. The reason for this is that building energy performance depends on various factors including, inter alia, weather conditions, building envelope thermal properties, occupant behaviour, controllable loads (e.g., HVAC system), and non-controllable loads (e.g., lighting, wet appliances, etc.) [10]. Building energy simulation approaches can be categorised into three broad categories: physics-based approaches (white-box models), hybrid approaches (grey-box models), and machine learning (ML) approaches (black-box models). White-box models predict building energy consumption by using numerical equations based on detailed physical properties of building materials and characteristics. Although these models can provide a physics-based interpretation of building phenomena, they are accompanied by several disadvantages including requirement of expertise and difficulties in making proper assumptions [11]. Consequently, these models need to be manually tuned to match the measured data from the building resulting in a significant gap between the modelled predictions and the real building performance [12]. Moreover, the feasibility of white-box models could be compromised especially when considering building load aggregation challenges that are likely to increase the computational burden [13]. Grey-box models combine physical models and data-driven approaches to simulate building energy. With appropriate calibration measures they can provide reasonable prediction accuracy while reducing the computational cost of physical models [14], however, these models still require building engineering expertise and a considerable amount of information about the physical system [15]. Black-box models can tackle the challenges associated with dynamic environments by making use of past recorded data. The evolution of information and communication technology along with the availability of building data by smart metres (including data associated with the HVAC system, thermal comfort, occupancy, and weather) is likely to boost increased uptake and broader application of data-driven algorithms in the building industry [16].

Despite their potential benefits in the building energy sector, data-driven models are associated with various limitations including, inter alia, non-interpretable parameters [17], the requirement for large datasets (related to indoor and outdoor temperatures, HVAC power consumption as well as thermostatic setpoints) [18], and poor performance generalisation under variable operation strategies, weather conditions, or occupant behaviour. The latter issue can be addressed by applying suitable updating strategies that are likely to enable scalable and customisable energy consumption models [11]. This feature is important since the diversity in appliance types and the interactional behaviours could cause increased uncertainty in evaluating the DR potential across different households [19] and ultimately jeopardise system reliability by generating another unforeseen peak (i.e., a rebound effect) [20]. These barriers can be addressed by identifying and engaging end-users with high flexibility potential [21]. Considering the above, a scalable and user-tailored framework to evaluate the flexibility of different thermal and electrical systems on an integrated common basis would allow electricity aggregators to assess a portfolio of residential buildings, not only within the end-user pre-qualification process but also in operational environments. In order to attain optimal bidding strategies, multi-settlement markets are utilised with planning horizons that range from day-ahead to the real time [22]. In several markets, however, market participants can also submit hourly bids apart from real-time settlements [23]. Thus, depending on the bidding strategy, day-ahead or hour-ahead prediction models are required. In this context, multistep-ahead predictions could enable aggregators to estimate power demand in advance of real-time and ultimately evaluate all possible flexibility measure combinations. In this context, multistep-ahead predictions could enable aggregators to estimate power demand in advance of real-time and ultimately evaluate all possible flexibility measure combinations.

To date, the lack of practicable methodologies for realistically assessing energy flexibility of individual buildings [24] coupled with the inadequacy of current characterisation methodologies (due to the different interpretations, and requirements associated with an energy flexible building) [25] impedes the accurate assessment of their DR potential. Although both commercial and residential buildings have been recognised as suitable candidates for providing flexibility to the grid [26], most studies have mostly focused on commercial buildings [27]. This is not only due to the associated sensor-based data scarcity but also to the relatively higher occupant behaviour variability compared to the commercial sector [28]. The prediction of baseline building energy consumption by data-driven models has been extensively studied in the literature [29–31], however, the suitability and performance of the developed models in the context of flexibility assessment have not been investigated.

A series of studies have developed data-driven energy flexibility assessment methods for residential buildings to evaluate the associated DR potential on an aggregate basis (e.g., district level, sector level) [32–34]. On the other hand, other studies in the literature focused on the evaluation of the flexibility of individual buildings and energy systems. For example, Bashir et al. [35] used a physics-based model to develop an optimal control scheme; the scope of the control strategy was to maximise building self-consumption by harnessing the flexibility potential of passive and active TES systems. In [36], a model predictive control is developed to investigate the flexibility potential of passive and active TES systems by following a grey-box modelling approach. Bunning et al. [37] developed a data predictive control approach based on random forests with affine functions to optimise cooling energy. Cotrufo et al. [38] used various ML models, namely, artificial neural networks, gauss process regression, support vector machine, decision trees, and random forests to optimise heating demand and gas boiler consumption. In [39], a model predictive control scheme based on deep time delay neural networks and regression trees is utilised to minimise energy usage. Smarra et al. [40] developed a data-driven predictive control based on random forests and regression trees to minimise energy usage. However, these research efforts assess building energy flexibility implicitly by developing various optimisation schemes to attain certain objectives. Furthermore, previous work has mainly focused on analysing energy flexibility in an implicit manner by determining the contribution of different control schemes to achieving specific objectives. The dependency on the underlying algorithms and the associated optimisation objectives [41] prevents electricity aggregators from accurately assessing the flexibility potential of their customers and this hampers the optimisation of their portfolios.

Nevertheless, other relevant works in the literature have utilised market-independent approaches that characterise the flexibility of individual buildings by proposing generic metrics/indicators. For example, Stinner et al. [42] used a low-order model developed in Modelica to quantify the power and the energy flexibility of heat pumps and combined heat and power plants when combined with thermal energy storage by calculating the associated forced and delayed operation times. Reynders et al. [43] use a detailed building energy simulation model for typical building typologies to quantify the flexibility of passive TES systems based on their power shifting capability, the energy increase, and associated energy costs with respect to the energy shifted

during a DR action. Foteinaki et al. [44] utilise building performance simulation software to quantify the flexibility of passive TES with respect to the energy that can be shifted and the associated energy savings/costs during a DR event. Le Dreau et al. [45] use a physics-based model based on measured and statistical data to assess the building heating system capability to shifting its demand away from high price periods. Masy et al. [46] use a detailed white-box model and a simplified grey-box model to evaluate flexibility arising from passive TES systems with respect to the load volumes shifted and the procurement costs avoided during a DR action. Kathirgamanathan et al. [47] utilise various building types and simulation software to quantify the passive TES flexibility considering the energy added/curtailed, and the associated energy costs/savings with respect to the energy shifted. In [48], a MATLAB-based simulation framework is used to quantify the active TES flexibility considering the energy shifted, the associated energy costs/savings with respect to the energy shifted, and the electricity cost. Zhou et al. [49] use a physics-based model to quantify the flexibility of passive and active TES systems by considering the period during which the energy consumption can be delayed or anticipated in the context of local RES. Although these studies assess various flexibility characteristics (e.g., power and energy shifting capability, energy costs, etc.) [42–49], they all follow white-box or grey-box approaches that do not investigate the developed model performance under dynamic environments — such as those resulting from DR events.

In contrast to the above-mentioned studies, other research efforts used data-driven methodologies to evaluate the flexibility of residential buildings on a one-by-one basis. For instance, Sadat-Mohammadi et al. [50] utilise a multilayer perceptron neural network in order to optimise the operation schedules of wet appliances and electric vehicles without, however, providing a flexibility quantification framework for the wet appliances. In [51], a neural network-based model predictive control strategy has been developed to optimise the flexibility of the passive TES with respect to various factors namely a flexibility factor, a supply cover factor, and a load cover factor. In [52], the flexibility of passive and active TES systems as well as EVs is quantified by determining a variety of possible power modulations during a DR action based on statistical analysis. Although these studies [50–52] propose explicit flexibility quantification frameworks in the context of the considered case studies, they do not investigate the associated flexibility arising from secondary or later effects (i.e., rebound effects). On the other hand, Balint and Kazmi [53] use stochastic gradient descent to quantify the passive TES energy shifting capability (storage capacity) by also assessing the associated energy costs (storage efficiency); however, this approach is based on offline training. The aforementioned reviewed studies do not assess building flexibility from an integrated system perspective but mainly focus on a single energy system type (electrical [50] or thermal [51]) or they do not assess building flexibility in an operational environment [52,53].

In the context of energy flexibility assessment, power consumption patterns vary not only because of potential DR actions but also due to consumer behaviour volatility [54], weather [55], etc. To this respect, additional information could improve the prediction model performance such as household size, income, appliance inventory, usage information, etc. [56], however, this information is difficult to obtain and is associated with user privacy challenges. A potential solution to tackle the limited-information challenge is the use of ensemble models; ensemble or integration models can be defined as frameworks that use multiple learning algorithms to exhibit better predictive performance than that achieved by any of the constituent algorithms [57]. Ensemble models can automatically manage the strengths and weaknesses of the constituent models and eventually exhibit higher generalisation performance. This feature can be particularly advantageous especially in dynamic environments where the performance of the individual learners (data-driven algorithms) is likely to change over time [58]. What is more is that ensemble models are likely to generalise better than single models as the currently available datasets may provide insufficient information for selecting a single best ML algorithm [15].

Ensemble learning methods can be classified into homogeneous and heterogeneous based on the base model selection. Specifically, homogeneous ensemble models use the same data-driven algorithm by considering different distributions of the training set (e.g., bagging and boosting). On the other hand, heterogeneous ensemble learning methods involve the development of different data-driven models (base-learners) that are trained by using the same dataset. Heterogeneous ensemble models alleviate the necessity for pre-selecting a suitable data-driven model, as no algorithm outperforms the other in predicting building energy use for all cases. Moreover, they are less likely to exhibit poor prediction performance as it is improbable that all the base models show low accuracy at the same time [15]. Although ensemble learning has been used in the literature for building energy consumption prediction [59–61], to date this framework has not been applied for evaluating building energy flexibility potential.

Existing research efforts into building flexibility evaluation considering market-independent frameworks are predominantly based on either physics-based or grey-box approaches without further investigating the performance of their methodologies under dynamic environments — such as those resulting from DR events [62]. Machine learning-based methodologies that evaluate the flexibility potential of individual residential buildings have mainly focused on a single energy system type without assessing building flexibility from an integrated system perspective or without analysing associated flexibility arising from later effects (i.e., rebound effects). Consequently, it is necessary to develop a data-driven energy flexibility assessment framework that not only dynamically adapts to new data patterns but also captures the DR potential of multicomponent thermal and electrical systems based on multistep-ahead predictions. A machine learning-based methodology that quantifies the energy flexibility of residential buildings on a one-by-one basis from such an integrated energy systems perspective, has not been reported in the literature to date [62] and hence is the scope of this study. A data-driven flexibility evaluation framework that quantifies the DR potential of thermal and electrical systems precisely and consistently will allow electricity aggregators to assess different flexibility measures in a user-tailored manner. This will facilitate the quantitative comparison across the various flexibility options available in residential buildings not only within the end-user pre-qualification process but also in operational environments.

In the current paper, an end user-tailored methodology is developed to assess the energy flexibility of both thermal and electrical building energy systems commonly found in residential buildings by considering the load shifting potential of individual DR actions as well as their net energy cost in the context of onsite electricity generation. To this end, a series of target variables associated with harnessing building DR potential (HVAC system load, zone temperature, non-controllable loads, PV electricity generation) are predicted based on dynamic training and ensemble learning. Specifically, a series of data-driven models are trained and periodically updated based on a sliding window method in order for predictions to adapt to changing conditions (building thermal characteristics, weather, and occupancy changes). This feature is particularly important considering the high variability of residential load profiles that exhibit high variation compared to commercial building load profiles [63]. The prediction models are developed by using features that can be realistically collected by a residential energy management system (e.g., historical data from the target variables and weather prediction data) as well as statistical properties related to data from the previous prediction period. For each target variable, the input variable pool is constructed by considering the most relevant historical data (selected with respect to the autocorrelation functions of the training sets), and the remainder of the candidate variables; the final feature set is selected by using the Spearman correlation coefficient. Subsequently, a heterogeneous ensemble prediction model is built consisting of a series of base models, namely random forests

(RF), multilayer perceptron neural network (MLP), support vector machine (SVM), and extreme gradient boosting (XGB) for short and very short-term forecasting (namely day-ahead and hour-ahead predictions). Day-ahead predictions enable not only the energy shifting capability assessment of the various building energy systems, but also the quantification of the energy costs arising from the activation of this flexibility. On the other hand, hour-ahead predictions allow for more accurate estimations of the energy shifting capability as well as any resulting temperature deviations. The forecasted variables are used to assess the energy flexibility potential of passive thermal energy storage (TES) and electric energy storage systems on a common basis by also considering their possible interactions as well as any potential energy costs and thermal comfort deviations resulting from harnessing this flexibility. Specifically, a series of DR performance indicators are used, namely, storage capacity, storage efficiency, and self-consumption, to acquire the daily energy flexibility mappings; the latter can facilitate the quantitative comparison across different flexibility options available in residential buildings and the evaluation of various building energy flexibility types.

The available building energy flexibility potential depends not only on the applied DR signals but also on occupant behavioural aspects that are usually random in residential buildings [63]. Data-driven models can be attuned to incoming smart metre data and adapt to dynamic environments (DR applications) without manual intervention or prior knowledge about the system dynamics such as required in physics-based models. Although data-driven methodologies have been extensively applied in the context of building controls, a machine learning framework that explicitly quantifies and characterises residential building flexibility from an integrated system perspective on a one-by-one basis has not been established to date [62]. To address this knowledge gap, an ensemble machine learning framework is developed based on periodically updated prediction models and a sliding window method considering day-ahead and hour-ahead predictions. Finally, a set of flexibility indicators [25] are used to quantify the flexibility of both thermal and electrical systems commonly found in residential buildings concisely and consistently. The proposed data-driven energy flexibility quantification framework can be of interest to electricity aggregators by allowing them to assess various flexibility measure combinations, and ultimately evaluate or optimise building portfolios in an end-user tailored manner. This methodology can be used to optimally harness the flexibility of each electrical component and to shift peak demand consumption to off-peak periods or periods of excess onsite electricity generation. This feature can facilitate the penetration of renewable energy sources, and to alleviate the potential production and distribution capacity issues.

This paper is organised as follows: Section 2 describes the overall methodology including the feature section process (Sections 2.1–2.3), the development of the base models (Section 2.4) and the ensemble learning framework (Section 2.5) as well as the energy flexibility quantification and characterisation methodology (Section 2.7). The case study building as well as its energy systems are described in Section 3, while Section 4 comprises the obtained simulation results. Finally, the discussion of the results and the conclusions are provided in Sections 5 and 6, respectively.

2. Methodology

As shown in Fig. 1 the proposed methodology consists of the following steps: generation of the synthetic database, determination of the candidate features, the feature selection methodology, the development of the base models and the ensemble learning framework, and finally, the DR potential evaluation for both day-ahead and hour-ahead predictions.

2.1. Training and test set determination

The training and testing sets are determined based on the sliding window horizon technique [64]: for each training process, a fixed number of previous observations are considered as model inputs while the output is the predicted variables of the time series. Specifically, when new data is collected and added to the most recent dataset, the oldest data is removed so that the training set size remains constant. Given the non-stationary nature of DR applications, this technique allows the models to be updated by discarding older and potentially less relevant observations. The corresponding prediction horizons (hour-ahead, day-ahead, etc.) depend on the DR market and scheme (time of use, real-time pricing, etc.), thus, the size of the test set is determined by the given prediction horizon.

2.2. Feature engineering

In this section, the candidate features for each target variable are described. Historical data could significantly increase the size of the training sets with redundant features thus rendering the prediction model computationally expensive. Herein, historical data are selected with respect to the top N most dominant lag terms based on the autocorrelation function (ACF) of the variables of interest. ACF is the linear dependence of a variable with itself at two points in time and it is used to identify repeating patterns and, ultimately, select suitable input variables. Since the training set is periodically updated, the autocorrelation – and thus the most dominant lag terms – changes over time. In this way, lag terms exhibiting low autocorrelation magnitude are filtered to mitigate the computational burden for the feature selection algorithm.

2.2.1. HVAC system load and zone temperature

The HVAC system load and the zone temperature are influenced not only by prevailing weather conditions but also by occupancy. Variables associated with calendar information (e.g., hour of the day, day of the week) as well as historical data, and non-controllable load predictions are potentially highly correlated to the target variables due to the periodicity of building occupancy patterns and, by association, energy consumption. Moreover, lag terms of the ambient temperature are also considered as candidate features; the reason is that a trend of change in weather variables will occur with a certain delay in the internal building due to the thermal inertia of the building envelope and building furnishings. Another paramount feature for DR applications is the zone thermostatic setpoint. The thermostatic setpoint is utilised as input not only because it reflects occupant thermal comfort preferences, but also because it is the main driver of the heating system performance, and thus, the flexibility of the passive thermal energy storage. Specifically, the zone temperature setpoint can be suitably modulated not only to harness the flexibility of the structural thermal mass but also to investigate any thermal comfort deviations arising from its activation. Finally, a set of auxiliary variables are considered including statistical properties from the previous prediction figures (e.g., average, minimum, and maximum values) and ratios involving these variables. The auxiliary variables proposed can give insights into the shape and dynamics of the target variables and eventually, improve the forecasting performance of the data-driven model. The non-controllable loads and the zone temperature are used as both predicted variables and candidate input variables.

2.2.2. Non-controllable loads

In this study, all building loads excluding the HVAC system load are considered non-controllable loads. For this load category, only historical data of the target variable are considered as well as calendar information and several auxiliary variables related to the shape and dynamics of the target variable, for instance, statistical properties from the previous prediction figures as well as ratios related to these variables.

