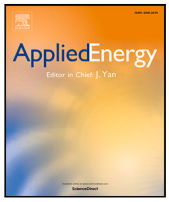
Applied Energy 348 (2023) 121576

Contents lists available at ScienceDirect

Applied Energy

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



A Bayesian deep-learning framework for assessing the energy flexibility of residential buildings with multicomponent energy systems

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

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

ARTICLE INFO

Keywords:
Energy flexibility
Flexibility indicators
Residential sector
Bayesian deep-learning
Probabilistic forecasting

ABSTRACT

This paper addresses the challenge of assessing uncertainty in energy flexibility predictions, which is a significant open question in the energy flexibility assessment field. To address this challenge, a methodology that quantifies the flexibility of multiple thermal and electrical systems is developed using appropriate indicators and considers the different types of uncertainty associated with building energy use. A Bayesian convolutional neural network is developed to capture aleatoric and epistemic uncertainty related to energy conversion device operation and temperature deviations resulting from exploiting building flexibility. The developed prediction models utilise residential occupancy patterns and a sliding window technique and are periodically updated. The energy systems evaluated include a heat pump, a photovoltaic system, and a stationary battery, and use synthetic datasets obtained from a calibrated physics-based model of an all-electric residential building for two occupancy profiles. Simulation results indicate that building flexibility potential predictability is influenced by weather conditions and/or occupant behaviour. Furthermore, the day-ahead and hour-ahead prediction models show excellent performance for both occupancy profiles, achieving coefficients of determination between 0.93 and 0.99. This methodology can enable electricity aggregators to evaluate building portfolios, considering uncertainty and multi-step predictions, to shift electricity demand to off-peak periods or periods of excess onsite renewable electricity generation in an end-user-customised manner.

1. Introduction

1.1. Motivation

Electricity systems require balancing supply and demand within strict real-time limits, which is traditionally achieved by ensuring that the electricity generation is sufficient to cover the projected demand, plus a margin for prediction errors and equipment failures [1]. The wide penetration of renewable energy sources (RES), often intermittent, into the electricity grid, along with the electrification of heat and transport, will provide further challenges to the power system [2]. Smart meter rollout combined with new information and communications technology as well as remote load control technologies have highlighted the role that the demand-side can play to improve the operation of electricity markets [3]. Buildings may play a key role in the future smart electricity grid as they account for approximately 36% of global energy consumption [4]. Moreover, the increasing electrification of buildings by the wider adoption of heat pumps, on-site electricity generation, storage technologies, and electric vehicles is likely to make them promising flexibility resources. IEA Annex 67 defines the energy flexibility of buildings as the ability to manage their demand and generation according to local climate conditions, user needs and grid requirements. Building energy flexibility will thus allow for demand-side management/load control and thereby demand response (DR) based on the requirements of the surrounding grids and the availability of RES to minimise the CO_2 emissions [5].

Building energy performance evaluation is characterised by increased uncertainty due to various factors including occupancy patterns and occupant behaviour, building operation strategies (e.g., lighting control, plug-in equipment control, supply air temperature control, etc.), internal gains, infiltration rates, weather conditions, HVAC system uncertainty, indoor air temperature distributions etc. [6,7]. Uncertainty can be distinguished into two categories, namely, aleatoric and epistemic uncertainty [8]: aleatoric uncertainty is a result of inherently random or natural processes associated with the system under study whereas epistemic uncertainty is caused by the lack of knowledge [8,9]. Based on the aforementioned definitions, the parameters influencing

* Corresponding author.
E-mail address: adamantios.bampoulas@ucd.ie (A. Bampoulas).

https://doi.org/10.1016/j.apenergy.2023.121576
Received 18 April 2023; Received in revised form 15 June 2023; Accepted 8 July 2023
Available online 22 July 2023

building energy analysis can be categorised with respect to the nature of uncertainty. For example, aleatoric uncertainties include weather variables because of their random nature whereas epistemic uncertainties include internal gains (due to lighting, use of appliances etc.) and infiltration rates [8]. As opposed to aleatoric uncertainty, epistemic uncertainty stems from ignorance, thus, it is potentially reducible on the basis of additional data. In the context of residential building energy analysis, additional information includes, inter alia, household size, income, appliance inventory, usage information, etc. [10], however, this data can be difficult to acquire and is often limited by user privacy issues. Uncertainty can be introduced in building energy analysis by considering various methods based on various criteria, such as data availability and uncertainty nature. Data availability classification identifies data associated with the statistics of the uncertain input variable. Measured data can be used to directly identify probability distribution functions related to aleatory uncertainty or fuzzy sets associated with epistemic uncertainty. In the absence of measured data, input uncertainty is quantified using previous literature. The nature of uncertainty classification separates the parameters that cause aleatory uncertainties from those that cause epistemic uncertainties [8].

Deterministic building energy analysis, and, hence, building flexibility assessment are based on a single output per time step, thus, the associated building flexibility potential characteristics (e.g., energy shifting capability, energy costs, response time, etc.) depend only on the expected values. The multiple uncertainties induced by the various energy systems used, as well as the lack of sufficient data can render deterministic building flexibility evaluation obsolete due to the false sense of validity in the estimated values. In this respect, evaluation of energy flexibility within prediction intervals will enhance the robustness of prediction results and ultimately provide uncertainty bounds for decision-making. Uncertainty quantification in building performance analysis can be broadly categorised into inverse and forward methods [7]. Forward uncertainty is quantified by assessing the output uncertainty with respect to the input uncertainty, whereas inverse uncertainty is assessed by obtaining input uncertainty from measurement data [7]. Inverse uncertainty quantification approaches can be categorised into Bayesian and frequentist methods [7]. Bayesian approaches include a series of methodologies to estimate posterior parameter distributions, for a given model and data by estimating model parameters and by quantifying the associated uncertainty [11]. Bayesian deep learning (BDL) in particular allows for uncertainty representation and reliable prediction resulting in more interpretable deep neural networks in the context of probability theory [12]. Bayesian deep learning neural networks (BDL NN) can capture both aleatoric and epistemic uncertainties in building energy performance, making them suitable for probabilistic load analysis. Dropout can be applied in the weight layers to approximate the posterior density using a Monte Carlo sampling approach, creating an ensemble of estimates. This ensemble of individual models can provide superior prediction performance compared to single models, and the mean of the posterior process can be used as the prediction function with the resulting variance serving as a measure of model uncertainty [13].

The energy flexibility of buildings is likewise uncertain because it depends not only on the modulated and reference demand but also on the DR programmes, the various energy systems of the installation as well as the onsite electricity generation — if it is available [14]. The energy flexibility evaluation across different households, in particular, is characterised by increased uncertainty due to the diversity in appliance types as well as interactional behaviours [15] that are likely to compromise system reliability by generating another unforeseen peak (i.e., a rebound effect) [16]. To tackle these challenges, it is desirable to identify and engage end-users that exhibit high flexibility potential [17]. It is therefore necessary to assess the energy flexibility of the various energy conversion systems available in residential buildings on an integrated common basis and in a scalable and end-user tailored manner. This would allow electricity aggregators to evaluate residential building portfolios within both the prequalification process and building operation. Moreover, load prediction uncertainty can compromise the ability of DR strategies to attain their optimisation targets, thus, significantly reducing the associated cost saving benefits [18]. In this respect, in order for DR strategies to achieve adequate performance, the impacts of load prediction uncertainty should be also considered [19]. Further to this, uncertainty quantification is likely to bolster engineering decision making by allowing the comparison across different analysis options (e.g., energy flexibility measures) as well as the allocation of the pertinent resources for uncertainty reduction through additional data collection and/or model refinement [20].

Although physics-based models demonstrate excellent performance in predicting building energy consumption, they require building energy simulation expertise, detailed knowledge of building physical properties and access to this information that can be challenging to obtain, especially for existing buildings [12]. Buildings are subject to uncertainties and variabilities, including occupant behaviour, weather conditions, and equipment performance variations. Physics-based models rely on detailed knowledge of building physical properties and assumptions, which may become redundant or outdated over time and face difficulties in accurately incorporating uncertainties without manual intervention. Additionally, these models demand extensive calibration and validation efforts, which are time-consuming and resource-intensive. In recent years, there has been a significant increase in the availability of building-related data, including sensor data, energy consumption data, and building operational data. With the recent emergence of data-driven models, such models, once developed and calibrated can potentially leverage this wealth of data to train and improve their performance. Such data availability now makes the use of data driven models more feasible for specific applications and thus it is an active area of current research [21]. Moreover, data-driven models continuously learn from updated data, adapt to changing conditions, and capture underlying patterns to make accurate predictions and optimisations considering uncertainty [22]. Specifically, Bayesian deep learning can facilitate uncertainty quantification by creating estimates based on model parameter sampling without the need to resort to multiple input scenarios [12]. Bayesian deep learning methodologies exhibit higher generalisation ability even when smaller datasets are used through the prior distribution over the model parameters [23].

1.2. Related research and contributions

Despite the growing interest in uncertainty quantification by using Bayesian deep learning approaches in various fields (e.g., image processing, medical applications, etc.) [24] few research efforts have developed BDL approaches for electric load or electricity generation forecasting. In [25], Bayesian neural networks and Gaussian process models are developed to forecast a series of annual building energy performance outputs, including, inter alia, HVAC system related variables, PV generation, interior equipment, etc., considering only epistemic uncertainty. Lork et al. [26] develop a Bayesian convolutional neural network (CNN) to capture the aleatoric uncertainties associated with residential air conditioning energy management. On the other hand, Xuan et al. [27] quantify the aleatoric uncertainties related to building electricity consumption by developing a hybrid network based on a CNN and a GRU. In [14,28], day-ahead prediction models were developed to capture both aleatoric and epistemic uncertainty based on Bayesian mixture density network and deep residual networks, respectively. In [29], a series of deep learning algorithms are developed (e.g., concrete dropouts, deep ensembles, Bayesian neural networks, deep Gaussian processes, and functional neural processes) to develop hour-ahead and day-ahead prediction models for electric loads considering both aleatoric and epistemic uncertainty. Sun et al. [12] developed day-ahead net load prediction models to capture both epistemic and aleatoric uncertainty based on Bayesian deep long short-term

memory (LSTM) network. In [30], three deep neural networks (recurrent neural network, LSTM, and gated recurrent unit GRU) are utilised to provide day-ahead predictions and quantify both epistemic and aleatoric uncertainties for building electric loads. Yang et al. [31] develop a multitask Bayesian neural network to quantify both aleatoric and epistemic uncertainties associated with building electricity consumption. The above-mentioned studies develop BDL methodologies to capture either aleatoric [26,27] or epistemic uncertainty [25] or focus on electric load forecasting without extending their analysis to individual energy conversion systems commonly found in buildings, namely thermostatically controlled loads, non-controllable loads, onsite electricity generation, etc. [12,14,28–31]. Thus, there is a need for the development of suitable probabilistic models on an individual energy system level which will allow, not only the evaluation of building energy performance under uncertainty, but also the exploitation of the associated flexibility from an integrated system perspective considering the various uncertainty types. The aforementioned research efforts have utilised Bayesian deep learning neural networks to forecast building energy use, however, without extending their analysis to the estimation of building flexibility potential.

Previous work has predominantly focused on implicit energy flexibility evaluation strategies by determining the contribution of different control schemes to achieving specific objectives, e.g., cost or/and carbon emissions minimisation, etc., [32–34]. On the other hand, a series of studies explicitly assess energy flexibility by using standardised methodologies that are independent of the underlying load control algorithms and the accompanying optimisation goals [35]. These methodologies are mainly based on either physics-based or machine learning-based models and focus on passive thermal energy storage (TES) systems; conversely, fewer research efforts quantified the flexibility of active TES or combinations of passive and active TES, as well as electrical storage systems. For example, in [36], the flexibility of passive TES systems is assessed by quantifying the energy that can be added or curtailed and the associated energy costs/savings with respect to the energy shifted based on various building types and simulation software. Foteinaki et al. [37] evaluate the flexibility of passive TES systems considering the energy shifting capability as well as the associated energy costs/savings during a DR action by utilising building energy simulation software. Zhang and Kummert [38] assessed the energy flexibility of passive TES systems by quantifying the flexible energy, the rebound energy, the flexible energy efficiency, and the maximum flexible power based on a grey-box model. Finck et al. [39] evaluate the active TES flexibility by calculating the energy shifting capability, the related energy costs/savings with reference to the energy shifted, as well as the electricity cost by utilising a MATLAB-based simulation framework. Zhou and Cao [40] assessed the flexibility of passive and active TES systems by calculating the period during which the energy consumption can be delayed or anticipated based on a physics-based model that also considers onsite electricity generation. Balint and Kazmi [41] used the indicators developed in [42] to evaluate the flexibility of passive TES by developing a stochastic gradient descent model. In [43], the passive TES flexibility is optimised considering a series of indicators, namely a flexibility factor as well as load and supply cover factors by developing a model predictive control based on neural networks. In [44], the net energy flows from/into the grid and the associated costs are quantified by optimising the operation schedules of electric vehicles and wet appliances based on a multilayer perceptron neural network. Ren et al. [45] utilise a physics-based model to evaluate the flexibility of a solar-assisted air conditioning system with thermal energy storage by quantifying the PV self-consumption and self-sufficiency as well as the heat pump solar contribution. In [46], the flexibility index is utilised in order to evaluate the flexibility potential of passive and active TES systems based on a grey box model.