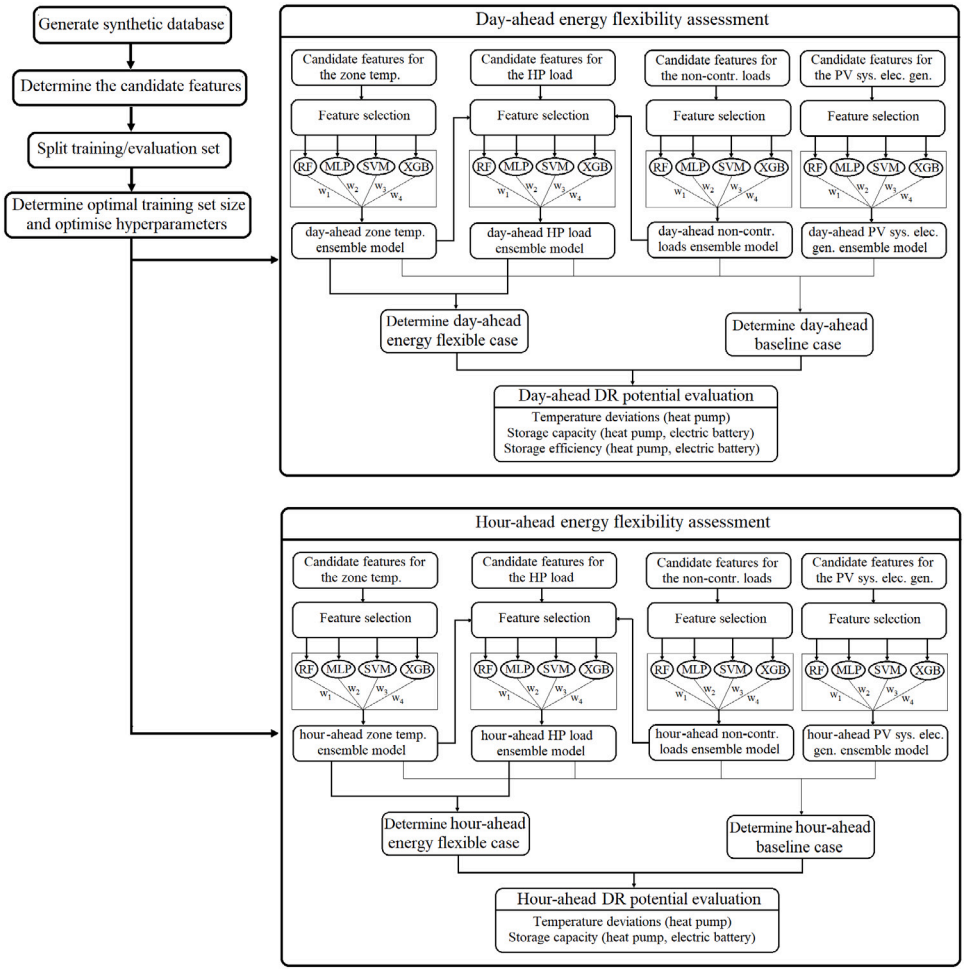


Fig. 1. Flowchart of the proposed day-ahead and-hour ahead energy flexibility assessment framework.

Table 1
Candidate features for each target variable.

	Target variables			
Candidate Features	HVAC system load	Zone temp.	Non-contr. loads	PV power
Weather variables	T_{out}, I_{tot}, RH, WS	T_{out}, I_{tot}, RH, WS	–	T_{out}, I_{tot}, RH
Calendar information	WT, DoW, MoD	WT, DoW, MoD	WT, DoW, MoD	MoD
Historical data	P_{hp} , T_z, T_{out}, T_{sp}, $P_{hp,O/O}$	P_{hp} , T_z, T_{out}, T_{sp}, $P_{hp,O/O}$	P_{nc}	P_{pv}
Auxiliary variables	$P_{hp,av}$, $P_{hp,max}$, $P_{hp,min}$, $R_{hp,av/max}$, $R_{hp,min/av}$, $T_{z,av}$, $T_{z,max}$, $T_{z,min}$, $R_{z,av/max}$, $R_{z,min/av}$	$P_{hp,av}$, $P_{hp,max}$, $P_{hp,min}$, $R_{hp,av/max}$, $R_{hp,min/av}$, $T_{z,av}$, $T_{z,max}$, $T_{z,min}$, $R_{z,av/max}$, $R_{z,min/av}$	$P_{nc,av}$, $P_{nc,max}$, $P_{nc,min}$, $R_{nc,av/max}$, $R_{nc,min/av}$	–
Miscellaneous	T_{sp}, P_{nc}, T_z	T_{sp}, P_{nc}	–	–

2.2.3. PV system

The electricity generation of PV plants is predominantly influenced by solar irradiation with other variables like outdoor temperature and humidity also affecting their performance [65]. To account for the PV system dynamics and for potential degradations in the PV system efficiency historical data are also included. Given that the PV output changes with time in a periodic manner, an additional proxy variable related to the time of day is also considered. Table 1 summarises the candidate input variables considered for each target variable.

2.3. Feature selection

The identification of the most important variables influencing the energy consumption patterns is an important part of the prediction process which can improve the model performance in terms of prediction accuracy and computational complexity [66]. For every new prediction, the training set is determined, and the optimal feature set is selected by considering the currently available training set and target variable values. In this study, the optimal number of features is selected by applying a correlation analysis.

The strength and the direction of the correlation between the predictor and the target variables can be obtained by using the Pearson or

the Spearman correlation coefficient. The Pearson coefficient measures the linear relationship between two continuous variables by assuming normal data distribution, whereas the Spearman coefficient is suitable for data that do not follow a normal distribution, and it captures the non-linear relationship between two variables. The correlation coefficient ranges from −1 to 1; values near 1 or −1 indicate highly positive or negative correlation, respectively, while 0 represents no correlation. Nevertheless, since the feature set also includes data following a non-normal distribution (e.g., relative humidity, solar radiation, wind speed), the features are selected based only on the Spearman coefficient [67]. In this analysis, candidate features are used as inputs to the predictive models when the associated correlation is above the threshold value of 0.5, based on the guide of the absolute values, as per [68].

2.4. Base model development

For each target variable (zone temperature, HVAC system load, non-controllable loads, PV panel electricity generation) a base model is trained. When the zone temperature and non-controllable load predictions become available, they are incorporated in the HVAC system training set as inputs. In this study, four regression models are used, namely random forests (RF), a support vector machine (SVM), a multilayer perceptron (MLP), and extreme gradient boosting (XGB). These models were chosen as they have demonstrated increased predictive performance compared to other regression techniques, especially for building load forecasting [30,59].

2.4.1. Regression algorithms

The random forests model builds a series of de-correlated trees and averages them. Each tree is built on a subsample from the training data. To enhance the diversity, the number of variables used to fit the data in each tree is randomly selected from the total number of variables. The correlation of trees will be eventually lessened since the algorithm randomly selects predictors at each split [69]. Support vector machines can be computationally advantageous in high dimensional feature spaces since they only depend on an input data subset. This is because they minimise a cost function that is less sensitive to points within a certain distance from the prediction. SVM can handle nonlinear input–output relationships by using a kernel function [70]. Multi-layer perceptron (MLP) is a feed-forward artificial neural network that can capture nonlinear relationships between input and output by using a nonlinear activation function. It normally consists of several nodes organised in several layers. The weights of each node are adaptively adjusted using the back-propagation technique [71]. The extreme gradient boosting is a gradient boosted decision tree where the final prediction is made by creating new models that predict the errors of previous models. First, the regression performance is measured, and the associated feature importance is ranked with respect to the relative contribution of each feature to the model. Secondly, features are recursively eliminated from the least important to the most important ones and the model is built with the remaining features until the optimal number of features is selected [72].

2.4.2. Hyperparameter optimisation

To achieve optimal performance for the selected ML algorithms, it is necessary to determine their configuration by optimising their hyperparameters. In this study, the model hyperparameters are tuned based on a random search. This technique applies a randomised search across hyperparameters over all possible parameter values. The method is performed iteratively until the desired accuracy is reached or the predetermined computational budget is exhausted [73].

One of the widely implemented cross-validation (CV) strategies for supervised regression is the k-fold method in which the dataset is randomly split into k folds, where k-1 of them are used to train the model, whereas the remaining fold is used to test its performance. This problem can be addressed by applying a time series CV model on a rolling basis. This signifies that the CV method used should be customised to the training strategy; the training and test sets slide forward in steps equal to the test set size.

2.4.3. Data pre-processing

Considering the different magnitude scales of the various variables involved, features with higher values are likely to be more influential on the model development, thus degrading their accuracy [74]. To tackle this problem, all variables are normalised within the range of [0,1]. Predicted values below 0 or above 1 are considered prediction errors, and thus they are set equal to 0 and 1, respectively, before they are denormalised.

2.5. Ensemble learning framework

An ensemble model can be defined as a strategy that considers a series of data-driven models to attain better predictive accuracy than that obtained by any of the constituent base models. Ensemble models can be categorised into homogeneous and heterogeneous based on the base model selection. The homogeneous model uses the same base learner by using different distributions of the training set (e.g., bagging and boosting), whereas the heterogeneous ensemble model consists of different ML models that are developed by using the same data set. Heterogeneous ML models can benefit from the irrelevance of the constituent models (each characterised by different assumptions) and eventually enhance the predictive performance. Moreover, it is not necessary to predetermine the optimal ML model since the final prediction results from updating combination schemes [57]. The ensemble model assigns a different weight to each base model based on its forecast accuracy, thus, the ML model that attains the best accuracy receives the highest weight.

In this study, heterogeneous ensemble models are developed for all target variables of interest. The final prediction is based on an aggregation rule that combines the individual models considered (i.e., RF, MLP, SVM, and XGB) to identify the optimal combination of the constituent models. Considering a prediction horizon of M timesteps, the ensemble model predicts the value of the target variable by combining the N base models. At the ith algorithm iteration, each base model will return a vector $x_{n,m}^{(i)} = \{x_{n,1}^{(i)}, x_{n,2}^{(i)}, \dots, x_{n,M}^{(i)}\}, \mathrm{n} \in \mathrm{N}$. The final forecasted vector is a linear combination of each base model prediction and is given by Eqs. (1) and (2).

$$\hat{Y}^{(i)} = \{\hat{y}_1^{(i)}, \hat{y}_2^{(i)} \dots \hat{y}_M^{(i)}\} \tag{1}$$

$$\hat{y}_m^{(i)} = \sum_{n=1}^{N} w_n^{(i)} x_{n,m}^{(i)} \tag{2}$$

Where $w_n^{(i)}$ is the weight of the nth base model at the ith iteration. When the actual values $Y^{(i)} = \{y_1^{(i)}, y_2^{(i)} \dots y_M^{(i)}\}$ become available (at the end of the prediction horizon), a convex loss function is used to quantify each base model performance. In this study, the considered loss function is the squared error difference ($l_n^{(i)} = \sum_{m=1}^{M} (y_{m,n}^{(i)} - x_{m,n}^{(i)})^2$), however, this methodology can be extended for various convex loss functions [75].

Various types of aggregation rules have been developed in the literature, including inter alia: the exponentially weighted average forecaster, the fixed share forecaster, the ridge regression forecaster, etc. [76]. The exponentially weighted average forecaster [77] adopted in this study is an online convex aggregation rule that assigns weights to the various base models considered with a view to minimising the ensemble model regret. The regret is defined as the difference between the ensemble model accumulated loss and that of the base model n over the ith iteration. Initially, it is assumed that each base model achieves equally good performance, thus each model receives the same

weight $w_n^{(0)} = 1/N$. For i>0, the weight of the nth base model at the ith iteration is given by Eq. (3).

$$w_n^{(i)} = \exp(-\eta \sum_{s=0}^{i-1} l_n^{(s)}) / \sum_{k=1}^{N} \exp(-\eta \sum_{s=0}^{i-1} l_k^{(s)}) \quad (3)$$

Where η is a learning parameter that should be tuned to achieve a small average regret with respect to the best base model [76].

2.6. ML model performance metrics

The predictive performance of the developed ML models is evaluated by utilising the coefficient of variation of the root mean squared error (CV-RMSE) (Eq. (2)) and the normalised mean bias error (NMBE) (Eq. (3)) for the power-related target variables (HVAC system power consumption, non-controllable loads power consumption, and PV system electricity generation). For the zone temperature prediction model, the RMSE (Eq. (4)) and the MBE (Eq. (5)) are used. The NMBE and the CV-RMSE have been recommended as evaluation criteria for building energy prediction models by ASHRAE guidelines [78]. As regards the NMBE, ASHRAE establishes a limit of +/-5% and +/-10% for calibration using monthly and hourly data, respectively. Regarding the CV-RMSE, the acceptable tolerances are 15% and 30% for monthly and hourly data, respectively. There are nonetheless no standardised calibration thresholds for models with higher resolution [79]. Finally, the predictive performance of the developed ML models as to the energy flexibility related characteristics is evaluated by utilising the coefficient of determination (Eq. (8)).

$$CV-RMSE = \frac{\sqrt{\sum_{i=1}^{N} (\hat{Y}_i - Y_i)^2 / N}}{\sum_{i=1}^{N} Y_i / N} \quad (4)$$

$$NMBE = \frac{\sum_{i=1}^{N} (\hat{Y}_i - Y_i)}{\sum_{i=1}^{N} Y_i} \quad (5)$$

$$RMSE = \sqrt{\sum_{i=1}^{N} (\hat{Y}_i - Y_i)^2 / N} \quad (6)$$

$$MBE = \sum_{i=1}^{N} (\hat{Y}_i - Y_i) / N \quad (7)$$

$$R^2 = \frac{\left[\sum_{i=1}^{N} \left(\hat{Y}_i - \sum_{i=1}^{N} \hat{Y}_i/N\right) \cdot \left(Y_i - \sum_{i=1}^{N} Y_i/N\right)\right]^2}{\sum_{i=1}^{i=N} \left(\hat{Y}_i - \sum_{i=1}^{N} \hat{Y}_i/N\right) \cdot \sum_{i=1}^{i=N} \left(Y_i - \sum_{i=1}^{N} Y_i/N\right)} \quad (8)$$

Where $\hat{Y}_i$ and Y_i stand for the predicted and actual value i, respectively, and N represents the size of Y_i.

2.7. Quantification and characterisation of building energy flexibility

To assess the energy flexibility potential, the load shifting capability of various DR actions in the context of locally produced electricity is quantified for various building electrical components, namely HVAC systems and electric batteries. The flexibility of these systems can be activated by considering two energy flexibility types: downward (down-flex) and upward flexibility (up-flex) [43]. In down-flex, energy is curtailed during the DR action, whereas in up-flex the energy can be stored in the considered storage medium by suitably modulating the control setpoint. In each case, the control setpoint is restored after the DR action so that the storage medium returns to the previous state before the DR action. The actual cost of DR actions can be reduced by the corresponding onsite electricity generation amount — if it is available. The flexibility indicators used in this study for downward and upward flexibility actions are as follows: storage capacity and storage efficiency.

Table 2
Limits on Temperature Drifts and Ramps [80].

Time Period	0.5 h	1 h	2 h
Maximum allowed operative temperature change	1.1 °C	2.2 °C	2.8 °C

2.7.1. Storage capacity

The available storage capacity (C_{DR}) is defined as the energy amount that can be added to or removed from a storage medium in the context of a DR event considering the boundary conditions. The storage capacity is given by Eqs. (9) and (10) for down-flex and up-flex actions, respectively.

$$C_{DF} = \int_0^{\infty} \left| (P_{\mathrm{mod}} - \mathrm{P}_{ref})^- \right| dt \quad (9)$$

$$C_{UF} = \int_0^{\infty} (P_{\mathrm{mod}} - \mathrm{P}_{ref})^+ dt \quad (10)$$

Where the "+ " and "-" superscripts are interpreted as follows: $x+ = max(x,0)$, $x- = max(-x,0)$, and P_{mod} and P_{ref} represent the total modulated and reference building load, respectively, while P_{RES} stands for the onsite electricity generation.

The evaluation of the HVAC system flexibility is harnessed based on on/off strategies for a duration of t_{dr}. To limit thermal comfort deviations, the zone temperature should be kept within an acceptable range as per ASHRAE standards [80]. The limits on temperature drifts and ramps are summarised in Table 2.

The electrical energy storage flexibility quantification depends on the underlying control algorithms. Given that battery charging and discharging power levels can be easily controlled by inverter-based systems, only down-flex actions are considered for electrical energy storage units. In this study, it is assumed that the maximum flexibility potential of a DR action is determined by considering zero electricity flows to the mains. Thus, the electrical storage system DR potential is exploited by considering a discharging power equal to the overall building load, P_b reduced by the onsite electricity generation P_{RES} for a certain duration (t_{dr}). In this case, the available storage capacity is given by Eq. (11).

$$C_{DF} = \int_0^{t_{DR}} \max(P_b - P_{RES}, 0) dt \quad (11)$$

2.7.2. Storage efficiency

The exploitation of the energy flexibility of an energy system results in a different power level after the end of the DR event. For example, in downward flexibility, the demand is expected to increase, in upward flexibility, this demand is expected to decrease. The interpretation of storage efficiency nonetheless varies depending on the type of energy flexibility considered. For instance, in down-flex, storage efficiency can be defined as the fraction of the energy cost of an active DR event with respect to the achieved energy curtailment. On the other hand, in up-flex, storage efficiency can be defined as the fraction of the energy consumed during the DR action and can be subsequently utilised to curtail the power required in order for the energy system to return to the state prior to the DR action. The storage efficiency associated with down-flex and up-flex scenarios are calculated as per equations (12) and (13), respectively.

$$\eta_{DF} = 1 - \frac{\int_0^{\infty} \max(P_{\mathrm{mod}} - \max(\mathrm{P}_{ref}, P_{RES}), 0) dt}{\int_0^{\infty} \left| (P_{\mathrm{mod}} - \mathrm{P}_{ref})^- \right| dt} \quad (12)$$

$$\eta_{UF} = \frac{\int_0^{\infty} \left| (P_{\mathrm{mod}} - P_{ref})^- \right| dt}{\int_0^{\infty} \left| (P_{\mathrm{mod}} - P_{ref})^- \right| dt + \int_0^{\infty} \max(P_{\mathrm{mod}} - \max(\mathrm{P}_{ref}, P_{RES})) dt} \quad (13)$$

The maximum flexibility potential of a battery during a DR action can be determined by considering zero power flow to the utility. This

Fig. 2. Picture and 3D rendering of testbed house [81].

means that the battery discharging power is equal to net building power demand. When local RES generation is higher than the building load, the battery remains inactive. Since battery loads can be controlled by inverter-based systems, the energy consumed or generated by a battery can be determined by the associated control algorithm. In contrast to thermal energy storage systems, in electrical energy storage systems, the amount of energy required to restore the energy spent for covering the building net energy demand during a DR action depends on the charging and discharging battery efficiencies as well as the reference building load and onsite electricity generation. This means that the rebound resulting from harnessing the DR potential of a battery can occur at any time without affecting the battery performance. Consequently, only down-flex actions are considered for the electric battery.

Considering that t_c and τ_{id} are the rebound starting time and total time of increased demand, respectively, Eq. (12) can be customised to calculate the battery storage efficiency:

$$\eta_{DF} = 1 - \max\left(\frac{1}{\eta_c \eta_d} + \frac{1}{C_{DF}} \int_{t_c}^{t_c+\tau_{id}} P_b - \max(P_b, P_{RES})dt, 0\right) \quad (14)$$

where η_d and η_c stand for the battery average discharging and charging efficiencies. Further discussion on the adopted indicators can be found in [25].

3. Case study

In this section, the virtual testbed and the associated components are described along with the adopted occupancy profiles. Further, the implementation and configuration of each ML used is described.

3.1. Building description and energy conversion components

The flexibility of the various building energy systems is evaluated by using a synthetic database that is generated from a calibrated white-box model of a residential building. Calibrated physics-based models not only exhibit realistic performance but also allow model development without accounting for data quality issues associated with data collection from buildings. The building model was developed by using EnergyPlus V.9.1 and calibrated based on measurements from an existing all-electric house located in eastern Ireland [25]. Fig. 2 illustrates the case study building and its modelled geometry. The thermal envelope, the HVAC system, and the PV system were calibrated with respect to the ASHRAE guidelines [78].

3.1.1. Thermal and geometrical properties

The single storey detached bungalow house that is utilised as the virtual testbed is the most common single building category representing approximately 40% of the Irish building stock [80]. The building opaque elements demonstrate increased insulation specifications compared to contemporary standards. As shown in Fig. 3, the building has twelve rooms and an uninhabited attic space at roof level. The U-values of the building floor, walls, windows, and roof are 0.21, 0.21, 1.7, and 0.21 W/m^{2}K, respectively. Further, because of its construction (i.e., two-leaf concrete wall with cavity insulation), it also demonstrates passive thermal energy storage capacity. The exterior wall total surface area (excluding fenestration) is 187 m^2, while the slate roof surface area is 279 m^2. Moreover, the roof does not include insulation and acoustic tiles which cover the ceiling to ensure both acoustic and thermal insulation. A 200 mm layer of fibreglass insulation with a thermal conductivity of 0.04 W/mK is located on top of the acoustic tiles and ensures high thermal resistance. Finally, the floor area is 208 m^2, and the overall window to wall ratio is 15%, with a 22% and 10% ratio on the south and north facades, respectively. The thermal envelope calibration was carried out during occupant absence achieving NMBE and the CV-RMSE average values below the hourly ASHRAE standards (4.41% and 3.28%, respectively) [82].

3.1.2. HVAC system

The space heating system of the dwelling (depicted in Fig. 4) is a ground source heat pump (GSHP) and has a rated thermal output of 12 kW. The GSHP is equipped with a hot water storage tank of 0.8 m^3 to provide thermal energy storage. The calibration was performed by considering data from the heating season and attained average CV-RMSE and NMBE values of 3.78% and −0.61%, respectively [81].

3.1.3. PV system

The installed PV system has a nominal power of 6 kWp and it is connected to the grid through a single-phase inverter with an efficiency of 95%. The PV system is located 30 metres from the building and it has a southerly orientation. The associated model was calibrated. The calibration was performed with respect to data for the months February through September and attained average CV-RMSE and NMBE values of 12.5%, and 3.6%, respectively [81].

3.1.4. Stationary battery

The building under study does not include an electrical energy storage system. However, a battery with a capacity of 12 kWh, a maximum power of 8 kW, and an efficiency of 90% was considered in the analysis for completeness.

3.2. Occupancy profiles

Two indicative occupancy profiles are adopted based on daily average occupancy profiles resulting from a Time Use Survey [83]. These occupancy profiles utilise two clusters that result from categorising residential weekday diaries and constitute 23% and 34% of the survey sample [84]. With regard to weekday schedules, the first occupancy profile (OC1) includes an absence from the building arising from work attendance (between 08:20 h and 18:10 h) and considers occupant

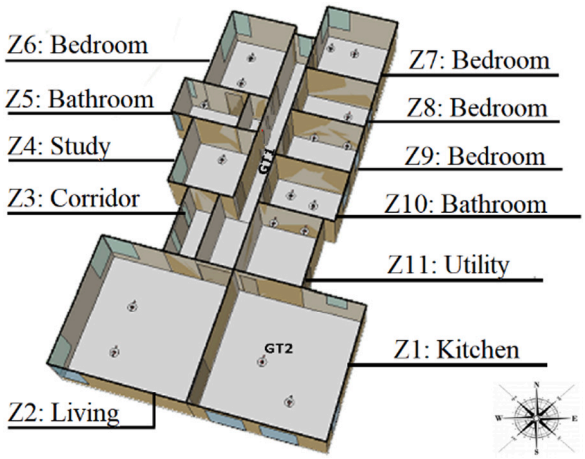


Fig. 3. Representation of the building with the ground floor thermal zones and orientation [81].

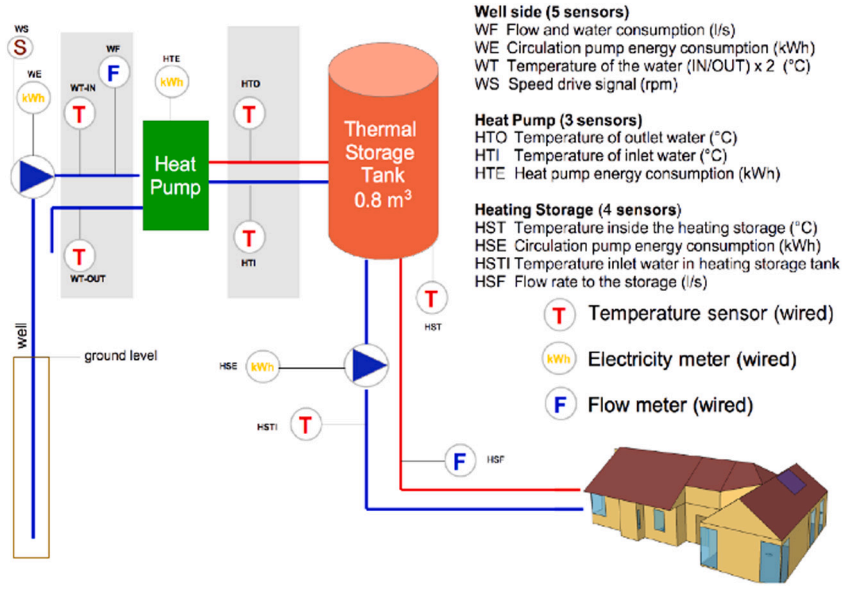


Fig. 4. Heat system design and sensor metring [82].

Table 3
Thermostatic setpoints for OC1 and OC2.

Weekdays				Weekend	
OC1		*OC2*		*OC1 and OC2*	
Time slots	Therm. Setpoint	Time slots	Therm. Setpoint	Time slots	Therm. Setpoint
23:10–06:40 06:40–08:20	16 20	07:10–23:20	21	05:40–01:10	21
08:20–18:10 18:10–23:10	16 21	23:20–07:10	16	01:10–05:40	16

activity according to three categorisations, i.e., active, non-active, and absent. The second occupancy profile (OC2) is comprised of active and non-active household states. Finally, both occupancy profiles share the same occupant activity during weekends. The zone thermostat setpoint is set to 20/21°C during occupant activity periods, whereas a setback of 16°C is considered during periods of occupant absence or inactivity. Table 3 summarises the adopted thermostatic setpoints.

3.3. Model setting

The synthetic database was generated by using data based on the heating season (extending from 01 September to 30 April) with a simulation time-step of 15 min. The selected splitting ratio of the database for the development and the evaluation set is 50%/50% [85]. The development set is utilised to determine the optimal number of training days for each ML model and to optimise the respective

Table 4
Methodology diagram.

Step	Section	Methodology			
1	3.3	Generate a synthetic database from a calibrated all-electrical smart grid ready residential building model			
2	2.1	Determine the optimal training set size based on the sliding window method			
3	2.2	Determine the candidate features (Table 1)			
4	2.3	Update the training set and the selected features based on the sliding window method and the Spearman correlation coefficient, respectively			
5	2.4	Build a base model based on the ML algorithms considered (RF, MLP, SVM, XGB) for each target variable (HVAC system load, zone temperature, non-controllable loads, PV system electricity generation) and prediction horizon (day-ahead, hour ahead)			
6	2.5	Develop an ensemble learning forecaster considering the base models build in step 5 for each target variable and prediction horizon considered			
7	2.6	Assess the predictive performance of the ensemble learning forecaster for each target variable and prediction horizon considered			
8	3.2	Select an indicative day from each occupancy profile considered			
9	3.4	Impose hourly independent DR actions both for down-flex and up-flex subject to the respective boundary conditions for all energy systems considered (heat pump, electrical energy storage)			
		Commencement times: 00:00,01:00,. . . 23:00			
10	2.7	Evaluate the flexibility potential by using the following indicators:			
			Heat pump		Electrical energy storage
			Down-flex	Up-flex	Down-flex
		Storage Capacity	Eq. 9	Eq. 10	Eq. 11
		Storage Efficiency	Eq. 12	Eq. 13	Eq. 14
11	3.4	Create the daily flexibility profile for all cases in Step 8 by using the indicators in Step 9			
12	2.6	Assess the predictability of the flexibility potential of the various energy systems considered			

hyperparameters; the evaluation set is used to evaluate the performance of the developed models in terms of predictive accuracy based on cross-validation. In order to evaluate the DR potential of the virtual testbed, daily random DR actions are introduced in the evaluation set. The proposed methodology is implemented by using Python 3.7 on a server machine with an Intel i7-7500 CPU 2.7 GHz and 8 GB of RAM. The developed data-driven models were implemented by using the scikit-learn library [86].

3.4. Energy flexibility mapping

To assess the building energy flexibility potential, one day from each occupancy profile is selected as a baseline case to create the daily energy flexibility mapping. The baseline cases were selected so that the ensemble model predictive performance (in terms of RMSE/CV-RMSE and MBE, NMBE) for each baseline case (for all target variables and prediction horizons) is comparable to that of the ensemble models for the evaluation set. To this end, consecutive and independent DR events are imposed, over a 24-hour period, for each energy system considered. Specifically, the DR action starting times were based on an on-the-hour basis, i.e., 00:00 hr, 01:00 hr, etc. For each case study, it is considered that there is a single baseline consumption curve, all applied DR actions are independent, and any specific DR action is not affected by the previous ones. In order to practically utilise these indicators, it is assumed that the period between DR actions is long enough (so that the rebounds of the previous DR actions do not affect the following ones). Thus, the heatmaps presented assume 24 different and independent DR actions each of duration of one hour, where it is assumed that only a single DR action occurs each day. The resulting energy flexibility profiles can be used to select the most suitable DR strategies – independently or in combination – considering the requested energy amount to be shifted. The methodological steps of this study are outlined in Table 4.

4. Results

In this section, the most important features are identified and subsequently the base and the ensemble learning models are developed for all target variables considered. Finally, the downward and upward flexibility potential of the various building energy systems is assessed by using the developed ensemble learning-based framework, and referenced against the physics-based model.

4.1. Input variable selection

The candidate input variables (as summarised in Table 1) for each target variable and prediction horizon (day-ahead and one hour-ahead predictions) are evaluated by using the Spearman correlation coefficient. For every short-term prediction, the selected features are updated by considering the sliding window horizon method. This section focuses on the HVAC system-related target variables (zone temperature, heating load) since they involve all types of features (i.e., calendar information, lag terms, weather variables, etc.). As a case study, in this section the candidate features for each target variable are evaluated by considering datasets that exclusively include observations from a single occupancy profile to identify potential feature selection differences arising from different data patterns. The autocorrelation function of these datasets is depicted in Fig. 5 for the zone temperature (Figs. 5a and 5b) and the heating load (Figs. 5c and 5d). This analysis aims to identify potential feature selection differences arising from various data patterns. The most recent historical datasets are always the most relevant ones for the zone temperature predictions regardless of occupancy profiles (Figs. 5a and 5b). On the other hand, different occupancy profiles result in different autocorrelation coefficients for the heating system power consumption (Figs. 5c and 5d). For example, OC1 exhibits weekly periodicity because it includes two distinct occupancy profiles (weekday and weekend), whereas OC2 shows daily periodicity since the weekday and weekend occupancy profiles are approximately similar.

Table 5
Spearman correlation results, zone temperature, day ahead predictions.

Definition	Symbol	OC1	OC2
Zone thermostatic setpoint	T_{sp}	x	x
	$T_{sp}(t-96)$	–	x
Zone temperature	$T_z(t-96) - T_z(t-101)$	x	x
	$T_z(t-102)$	–	x
	$T_z(t-103)$	–	x
	$T_z(t-104)$	–	x
Heat pump on/off operation	$P_{hp,O/O}$	x	x
Workday type	WT	x	–

Table 6
Spearman correlation results, heating load, day ahead predictions.

Definition	Symbol	OC1	OC2
Heat pump load	$P_{hp}(t-96) - P_{hp}(t-98)$	x	x
	$P_{hp}(t-99) - P_{hp}(t-100)$	–	x
	$P_{hp}(t-187) - P_{hp}(t-193)$	–	x
	$P_{hp}(t-285) - P_{hp}(t-289)$	–	x
	$P_{hp}(t-383)$	–	x
	$P_{hp}(t-574) - P_{hp}(t-577)$	x	–
	$P_{hp}(t-666) - P_{hp}(t-675)$	x	–
	$P_{hp}(t-766) - P_{hp}(t-767)$	x	–
Zone thermostatic setpoint	T_{sp}	x	x
	$T_{sp}(t-96)$	x	x
Heat pump on/off operation	$P_{hp,O/O}$	x	x
Zone temperature	T_z	x	x

4.1.1. Day-head prediction models

Table 5 shows the selected features for the indoor temperature and the two occupancy profiles considered. For each occupancy profile, a different number of lag terms exhibits high correlation (> 0.5), however, the thermostatic setpoint, and the heating system operation times from the previous day are selected for both occupancy profiles. Nevertheless, the thermostatic setpoints of the previous day exhibit high correlation only for OC2, whereas the binary variable DoW is a relevant feature only for OC1. This is because the occupancy schedules for weekdays and weekends are similar for OC2 thus making DoW a less significant input variable.

Table 6 illustrates the selected features for the heating load and the two occupancy profiles considered. As regards OC1, lag terms associated with the heating load from the previous day as well as six and seven previous days are selected. On the other hand, the lag terms selected for OC2 are associated with data up to the previous four days. This stems from the fact that OC1 exhibits a weekly periodicity (because the weekday and weekend occupancy profiles are distinctly different) and OC2 shows daily periodicity (because the weekday and weekend occupancy profiles are approximately the same).

4.1.2. Hour-ahead prediction models

Table 7 summarises the selected features for the zone temperature and the two occupancy profiles considered. In contrast to day-ahead predictions, statistical properties from the previous hour are highly correlated with the values from the next hour for both occupancy profiles. It should be noted that statistical properties of the heat pump power exhibit a high correlation with the zone temperature only for OC1. Table 8 shows the selected features for the heating load and the two occupancy profiles considered. As regards OC1, lag terms associated with the heating load from the previous day as well as six to eight previous days are selected. On the other hand, the lag terms selected for OC2 are associated with data up to the previous three days. In contrast to day-ahead predictions, statistical properties from the previous predictions (previous hour) are also selected.

Table 7
Spearman correlation results, zone temperature, hour ahead predictions.

Definition	Symbol	OC1	OC2
Zone thermostatic setpoint	T_{sp}	x	x
	$T_{sp}(t-4)$	x	x
Zone temperature	$T_z(t-4) - T_z(t-11)$	x	x
	$T_z(t-12)$	–	x
Maximum zone temperature	$T_{z,max}$	x	x
Average maximum zone temperature	$T_{z,av}$	x	x
Minimum zone temperature	$T_{z,min}$	x	x

Table 8
Spearman correlation results, heating load, hour ahead predictions.

Definition	Symbol	OC1	OC2
Zone thermostatic setpoint	T_{sp}	x	x
	$T_{sp}(t-4)$	x	x
Zone temperature	T_z	x	x
Heat pump load	$P_{hp}(t-4)$	x	x
	$P_{hp}(t-89) - P_{hp}(t-92)$	–	x
	$P_{hp}(t-93) - P_{hp}(t-97)$	x	x
	$P_{hp}(t-98) - P_{hp}(t-99)$	–	x
	$P_{hp}(t-186) - P_{hp}(t-193)$	–	x
	$P_{hp}(t-284) - P_{hp}(t-287)$	–	x
	$P_{hp}(t-574) - P_{hp}(t-576)$	x	–
	$P_{hp}(t-666) - P_{hp}(t-675)$	x	–
	$P_{hp}(t-765) - P_{hp}(t-766)$	x	–
Average hp load	$P_{hp,av}$	x	x
Maximum hp load	$P_{hp,max}$	x	x
Minimum hp load	$P_{hp,min}$	x	x
Heat pump on/off operation	$P_{hp,O/O}$	x	x

4.2. Base model initialisation

To establish the ML models for the day-ahead and hourly predictions, the optimal training set is first determined. The power-related target variables (heating load, non-controllable loads, PV plant power output) were assessed based on the CV-RMSE and the NMBE whereas the zone temperature by using the RMSE and the MBE.