The aforementioned research efforts have focused on the development of suitable indicators to quantify building energy flexibility. However, these studies have mostly used deterministic approaches without considering uncertainties associated with building operation. This can lead to misestimation of energy shifting capability and thermal comfort deviations, resulting in extra costs and occupant discomfort [47]. Therefore, there is a need for probabilistic key performance indicators (KPIs) that can account for uncertainty and determine confidence levels for flexibility characteristics in residential buildings. These metrics will allow for the comparison and evaluation of different flexibility options and the uncertainty levels of individual DR strategies.

In contrast to these studies, a limited number of studies utilise market-independent indicators to quantify building flexibility considering uncertainty, however, the solely focus on passive TES systems. Hu and Xiao [6] evaluate building flexibility by introducing a probabilistic energy flexibility indicator based on a low order model; however, this considers only occupancy profile stochasticity to quantify the uncertainty associated with building flexibility. Martinez et al. [48] quantify the average shiftable power of passive TES systems considering uncertainty in a residential district by permuting various building typologies, occupant behaviours, and appliances based on linear regression. On the other hand, Amadeh et al. [49] evaluate the energy flexibility of passive TES in terms of storage capacity and efficiency and quantify uncertainty by analysing historical weather data and by placing a white noise distribution on the reduced-order parameters. Tables A.1 and A.2 (Appendix A) provide a concise summary of all reviewed papers, offering a useful overview of the literature references cited in the current paper.

There is currently no scalable methodology to evaluate the flexibility of individual residential buildings with multiple thermal and electrical systems on an integrated basis, considering uncertainty types related to energy demand and onsite electricity generation. There is also a lack of suitable KPIs to provide confidence intervals for flexibility characteristics. A data-driven, bottom-up approach for estimating building energy use and associated uncertainty can facilitate customised flexibility assessment across residential buildings, providing more realistic predictions of building flexibility potential. Probabilistic building flexibility assessment can facilitate the quantification of associated uncertainties and provide a framework for evaluating short-term flexibility prediction robustness and reliability. An uncertainty-aware flexibility evaluation methodology can provide confidence bounds for DR strategies, facilitating decision-making and optimal bidding in electricity markets [12]. Accurate flexibility evaluation can facilitate the selection of residential buildings for flexibility programs during the prequalification process and the implementation of optimal DR strategies during building operation.

To address the lack of bottom-up approaches for evaluating building flexibility by considering uncertainty, a probabilistic energy flexibility assessment framework is developed to evaluate the DR potential of both thermal and electrical systems commonly found in residential buildings. The proposed framework is capable of quantifying both epistemic and aleatoric uncertainties associated with the operational flexibility of buildings considering multistep-ahead predictions. 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 periodically updated models, a sliding window method, and Bayesian deep learning neural networks. The data-driven models developed utilise input variables that can be realistically collected by a residential energy management system (e.g., weather prediction data and historical data related to the target variables). The final set of input variables is selected based on the Spearman correlation coefficient. Subsequently, for each target variable, a Bayesian feedforward convolutional neural network [50–52] is developed by using the Keras Tensorflow framework [53]. The performance of the developed machine-learning models is examined through multiple case studies including hour-ahead day-ahead predictions, which can be utilised to schedule DR measures in response to grid signals. The effectiveness of these forecasting techniques for specific DR programs, such as day-ahead scheduling and secondary reserve

time resolution, is evaluated by comparing both their deterministic and probabilistic performance [54].

Given that multiple predictions are built at each time step based on the Monte Carlo sampling approach, suitable flexibility indicators should be used to account for the uncertainty associated with the energy flexibility potential of passive TES and electric energy storage systems. Specifically, the flexibility quantification framework developed involves the estimation of both the energy shifting capability of the energy systems considered and any zone temperature deviations arising from activating this flexibility. To this end, the notion of storage capacity is extended beyond its deterministic definition and is quantified in terms of the associated mean and variance values. In this way, the flexibility of various thermal and electrical systems is evaluated on a common basis by also considering the different uncertainties characterising building energy consumption and onsite electricity generation. The proposed Bayesian deep learning framework allows for the comparison of flexibility options in residential buildings by considering both inherent randomness and lack of knowledge related to building demand and onsite electricity generation. It also allows for the development of prediction models for individual energy systems and zone temperatures to quantify building flexibility potential and the potential impact on occupant thermal comfort. The proposed KPIs are independent of control strategies and market structure, making them suitable for use in dynamic environments such as DR applications. This framework can be used by electricity aggregators to evaluate building portfolios and the uncertainty involved, leading to more accurate estimates of building flexibility potential. The proposed methodology can not only quantify the energy shifting capability and the potential thermal comfort deviations but also provide a framework for capturing the pertinent uncertainties. It allows for the evaluation or optimisation of building portfolios within prediction intervals, both during the end-user prequalification process and during building operation. It can also be used to optimise the exploitation of flexibility potential in various energy systems to shift demand to off-peak periods or periods of excess onsite electricity generation, increase the share of renewable energy, and mitigate potential electricity generation and distribution capacity issues. The remainder of this paper is organised as follows: Section 2 describes the overall methodology. The case study building as well as its energy systems are described in Section 3, while Section 4 includes the obtained simulation results. Finally, a discussion on the obtained results and the conclusions are provided in Sections 5 and 6, respectively.

2. Methodology

Section 2 summarises the methodological steps of this study. Specifically, Section 2.1 describes the updating strategy of the Bayesian convolutional neural network while Section 2.2 describes the candidate features for each target variable considered as well as the feature selection methodology. Section 2.3 outlines the structure of the developed BCNN, the optimisation of the hyperparameters as well as the data preprocessing techniques utilised. Section 2.4 describes the quantification of the epistemic and aleatoric uncertainties by using the BCNN, while Section 2.5 summarises the deterministic and the probabilistic predictive performance metrics utilised in this study. Finally, Section 2.6 describes the energy flexibility quantification and characterisation methodology by considering both deterministic and probabilistic approaches.

2.1. Training and test set determination

The target variable values for a specified prediction horizon are forecasted considering the sliding or moving window method [55]. This approach involves the prediction of the next m values of a time series, considering the previous n observations as model inputs. After each prediction, the oldest m observations are removed from the rear of the window and the m most recent values (collected from real-time data) are added to the front of the window so that the training set size remains equal to n. This method allows the machine learning (ML) models to be updated by considering real-time observations and by omitting older observations that are likely to be less relevant. Accordingly, the test set size is determined by the selected prediction horizon (hour-ahead, day-ahead, etc.) which depends on the DR market and the bidding strategy.

2.2. Input variable selection

This section describes the candidate features for each target variable (summarised in Table 1) as well as the feature selection method. The definitions of the variables utilised in this study are summarised in the Nomenclature table.

The input variables considered for the HVAC system and the zone temperature include weather conditions (outdoor temperature (Tout), total solar irradiance (I_{out}), relative humidity (RH), wind speed (WS)), the thermostatic setpoint (T_{sp}), the remainder of the building loads (P_{nc}), calendar data, statistical properties associated with previous prediction figures as well as lagged values associated with the thermostatic setpoint, the target variables, and the weather variables. Calendar information (e.g., min of the day (MoD), day of the week (DoW)), non-controllable load predictions, and historical data are used as candidate features since they potentially exhibit a high correlation with the target variables due to the building occupancy pattern periodicity and, therefore, energy consumption. Furthermore, lag terms associated with weather variables are also considered as candidate features. This is because there is a degree of inertia in the heating load and the zone temperature due to the building thermal mass and furnishings when weather conditions change. Moreover, a series of statistical properties from the previous prediction figures (e.g., average, minimum, and maximum values) and the ratios associated with these variables are considered as candidate input variables since they can provide information about the target variable dynamics. Finally, the thermostatic setpoint is considered not only because it gives insights into occupancy patterns and occupant thermal comfort but also because it can be suitably modulated to harness the HVAC system flexibility potential.

The non-controllable loads (all building loads excluding the heating system power consumption) are predicted based on historical data, calendar information as well as statistical properties from the previous prediction figures and ratios involving these variables. The candidate features selected for the PV system electricity generation P_{pv} include historical data from the target variable, weather variables, and calendar information. Variables associated with weather conditions are selected since the PV plant power output is mainly influenced by solar irradiation as well as other variables including outdoor temperature and humidity RH [56]. Historical data of the target variable are considered candidate features to consider the dynamics and potential performance degradation of the PV system. Finally, to account for the periodicities in the PV power output, the time of day is also considered.

Since building load is influenced by several factors (e.g., occupancy, weather, etc.), features should be periodically selected. Specifically, the updated training set is determined for each new prediction by considering the sliding window method. Subsequently, the optimal feature set is selected based on the training set and target variable values that are currently available. In this study, a correlation analysis is applied to determine the optimal feature set. The Pearson or the Spearman correlation coefficients can be utilised to determine the strength and the direction of the correlation between the input and the target variables. The Pearson coefficient assumes normal data distribution to obtain the linear relationship between two continuous variables. Alternatively, the Spearman coefficient can capture the non-linear relationship between two variables, thus, it is suitable for data that is not normally distributed. In this study, the input variables are selected by using the Spearman coefficient [57] since the candidate input variables include

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
Calendar information	WT, DoW, MoD			MoD
Historical data	P_{hp} , T_z, T_z, T_{out}, T_{sp}, $P_{hp,O/O}$		P_{nc}	P_{pv}
Auxiliary variables	$P_{hp,av}$, $P_{hp,max}$, $P_{hp,min}$, $R_{hp,av/max}$, $R_{hp,min/av}$, $T_{z,av}$, $T_{z,max}$, $T_{z,min}$, $R_{z,av/max}$, $R_{z,min/av}$		$P_{nc,av}$, $P_{nc,max}$, $P_{nc,min}$, $R_{nc,av/max}$, $R_{nc,min/av}$	–
Miscellaneous	T_{sp}, P_{nc}, T_z	T_{sp}, P_{nc}, T_z	–	–

data that is not necessarily normally distributed (e.g., relative humidity, solar radiation, wind speed). In the context of this study, the selected correlation threshold is 0.5, considering the absolute value guide as per [58].

2.3. Bayesian convolutional neural network

The machine learning model developed to provide predictions and to quantify the uncertainties for the various target variables considered is a convolutional feedforward neural network (CNN) implemented by using the Keras Tensorflow framework [53]. Convolutional neural networks have been successfully used in various applications, including, inter alia, time series prediction (electrocardiogram time series, weather forecast, traffic flow prediction, etc.), signal identification, image classification, object detection, face recognition, human action recognition, etc. [50]. One of the promising forthcoming approaches for dealing with high-dimensional data is to learn the latent input representation automatically through output supervision. This is the fundamental principle of deep neural networks, notably convolutional neural networks, which stack (deeper) layers of linear convolutions with nonlinear activations to automatically extract the multi-scale features or concepts from high-dimensional input. This reduces the need for manual feature engineering, such as looking for the appropriate set of basis functions or relying on expert knowledge [59]. In convolutional feedforward NNs, the information flows from any layer to all subsequent layers. The lth layer obtains the feature maps from the previous layers as input such that $x_l = H_l\left(\left[x_0, x_1 \ldots, x_{l-1}\right]\right)$, where $[x_0, x_1 ..., x_{(l-1)}]$ represents the feature maps from layers 0 to $l-1$ [51]. A rectified linear function (ReLU) is utilised to activate the BCNN layers. The hyperparameters of the Bayesian convolutional neural network (BCNN) are tuned based on a random search. By utilising this method, a randomised search across hyperparameters over all possible parameter values is conducted. A random search is iteratively carried out until the ML model achieves the desired accuracy or the predetermined computational budget is exhausted [60]. The hyperparameter selection is based on a time series cross-validation (CV) technique that is applied on a rolling basis. Given that the CV method follows the model training strategy (sliding window), the training and test sets slide forward in steps equal to the test set size. The hyperparameters tuned in this study include the number of hidden layers, the number of hidden neurons in each layer, the dropout rate, the batch size, the number of epochs, and the window size. Moreover, features with higher values can influence the development of the ML models, thus deteriorating their accuracy [61]. This problem is addressed by normalising all variables within the range of [0, 1]. Before de-normalising the prediction output, forecasted values that lie below 0 or above 1 are set equal to 0 and 1, respectively, since they are prediction errors. Table 2 summarises the methodological steps for the development of the BCNN model.

2.4. Uncertainty quantification

To capture both epistemic and aleatoric uncertainties for the target variables considered, a Bayesian convolutional neural network (BCNN) is developed. Specifically, epistemic uncertainty is modelled by considering a prior distribution for the weights of the model and by capturing the variation of these weights. On the other hand, aleatoric uncertainty is modelled by considering a distribution over the model output by learning the variance of the models as a function of the various model inputs [62].