4.2.1. Day-ahead prediction models

The accuracy of each ML model was assessed for various training set sizes from 1–12 weeks with increments of 1 week. Fig. 6 illustrates the RMSE and CV-RMSE for each target variable and ML model with respect to the training set size. The presentation of the NMBE and MBE was omitted since the respective graphs exhibit similar trends to the CV-RMSE and the NMBE, respectively. Consequently, the optimal number of training weeks is selected with respect to the RMSE/CV-RMSE. Fig. 6 shows that the prediction error for each graph is reduced rapidly until a certain number of weeks and subsequently it either remains approximately constant or it increases. This means that training the models with larger datasets will increase the computational time resulting in similar or worse model accuracy. The training set size for each target variable and ML model is selected by identifying the above-mentioned cut-off or minimum point. The predictive performance of each target variable varies with the ML algorithm selected and the training set size. For example, as regards the zone temperature, all ML algorithms considered exhibit a similar accuracy after a certain number of training weeks. On the other hand, regarding the heating and the non-controllable loads, the RF and the XGB models show similar accuracy outperforming the SVR and the MLP models in respect to the heating load and the non-controllable load predictions for any training size. Finally, as regards, the PV electricity generation, the RF, the SVM, and the XGB exhibit similar accuracy when trained with only two weeks of data. The optimal number of training weeks is outlined in Table 9.

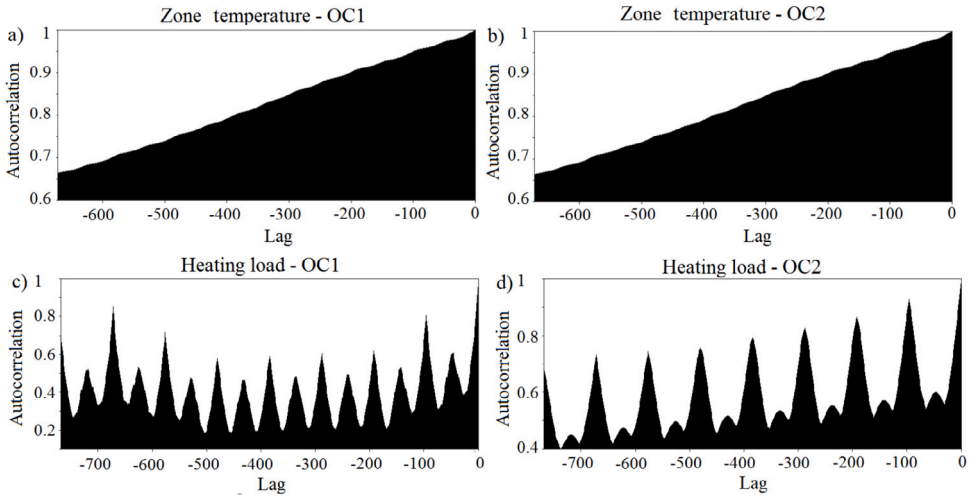


Fig. 5. Autocorrelation, (a) zone temperature (OC1), (b) zone temperature (OC2), (c) heating load (OC1), (d) heating load (OC2).

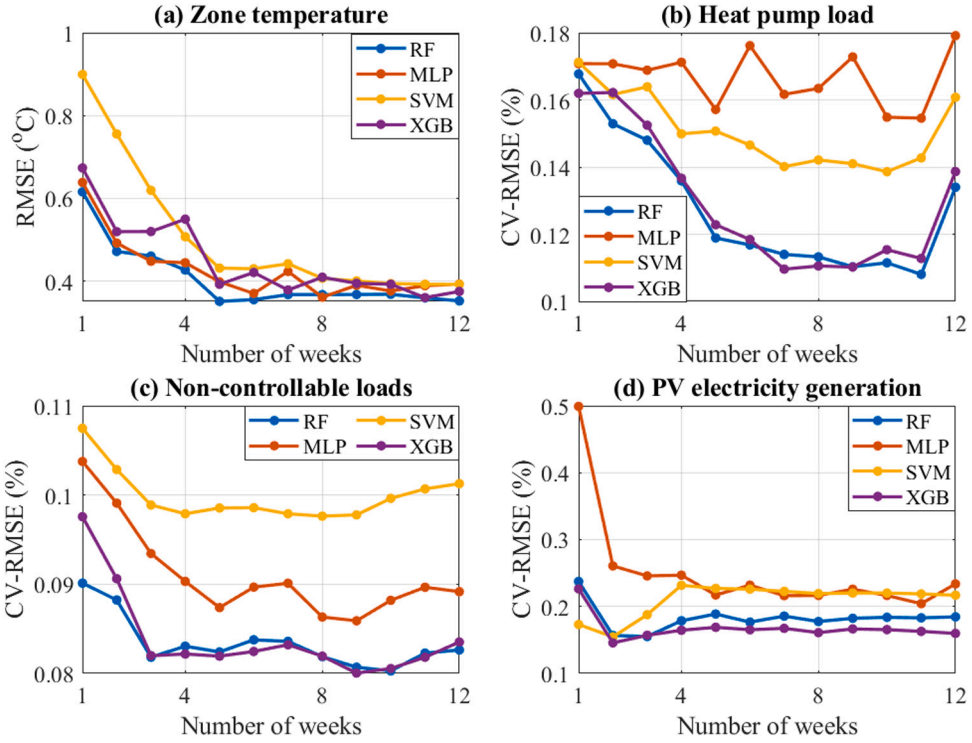


Fig. 6. RMSE and CV-RMSE for each ML model with respect to the training set size for day-ahead predictions, (a) zone temperature, (b) heat pump load, (c) non-controllable loads, (d) PV electricity generation.

Table 9
Selected training set sizes for each target variable and ML model, day-ahead prediction models.

ML model	Number of weeks			
	Zone temp.	Heat pump load	Non-contr. loads	PV system electr. gen.
RF	5	8	4	2
MLP	6	5	5	5
SVM	5	7	4	2
XGB	7	7	3	2

4.2.2. Hour-ahead prediction models

As regards the hourly predictions, the prediction performance for each data-driven model was assessed for various training set sizes from 1–20 days with increments of 1 day. Fig. 7 illustrates the RMSE and CV-RMSE for each target variable and ML model with respect to the training set size. As in day-ahead predictions, the accuracy of each ML algorithm varies with the ML algorithm selected and the training set size. The prediction error for each graph is reduced rapidly until a certain number of days and then it is reduced slowly after this point. Given that the MBE/NMBE graphs exhibit similar trends to the RMSE/CV-RMSE graphs, their presentation is omitted, and the optimal number of days is selected with respect to the RMSE/CV-RMSE. The optimal number of training days is defined by the above-mentioned

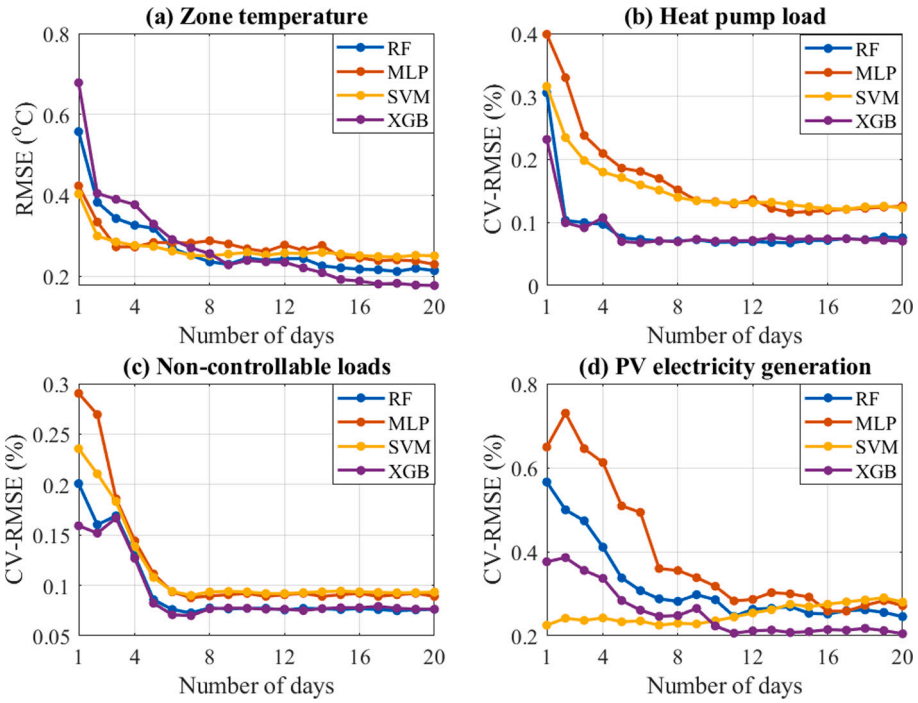


Fig. 7. RMSE and CV-RMSE for each ML model with respect to the training set size for hour-ahead predictions, (a) zone temperature, (b) heat pump load, (c) non-controllable loads, (d) PV electricity generation.

Table 10
Selected training set sizes for each target variable and ML model, hour-ahead prediction models.

ML model	Number of days			
	Zone temp.	Heat pump load	Non-contr. load	PV system elec. gener.
RF	9	5	6	11
MLP	3	14	7	11
SVM	7	9	7	7
XGB	15	5	6	11

point, as outlined in Table 10. The SVM model exhibits its optimal predictive performance for the heating load, the zone temperature, and the PV system electricity generation for smaller training set sizes. Furthermore, the XGB model outperforms the rest of the regression techniques considered for larger training set sizes.

4.3. Ensemble model deployment

The configuration of the individual base models is followed by the establishment of the final forecast by using the ensemble learning model for both day-ahead and hour-ahead predictions.

4.3.1. Day-ahead prediction models

Tables 11 and 12 summarise the RMSE/CV-RMSE and the MBE/NMBE, respectively, for each target variable and base model considered as well as the resulting ensemble model forecaster. The bar chart in Fig. 8 illustrates the RMSE/CV-RMSE (Fig. 8a) and the MBE/NMBE (Fig. 8b) for all target variables considered.

The RF and the XGB models outperform the MLP and the SVR models for all target variables considered both in terms of RMSE/CV-RMSE and MBE/NMBE. In addition, the ensemble models developed for the zone temperature, the heating load, and the non-controllable loads attain lower RMSE and CV-RMSE compared to that of the pertinent base models. As regards the PV system electricity generation, the CV-RMSE achieved by the ensemble model is 1.1% higher than that of the

Table 11
Comparison between various base models and the ensemble model forecaster for each target variable by RMSE(°C) (zone temperature) and CV-RMSE(%) (heat pump load, non-controllable loads, PV system electricity generation)- day-ahead prediction models.

	Zone temp. (°C)	Heat pump load (%)	Non-contr. loads (%)	PV system elec. gener. (%)
RF	0.331	14.8	8.2	16.7
MLP	0.413	19.2	8.9	32.1
SVM	0.414	17.1	10	28.6
XGB	0.339	14.3	8.3	18.5
Ensemble model	0.3097	14.3	8.1	17.8

Table 12
Comparison between various base models and the ensemble model forecaster for each target variable by MBE (%)(zone temperature) and NMBE (%)(heat pump load, non-controllable loads, PV system electricity generation)- day-ahead prediction models.

	Zone temp. (°C)	Heat pump load (%)	Non-contr. loads (%)	PV system elec. gener. (%)
RF	−0.0066	−1.75	−1.9	2.75
MLP	0.0563	−1.13	6.7	10.92
SVM	0.047	−2.38	−7.5	−7.22
XGB	0.0503	−0.56	1.3	3.19
Ensemble model	0.0268	−1.37	−0.5	−1.96

RF model (16.7%) but considerably lower than the CV-RMSE values attained by the rest of the base models. The MBE and NMBE achieved by the ensemble models lie within acceptable limits (as per the ASHRAE criteria) even when the associated base models exhibit poorer performance. For example, for the PV system electricity generation, although the SVM and the MLP models exhibit non-acceptable NMBE (−7.22% and 10.92%), the ensemble model achieves an NMBE of −1.96%.

The weight evolution over time of the various base models considered for 40 days is illustrated in Fig. 9a, 9b, 9c, 9d for the zone temperature, the heating load, the non-controllable loads, and the PV system electricity generation, respectively. Fig. 9a shows that the contribution of the RF and the XGB models is higher compared to that of the MLP and the SVR models for the zone temperature ensemble forecaster. Nevertheless, there is no single best base model over time,

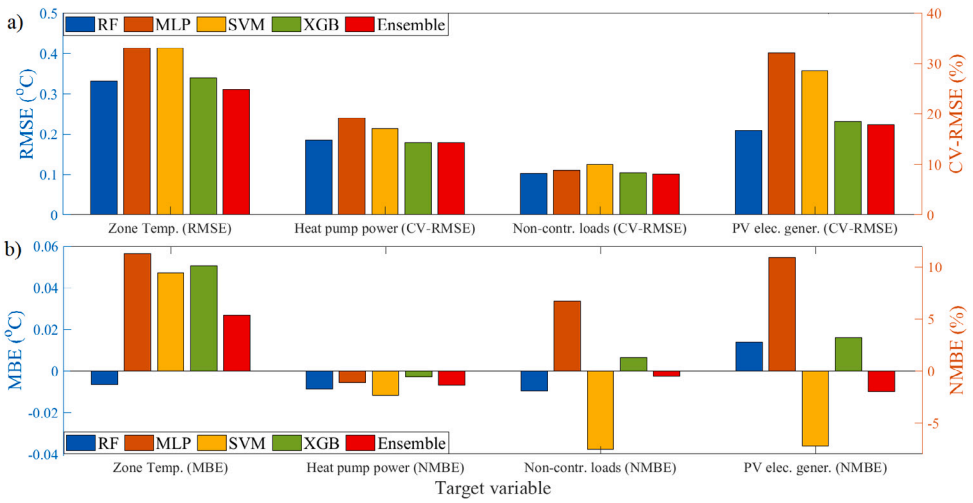


Fig. 8. (a) RMSE and CV-RMSE and (b) MBE/NMBE for the various base models (RF, MLP,SVM,XGB) and the ensemble model forecaster for each target variable (zone temperature, heat pump load, non-controllable loads, PV system electricity generation), day-ahead prediction models.

Table 13
Comparison between various base models for each target variable by execution time (sec), day-ahead prediction models (Intel® i7-7500 CPU 2.7 GHz, 8 GB RAM).

	Zone temp.	Heat pump load	Non-contr. loads	PV system elec. gener.
RF	6.02	2.58	1.12	0.76
MLP	3.02	2.35	2.82	0.96
SVM	1.42	0.9	0.4	0.57
XGB	7.29	3.94	1.7	2.34
Ensemble model	17.76	9.78	6.05	4.64

Table 14
Comparison between various base models for each target variable by RMSE(°C) (zone temperature) and CV-RMSE(%) (heat pump load, non-controllable loads, PV system electricity generation)- hour-ahead prediction models.

	Zone temp. (°C)	Heat pump load (%)	Non-contr. loads (%)	PV system elec. gener. (%)
RF	0.208	6.7	7.48	24.1
MLP	0.217	10	8.56	19.82
SVM	0.242	12.6	9.15	22.07
XGB	0.179	7.7	7.68	20.59
Ensemble model	0.178	6.7	7.39	16.9

Table 15
Comparison between various base models for each target variable by MBE (°C) (zone temperature) and NMBE(%) (heat pump load, non-controllable loads, PV system electricity generation)- hour-ahead prediction models.

	Zone temp. (°C)	Heat pump load (%)	Non-contr. loads (%)	PV system elec. gener. (%)
RF	0.0012	−0.44	−0.142	−0.23
MLP	0.0044	−1.2	−0.0416	−2.03
SVM	−0.058	−4.26	−1.676	−1.39
XGB	−0.0022	0.11	−0.16	0.47
Ensemble model	−0.005	−0.63	−0.334	−0.79

as the suitability of the XGB and the RF models (in terms of weight magnitude) alternates over time. It should be noted that for days 37–39 the ensemble model performance is identical to that of the RF model since it receives a weight equal to 1. This means that the ensemble model is likely to discard one or more constituent models (by assigning weight equal to zero) in case there is a single model that strongly outperforms the rest. Similar conclusions can be drawn for the ensemble model developed for the non-controllable loads and the PV system electricity generation; each base model is likely to have a higher contribution depending on the simulation day. The allocated weights are a measure of the predictive performance of each model (in terms of cumulative error), thus, the performance of each base model changes over time. This means that it is not always possible to select a single best model for each target variable in advance. On the other hand, as regards the heating load ensemble model, the relative contribution of each base model does not change significantly over time. Table 13 summarises the average execution times for each target variable as well as the ensemble model forecaster for day-ahead predictions. The execution times range between 4.64 to 17.79 s depending on the target variable. SVM and XGB exhibit the lowest and highest execution times, respectively.

4.3.2. Hour-ahead prediction models

Tables 14 and 15 summarise the RMSE/CV-RMSE and the MBE/NMBE, respectively, for each target variable and base model considered as well as the resulting ensemble model forecaster. The bar chart in Fig. 10 illustrates the RMSE/CV-RMSE (Fig. 10a) and the MBE/NMBE (Fig. 10b) for all target variables considered. By comparing Tables 11 and 12 with Tables 14 and 15 it can be concluded that hour-ahead prediction models for the HVAC system-related variables exhibit higher overall accuracy compared to the day ahead prediction models, especially for the heating system-related variables (heating load and zone temperature). All ensemble models developed attain lower RMSE/CV-RMSE values than those achieved by the constituent models. The MBE/NMBE lie within acceptable limits as per ASHRAE criteria [80].