The developed BCNN captures the epistemic uncertainty by placing a Gaussian prior distribution over the model weights. For a given dataset (X_{train}, Y_{train}), the posterior distribution over the weights is calculated $p(W|X_{train}, Y_{train})$ based on Bayesian inference, where $f^w(x)$ is a random output of the BCNN and W denotes the model parameters. It is noted that $X_{\text{train}} \in \mathbb{R}^{N\times d_x}$ and $Y_{\text{train}} \in \mathbb{R}^{N\times 1}$, where N is the sample size and d_x represents the input dimensions. The model likelihood is defined as a Gaussian distribution with the mean determined by the model output and an observation noise σ $(p(Y_{train}|f^w(X_{train})) = \aleph(f^w(X_{train}), \sigma^2))$. Since the marginal probability $p(X_{train}, Y_{train})$ cannot be analytically assessed to evaluate the posterior $p(W|X_{train}, Y_{train}) = p(Y_{train}|X_{train}, W)p(W)/p(Y_{train}, X_{train})$, the inference is implemented by considering the Monte Carlo (MC) dropout. To carry out this inference, the model is trained with dropout before every weight layer; additionally, the dropout is performed at test time based on samples from the approximate posterior [62]. This posterior allows predictions on unseen data by obtaining sets of all possible model parameters [25].

The posterior $p(W|X_{train}, Y_{train})$ is fitted with a distribution $q_\theta(W)$ by minimising the Kullback–Leibler divergence between $q_\theta(W)$ and $p(W|X_{train}, Y_{train})$ so that the optimal distribution $q^*_\theta(W)$ can represent $p(W|X_{train}, Y_{train})$. The minimisation objective (Eq. (1)) is obtained by implementing dropout as a variational Bayesian approximation with a probability equal to p. Considering T samples, the predictive mean and variance are given by Eqs. (2) and (3), respectively [52].

$$\mathfrak{L}(\theta, p) = \frac{1}{2N\sigma^2}\sum_{i=1}^{T}\left\|y_i - f^{\widehat{W}_i(x_i)}\right\|^2 + \frac{\log\sigma^2}{2} + \frac{1-p}{2N}\|\theta\|^2 \quad (1)$$

$$E(y) = \frac{1}{T}\sum_{t=1}^{T} f^{\widehat{W}_t}(x) \quad (2)$$

$$\text{Var}(ep)(y) = \frac{1}{T}\sum_{t=1}^{T} f^{\widehat{W}_t}(x)^T f^{\widehat{W}_t}(x) - \frac{1}{T^2}\sum_{t=1}^{T} f^{\widehat{W}_t}(x)^T \sum_{t=1}^{T} f^{\widehat{W}_t}(x) + \sigma^2 \quad (3)$$

Table 2
Methodology table.

Step	Methodology section	Description
1	2.1	Determine the optimal training set size
2	2.1	Determine the candidate features (Table 1)
3	2.2	Update the training set and the selected features
4	2.3	Build a BCNN for each target variable
5	2.4	Calculate the mean and standarddeviation for each BCNN model

To account for aleatoric uncertainty, it is considered that the observation noise is data dependent. The resulting minimisation objective (Eq. (4)) is derived by expressing the observation noise as a function of the data.

$$\mathfrak{L}(\theta) = \frac{1}{2T}\sum_{i=1}^{T}\frac{1}{\sigma\left(x_i\right)^2}\left\|y_i - f\left(x_i\right)\right\|^2 + \frac{\log\sigma\left(x_i\right)^2}{2} \tag{4}$$

To combine both the epistemic and the aleatoric uncertainty, a prior distribution over the weights and bias is placed. Subsequently, the posterior distribution is inferred based on the given data. To combine both the epistemic and the aleatoric uncertainty, a distribution over the weights is placed. The model output (comprised of both the predictive mean and the predictive variance) is based on the approximate posterior $\hat{W} \sim q(W)$, as given in Eq. (5). Further, a normal likelihood is considered to model the aleatoric uncertainty. The resulting minimisation objective and the predictive variance (composed of both the epistemic and the aleatoric uncertainty) are given in Eqs. (6) and (7), respectively [25].

$$\left[\hat{y}, \hat{\sigma}^2\right] = f^{\widehat{W}}(x) \tag{5}$$

$$\mathfrak{R}_{NN} = \frac{1}{2T}\sum_{i=1}^{T}\frac{1}{\hat{\sigma}_i^2}\left\|y_i - \hat{y}_i\right\|^2 + \frac{\log\sigma_i^2}{2} \tag{6}$$

$$\mathrm{Var}(y) = \frac{1}{T}\sum_{i=1}^{T}\hat{y}^2 - \left(\frac{1}{T}\sum_{i=1}^{T}\hat{y}^2\right)^2 + \frac{1}{T}\sum_{i=1}^{T}\hat{\sigma}_t^2 \tag{7}$$

where and stand for the predicted and actual value i, respectively.

2.5. BCNN performance evaluation

The predictive performance of the BCNN model is assessed by evaluating both the model deterministic accuracy and the accuracy of the uncertainty estimate. Considering the deterministic accuracy evaluation, the coefficient of variation of the root mean squared error (CV-RMSE) (Eq. (8)) and the normalised mean bias error (NMBE) (Eq. (9)) are used for the power-related target variables (HVAC system power consumption, non-controllable loads, and PV system electricity generation). The zone temperature BCNN model is evaluated based on the RMSE (Eq. (10)) and the MBE (Eq. (11)). According to the ASHRAE guidelines [63], the acceptable tolerances for calibration using hourly data for the CV-RMSE and the NMBE are +/−10% and 30%, respectively. Nevertheless, no standardised calibration thresholds have been established for prediction models with higher resolution [64]. The predictability of the energy flexibility related characteristics is assessed by using the coefficient of determination (Eq. (12)).

To measure the percentage of the observed values that fall within the prediction interval (PI), the prediction interval coverage probability (PICP) (Eq. (13)) is utilised. To quantify the width of the prediction interval, the prediction interval normalized mean width (PINMW) (Eq. (14)) and the prediction interval mean width (PIMW) (Eq. (15)) are utilised for the power-related variables and the zone temperature, respectively. The PINMW is a normalised version of the PIMW and is used to compare the PI of target variables exhibiting different magnitudes. The PICP and the PINMW (or PIMW) should be collectively used since the PICP can give misleading results; this is because a high/low PICP can be achieved for wide/narrow PIs [54].

$$CV-RMSE = \frac{\sqrt{\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2}}{\bar{Y}} \tag{8}$$

$$PICP = \frac{1}{N}\sum_{i=1}^{N} c_i, c_i = \begin{cases} 1, l_i \le Y_i \le u_i \\ 0, Y_i < l_i \text{ or } Y_i > u_i \end{cases} \tag{9}$$

$$NMBE = \frac{\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)}{\sum_{i=1}^{N} Y_i} \tag{10}$$

$$RMSE = \sqrt{\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2} \tag{11}$$

$$MBE = \frac{\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)}{N} \tag{12}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2} \tag{13}$$

$$\mathrm{PINMW} = \frac{1}{NR}\sum_{i=1}^{N}\left(u_i - l_i\right) \tag{14}$$

$$PIMW = \frac{1}{N}\sum_{i=1}^{N}\left(u_i - l_i\right) \tag{15}$$

where $\bar{Y}$ is the mean value of the actual values, N and R represent the size of y_i and the range of the measured value, respectively, and l_i and u_i denote the lower and upper bounds of the ith PI, respectively. To assess the significance of each uncertainty type, the PICP and the PIMW, PINMW can be calculated by considering only one type of uncertainty (either aleatoric or epistemic) and the total model uncertainty. These metrics depend on the lower and upper prediction bounds, which are a function of the predictive variance. Thus, the consideration of different uncertainty types (epistemic, aleatoric, total) will result in different PICP and PIMW/PINMW values and also depend on the predictive variance considered. Therefore, $PICP_{ep}$, $PICP_{al}$, and $PICP_{tot}$ correspond to the percentage of the observed values that fall within a prediction interval that results from considering the epistemic, the aleatoric, and the total uncertainty, respectively.

2.6. Energy flexibility evaluation

2.6.1. Deterministic energy flexibility quantification

The building energy flexibility potential is assessed by quantifying the load shifting capability for various energy conversion systems, namely space heating systems and electric storage units. Building flexibility is evaluated by implementing various DR actions by considering locally produced electricity and can be activated on the basis of two types of flexibility: downward (down-flex) and upward flexibility (up-flex) [42]. Energy is curtailed during down-flex actions, whereas energy is stored in the storage medium during up-flex actions by implementing suitable control setpoint modulations. In this context, the energy shifting capability can be quantified by considering the notion of storage capacity. In the context of a DR action, the available storage capacity can be defined as the amount of energy that can be added or removed

Fig. 1. Picture and 3D rendering of testbed house [66].

Table 3
Limits on temperature drifts and ramps [65].

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

by a storage medium based on the respective boundary conditions. The storage capacity for down-flex and up-flex actions is given by Eqs. (16) and (17), respectively.

$$C_{DF} = \int_0^{\infty} \left| P_{\mathrm{mod}}(t) - P_{ref}(t) \right|^{-} dt \tag{16}$$

$$C_{UP} = \int_0^{\infty} \left(P_{\mathrm{mod}}(t) - P_{ref}(t) \right)^{+} dt \tag{17}$$

In down-flex, the HVAC system flexibility is determined by considering on/off strategies, whereas, in up-flex, by applying suitable thermostatic setpoint modulations. To avoid unacceptable occupant thermal comfort deviations during DR actions, the operative temperature changes should remain within certain limits as per ASHRAE standards [65] (Table 3).

To determine the maximum flexibility potential of a battery during a down-flex action, it assumed that the power flow to the utility is zero. Thus, the battery power shifting capability is determined by considering a discharging power equal to the net building load. Thus, a discharging power equal to the net building load ($P_b - P_{RES}$) is considered for a specified duration, t_{DR}, where P_b and P_{RES} are the building load and the onsite electricity generation, respectively. The resulting storage capacity is given by Eq. (18). In contrast to down-flex, during an up-flex action, the battery upward flexibility is calculated by considering a charging power equal to the battery maximum charge rate, $P_{bat,max}$. Given that the battery charging power is constant, the storage capacity in up-flex is equal to $P_{bat,max}.t_{DR}$.

$$C_{DF} = \int_0^{t_{DR}} \max\left(P_b - P_{RES}, 0\right) dt \tag{18}$$

2.6.2. Probabilistic energy flexibility quantification

In this section, the deterministic flexibility indicators, as described in Eqs. (16)–(18), are extended so that the energy shifting capability of the various energy systems considered can be described by prediction intervals. If a DR event with a duration of t_{DR} corresponds to K timesteps, Eq. (16) can be restated as:

$$C_{DF} = \frac{t_{DR}}{K} \sum_{k=1}^{K} P_{HVAC,ref}^{(k)} - \frac{t_{DR}}{K} \sum_{k=1}^{K} P_{HVAC,\,\mathrm{mod}}^{(k)} \tag{19}$$

Regarding the HVAC system, if the reference and the modulated power demand follow Gaussian distributions (as described in Section 2.4) such that $P_{HVAC,mod}^{(k)} \sim N\left(\mu_{HVAC,mod}^{(k)}, \sigma_{HVAC,mod}^{2}\right)$ and $P_{HVAC,ref}^{(k)} \sim N\left(\mu_{HVAC,ref}^{(k)}, \sigma_{HVAC,ref}^{2}\right)$, then the resulting storage capacity in down-flex and up-flex will also follow a Gaussian distribution as per Eqs. (20) and (21), respectively.

$$C_{DF} \sim N\Big(\frac{t_{DR}}{K} \sum_{k=1}^{K} \mu_{HVAC,ref}^{(k)} - \mu_{HVAC,\,\mathrm{mod}}(k), \frac{t_{DR}}{K}^{2} \sum_{k=1}^{K} \sigma_{HVAC,ref}^{2\,(k)} + \sigma_{HVAC,\,\mathrm{mod}}^{2\,(k)}\Big) \tag{20}$$

$$C_{UF} \sim N\Big(\frac{t_{DR}}{K} \sum_{k=1}^{K} \mu_{HVAC,\,\mathrm{mod}}^{(k)} - \mu_{HVAC,ref}(k), \frac{t_{DR}}{K}^{2} \sum_{k=1}^{K} \sigma_{HVAC,\,\mathrm{mod}}^{2} + \sigma_{HVAC,ref}^{2}\Big) \tag{21}$$

where k denotes the corresponding timestep. Assuming that the indoor temperature prediction follows a Gaussian distribution such that $T \sim N\left(\mu_z^{(k)}, \sigma_z^{2(k)}\right)$, the temperature deviation during a DR action will also follow a Gaussian distribution as per Eq. (22).

$$\Delta T \sim N\left(\left|\mu_z^{(k_i)} - \mu_z^{(k_j)}\right|, \sigma_z^{2(k_i)} + \sigma_z^{2(k_j)}\right) \tag{22}$$

where k_i and k_j denote the timesteps for which the quantity $\left|\mu_z^{(k_i)} - \mu_z^{(k_j)}\right|$ is maximised. The battery flexibility potential is quantified by assuming a DR action with a duration of t_{DR} that corresponds to K timesteps. Considering that the building load (P_b) equals the sum of the HVAC load (P_{HVAC}) and the non-controllable loads. Eq. (18) can be rewritten as:

$$C_{DF} = \max\left(\frac{t_{DR}}{K} \sum_{k=1}^{K} P_{HVAC}(k) + \frac{t_{DR}}{K} \sum_{k=1}^{K} P_{nc}^{(k)} - \frac{t_{DR}}{K} \sum_{k=1}^{K} P_{RES}^{(k)}, 0\right) \tag{23}$$