The weight evolution of the various base models considered for 40 days is illustrated in Fig. 11a, 11b, 11c, 11d for the zone temperature, the heating load, the non-controllable loads, and the PV system electricity generation, respectively. Although Figs. 9 and 11 depict the weight evolution of the same period, the weight allocation is considerably different. This suggests that the prediction horizon strongly affects the performance of the base models. For example, considering the PV plant power output, the associated ensemble model achieves considerably lower CV-RMSE than the rest of the base models. Moreover, each constituent model becomes more influential than the rest depending on the simulation period. Table 16 summarises the average execution times for each target variable as well as the ensemble model forecaster for hour-ahead predictions. The execution times range between 2.8 to 3.44 s depending on the target variable.

4.4. Energy flexibility assessment

In order for the predictive models to be integrated into energy management systems, it is necessary to assess their performance not

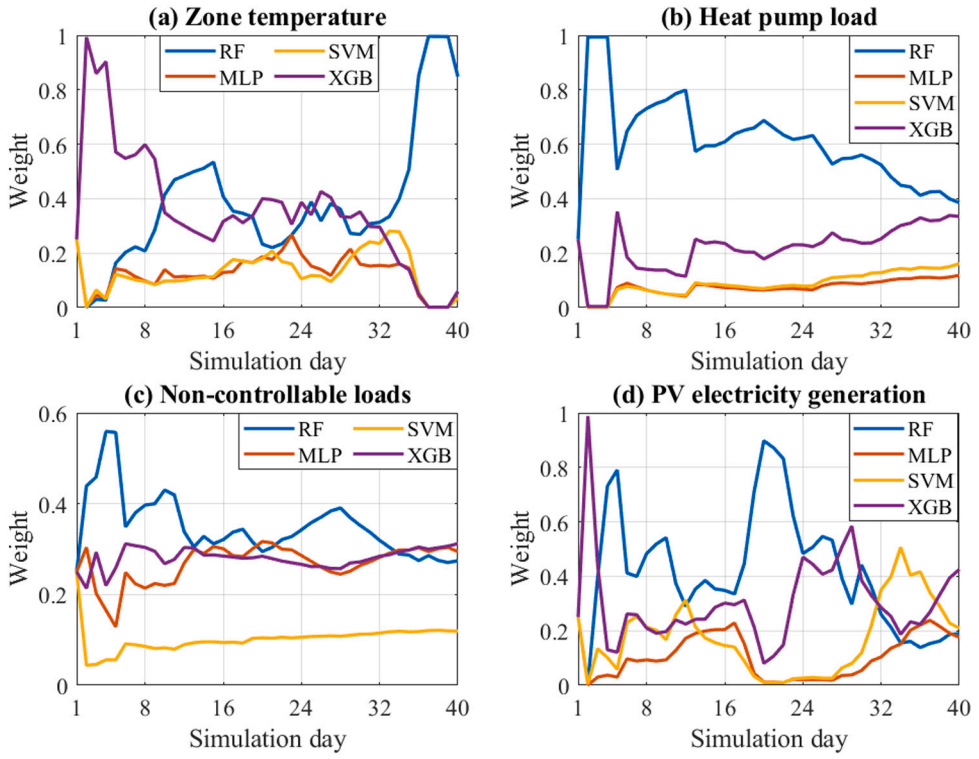


Fig. 9. Weight evolution for each ML model for day-ahead predictions, (a) zone temperature, (b) heat pump load, (c) non-controllable loads, (d) PV electricity generation.

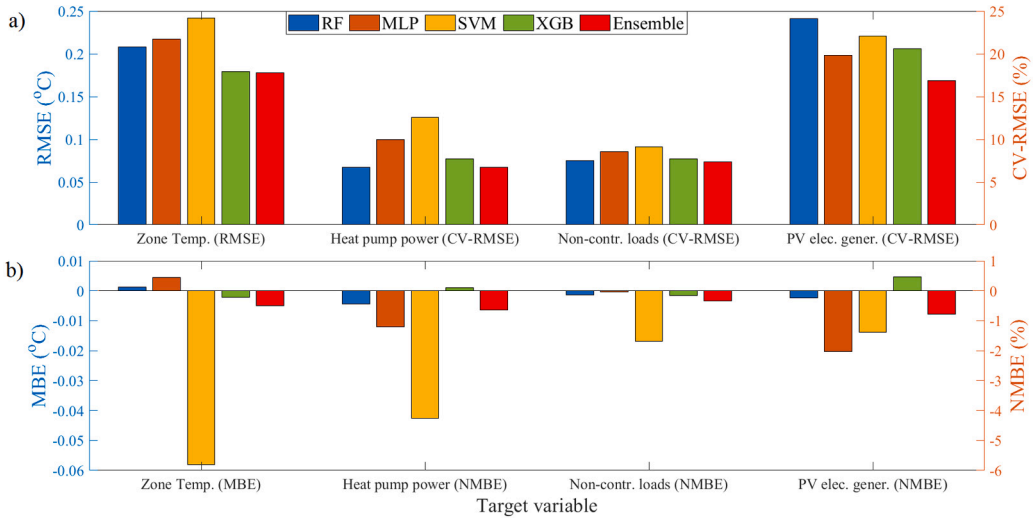


Fig. 10. RMSE and CV-RMSE and (b) MBE/NMBE for the various base models (RF, MLP,SVM,XGB) and the ensemble model forecaster for each target variable (zone temperature, heat pump load, non-controllable loads, PV system electricity generation), hour-ahead prediction models.

Table 16
Comparison between various base models for each target variable by execution time (sec), hour-ahead prediction models (Intel® i7-7500 CPU 2.7 GHz, 8 GB RAM).

	Zone temp.	Heat pump load	Non-contr. load	PV system elec. gener.
RF	0.29	0.5	0.88	0.24
MLP	0.47	1.52	1.04	0.89
SVM	0.06	0.38	0.1	0.07
XGB	2.6	0.75	0.76	1.14
Ensemble model	3.44	3.15	2.8	2.34

only for baseline consumption patterns but also during DR actions. To evaluate the energy flexibility potential of the virtual testbed, two different baseline cases are selected from the evaluation set, each based on the two occupancy profiles considered as shown in Table 3. Figures 12a and 12c depict the zone thermostat setpoint (black) as well as the actual (continuous blue) and predicted values (day-ahead prediction: dashed blue, hour-ahead prediction: dotted blue) for the first (OC1) and the second occupancy profile (OC2), respectively. Figs. 12b and 12d illustrate both the actual and the predicted values of the building power demand (heat pump power and non-controllable loads) and self-generation for OC1 and OC2, respectively.

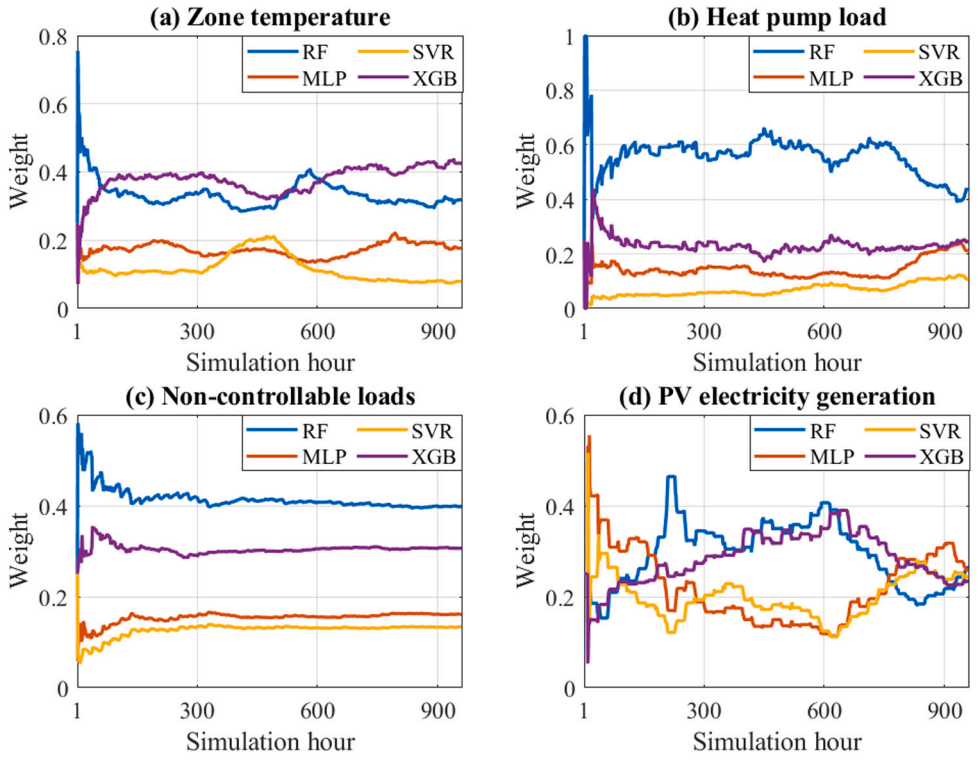


Fig. 11. Weight evolution for each ML model for hour-ahead predictions, (a) zone temperature, (b) heat pump load, (c) non-controllable loads, (d) PV electricity generation.

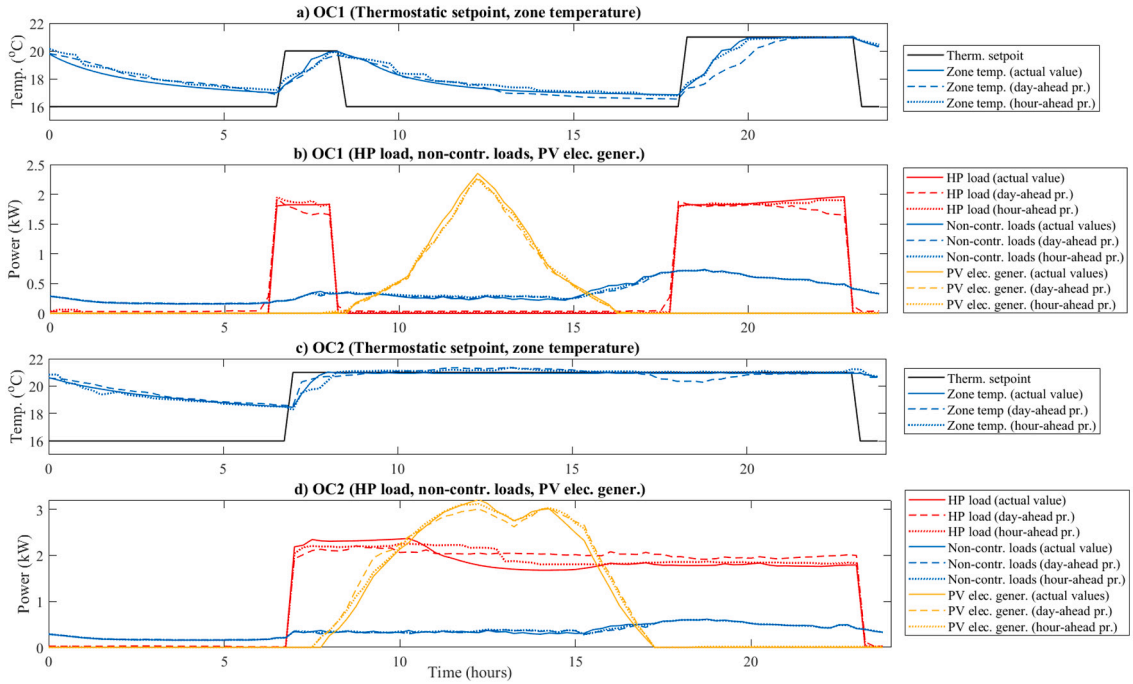


Fig. 12. Predicted (ensemble) and actual (white-box) values: (a) Thermostatic setpoint and zone temperature, OC1, (b) heat pump load, non-controllable loads, PV electricity generation, OC1, (c) Thermostatic setpoint and zone temperature, OC2, (d) heat pump load, non-controllable loads, PV electricity generation, OC2.

4.4.1. Downward flexibility

In this section, the downward flexibility potential of the heat pump and the battery is evaluated for the two occupancy profiles considered by using the white-box model (actual values) as well as the ensemble-based learning model for day-ahead and hour-ahead predictions. It is assumed that the imposed hourly DR actions are independent. The heat pump storage capacity (C_{DF}) (Eq. (9)), resulting from hourly zone thermostat reductions, is presented in Figs. 13a (OC1) and 13b (OC2). To assess whether the temperature drifts during a DR action lie within acceptable limits as per ASHRAE standards (Table 2), the temperature reductions arising from the applied DR actions are calculated. Figs. 14a (OC1) and 14b (OC2) depict the temperature deviations during the applied DR actions. Given that rebound effects resulting from the activation of the passive thermal energy storage exhibit a

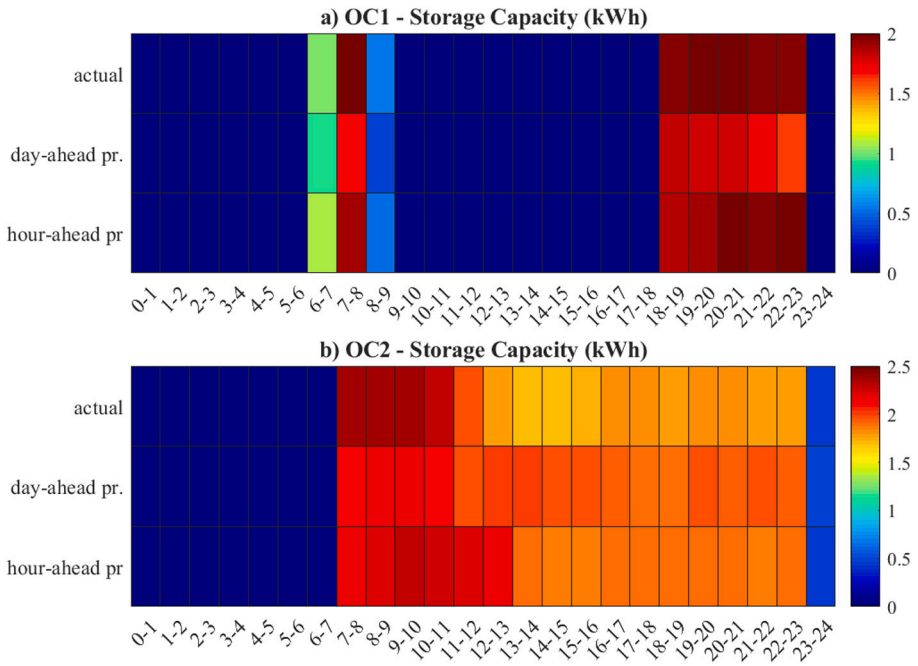


Fig. 13. Ensemble predicted heat pump storage capacity (C_{DF}) (hourly down-flex DR actions): (a) OC1, (b) OC2.

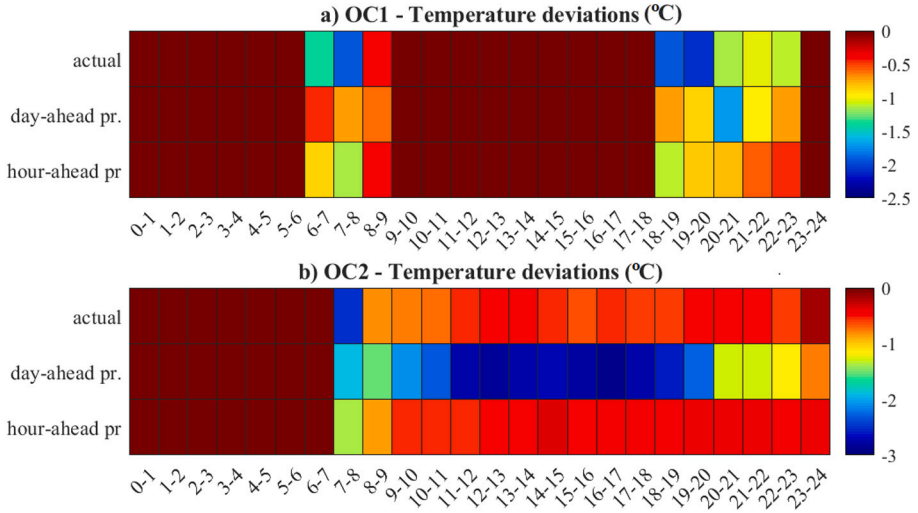


Fig. 14. Ensemble predicted temperature deviations (hourly down-flex DR actions): (a) OC1, (b) OC2.

duration of approximately six hours, it is possible to assess the storage efficiency by considering only day-ahead predictions. Hourly down-flex actions do not result in any rebound effects for the first occupancy profile. This is because OC1 is characterised by long periods of heat pump inactivity, thus resulting in lower water tank temperatures that limit the heat pump capability to increase its power output. Thus, the storage efficiency (η_{DF}) (Eq. (12)) is only evaluated for the second occupancy profile, as illustrated in Fig. 15. The prediction (ensemble) model capability to forecast the hourly attributes of the heat pump energy flexibility (as illustrated in Figs. 13–15) is evaluated by using suitable regression charts. Fig. 16 depicts the actual and the predicted values as well as the coefficient of determination (R^2) for the storage capacity (Fig. 16a-OC1, Fig. 16b-OC2), the temperature deviations (Fig. 16c-OC1, Fig. 16-OC2), and the storage efficiency (Fig. 16e-OC2).