The HVAC system load, the non-controllable loads, and the onsite electricity generation follow Gaussian distributions such that $P_{HVAC}^{(k)} \sim N\left(\mu_{HVAC}^{(k)}, \sigma_{HVAC}^{2}(k)\right)$, $P_{nc}^{(k)} \sim N\left(\mu_{nc}^{(k)}, \sigma_{nc}^{2(k)}\right)$, $P_{RES}^{(k)} \sim N\left(\mu_{RES}^{(k)}, \sigma_{RES}^{2}(k)\right)$, respectively. Eq. (24) provides the storage capacity that corresponds to the distribution of the resulting building net load C_{net}. The positive values of C_{net} correspond to the battery storage capacity, whereas the pertinent negative values correspond to the excess energy amount from renewables during the K step period. In upward flexibility, the battery storage capacity is determined by the associated maximum charging power, and it is therefore constant. In this context,

uncertainty analysis is not applicable to up-flex actions.

$$C_{net} \sim N\Big(\frac{t_{DR}}{K}\sum_{k=1}^{K}\mu_{HVAC}{}^{(k)} + \mu_{nc}^{(k)} - \mu_{RES}(k), \frac{t_{DR}}{K}^2\sum_{k=1}^{K}\sigma^2_{HVAC}(k) + \sigma^{2(k)}_{nc} + \sigma^2_{RES}(k)\Big) \quad (24)$$

3. Case study

The proposed methodology is exemplified by considering an existing all-electric house located in eastern Ireland [66]. In this section, the main energy conversion systems of the virtual testbed along with the adopted occupancy profiles are described. Moreover, this section includes the Bayesian learning model implementation and configuration process.

3.1. Building description and energy conversion components

To evaluate the flexibility of the various energy conversion systems involved, a synthetic database is created from a physics-based model of a residential building developed by using EnergyPlus V.9.1 [67] and calibrated based on onsite measurements. Calibrated white-box models can provide adequate performance data and facilitate the development of ML models without considering issues associated with the size and the quality of sensor-based data collected from buildings [68]. Further to this, the utilisation of synthetic data can facilitate the performance evaluation of existing ML algorithms under varying conditions and case studies [69]. The case study dwelling and the associated modelled geometry are illustrated in Fig. 1 [66]. The thermal envelope, the ground source heat pump, and the PV system were calibrated to meet the ASHRAE criteria [63]. The following subsections include the descriptions of the building physics, the HVAC system, the PV plant, and the stationary battery.

3.1.1. Building description and thermal properties

The considered test-bed is a single-storey detached house, demonstrates a significant passive thermal energy storage capacity as a result of its construction (i.e., two-leaf concrete wall with cavity insulation), and represents approximately 40% of the Irish building stock [66]. The building is comprised of twelve rooms and an uninhabited attic space at roof level (Fig. 2). The building exhibits increased thickness of insulation materials in its opaque elements compared to the contemporary standards. The roof does not include insulation and is covered with slate, while the ceiling is covered with acoustic tiles to ensure both acoustic and thermal insulation. A 200 mm layer of fibreglass insulation with thermal conductivity of 0.04 W/m K is located on top of the acoustic tiles and ensures high thermal resistance. The overall window to wall ratio is 15%, with a 22% and 10% ratio on the south and north facades, respectively. The total surface area of the exterior wall (excluding fenestration) is 187 m^2, while the slate roof surface area is 279 m^2. The U-values of the roof, building floor, walls, and windows are 0.21, 0.21, 0.21, and 1.7 W/m^2 K, respectively. The thermal envelope was calibrated during a period of occupant absence [64]. The NMBE and the CV-RMSE exhibited average values of 4.41% and 3.28%, respectively, thus, lying within acceptable limits as per the hourly ASHRAE standards.

3.1.2. HVAC system

The building space heating system is a ground source heat pump (GSHP) and has a rated thermal output of 12 kW. As illustrated in Fig. 3, the GSHP is coupled with a hot water storage tank of 0.8 m^3 to provide thermal energy storage. The GSHP was calibrated by considering data from the heating season and resulted in average CV-RMSE and NMBE values of 3.78% and −0.61%, respectively [66].

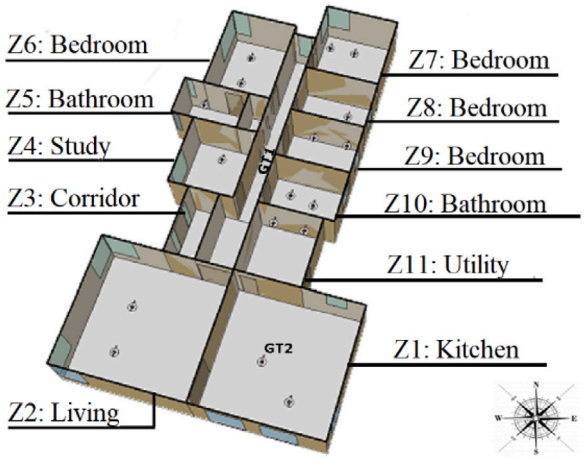


Fig. 2. Representation of the building with the ground floor thermal zones and orientation [66].

3.1.3. PV system

The electrical installation comprises of a south-facing PV system with a nominal power of 6 kW. The PV plant is located 30 metres from the house and is connected to the grid through a single-phase inverter with an efficiency of 95%. The associated model was calibrated based on data for the months February through September and achieved average CV-RMSE and NMBE values of 12.5%, and 3.6%, respectively, which meet the ASHRAE criteria [66].

3.1.4. Stationary battery

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

3.2. Occupancy profiles

In this study, two occupancy profiles are considered that utilise daily average occupancy profiles based on a Time Use Survey [70]. The adopted occupancy profiles use two indicative clusters resulting from the categorisation of residential weekday diaries and represent 23% and 34% of the survey sample [71]. Both occupancy profiles considered are based on weekday and weekend schedules. The weekday schedule of the first occupancy profile (OC1) considers three types of occupant activity (i.e., active, non-active, and absent); specifically, OC1 is characterised by an absence from the dwelling resulting from work attendance (between 08:20 h and 18:10 h). The second occupancy profile (OC2) includes active and non-active household states. The weekend occupant activity is the same for both occupancy profiles. During occupant activity periods, two thermostatic setpoints are adopted (i.e., 20/21 °C), whereas during periods of occupant absence or inactivity, a setback of 16 °C is utilised. The thermostatic setpoints considered for both occupancy profiles are outlined in Table 4.

3.3. Model setting

One heating season (extending from 01 September to 30 April) is considered in order to generate the synthetic database from the virtual testbed. The selected simulation time-step is 15 min, thus, the case study is based on 23,232 data points. Further, the developed dataset is split into the development and the evaluation set with a ratio of 50%/50% [72]. The optimal number of training days for each target variable as well as the hyperparameter optimisation are determined based on the development set, whereas the evaluation set is utilised

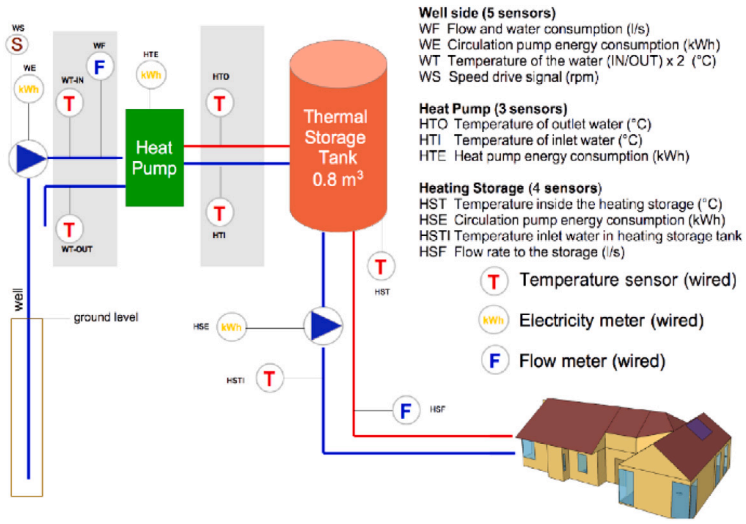


Fig. 3. Heat system design and sensor metering [66].

Table 4
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	1:10–05:40	16

to assess the performance of the various ML models with respect to the predictive accuracy considering cross-validation. To assess the flexibility potential of the various energy system included in the virtual testbed, daily random DR events are applied in the evaluation set. The Bayesian deep learning neural networks are developed by using Python 3.7, the Tensorflow 2.8.0 open-source framework [53], and the scikit-learn library [73] on an INTEL Core i7-7500U processor @2.90 GHz system with 16 GB RAM and an Intel HD Graphics 620 running Windows 10 operating system.

3.4. Energy flexibility mapping

The energy flexibility potential of the virtual testbed is evaluated by calculating the daily energy flexibility mapping based on the selection of one day from each occupancy profile as baseline cases. The latter are selected so that the predictive performance of the BCNN (in terms of RMSE/CV-RMSE, MBE/NMBE, PICP, PIMW/PINMW) for each baseline case (considering all target variables and prediction horizons) is comparable to that of the BCNN models for the evaluation set. The flexibility mappings are created based on consecutive and independent DR actions that are applied, over a 24-hour period, for each energy system considered. The starting times of each DR action are based on an on-the-hour basis, i.e., 00:00 h, 01:00 h, etc. It is assumed that all applied DR actions are independent, so that each individual DR action is not affected by prior DR actions. It is, thus, considered that there is a single baseline consumption curve for each case study; this can be practically achieved by assuming that the period between DR actions is long enough (so that every next DR action is not affected by the potential rebounds resulting from the previous DR actions). The developed flexibility profiles can be utilised to select the most suitable DR strategies considering not only the requested energy amount to be shifted but also the predictability and the uncertainty associated with each DR action.

4. Results

This section describes the assessment of the developed Bayesian deep learning neural networks for various target variables and prediction horizons considered. The BCNN-based approach is capable of evaluating both the deterministic model accuracy (by using the prediction mean) and the probabilistic model accuracy (by using the model uncertainty). Moreover, the flexibility potential of the various building energy conversion systems is evaluated by utilising the developed BDL framework, and it is benchmarked against the physics-based model. Table 5 summarises the structure of th simulation results of this study.

4.1. Predictive performance evaluation

In this section, the predictive performance of the day-ahead and the hour-ahead BCNN is assessed for each target variable (zone temperature, heat pump load, non-controllable loads, PV system electricity generation) by using the evaluation set (Section 3.3). Table 6 summarises the number of hidden layers, the number of hidden neurons in each layer, the dropout rate the batch size, the number of epochs, and the sliding window size (the number of days) for each target variable and prediction horizon considered. Tables 7 and 8 summarise the RMSE/CV-RMSE, the MBE/NMBE, the PIMW/PINMW, and the PICP for each target variable, for day-ahead and hour-ahead predictions, respectively. Figs. 4 and 5 give a graphic overview of the metrics presented in Tables 7 and 8. Specifically, in order to evaluate the effect of each type of uncertainty on each target variable, the PIMW/PINMW, and the PICP are calculated by considering the epistemic, the aleatoric, and the total uncertainties.

Considering the model deterministic accuracy (CV-RMSE/RMSE and NMBE /MBE), hour-ahead prediction models outperform the pertinent day-ahead models for all target variables considered. Furthermore, the CV-RMSE and the NMBE values can be observed to lie within

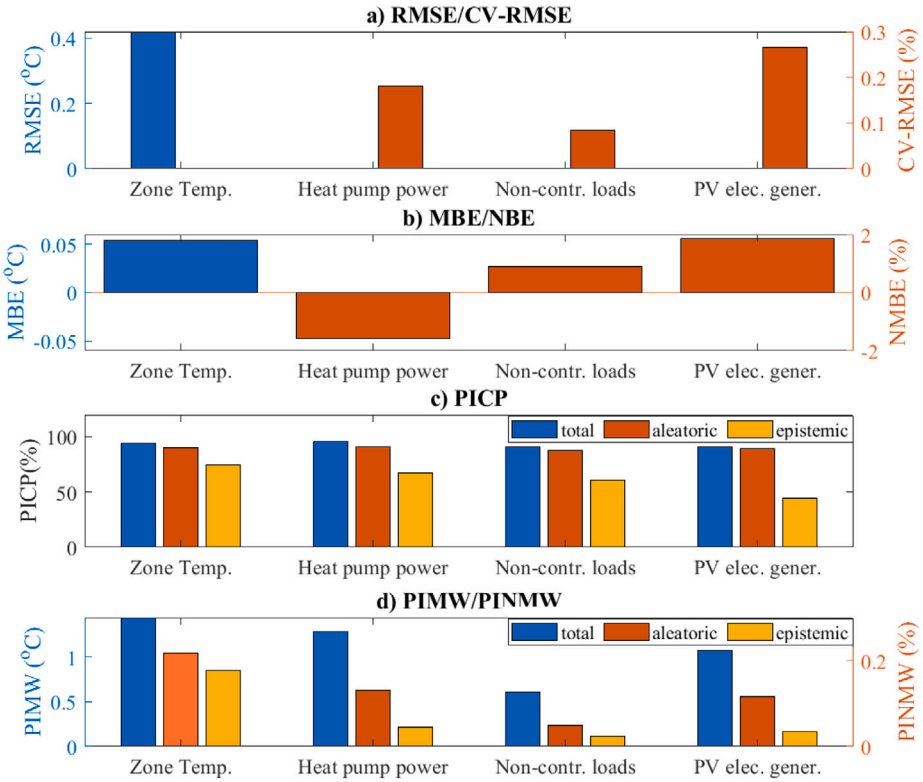


Fig. 4. (a) RMSE/CV-RMSE, (b) MBE/NMBE, (c) PICP, (d) PIMW/PINMW for each target variable (zone temperature, heat pump load, non-controllable loads, PV system electricity generation), day-ahead prediction models.