During periods of occupant inactivity, the heat pump is switched off and it exhibits zero storage capacity. Considering the first occupancy profile, both day-ahead and hour-ahead prediction models exhibit very good performance, achieving R^2 values equal to 0.97 and 0.998, respectively. The storage capacity is consistently underestimated both by day-ahead and hour-ahead prediction models, however, the hour-ahead prediction model achieves a marginally better performance than that of the pertinent day-ahead model. The temperature reductions during DR actions lie within acceptable limits as per ASHRAE standards [80]. Further, the hour-ahead prediction model can capture the temperature deviations arising from DR actions better compared to the day-ahead prediction model. The heat pump storage capacity exhibited by the second occupancy profile is accurately predicted for both prediction horizons. The hour-ahead prediction model can adequately forecast the temperature deviations arising from the zone thermostat modulations, whereas the day-ahead prediction model consistently fails to capture them. In addition, the storage efficiency is consistently overestimated by the day-ahead prediction model achieving a coefficient of determination equal to 0.902. This is due to the underestimation of all rebound effects arising from the applied DR actions.

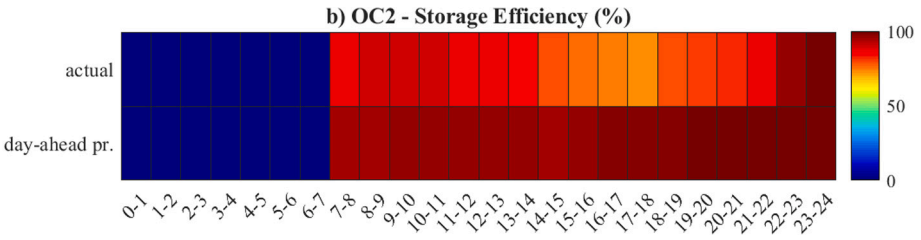


Fig. 15. Ensemble predicted heat pump storage efficiency (η_{DF}) (hourly down-flex DR actions) for OC2.

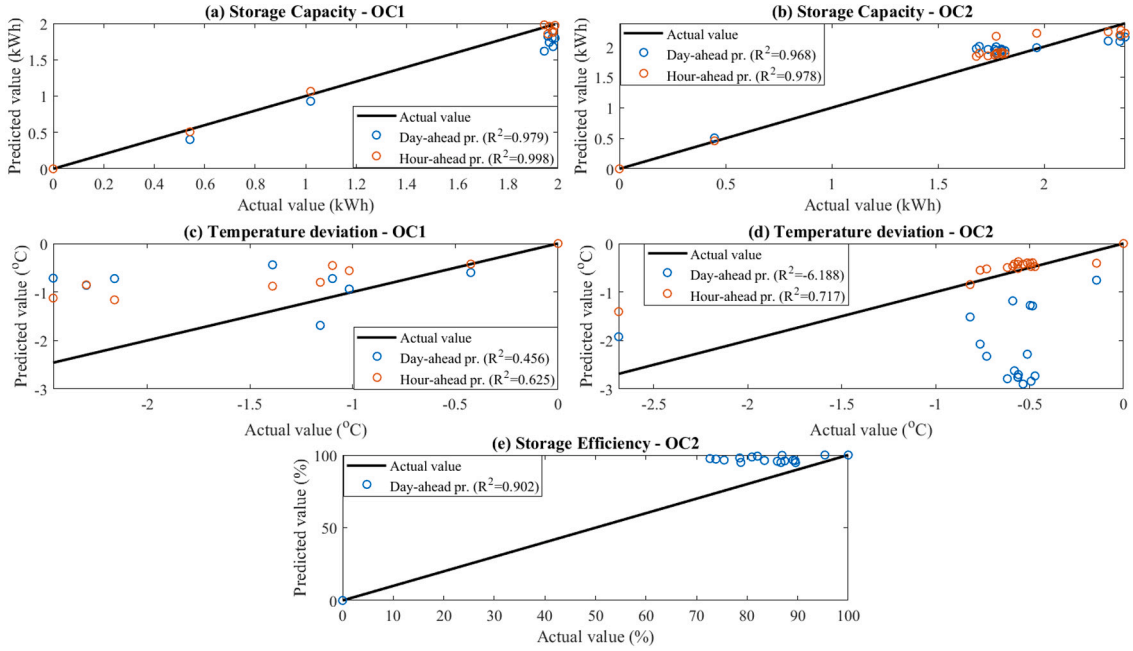


Fig. 16. Actual (white-box) and predicted (ensemble) values (thermostatic setpoint hourly down-flex DR actions): (a) storage capacity — OC1, (b) storage capacity — OC2, (c) temperature deviations — OC1, (d) temperature deviations — OC2, (e) storage efficiency — OC2.

The battery storage capacity (C_{DF}) (Eq. (11)), resulting from hourly zone temperature reductions is presented in Figs. 17a (OC1) and 17b (OC2). The storage efficiency (η_{DF}) (Eq. (12)) resulting from the activation of the battery flexibility depends not only on the action per se but also on the rebound starting times (t_c) and duration (τ_{id}). Figs. 18a (actual values — OC1), 18b (day-ahead predictions-OC1), 18c (actual values — OC2), and 18d (day-ahead predictions-OC2) provide the storage efficiency (η_{DF}) (Eq. (14)) resulting from the use of electrical storage for 1 h down-flex DR actions, a rebound duration of one hour and various battery charging commencement times. Thus, in Fig. 18, the X axis corresponds to the occurrence time of a given DR event, while the Y axis corresponds to the occurrence time of the associated rebound. Given that the PV is likely to generate electricity for several hours during the day, the battery efficiency can be assessed by considering only day-ahead predictions. This indicates that the storage efficiency can be only assessed for rebounds occurring within one day of the prediction. The regression chart in Fig. 19 depicts the actual and the predicted values as well as the coefficient of determination (R^2) for the battery storage capacity (Fig. 19a- OC1, Fig. 19b- OC2) and efficiency (Fig. 19c- OC1, Fig. 19d- OC2).

Since the battery is discharged to cover the building load, its storage capacity depends on the difference between the building load and the PV system electricity generation. This means that during periods of high solar irradiance, the storage capacity is minimised, whereas when the heat pump is activated the battery storage capacity increases. If solar power is unavailable or less than the building load, the storage efficiency only depends on the charging and discharging battery efficiency and is given by $\eta_{DF} = 1 - 1/\eta_d\eta_c = 1 - 1/0.92 = -23.5\%$. The storage efficiency could take negative values for a given DR action when the associated energy shifting potential is lower than the energy consumed during the rebound. On the contrary, the storage efficiency reaches 100% when the battery is charged during periods of high electricity generation by the PV system. When the building load is lower than the PV system electricity generation, no DR actions take place (grey tile areas). The storage capacity predictions exhibit high accuracy for all occupancy profiles and prediction horizons considered. On the other hand, the storage efficiency predictions exhibit lower accuracy for both occupancy profiles due to the misestimation of the PV electricity generation and the heat pump load by the associated day-ahead models.

4.4.2. *Upward flexibility*

In this section, the upward flexibility potential of the heat pump and the battery is evaluated for the two occupancy profiles considered by using the white-box model (actual values) as well as the ensemble-based learning model for day-ahead and hour-ahead predictions. The heat pump storage capacity (C_{UF}) (Eq. (10)), resulting from hourly zone temperature modulations is presented in Figs. 20a (OC1) and 20b (OC2). Figs. 21a (OC1) and 21b (OC2) depict the actual value, the day-ahead predictions, and the hour-ahead predictions for the temperature deviations during a DR action. Hourly up-flex actions do not result in any rebound effects for the first occupancy profile, consequently, the ensuing storage efficiency is always 0. This is because the heat pump is deactivated at the end of the DR action during periods of occupant inactivity or absence, thus not allowing for load curtailment. In addition, during periods of occupant activity the heat pump power shows

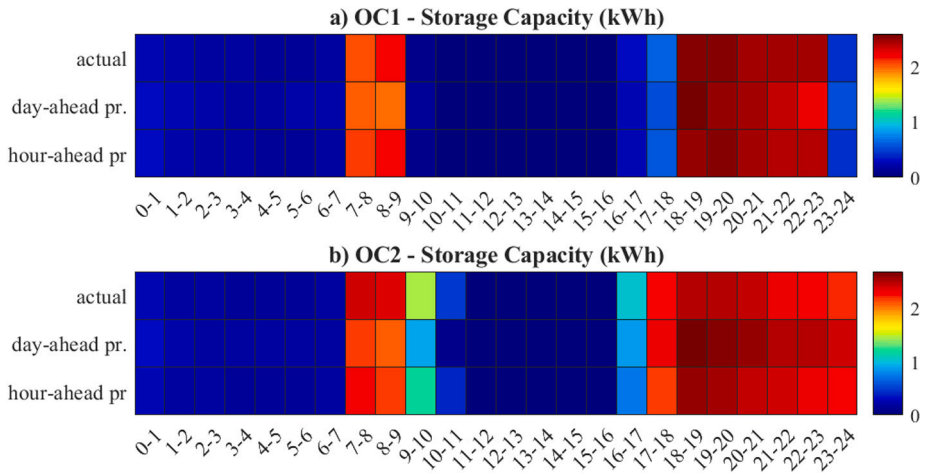


Fig. 17. Ensemble predicted heat pump storage capacity (C_{DF}) (hourly down-flex DR actions): (a) OC1, (b) OC2.

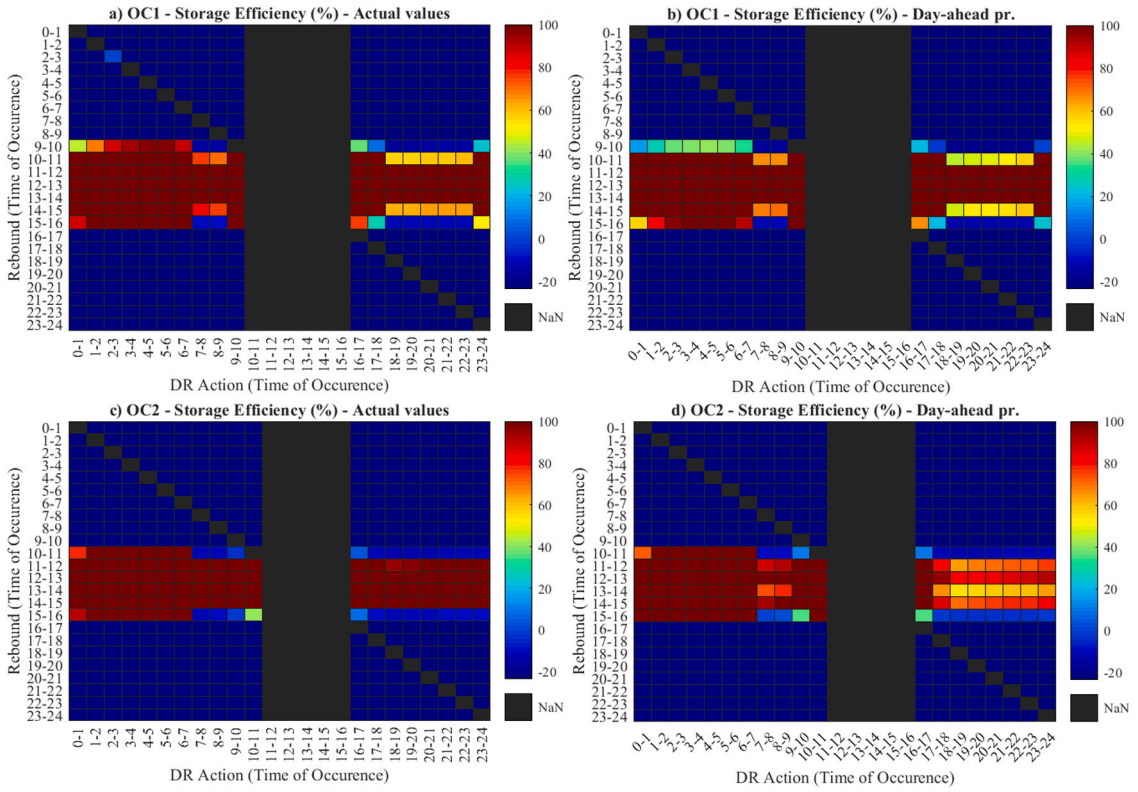


Fig. 18. Storage efficiency (η_{DF}) for the stationary battery, (hourly down-flex actions): (a) OC1 — actual values (white-box), (b) OC1 — day-ahead predictions (ensemble), (c) OC2 — actual values (white-box), (d) OC2 — day-ahead predictions (ensemble) X axis: DR event time, Y axis: Rebound time.

a minor increase that results in the absence of rebounds. Accordingly, the heat pump storage efficiency (η_{UF}) (Eq. (13)) is assessed only for the second occupancy profile (Fig. 22). Fig. 23 depicts the actual and the predicted values as well as the coefficient of determination (R^2) for the storage capacity (Fig. 23a-OC1, Fig. 23b-OC2), the temperature deviations (Fig. 23c-OC1, Fig. 23d-OC2), and the storage efficiency (Fig. 23e-OC2).

As regards the first occupancy profile, the storage capacity arising from exploiting the upward flexibility is considerably underestimated by day-ahead prediction models achieving an R^2 equal to 0.576; the hour-ahead prediction models nonetheless exhibit a significantly better accuracy attaining an R^2 equal to 0.961. Considering the second occupancy profile, both the day-ahead and the hour-ahead models fail to capture the associated storage capacity achieving R^2 values equal to −0.736 and −0.347, respectively. It is noteworthy that the storage capacity resulting from down-flex actions is likely to be more accurately captured by data-driven models. This is due to the fact that the heat pump is deactivated during down-flex DR actions and does not depend on the zone thermostatic setpoint. Conversely, in up-flex actions, the data-driven model must predict the heat pump dynamics under different zone thermostatic setpoints. Finally, considering the temperature deviations during DR actions, prediction models exhibit poor performance for both occupancy profiles, however, the hour-ahead prediction models exhibit better performance than the day-ahead prediction models. The storage efficiency for OC2 is consistently underestimated since all rebounds arising from up-flex actions are also underestimated.

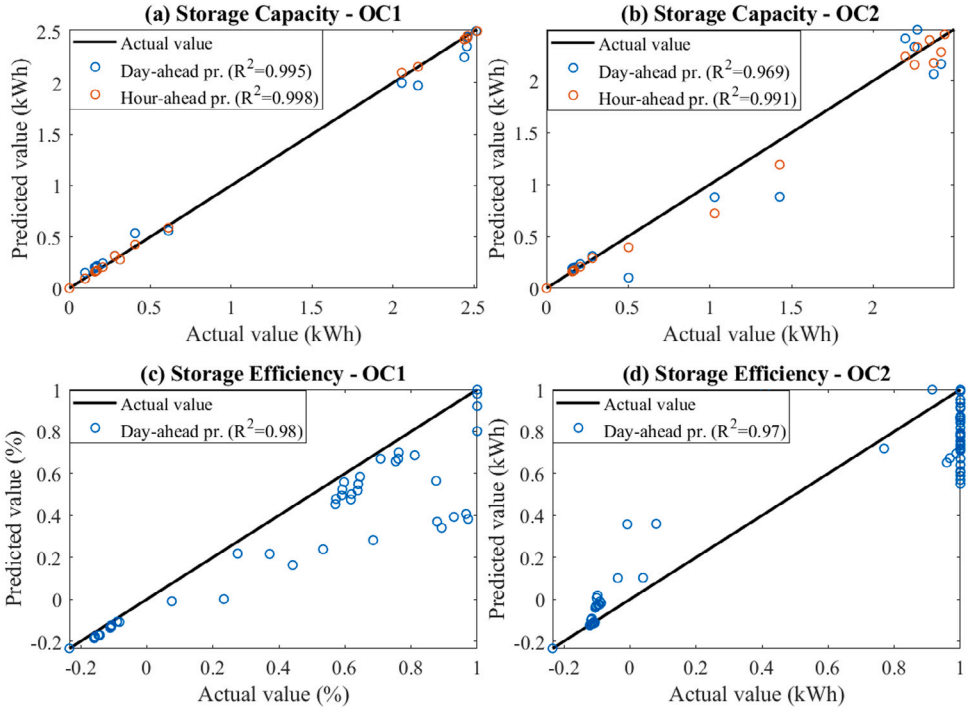


Fig. 19. Actual (white-box) and predicted (ensemble) values, battery (hourly down-flex DR actions): (a) storage capacity — OC1, (b) storage capacity — OC2, (c) storage efficiency — OC1, (d) storage efficiency — OC2.

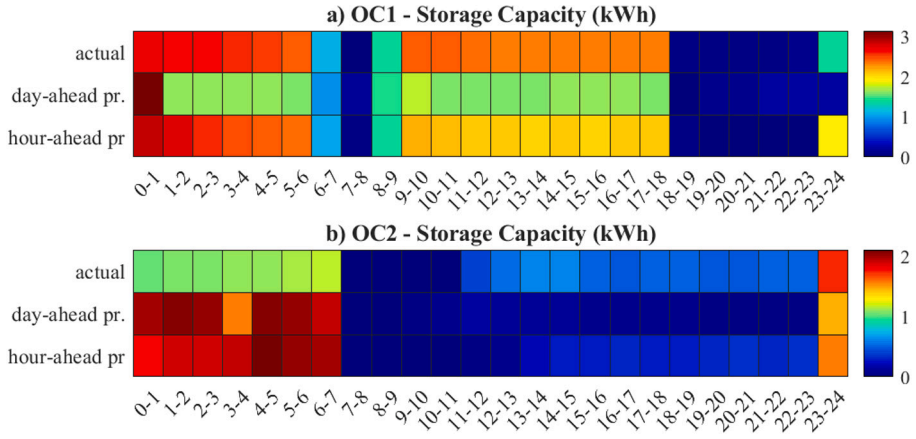


Fig. 20. Ensemble predicted heat pump storage capacity (C_{DF}) (hourly up-flex DR actions): (a) OC1, (b) OC2.

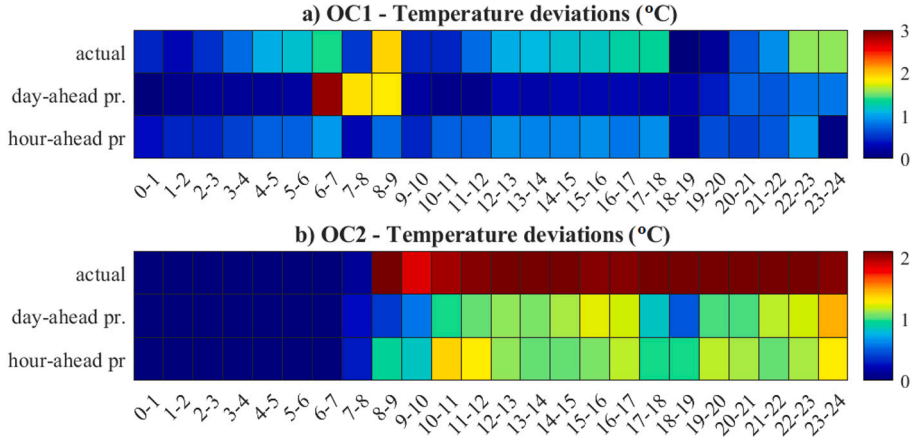


Fig. 21. Ensemble predicted temperature deviations (hourly up-flex DR actions): (a) OC1, (b) OC2.

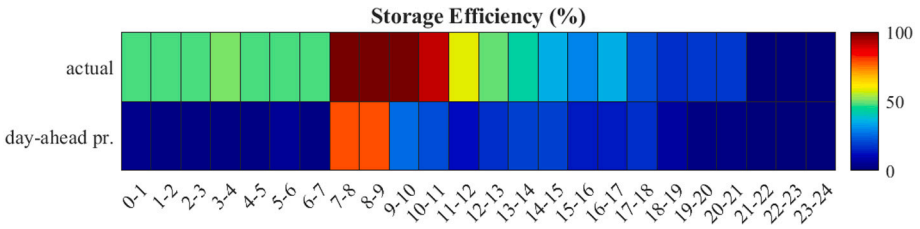


Fig. 22. Ensemble predicted heat pump storage efficiency (η_{DF}) (hourly up-flex DR actions) for OC2.

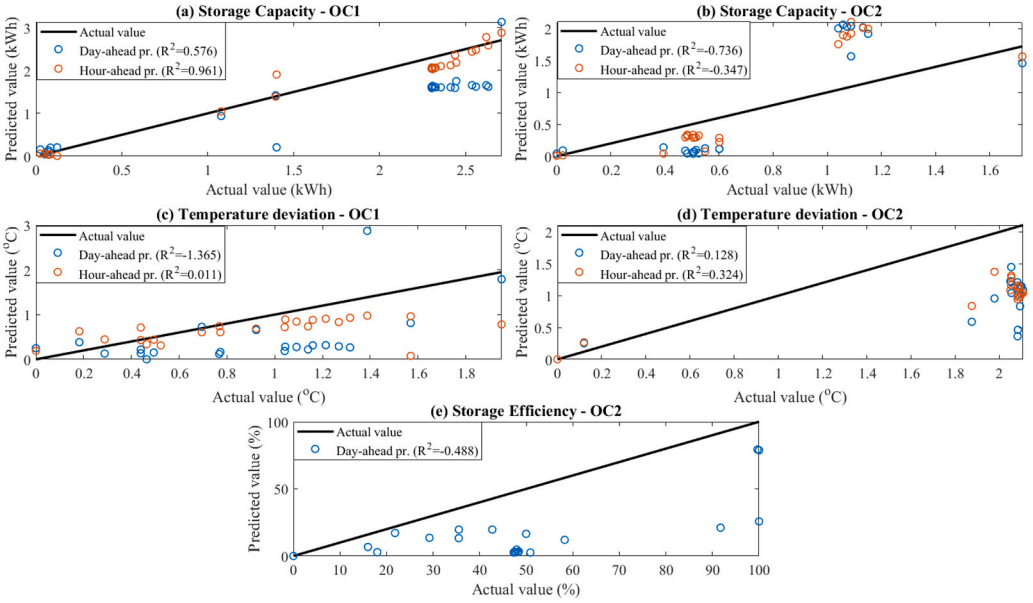


Fig. 23. Actual (white-box) and predicted (ensemble) values (thermostatic setpoint hourly up-flex DR actions): (a) storage capacity — OC1, (b) storage capacity — OC2, (c) temperature deviations — OC1, (d) temperature deviations — OC2, (e) storage efficiency — OC2.

5. Discussion

To date, the variety of definitions of building energy flexibility is accompanied by numerous flexibility evaluation frameworks. This is because the definitions are associated with diverse demand response applications that utilise various energy conversion systems based on different objectives [62]. In this study, this problem is addressed by developing a data-driven energy flexibility quantification framework to characterise the DR potential of the most common energy systems found in residential buildings, namely, building structural thermal mass and electrical energy storage units by using suitable indicators (i.e., storage capacity, storage efficiency). This is important since the flexibility potential of each system is a multifactor characteristic that depends not only on weather conditions and end-user preferences, but also on the power profiles of the various building energy conversion systems [25]. This methodology can be potentially applied to all residential building types, account for the contribution of onsite electricity generation to the energy cost of individual power modulation strategies, and illustrate the associated DR potential uniformly and comprehensively.

Existing market-independent methodologies for assessing the energy flexibility of residential buildings have followed either physics-based or grey-box approaches, without evaluating building flexibility in dynamic environments such as those found in DR applications. Previous data-driven-based flexibility evaluation frameworks for individual residential buildings have mainly focused on a single energy system category (either thermal or electrical) without analysing this flexibility from an integrated system perspective or without considering potential energy costs (rebound effects) arising from harnessing this flexibility. There is therefore a significant research gap in the literature related to the explicit energy flexibility assessment in residential buildings on a one-by-one basis including multi-component thermal and electrical systems. To address this gap, an ensemble learning framework for multi-step-ahead predictions has been developed based on a series of periodically trained and updated data-driven models and a sliding window method. The developed prediction models are updated based on dynamic feature selection and the selection of the most recent occupancy patterns. The ensemble learning model developed can automatically manage the benefits and shortcomings of the constituent ML models and ultimately generalise better than the constituent models. In this way, building energy flexibility can be evaluated in an end-user tailored manner not only by considering new patterns in data but also by assigning higher weights to individual learners that have achieved better performance by the prediction time. The obtained flexibility maps give insights, not only into the actual building flexibility potential, but also into the data-driven model capability to quantify the associated flexibility by using suitable indicators. The proposed flexibility evaluation framework can be extended for any thermostatically controlled load and battery-based system.

The proposed energy flexibility quantification framework can be applied to all residential building types, where at least one of the following energy conversion systems: thermostatically controlled thermal loads (active or passive), electrical energy storage units (batteries), or onsite electricity generation, is applicable. The developed prediction models use features that can be realistically collected by a residential energy management system (e.g., historical data from the heat pump, the remainder of the electrical loads, the PV electricity generation, and weather prediction data). To develop machine learning algorithms for

building energy performance evaluation, it is necessary to train and optimise various data-driven algorithms and ultimately select manually the most suitable ones. Heterogeneous ensemble models can automatically build optimal forecasts by assigning suitable weights to the constituent models for each prediction step. Moreover, the ensemble learning framework presented can be extended for various data-driven algorithms, depending on the building portfolio and the available computational resources. To reduce the computational burden (in terms of execution time and resources required) for each constituent data-driven model, it is necessary to use a limited number of features and samples for each model developed. To this end, the number of samples used as training sets is determined parametrically, whereas the optimal number of features by using both the autocorrelation function and the Spearman correlation coefficient. The developed framework lays a foundation for explicitly evaluating the energy flexibility potential of individual residential buildings, however, without further investigating its commercial potential. Nevertheless, the ongoing advances in computing systems and hardware capabilities will facilitate the commercial viability of more computationally demanding machine learning applications or larger building portfolios. Accordingly, a combination of further methodology enhancements, coupled with computing advances, has the potential to facilitate the further enhancement and refinement of the described methodology.

Simulation results show that the features selected for each target variable do not only depend on occupancy profiles but also depend on the model prediction horizon. Moreover, statistical properties from previous prediction periods are likely to be relevant features, especially for hour-ahead predictions models. As regards the ML model training set sizes, the considered data-driven models achieve optimal performance for a different number of training weeks (day-ahead forecasts) and training days (hour-ahead forecasts). Results show that there is no single best data-driven algorithm for every target variable as their predictive performance (and relative contribution) changes with time (due to weather and occupancy changes). Further, the ensemble learning framework is likely to discard one or more of the constituent models if the remainder of the ML models largely outperform them. Hour-ahead prediction models show considerably better performance than that of the day-ahead models, especially for the target variables related to the HVAC system; this is potentially due to the intricate dynamics of the zone temperature and the heating system power consumption that cannot be captured for longer prediction horizons.

The temperature deviations arising from both downward and upward DR actions are poorly estimated by day-ahead predictions, whereas hour-ahead predictions exhibit considerably better performance for both occupancy profiles considered. This indicates that shorter prediction horizons are required in order to capture accurate assessments of the impact of DR actions on occupant thermal comfort. Moreover, the prediction accuracy of the heat pump up-flex energy shifting capability and the associated temperature deviations varies depending on the occupancy profile. This indicates that different weather conditions and/or occupant behaviour are likely to influence the predictability of building flexibility potential. The underestimation of the rebound effects resulting from harnessing the heat pump flexibility results in the underestimation (of up-flex DR actions) and the overestimation (of down-flex DR actions) of the associated storage efficiency. This is because the accuracy of the heat pump storage efficiency predictions is limited by the failure of the day-ahead models to capture the complex dynamics of the heating system. Finally, the predictability of the battery storage efficiency strongly depends on the accuracy of the relationship between the PV electricity generation day-ahead predictions and the total building load day-ahead predictions. In this sense, the ability of data-driven models to forecast the energy costs associated with activating the flexibility potential of the various building energy systems is limited by the potentially longer prediction horizons required.

The obtained flexibility maps showed that the heat pump storage capacity resulting from down-flex actions can be more accurately predicted for both occupancy profiles and prediction horizons considered. This is because the heat pump downward flexibility can be harnessed by applying on–off signals, whereas the pertinent upward flexibility depends on zone thermostatic modulations. In this study, the HVAC system-related target variables (i.e., heat pump power and zone temperature) are predicted by considering only the associated historical data and statistical properties of these variables. However, the ground source heat pump performance is influenced by factors that have not been considered in the feature selection process and are likely to impact the associated prediction model accuracy. These factors include: temperatures and mass flowrates associated with the supply water in the evaporator and the condenser, as well as the heat pump compressor characteristics. Nevertheless, the establishment of a detailed machine learning model for the ground source heat pump energy consumption is beyond the scope of this analysis and is left as future work.

The datasets used to develop and evaluate the data-driven models, as well as to evaluate building energy flexibility, are generated by utilising a calibrated white-box model and average daily occupancy profiles deriving from categorising the household weekday diaries and represent 56% of the survey sample [84]. Nevertheless, the occupancy profile selection is indicative of exemplifying the performance of the proposed flexibility evaluation framework for various power consumption patterns. Although the use of deterministic occupancy profiles does not allow the diversity of consumption patterns to be taken into consideration, a cumulative building energy consumption calculation is beyond the scope of this paper. This study is still considered to be early-stage research, rather than mature application focused research. Hence, the thermostatic setpoint is utilised as input not only to investigate the flexibility potential of thermostatically controlled loads and their potential impact on occupant thermal comfort but also the predictability of energy flexibility related characteristics. Given that the current study is early-stage research, it does not explore any potential privacy implications arising from bi-directional communication between consumers and aggregators. Further research is required to identify suitable system architectures and specifications that will be aligned with the evolving data-privacy technology.

6. Conclusions

In this paper, a generic data-driven-based energy flexibility quantification framework has been developed to characterise the flexibility potential of thermal and electrical energy systems commonly found in residential buildings (passive thermal energy storage and stationary batteries). Building DR potential is assessed by using an ensemble learning framework based on dynamic feature selection and machine learning model development. Specifically, the ensemble learning-based framework described utilises four base models (namely, random forests, multilayer perceptron neural network, support vector machine, extreme gradient boosting) and builds its prediction based on the cumulative error of the constituent models. Simulation results indicate that the ensemble prediction model outperforms the constituent base learners and that the suitability of each learner changes over time. Furthermore, shorter prediction horizons are more suitable to forecast any thermal comfort deviations arising from the activation of the passive TES. Finally, the predictability of the potential energy costs arising from activating the flexibility of the various systems is limited by the day-ahead model failure to capture the intricate dynamics of the various systems involved.

The developed methodology can be potentially applied to all residential building types, depict the flexibility potential of individual strategies as well as their occupant thermal comfort impacts, and ultimately illustrate it in a precise and uniform way. The proposed data-driven framework allows electricity aggregators to evaluate various energy flexibility strategies in the context of multistep-ahead predictions by using data that can be realistically collected from residential buildings and eventually evaluate or optimise a portfolio of buildings in an end-user tailored manner. Regarding the commercial viability of the

current study, the proposed energy flexibility assessment framework can be potentially attractive to utilities and aggregators considering the increasing energy prices, the higher requirements for energy flexibility, the increasing computing resources, and the changing regulatory environment. In this context, the current work represents early research towards realising such flexibility in a rapidly changing sector.

CRediT authorship contribution statement

Adamantios Bampoulas: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Fabiano Pallonetto:** Conceptualization, Writing – review & editing, Supervision. **Eleni Mangina:** Conceptualization, Writing – review & editing, Funding acquisition, Project administration, Supervision. **Donal P. Finn:** Conceptualization, Writing – review & editing, Funding acquisition, Project administration, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has emanated from research conducted with the financial support of Science Foundation Ireland under the SFI Strategic Partnership Programme Grant Number SFI/15/SPP/E3125 and additional funding provided by the UCD Energy Institute. The opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Science Foundation Ireland.

Symbol	Definition
T_{out}	Outdoor temperature (°C)
I_{tot}	Total solar irradiance (W/m^2)
RH	Relative humidity (%)
WS	Wind speed (m/s)
WT	Workday type (binary)
DoW	Day of week
MoD	Minute of day
T_z	Zone temperature (°C)
T_{sp}	Zone thermostatic setpoint (°C)
P_{hp}	Heat pump (hp) load (W)
P_{nc}	Non-controllable (NC) load (W)
P_{pv}	PV system electricity generation (W)
$P_{hp,O/O}$	hp on/off operation (binary)
$P_{hp,av}$	Average hp load (W)
$P_{hp,max}$	Maximum hp load (W)
$P_{hp,min}$	Minimum hp load (W)
$R_{hp,av/max}$	Ratio 1: P_{hp_av}/P_{hp_max} (adim)
$R_{hp,min/av}$	Ratio 2: P_{hp_min}/P_{hp_av} (adim)
$T_{z,av}$	Average zone temperature (°C)
$T_{z,max}$	Maximum zone temperature (°C)
$T_{z,min}$	Minimum zone temperature (°C)
$R_{z,av/max}$	Ratio 3: T_{z_av}/T_{z_max} (adim)
$R_{z,min/av}$	Ratio 4: T_{z_min}/T_{z_av} (adim)
$P_{nc,av}$	Average NC load
$P_{nc,max}$	Maximum NC load
$P_{nc,min}$	Minimum NC load
$R_{nc,av/max}$	Ratio 5: T_{z_av}/T_{z_max} (adim)
$R_{nc,min/av}$	Ratio 6: T_{z_min}/T_{z_av} (adim)
t_{DR}	Duration of DR Event (hours)
τ_{id}	Total time of increased demand (hours)
SC_{DR}	Self-consumption during a DR action (adim)
P_{mod}	Modulated building load (W)
P_{ref}	Reference building load (W)
P_{RES}	Onsite electricity generation (W)
C_{DR}	Available storage capacity (kWh)
C_{DF}	Available storage capacity in down-flex (kWh)
C_{UF}	Available storage capacity in up-flex (kWh)
η_{DF}	Storage Efficiency in down-flex (adim)
η_{UF}	Storage Efficiency in up-flex (adim)
P_b	Building Load (W)
η_d	Battery discharging efficiency (adim)
η_c	Battery charging efficiency (adim)
t_c	Rebound commencement time (hours)

Abbreviation	Definition
RES	Renewable Energy Sources
DR	Demand Response
HP	Heat Pump
HVAC	Heat, Ventilation, and Air Conditioning
TES	Thermal Energy Storage
ACF	Autocorrelation Function
CV	Cross-Validation
MBE	Mean Bias Error
NMBE	Normalised Mean Bias Error
RMSE	Root Mean Error
CV-RMSE	Cumulative Variation Root Mean Error
GSHP	Ground Source Hear Pump
ML	Machine Learning
RF	Random Forests
MLP	Multilayer Perceptron Neural Network
SVM	Support Vector Machine
XGB	Extreme Gradient Boosting

References

[1] Demand side flexibility for power sector transformation. International Renewable Energy Agency; 2019.
[2] CEER advice on ensuring market and regulatory arrangements help deliver demand-side flexibility. 2014, CEER (Council of European Energy Regulators) (2014).
[3] Towards a smart energy system. 2015, Department of Energy Climate Change, London.
[4] Khan ZA, Jayaweera D. Smart meter data based load forecasting and demand side management in distribution networks with embedded PV systems. IEEE Access 2019;8:2631–44. http://dx.doi.org/10.1109/ACCESS.2019.2962150.
[5] Global status report for buildings and construction. 2019, Global Alliance for Buildings and Construction, International Energy Agency and the United Nations Environment Programme.
[6] Li PH, Pye S. Assessing the benefits of demand-side flexibility in residential and transport sectors from an integrated energy systems perspective. Appl Energy 2018;228:965–79. http://dx.doi.org/10.1016/j.apenergy.2018.06.153.
[7] Jensen SO, Marszal-Pomianowska A, Lollini R, Pasut W, Knotzer A, Engelmann P, et al. IEA EBC Annex 67 energy flexible buildings. Energy Build 2017;155:25–34. http://dx.doi.org/10.1016/j.enbuild.2017.08.044.
[8] Junker RG, Azar AG, Lopes RA, Lindberg KB, Reynders G, Relan R, et al. Characterizing the energy flexibility of buildings and districts. Appl Energy 2018;225:175–82. http://dx.doi.org/10.1016/j.apenergy.2018.05.037.
[9] Behl M, Smarra F, Mangharam R. DR-Advisor: A data-driven demand response recommender system. Appl Energy 2016;170:30–46. http://dx.doi.org/10.1016/j.apenergy.2016.02.090.
[10] Zhao H, Magoulès F. A review on the prediction of building energy consumption. Renewable Sustainable Energy Reviews 2012;16:3586–92. http://dx.doi.org/10.1016/j.rser.2012.02.049.