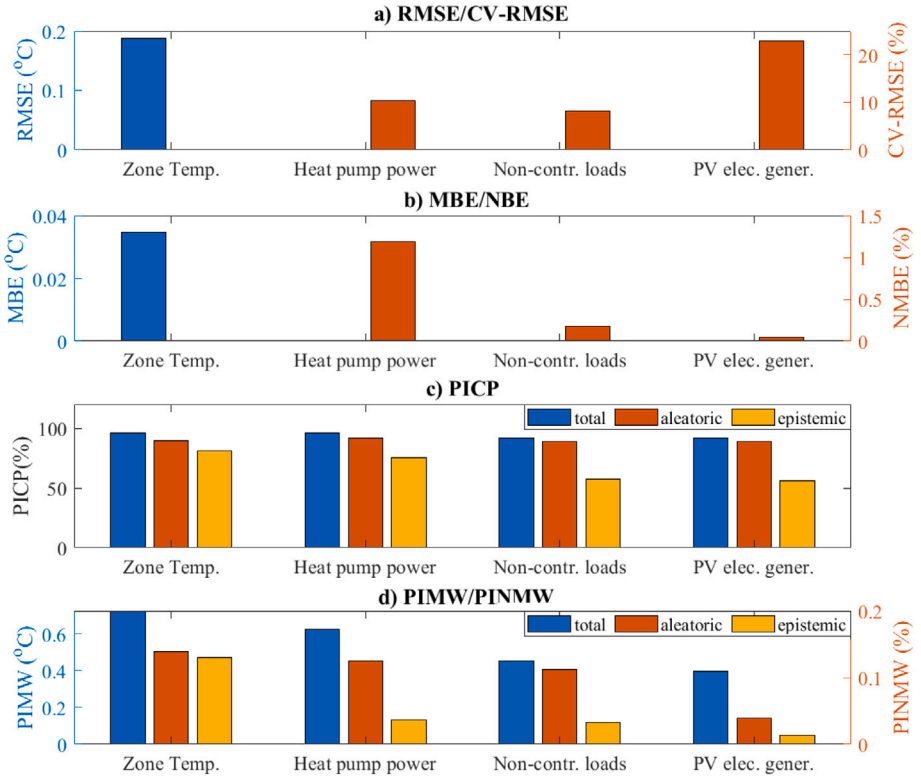


Fig. 5. (a) RMSE/CV-RMSE, (b) MBE/NMBE, (c) PICP, (d) PIMW/PINMW for each target variable (zone temperature, heat pump load, non-controllable loads, PV system electricity generation), hour-ahead prediction models.

Table 5
Structure of simulation results.

Step	Methodology Section	Results Section	Description			
1	2.5	4.1	Assess the predictive performance of the BCNN for each target variable and prediction horizon considered			
			Deterministic metrics		Probabilistic metrics	
			RMSE/ CV-RMSE	Eq. (10)/Eq. (8)	PICP	Eq. (13)
			MBE/ NMBE	Eq. (11)/Eq. (9)	PIMW/ PINMW	Eq. (14)/Eq. (15)
2	3.2	4.2.1	Select an indicative day for each occupancy profile considered and assess the associated predictive performance			
3	3.4	–	Impose hourly independent down-flex and up-flex actions for all energy systems considered			
			Commencement times: 00:00,01:00,. . . 23:00			
4	2.6.1	4.2.2	Evaluate the flexibility potential and the associated predictability by using deterministic flexibility indicators:			
				Heat pump		Electrical energy storage
				Down-flex	Up-flex	Down-flex
			Storage Capacity	Eq. (16)	Eq. (18)	Eq. (17)
5	2.6.2	4.2.3	Evaluate the flexibility potential and the associated predictability by using probabilistic indicators:			
				Heat pump		Electrical energy storage
				Down-flex	Up-flex	Down-flex
			Storage Capacity	Eq. (20)	Eq. (21)	Eq. (24)
			Temperature deviation	Eq. (22)		–

Table 6
BCNN model hyperparameters.

	Zone temp.		Heat pump load		Non-contr.loads		PV system electr. gener.	
	Day ahead	Hour ahead	Day ahead	Hour ahead	Day ahead	Hour ahead	Day ahead	Hour ahead
Hidden layers	2	2	3	2	2	3	3	2
Hidden neurons	16	40	48	48	48	30	80	32
Dropout rate	0.015	0.01	0.095	0.04	0.01	0.04	0.025	0.035
Batch size	26	18	12	16	14	12	26	10
Epochs	50	50	70	80	50	100	55	110
Window size	60	18	75	18	65	8	55	12

acceptable limits as per ASHRAE criteria [63]. Moreover, $PICP_{al}$ and $PIMW_{al}/PINMW_{al}$ are significantly close to $PICP_{tot}$ and $PIMW_{tot}/PINMW_{tot}$, respectively, for all target variables and predictioNMBEn horizons considered. This means that aleatoric uncertainty is the predominant type of uncertainty for all cases considered. Moreover, regarding the deterministic metrics (CV-RMSE/RMSE and NMBE/MBE), hour-ahead prediction models outperform the pertinent day-ahead models for the HVAC system related variables (heat pump load and zone temperature) and the PV system electricity generation. On the other hand, the prediction performance of the hour-ahead and the day-ahead models for the non-controllable loads is similar by considering both deterministic and probabilistic metrics. The same applies to $PIMW_{tot}/PINMW_{tot}$ for the day-ahead models which is significantly higher than that of the hour-ahead models, especially for the HVAC system-related variables. As regards the zone temperature, the RMSE and the $PIMW_{tot}$ exhibited by the hour-ahead model are approximately 50% lower than that of the pertinent day-ahead models. Given that the $PIMW_{tot}/PINMW_{tot}$ is a measure of the standard deviation, it can be concluded that hour-ahead models are characterised by less uncertainty compared to the corresponding day-ahead models. From the above it can be concluded that forecast models using shorter prediction horizons not only exhibit higher accuracy but they are also accompanied by reduced uncertainty. It should be noted, though, that the degree to which shorter prediction horizons provide superior forecasts depends on the target variable considered. Finally, Table 9 summarises the average execution times for each target variable for both day-ahead and hour-ahead predictions. The execution times range between 26.17 s to 51.75 s for day-ahead predictions, and 13.96 s to 19.23 s for hour-ahead predictions depending on the target variable. The execution time difference, especially in the case of day-ahead predictions, is a result of the different number of input variables selected for each target variable.

4.2. Energy flexibility assessment

In this section, the predictive performance of the BCNN is evaluated by selecting two different baseline cases from the evaluation set (Section 3.4), based on the two occupancy profiles considered (Table 4). Specifically, Section 4.2.1 evaluates building load demand by

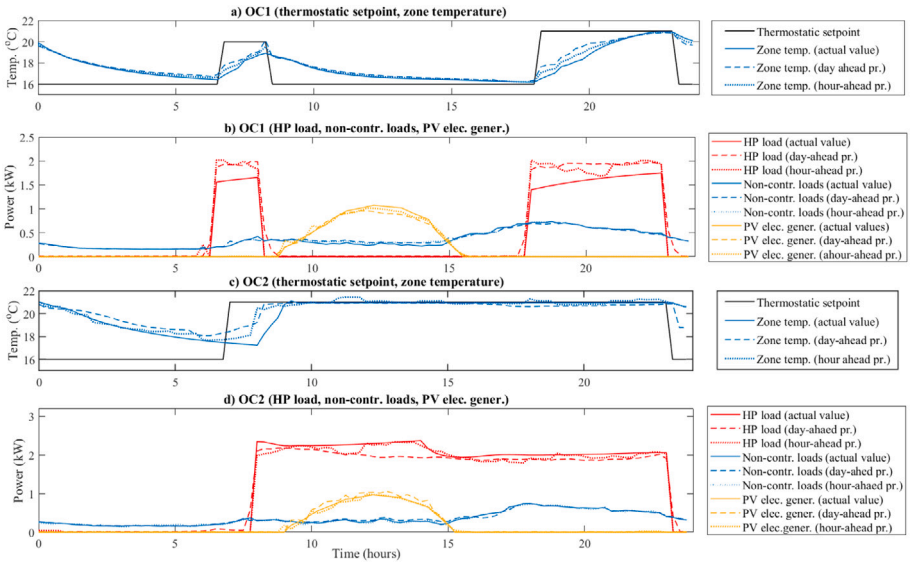


Fig. 6. Predicted (day-ahead (dashed), hour-ahead (dotted)) and actual (continuous) values: (a) Thermostatic setpoint and zone temperature, OC1, (b) heat pump load (red), non-controllable loads (blue), PV electricity generation (yellow), OC1, (c) thermostatic setpoint and zone temperature, OC2, (d) heat pump load (red), non-controllable loads (blue), PV electricity generation (yellow), OC2.

Table 7
Prediction performance of all target variables by RMSE (°C), MBE (°C), PIMW (°C) (zone temperature), CV-RMSE (%), NMBE (%), PINMW (%) (heat pump load, non-controllable loads, PV system electricity generation), and PICP (%) (zone temperature, heat pump load, non-controllable loads, PV system electricity generation) — day-ahead BCNN.

	Zone temp. (°C)	Heat pump load (%)	Non-contr. loads (%)	PV elec. gener. (%)
$RMSE/CV-RMSE$	0.4202	18.15	8.45	26.54
$MBE/NMBE$	0.0535	−1.57	0.89	1.86
$PICP_{tot}$	93.99	95.83	90.72	91.18
$PICP_{ep}$	74.49	67.72	60.86	44.09
$PICP_{al}$	90.39	91.23	87.7	89.14
$PIMW_{tot}/PINMW_{tot}$	1.4408	26.84	13.2	4.46
$PIMW_{ep}/PINMW_{ep}$	0.855	12.74	4.97	2.48
$PIMW_{al}/PINMW_{al}$	1.0399	22.38	11.64	3.43

Table 8
Prediction performance of all target variables by RMSE (°C), MBE (°C), PIMW (°C) (zone temperature), CV-RMSE (%), NMBE (%), PINMW (%) (heat pump load, non-controllable loads, PV system electricity generation), and PICP (%) (zone temperature, heat pump load, non-controllable loads, PV system electricity generation) — hour-ahead BCNN.