[11] Sun Y, Haghighat F, Fung BC. A review of the-state-of-the-art in data-driven approaches for building energy prediction. Energy Build 2020;221. http://dx.doi.org/10.1016/j.enbuild.2020.110022.

[12] New JR, Sanyal J, Bhandari M, Shrestha S. Autotune e+ building energy models. In: Proceedings of the 5th national SimBuild of IBPSA-USA. 2012.

[13] Yin R, Kara EC, Li Y, DeForest N, Wang K, Yong T, et al. Quantifying flexibility of commercial and residential loads for demand response using setpoint changes. Appl Energy 2016;177:149–64. http://dx.doi.org/10.1016/j.apenergy.2016.05.090.

[14] Dong B, Li Z, Rahman S, Vega R. A hybrid model approach for forecasting future residential electricity consumption. Energy Build 2016;117:341–51. http://dx.doi.org/10.1016/j.enbuild.2015.09.033.

[15] Wang Z, Srinivasan R. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. Renew Sustain Energy Rev 2017;75:796–808. http://dx.doi.org/10.1016/j.rser.2016.10.079.

[16] Yang Y, Li R, Huang T. Smart meter data analysis of a building cluster for heating load profile quantification and peak load shifting. Energies 2020;13(17). http://dx.doi.org/10.3390/en13174343.

[17] Deb C, ad J. Yang FZ, Shah SLK. A review on time series forecasting techniques for building energy consumption. Renew Sustain Energy Rev 2017;74:902–24. http://dx.doi.org/10.1016/j.rser.2017.02.085.

[18] Li X, Wen J. Review of building energy modeling for control and operation. Renew Sustain Energy Rev 2014;37:517–37. http://dx.doi.org/10.1016/j.rser.2014.05.056.

[19] Vanthournout K, Dupont B, Foubert W, Stuckens C, Claessens S. An automated residential demand response pilot experiment, based on day-ahead dynamic pricing. Appl Energy 2015;155:195–203. http://dx.doi.org/10.1016/j.apenergy.2015.05.100.

[20] Pipattanasomporn M, Kuzlu M, Rahman S. An algorithm for intelligent home energy management and demand response analysis. IEEE Trans Smart Grid 2012;3:2166–73. http://dx.doi.org/10.1109/TSG.2012.2201182.

[21] Guo P, Lam J, Li V. Drivers of domestic electricity users' price responsiveness: a novel machine learning approach. Appl Energy 2019;235:900–13. http://dx.doi.org/10.1016/j.apenergy.2018.11.014.

[22] Jiang DR, Powell WB. Optimal hour-ahead bidding in the real-time electricity market with battery storage using approximate dynamic programming. INFORMS J Comput 2015;27. http://dx.doi.org/10.1287/ijoc.2015.0640.

[23] Walawalkar R, Apt J, Mancini R. Economics of electric energy storage for energy arbitrage and regulation in New York. Energy Policy 2007;35:2558–68. http://dx.doi.org/10.1016/j.enpol.2006.09.005.

[24] Kazmi H, Suykens JAK, Driesen J. Large-scale transfer learning for data-driven modelling of hot water systems. IBPSA Building Simulation 2019, Rome, Italy. http://dx.doi.org/10.26868/25222708.2019.210352.

[25] Bampoulas A, Saffari M, Pallonetto F, Mangina E, Finn DP. A fundamental unified framework to quantify and characterise energy flexibility of residential buildings with multiple electrical and thermal energy systems. Appl Energy 2021;282. http://dx.doi.org/10.1016/j.apenergy.2020.116096.

[26] Maddalena ET, Lian Y, Jones C. Data-driven methods for building control — A review and promising future directions. Control Eng Pract 2020;95. http://dx.doi.org/10.1016/j.conengprac.2019.104211.

[27] Kathirgamanathan A, Rosa MD, Mangina E, Finn DP. Data-driven predictive control for unlocking building energy flexibility: A review. Renew Sustain Energy Rev 2021;135. http://dx.doi.org/10.1016/j.rser.2020.110120.

[28] Amasyali K, Gohary NME. A review of data-driven building energy consumption prediction studies. Renew Sustain Energy Rev 2018;81:1192–205. http://dx.doi.org/10.1016/j.rser.2017.04.095.

[29] Wang R, Lu S, Li Q. Multi-criteria comprehensive study on predictive algorithm of hourly heating energy consumption for residential buildings. Sustainable Cities Soc 2019;49. http://dx.doi.org/10.1016/j.scs.2019.101623.

[30] Gao W, Alsarraf J, Moayedi H, Shahsavar A, Nguyen H. Comprehensive preference learning and feature validity for designing energy-efficient residential buildings using machine learning paradigms. Appl Soft Comput 2019;84. http://dx.doi.org/10.1016/j.asoc.2019.105748.

[31] Somua N, Raman GMR, Ramamritham K. A hybrid model for building energy consumption forecasting using long short term memory networks. Appl Energy 2020;261. http://dx.doi.org/10.1016/j.apenergy.2019.114131.

[32] Du C, Li B, Yu W, Liu H, Yao R. Energy flexibility for heating and cooling based on seasonal occupant thermal adaptation in mixed-mode residential buildings. Energy 2019;189. http://dx.doi.org/10.1016/j.energy.2019.116339.

[33] Afzalan M, Jazizadeh F. Residential loads flexibility potential for demand response using energy consumption patterns and user segments. Appl Energy 2019;254. http://dx.doi.org/10.1016/j.apenergy.2019.113693.

[34] Oprea SV, Bara A, Ifrim G. Flattening the electricity consumption peak and reducing the electricity payment for residential consumers in the context of smart grid by means of shifting optimization algorithm. Comput Ind Eng 2018;122. http://dx.doi.org/10.1016/j.cie.2018.05.053.

[35] Bashir AA, Kasmaei MP, Safdarian A, Lehtonen M. Matching of local load with on-site PV production in a grid-connected residential building. Energies 2018;11. http://dx.doi.org/10.3390/en11092409.

[36] Golmohamadi H, Larsen KG, Jensen PG, Hasrat IR. Optimization of power-to-heat flexibility for residential buildings in response to day-ahead electricity price. Energy Build 2021;232. http://dx.doi.org/10.1016/j.enbuild.2020.110665.

[37] Bünning F, Huber B, Heer P, Aboudoni A, Lygeros J. Experimental demonstration of data predictive control for energy optimization and thermal comfort in buildings. Energy Build 2020;211. http://dx.doi.org/10.1016/j.enbuild.2020.109792.

[38] Cotrufo N, Saloux E, Hardy J, Candanedo J, Platon R. A practical artificial intelligence-based approach for predictive control in commercial and institutional buildings. Energy Build 2020;206. http://dx.doi.org/10.1016/j.enbuild.2019.109563.

[39] Drgoňa J, Picard D, Kvasnica M, Helsen L. Approximate model predictive building control via machine learning. Appl Energy 2018;218:199–216. http://dx.doi.org/10.1016/j.apenergy.2018.02.156.

[40] Smarra F, Jain A, Rubeis T, Ambrosini D, D'Innocenzo A, Mangharam R. Data-driven model predictive control using random forests for building energy optimization and climate control. Appl Energy 2018;226:1252–72. http://dx.doi.org/10.1016/j.apenergy.2018.02.126.

[41] Reynders G, Lopes RA, Marszal-Pomianowska A, Aelenei D, Martins J, Saelens D. Energy flexible buildings: An evaluation of definitions and quantification methodologies applied to thermal storage. Energy Build 2018;166:372–90. http://dx.doi.org/10.1016/j.enbuild.2018.02.040.

[42] Stinner S, Huchtemann K, Muller D. Quantifying the operational flexibility of building energy systems with thermal energy storages. Appl Energy 2016;181:140–54. http://dx.doi.org/10.1016/j.apenergy.2016.08.055.

[43] Reynders G, Diriken J, Saelens D. Generic characterization method for energy flexibility: Applied to structural thermal storage in residential buildings. Appl Energy 2017;198:192–202. http://dx.doi.org/10.1016/j.apenergy.2017.04.061.

[44] Foteinaki K, Rongling L, Heller A, Rode C. Heating system energy flexibility of low-energy residential buildings. Energy Build 2018;180. http://dx.doi.org/10.1016/j.enbuild.2018.09.030.

[45] Dréau JL, Heiselberg P. Energy flexibility of residential buildings using short term heat storage in the thermal mass. Energy 2016;111:991–1002. http://dx.doi.org/10.1016/j.energy.2016.05.076.

[46] Masy G, Georges E, Verhelst C, Lemort V, André P. Smart grid energy flexible buildings through the use of heat pumps and building thermal mass as energy storage in the Belgian context. Sci Technol Built Environ 2015;21:800–11. http://dx.doi.org/10.1080/23744731.2015.1035590.

[47] Kathirgamanathan A, Péan T, Zhange K, Rosa MD, Salom J, Kummert M, et al. Towards standardising market-independent indicators for quantifying energy flexibility in buildings. Energy Build 2020;220. http://dx.doi.org/10.1016/j.enbuild.2020.110027.

[48] Finck C, Lib R, Kramer R, Zeilera W. Quantifying demand flexibility of power-to-heat and thermal energy storage in the control of building heating systems. Appl Energy 2018;209:409–25. http://dx.doi.org/10.1016/j.apenergy.2017.11.036.

[49] Zhou Y, Cao S. Quantification of energy flexibility of residential net-zero-energy buildings involved with dynamic operations of hybrid energy storages and diversified energy conversion strategies. Sustain Energy Grids Netw 2020;21. http://dx.doi.org/10.1016/j.segan.2020.100304.

[50] Sadat-Mohammadi M, Nazari-Heris M, Nazerfard E, Abedi M, Asadi S. Intelligent approach for residential load scheduling. IET Gener Transm Distrib 2020;14(21):4738–45. http://dx.doi.org/10.1049/iet-gtd.2020.0143.

[51] Finck C, Li R, Zeiler W. Economic model predictive control for demand flexibility of a residential building. Energy 2019;176:365–79. http://dx.doi.org/10.1016/j.energy.2019.03.171.

[52] D'hulst R, Labeeuw W, Beusen B, Claessens S, Deconinck G, Vanthournout K. Demand response flexibility and flexibility potential of residential smart appliances: Experiences from large pilot test in Belgium. Appl Energy 2015;155:79–90. http://dx.doi.org/10.1016/j.apenergy.2015.05.101.

[53] Balint A, Kazmi H. Determinants of energy flexibility in residential hot water systems. Energy Build 2019;188–189:286–96. http://dx.doi.org/10.1016/j.enbuild.2019.02.016.

[54] Haben S, Ward J, Greetham DV, Singleton C, Grindrod P. A new error measure for forecasts of household-level, high resolution electrical energy consumption. Int J Forecast 2014;30(2):246–56. http://dx.doi.org/10.1016/j.ijforecast.2013.08.002.

[55] Beccali M, Cellura M, Brano VL, Marvuglia A. Short-term prediction of household electricity consumption: Assessing weather sensitivity in a Mediterranean area. Renew Sustain Energy Rev 2008;12(8):2040–65. http://dx.doi.org/10.1016/j.rser.2007.04.010.

[56] Alobaidia MH, Chebana F, Meguida MA. Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. Appl Energy 2018;212:997–1012. http://dx.doi.org/10.1016/j.apenergy.2017.12.054.

[57] Opitz D, Maclin R. Popular ensemble methods: An empirical study. J Artificial Intelligence Res 1999;11. http://dx.doi.org/10.1613/jair.614.

[58] Kolter JZ, Maloo MA. Dynamic weighted majority: An ensemble method. J Mach Learn Res 2007;8:2755–90.

[59] Fan C, Xiao F, Wang S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Appl Energy 2014;127:1–10. http://dx.doi.org/10.1016/j.apenergy.2014.04.016.

[60] Jovanovic RZ, Sretenovic AA, Zivkovic BD. Ensemble of various neural networks for prediction of heating energy consumption. Energy Build 2015;94:189–99. http://dx.doi.org/10.1016/j.enbuild.2015.02.052.
[61] Wang R, Lu S, Feng W. A novel improved model for building energy consumption prediction based on model integration. Appl Energy 2020;262. http://dx.doi.org/10.1016/j.apenergy.2020.114561.
[62] Li H, Wang Z, Hong T, Piette M. Energy flexibility of residential buildings: A systematic review of characterization and quantification methods and applications. Adv Appl Energy 2021;3. http://dx.doi.org/10.1016/j.adapen.2021.100054.
[63] Waseem M, Lin Z, Yang L. Data-driven load forecasting of air conditioners for demand response using Levenberg–Marquardt algorithm-based ANN. Big Data Cogn Comput 2019. http://dx.doi.org/10.3390/bdcc3030036.
[64] Shamisa A, Majidi B, Patra JC. Sliding-window-based real-time model order reduction for stability prediction in smart grid. IEEE Trans Power Syst 2019;34(1):326–37. http://dx.doi.org/10.1109/TPWRS.2018.2868850.
[65] Gómez L, Martínez AO, Pastoriza FT, Garrido LF, Álvarez EG, García JAO. Photovoltaic power prediction using artificial neural networks and NumericalWeather data. Sustainability 2019;12. http://dx.doi.org/10.3390/su122410295.
[66] Pirbazari AM, Chakravorty A, Rong C. Evaluating feature selection methods for short-term load forecasting. In: IEEE international conference on big data and smart computing. 2019, http://dx.doi.org/10.1109/BIGCOMP.2019.8679188.
[67] Gibbons J, Chakraborti S. Nonparametric statistical inference. Routledge, Taylor and Francis; 2020.
[68] Evans JD. Straightforward statistics for the behavioral sciences. Brooks/Cole Publishing Company; 1996.
[69] Kuhn M, Johnson K. Applied predictive modeling. New York, NY: Springer; 2013, http://dx.doi.org/10.1007/978-1-4614-6849-3.
[70] Alpaydin E. Introduction to machine learning. Adaptive computation and machine learning series, 2nd ed.. The MIT Press; 2009, http://dx.doi.org/10.1007/978-1-4614-6849-3.
[71] Bishop CM. Pattern recognition and machine learning. Springer; 2006.
[72] Kamel E, Sheikh S, Huang X. Data-driven predictive models for residential building energy use based on the segregation of heating and cooling days. Energy 2020;206:5. http://dx.doi.org/10.1016/j.energy.2020.118045.
[73] Bergstra J, Bengio T. Random search for hyper-parameter optimization. J Mach Learn Res 2012;13:281–305.
[74] Amber KP, Aslam MW, Mahmood A, Kousar A, Younis MY, Akbar B, et al. Energy consumption forecasting for university energy consumption forecasting for university. Energies 2017;10. http://dx.doi.org/10.3390/en10101579.
[75] Antoniadis A, Poggi JM, Brossat X. Modeling and stochastic learning for forcasting in high dimensions. Lecture notes in statistics, Springer; 2013.
[76] Cesa-Bianchi N, Lugosi G. Prediction, learning, and games. Cambridge University Press; 2006.
[77] Littlestone N, Warmuth M. The weighted majority algorithm. Inform and Comput 1994;8(2):212–61. http://dx.doi.org/10.1006/inco.1994.1009.
[78] Measurement of energy, demand, and water savings. ASHRAE Guideline 14-2014.
[79] Coakley D, Raftery P, Keane M. A review of methods to match building energy simulation models to measured data. Renew Sustain Energy Rev 2014;37. http://dx.doi.org/10.1016/j.rser.2014.05.007.
[80] Thermal environmental conditions for human occupancy. 2017, ANSI/ASHRAE Standard 55-2017.
[81] Pallonetto F, Oxizidis S, Milano F, Finn DP. The effect of time-of-use tariffs on the demand response flexibility of an all-electric smart-grid-ready dwelling. Energy Build 2016. http://dx.doi.org/10.1016/j.enbuild.2016.06.041.
[82] Pallonetto F. Demand response algorithms for home area networks utilising forecasted meteorological data and utility demand-supply profiles (Ph.D. thesis), University College Dublin; 2017.
[83] Buttitta G, Turner WJN, Neu O, Finn DP. Development of occupancy-integrated archetypes: Use of data mining clustering techniques to embed occupant behaviour profiles in archetypes. Energy Build 2019;198:84–99. http://dx.doi.org/10.1016/j.enbuild.2019.05.056.
[84] Buttitta G, Finn DP. A high-temporal resolution residential building occupancy model to generate high-temporal resolution heating load profiles of occupancy-integrated archetypes. Energy Build 2020;206(2020). http://dx.doi.org/10.1016/j.enbuild.2019.109577.
[85] Kapetanakis DS, Mangina E, Finn DP. Input variable selection for thermal load predictive models of commercial buildings. Energy Build 2017;137:13–26. http://dx.doi.org/10.1016/j.enbuild.2016.12.016.
[86] Pedregosa, et al. Scikit-learn: Machine learning in Python. JMLR 2011;12:2825–30.