	Zone temp. (°C)	Heat pump load (%)	Non-contr. loads (%)	PV elec. gener. (%)
$RMSE/CV-RMSE$	0.1883	10.39	8.24	22.94
$MBE/NMBE$	0.0346	1.19	0.17	0.05
$PICP_{tot}$	95.74	96.31	91.68	92.14
$PICP_{ep}$	80.9	75.26	57.8	56.16
$PICP_{al}$	89.4	91.84	88.87	89.13
$PIMW_{tot}/PINMW_{tot}$	0.725	17.3	12.45	3.68
$PIMW_{ep}/PINMW_{ep}$	0.4728	10.91	3.93	1.38
$PIMW_{al}/PINMW_{al}$	0.5058	12.52	11.26	3.28

considering the reference building demand (no DR actions considered), whereas Sections 4.2.2 and 4.2.3 assess the building flexibility potential as well as the predictability of the various flexibility attributes (storage capacity, temperature deviations). Specifically, the daily flexibility

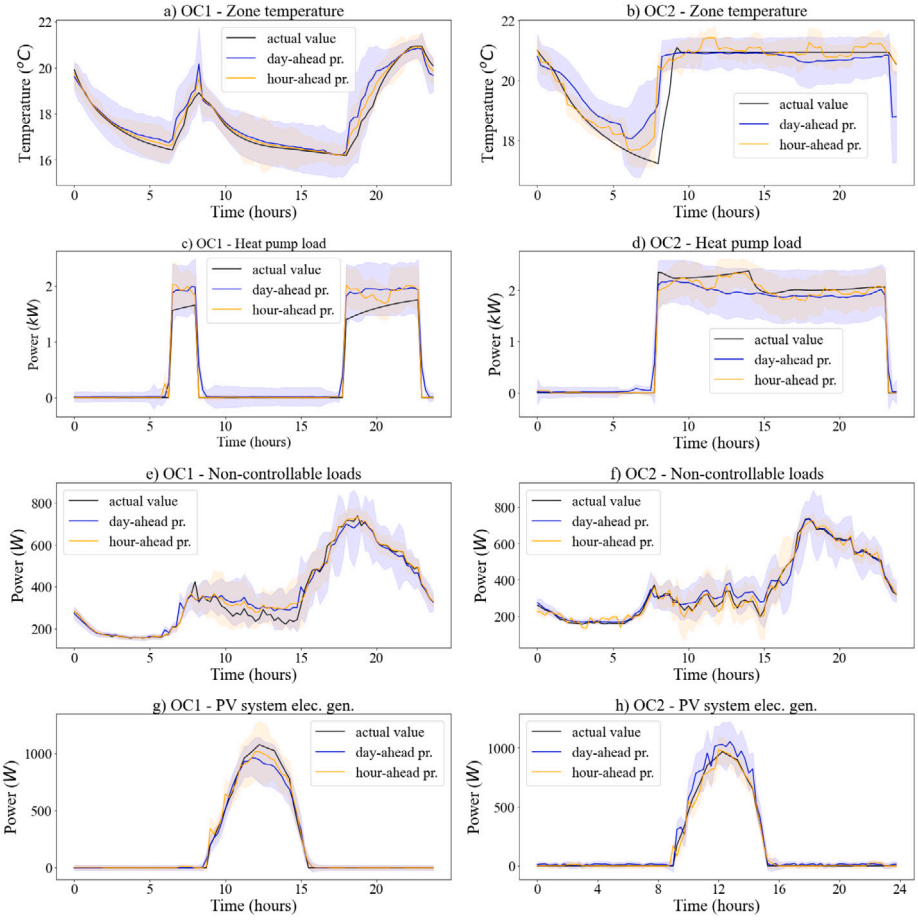


Fig. 7. Predicted (BCNN), 95% prediction intervals, and actual (white-box) values: Zone temperature: (a) OC1, (b) OC2, Heat pump load: (c) OC1, (d) OC2, Non-controllable loads: (e) OC1, (f) OC2, PV electricity generation: (g) OC1, (h) OC2.

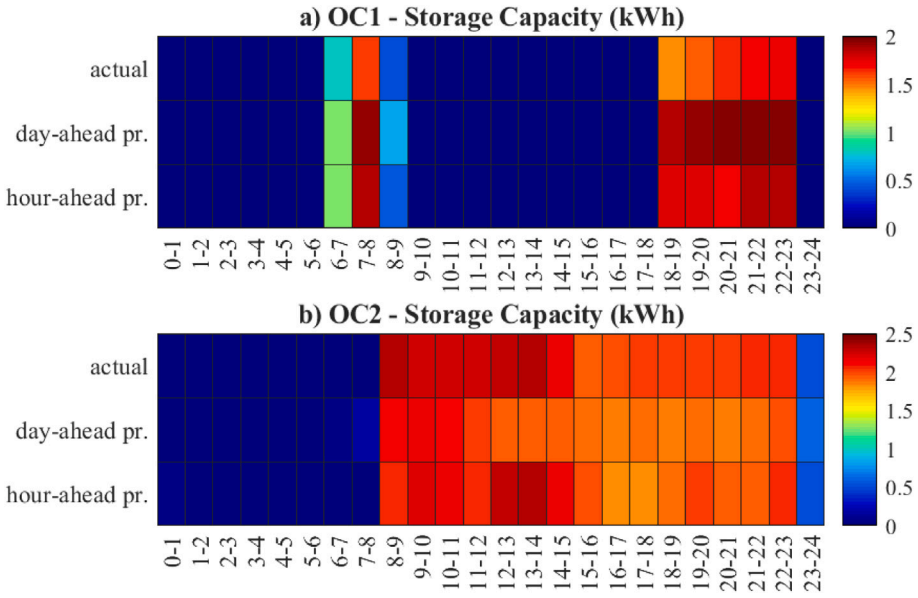


Fig. 8. BCNN predicted heat pump storage capacity (C_{DF}) (hourly down-flex DR actions): (a) OC1, (b) OC2.

mappings developed in Section 4.2.2 assess the energy flexibility predictability deterministically (based on the prediction mean), whereas Section 4.2.3 evaluates building flexibility by also considering the associated uncertainty.

4.2.1. Baseline cases

Fig. 6 assesses the predictive performance of the BCNN model in a deterministic fashion by measuring the error of the predictive means

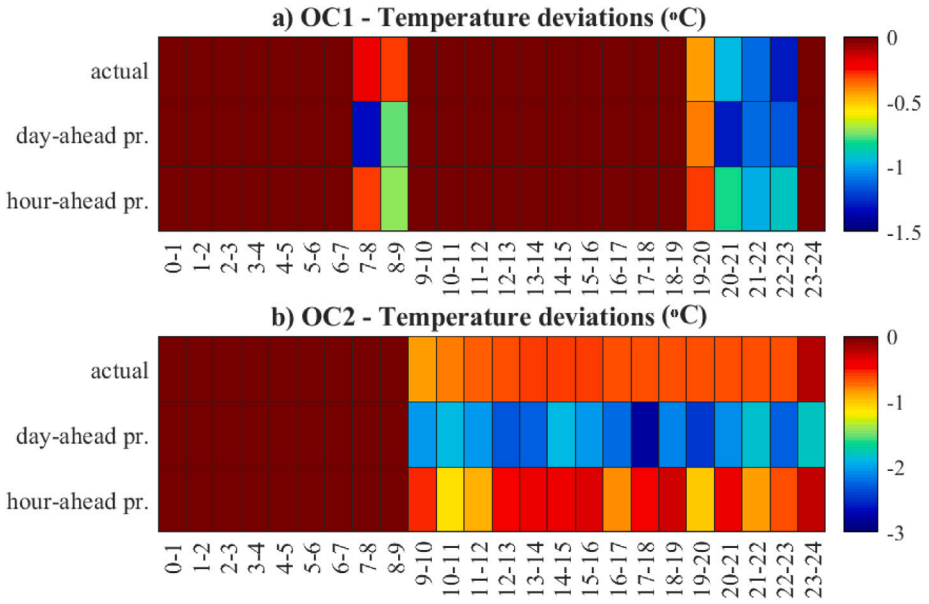


Fig. 9. BCNN predicted temperature deviations (hourly down-flex DR actions): (a) OC1, (b) OC2.

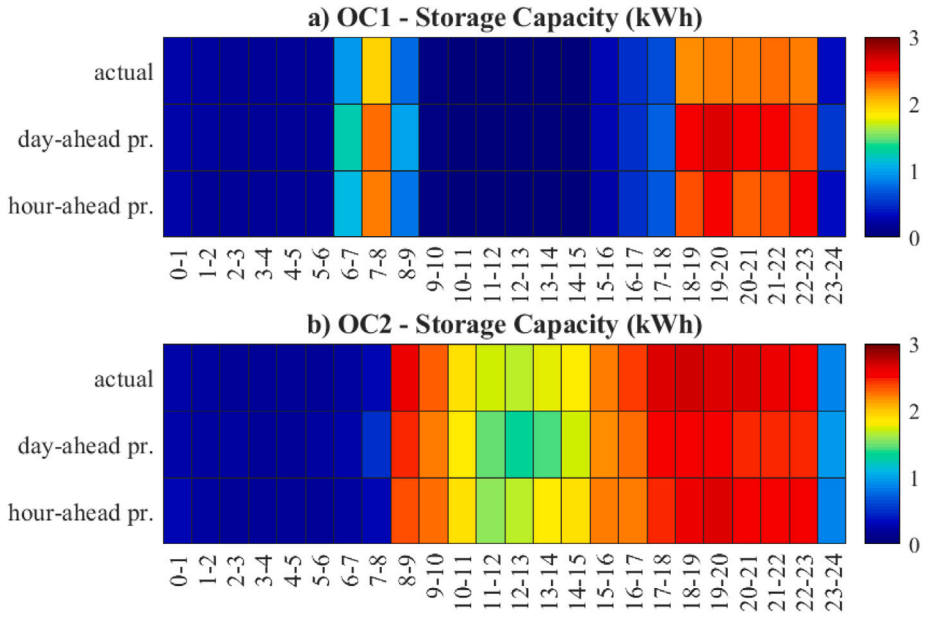


Fig. 10. BCNN predicted battery storage capacity (C_{DF}) (hourly down-flex DR actions): (a) OC1, (b) OC2.

Table 9
Average execution time (sec) for each target variable, day-ahead and hour-ahead prediction models (Intel® i7-7500 CPU 2.7 GHz, 8 GB RAM).

	Day-ahead pred. models	Hour-ahead pred. models
Zone temp.	26.17	13.97
Heat pump load	51.75	19.23
Non-contr. loads	33.65	13.96
PV elec. gener.	26.19	19.31

to estimate the white-box model simulation results for each target variable. Specifically, Figs. 6a and 6c illustrate the zone thermostat setpoint (black) as well as the zone temperature for the first (OC1) and the second occupancy profile (OC2), respectively. Figs. 6b and 6d depict the actual (continuous line), the day-ahead predictions (dashed line), and the hour-ahead predictions (dotted line) of the heat pump power (red), the non-controllable loads (blue) and the self-generation (yellow) for OC1 and OC2, respectively. The zone temperature is overestimated by both day-ahead and hour-ahead prediction models when the actual zone temperature is lower than the thermostatic setpoint for both occupancy profiles and prediction horizons considered. This is potentially because the BCNN cannot capture the dynamic response of the building envelope and/or it is influenced by the increased importance of the thermostat setpoint that is the main driver of HVAC system related variables. On the other hand, both day-ahead and hour-ahead BCNNs models are likely to overestimate or underestimate the heat pump load depending on the occupancy profile considered.

Fig. 7 shows the pertinent probabilistic forecasts for all occupancy profiles and prediction horizons evaluated by considering the prediction intervals that correspond to a confidence level of 95%. It is noteworthy that the prediction errors and the prediction interval for each target variable change over the day. For example, during

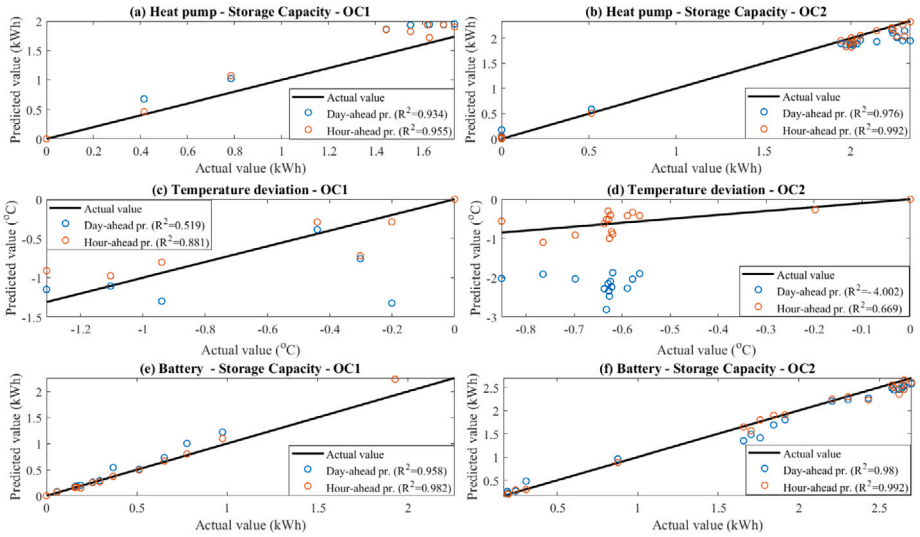


Fig. 11. Actual (white-box) and predicted values (BCNN), hourly down-flex DR actions: (a) heat pump storage capacity — OC1, (b) heat pump storage capacity — OC2, (c) temperature deviation — OC1, (d) temperature deviation — OC1, (e) battery storage capacity — OC1, (f) battery storage capacity — OC2.

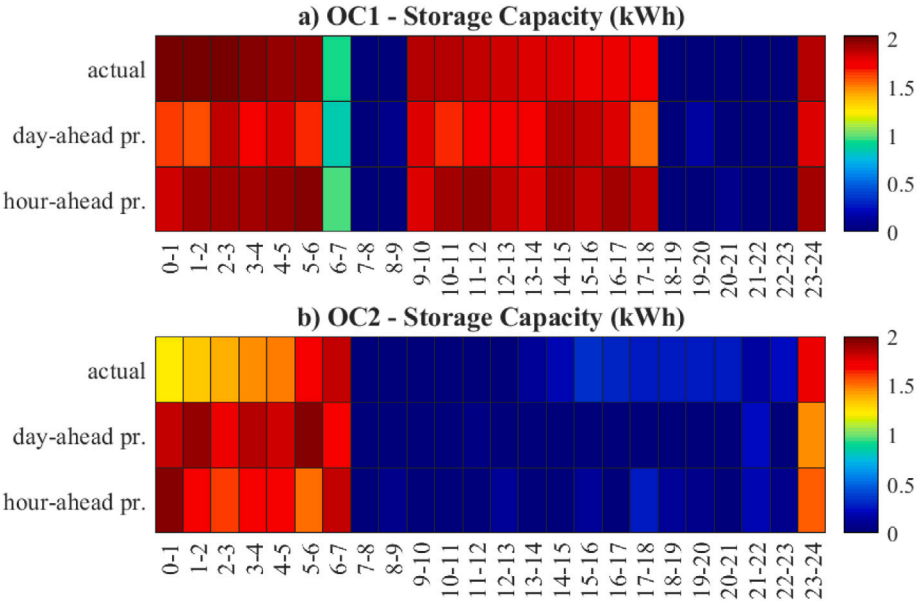


Fig. 12. BCNN predicted heat pump storage capacity (C_{UF}) (hourly up-flex DR actions): (a) OC1, (b) OC2.

periods of temperature setbacks, the heat pump is switched off and the associated load is zero. In this case, the heat pump load predictions are considerably accurate and corresponding uncertainty is reduced, whilst the zone temperature is consistently overestimated and is accompanied by increased uncertainty. During periods of increased non-controllable load levels (08:00–00:00), the developed ML models exhibit lower accuracy and higher uncertainty for all occupancy profiles and prediction horizons considered. Moreover, the BCNN developed for the PV system electricity generation exhibit increased prediction errors and uncertainty during periods of high power output. Finally, hour-ahead models not only provide more accurate predictions for the various target variables considered but also the associated prediction interval is considerably narrower.

4.2.2. Deterministic energy flexibility assessment

In this section, the flexibility potential of the heat pump and the battery is analysed in a deterministic manner by considering the prediction mean for each target variable and prediction horizon. The flexibility potential of the energy conversion systems considered is evaluated for both down-flex and up-flex actions for the two baseline cases considered by using the white-box model as ground truth (actual values) and the BCNN for day-ahead and hour-ahead forecasts. Figs. 8 to 14 assume 24 independent DR actions with a one-hour duration considering that only a single DR action occurs each day.

The heat pump storage capacity (C_{DF}) (Eq. (16)), resulting from hourly down-flex actions of the thermostatic setpoint, is depicted in Fig. 8a (OC1) and 8b (OC2). Fig. 9a (OC1) and 9b (OC2) show the zone temperature deviations resulting from the aforementioned DR actions. The calculation of the temperature reductions aims to investigate whether the activation of the heat pump flexibility results in temperature drifts that lie within acceptable limits as per ASHRAE standards (Table 3). In addition, Fig. 10a (OC1) and 10b (OC2) illustrate the storage capacity (C_{DF}) (Eq. (18)) arising from the activation of the stationary battery flexibility. Fig. 11 shows the actual (ground truth) and the predicted values (BCNN) along with the coefficient of determination (R^2) for the heat pump storage capacity (Fig. 11a-OC1, Fig. 11b-OC2),

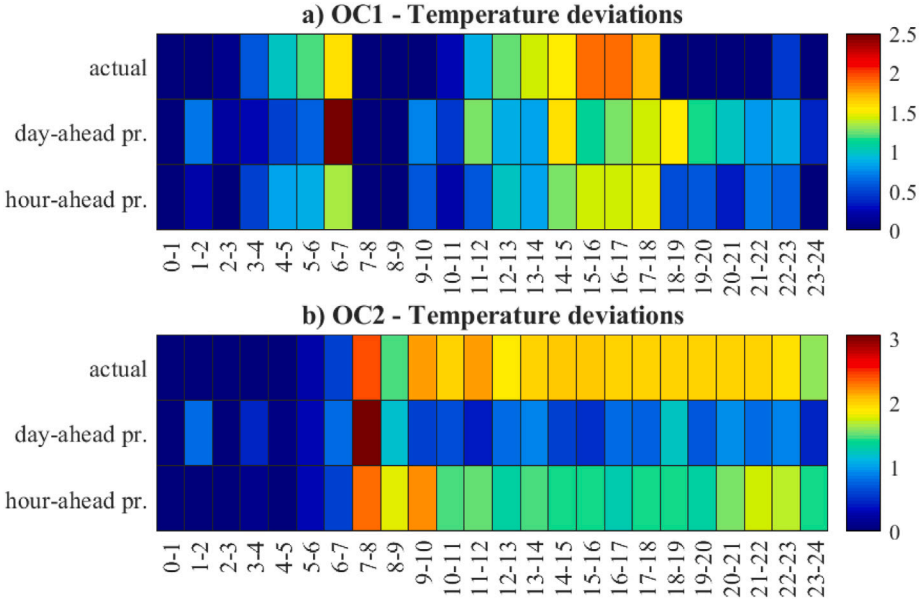


Fig. 13. BCNN predicted temperature deviations (hourly up-flex DR actions): (a) OC1, (b) OC2.

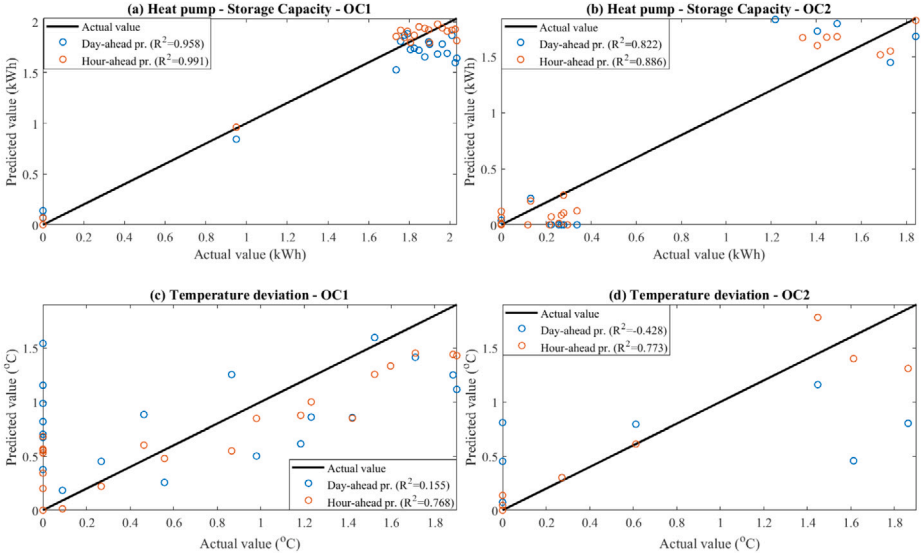


Fig. 14. Actual (white-box) and predicted values (BCNN), hourly up-flex DR actions: (a) heat pump storage capacity — OC1, (b) heat pump storage capacity — OC2, (c) temperature deviation — OC1.

the associated temperature deviations (Fig. 11c-OC1, Fig. 11d-OC2), and the battery storage capacity (Fig. 11e-OC1, Fig. 11f-OC2).

During periods of occupant inactivity or absence, the heat pump is switched off (due to the temperature setback) and it exhibits zero storage capacity. The storage capacity is overestimated (OC1) or underestimated (OC2) by both day-ahead and hour-ahead prediction models. This is due to the overestimation (OC1) or the underestimation (OC2) of the heat pump load by the pertinent prediction models. Both day-ahead and hour-ahead BCNN models show very good predictive performance for both occupancy profiles, attaining coefficients of determination between 0.934 and 0.992. It is noteworthy that the storage capacity can be more accurately predicted for hour-ahead predictions and the second occupancy profile. All temperature deviations arising from down-flex actions lie within acceptable limits as per ASHRAE standards [65], however, their predictability varies depending on the prediction horizon and the occupancy profile. Specifically, the hour-ahead predictions are consistently more accurate compared to the day-ahead prediction models. Specifically, the day-ahead predictions consistently overestimate the temperature reductions during DR actions exhibiting unacceptable coefficients of determination, especially for the second occupancy profile. This indicates that longer prediction horizons are likely to be inadequate to estimate thermal comfort deviations arising from the activation of the passive TES. Given that the battery is discharged to cover the building load, the associated storage capacity depends on the net building load. Thus, during periods of high onsite electricity generation, the battery storage capacity (C_{DF}) (Eq. (18)) is reduced. On the other hand, during periods of high thermostatic setpoints (20 °C–21 °C), the storage capacity of the battery increases due to the higher heat pump power consumption. The battery storage capacity can be accurately predicted by the prediction models for hour-ahead predictions and the second occupancy profile.

The heat pump storage capacity (C_{UF}) (Eq. (17)), resulting from hourly up-flex actions, is depicted in Fig. 12a (OC1) and Fig. 12b (OC2). Fig. 13a (OC1) and Fig. 13b (OC2) show the temperature deviations resulting from the aforementioned DR actions. The battery storage capacity along with the temperature deviations in up-flex increase

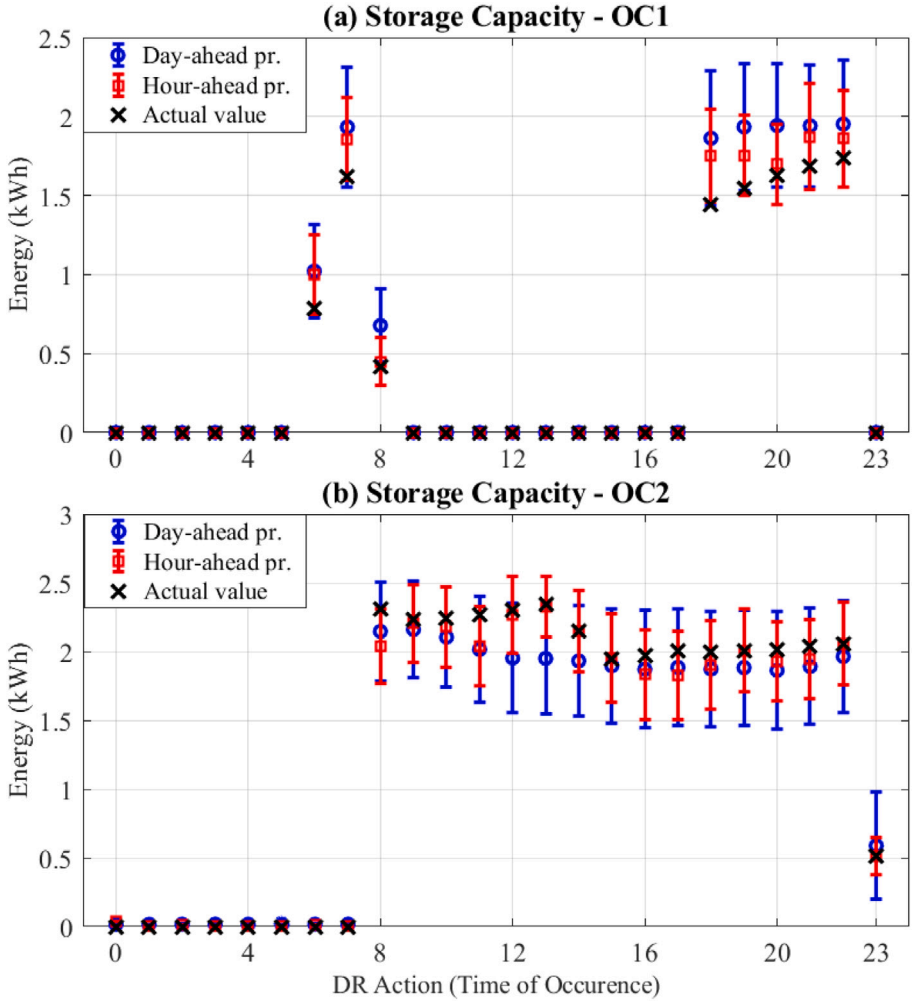


Fig. 15. Hourly uncertainty map of the heat pump storage capacity (hourly down-flex DR actions): (a) OC1, (b) OC2.

during periods of occupant absence or inactivity because of the temperature setback utilised. The hour-ahead prediction models provide more accurate forecasts for both occupancy profiles, while the storage capacity can be more accurately predicted for the first occupancy profile. Day-ahead predictions for the temperature deviations exhibit poor performance, whereas the corresponding hour-ahead prediction models provide considerably more accurate forecasts. As in down-flex actions, the BCNN models developed for the zone temperature show inferior performance when the actual value is different from the temperature setpoint. Therefore, a potential reason that the temperature deviations are poorly predicted is because the BCNN cannot capture the transient response of the zone temperature.

4.2.3. Probabilistic energy flexibility assessment

In this section, the flexibility potential of the various energy conversion systems considered is assessed in a probabilistic fashion by considering the associated prediction intervals for the two baseline cases (Section 4.2.1). The energy flexibility potential of the heat pump is assessed by considering both down-flex and up-flex actions, whereas the battery flexibility is calculated only for downward DR actions, as the pertinent DR potential in up-flex is constant. The energy flexibility potential of the heat pump and the battery is assessed by considering both down-flex and up-flex actions by utilising the BCNN approach; the latter is referenced against the white-box physics-based model for day-ahead and hour-ahead predictions. As in the deterministic flexibility evaluation, Figs. 15 to 19 assume 24 independent DR actions with a one-hour duration, considering that only a single DR action occurs each day. Specifically, they show: (i) the actual value, (ii) the 95% prediction interval, and (iii) the prediction mean, of each energy flexibility characteristic for the day-ahead and hour-ahead for the forecasting models.

Considering downward flexibility actions, the storage capacity (Fig. 15) resulting from the activation of the heat pump flexibility is consistently overestimated (OC1) or underestimated (OC2) for both day-ahead and hour-ahead prediction models; nevertheless, the prediction intervals associated with hour-ahead predictions are narrower for both occupancy profiles. In addition, the prediction interval of each DR action depends on the associated time of occurrence. This is because the heat pump load is accompanied by different levels of uncertainty for each prediction step. Moreover, the average width of the prediction intervals for the day-ahead model is 0.77 kWh (OC1) and 0.8 kWh (OC2), whereas the pertinent average width for the hour-ahead prediction models is 0.52 kWh (OC1) and 0.56 kWh (OC2). This means that the average prediction interval for day-ahead predictions is 38.2% and 41.5% higher than that of the hour-ahead prediction intervals for OC1 and OC2, respectively. It is noteworthy that the prediction intervals for the hour-ahead prediction models are considerably narrower for both occupancy profiles. Considering the temperature deviations, the prediction intervals vary depending on the occupancy profile. For example, the average prediction interval for the day-ahead model is 4.18 °C (OC1) and 5.02 °C (OC2), whereas the pertinent average width for the hour-ahead prediction models is 1.98 °C (OC1) and 1.55 °C (OC2). Therefore, the average prediction interval for day-ahead predictions is 110.1% and 223.6% higher than that of the

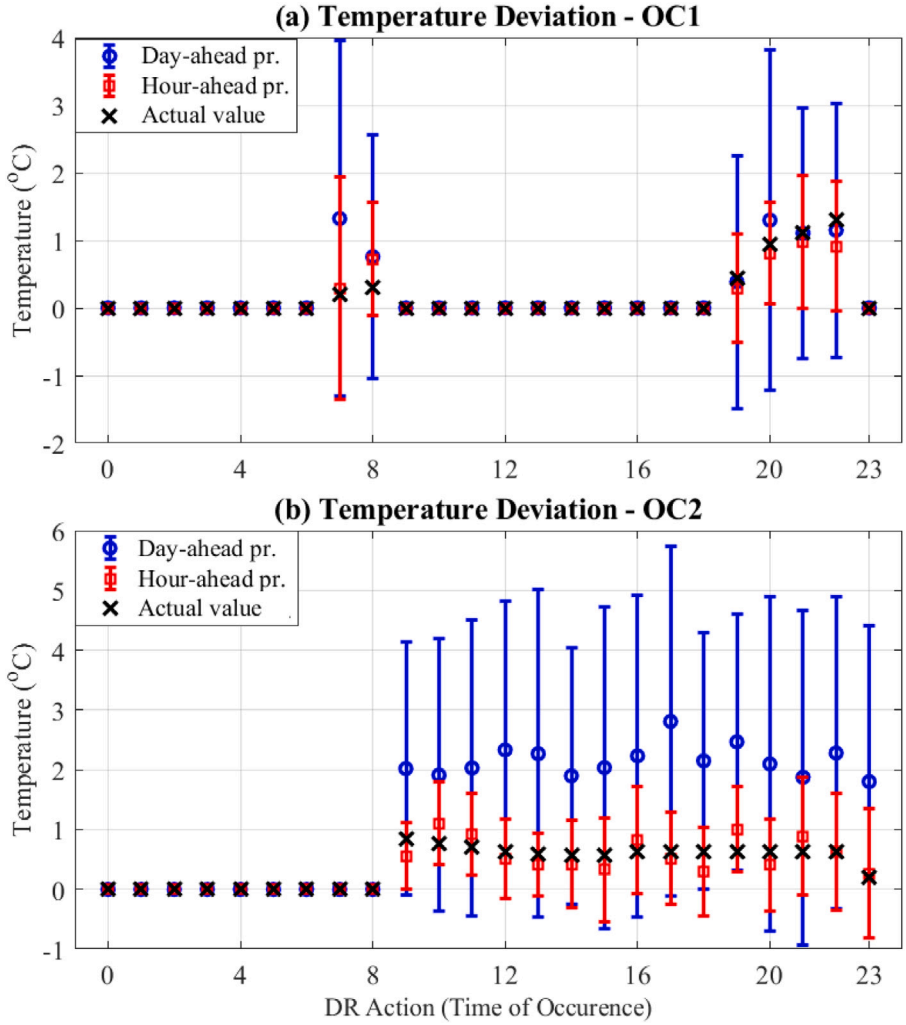


Fig. 16. Hourly uncertainty map of the temperature deviations (hourly down-flex DR actions): (a) OC1, (b) OC2.

hour-ahead prediction intervals for OC1 and OC2, respectively. The battery storage capacity can be assessed by developing the building net load prediction intervals. The positive values of the building net load correspond to the battery storage capacity, whereas the negative values indicate periods of zero storage capacity. Fig. 17 shows that the battery storage capacity is overestimated for OC1 and underestimated for OC2, whereas hour-ahead prediction models exhibit better accuracy compared to day-ahead models. For example, the average prediction interval is 0.44 kWh (OC1) and 0.62 kWh (OC2) for the day-ahead predictions and 0.29 kWh (OC1) and 0.43 kWh (OC2) for the hour-ahead predictions. This means that the average prediction interval for day-ahead predictions is 48.6% and 43.5% higher than that of the hour-ahead prediction intervals for OC1 and OC2, respectively.

Fig. 18 illustrates the heat pump storage capacity resulting from up-flex actions. The prediction intervals for hour-ahead predictions are narrower for both occupancy profiles. Specifically, the average prediction interval for the day-ahead model is 1.05 kWh (OC1) and 1.03 kWh (OC2), whereas the pertinent average width for the hour-ahead prediction models is 0.63 kWh (OC1) and 0.78 kWh (OC2). Consequently, the average prediction interval for day-ahead predictions is 66.3% and 32.7% higher than that of the hour-ahead prediction intervals for OC1 and OC2, respectively. It can be observed that the storage capacity resulting from down-flex actions (Fig. 15) can be more accurately predicted compared to the storage capacity arising from up-flex actions (Fig. 18). Further to this, the average prediction intervals for up-flex actions is considerably larger compared to the associated prediction intervals for down-flex actions. This is because the heat pump downward flexibility can be exploited by considering on–off signals, whereas the associated upward flexibility can only be harnessed by modulating the zone thermostat.

Fig. 19 shows the temperature deviations resulting from upward thermostatic modulations. Although hour-ahead prediction models exhibit a reasonable performance, day-ahead models consistently fail to capture them; this is also depicted in the relationship between the average prediction intervals for the day-ahead and the hour-ahead prediction models. For example, the average prediction interval for the day-ahead model is 3.42 °C (OC1) and 2.57 °C (OC2), whereas the pertinent average prediction interval for the hour-ahead prediction models is 1.95 °C (OC1) and 1.42 °C (OC2). This means that the average prediction interval for day-ahead predictions is 75.4% and 77.5% higher than that of the hour-ahead prediction intervals for OC1 and OC2, respectively. Consequently, the largest percentage difference between hour-ahead and day-ahead predictions is observed for the temperature deviation predictions for both downward and upward flexibility. This result is in agreement with the significant performance difference (in terms of RMSE and $PIMW_{tot}$) between the hour-ahead and day-ahead zone temperature prediction models (Tables 5 and 6).

5. Discussion

Previous research on building energy flexibility assessment methodologies has mainly focused on deterministic approaches without considering uncertainties. As a result, current flexibility indicators involve

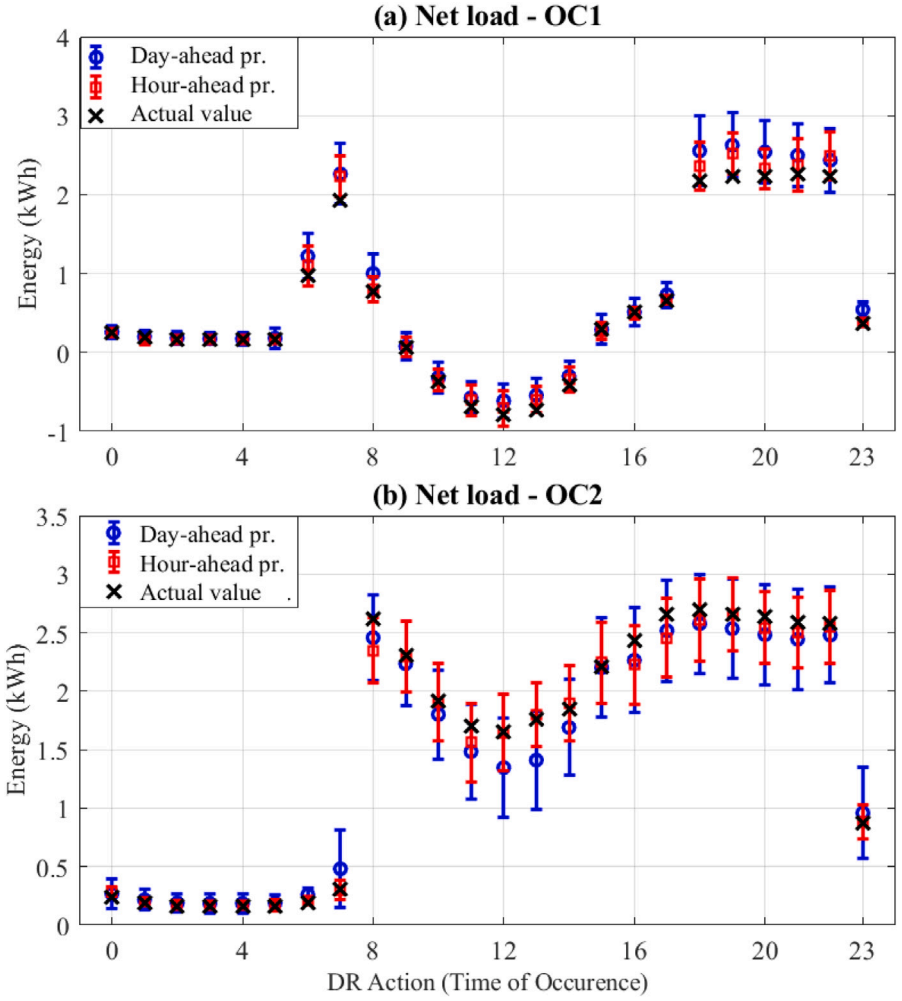


Fig. 17. Hourly uncertainty map of the battery storage capacity (hourly down-flex DR actions): (a) OC1, (b) OC2.

deterministic mathematical formulations that cannot account for confidence levels. There is a significant gap in evaluating building flexibility potential with confidence intervals [49]. To address this gap, a probabilistic energy flexibility assessment framework based on Bayesian feedforward convolutional neural networks has been developed to consider aleatoric and epistemic uncertainties in residential buildings. Bayesian deep learning can enable bottom-up assessments of building energy use without multiple input scenarios, but has not been used in the context of building flexibility evaluation. The proposed framework decouples building energy flexibility evaluation from case-specific DR strategies and control signals and assesses it at an individual energy system level by using suitable probabilistic indicators. Uncertainty quantification in the context of building energy flexibility analysis allows for a more accurate assessment of the building potential to provide energy flexibility services. The latter are characterised by potential risks and uncertainties associated with the energy shifting potential of individual DR actions and the resulting occupant thermal comfort deviations.

Simulation results show that aleatoric uncertainty is considerably higher than epistemic uncertainty for all target variables and prediction horizons considered. In contrast to storage capacity, the temperature deviations arising from both downward and upward DR actions are poorly estimated by day-ahead predictions and they are accompanied by particularly wide average prediction intervals (2.57 °C to 4.18 °C), depending on the DR activation method and the occupancy profile. Conversely, hour-ahead predictions not only exhibit better deterministic accuracy but also they are characterised by considerably less uncertainty for both occupancy profiles considered.

Moreover, the heat pump flexibility potential and the resulting temperature deviations are characterised by increased uncertainty compared to the battery flexibility. One possible reason for this is that the thermal inertia associated with the HVAC system (i.e., pipes, radiators, tank walls, heat pump components, ground loop, etc.) are not explicitly captured by the EnergyPlus model, thus leading to greater uncertainty. In addition, the machine learning models used, do not fully capture the complexity of the dependent variables related to the HVAC system, such as the heat pump load and the zone temperature. These models rely on historical data, weather predictions, and statistical properties from previous prediction periods, but do not consider other potentially important factors such as supply water temperatures and mass flow rates in the evaporator and condenser, the heat pump compressor characteristics, and the setpoint temperature of the water tank. The aforementioned discrepancies arise from the inability of EnergyPlus to accurately capture the transient response of HVAC system equipment, as well as other system behaviour, thereby amplifying the overall uncertainty. Future work will consider the utilisation of white-box models, such as Modelica, which are better suited for transient analysis, to address this limitation and enhance the reliability of the findings.

The weather variables utilised in this study are based on the EnergyPlus Weather File (EPW) Data Dictionary [74] and thus no uncertainty is attributed to this data. However, simulation results show a weak correlation between the weather variables (e.g., outdoor temperature, solar radiation, humidity, wind speed) and the HVAC system-related variables (heat pump load, zone temperature), and thus energy flexibility, due to the building high thermal inertia. Consequently, the weather

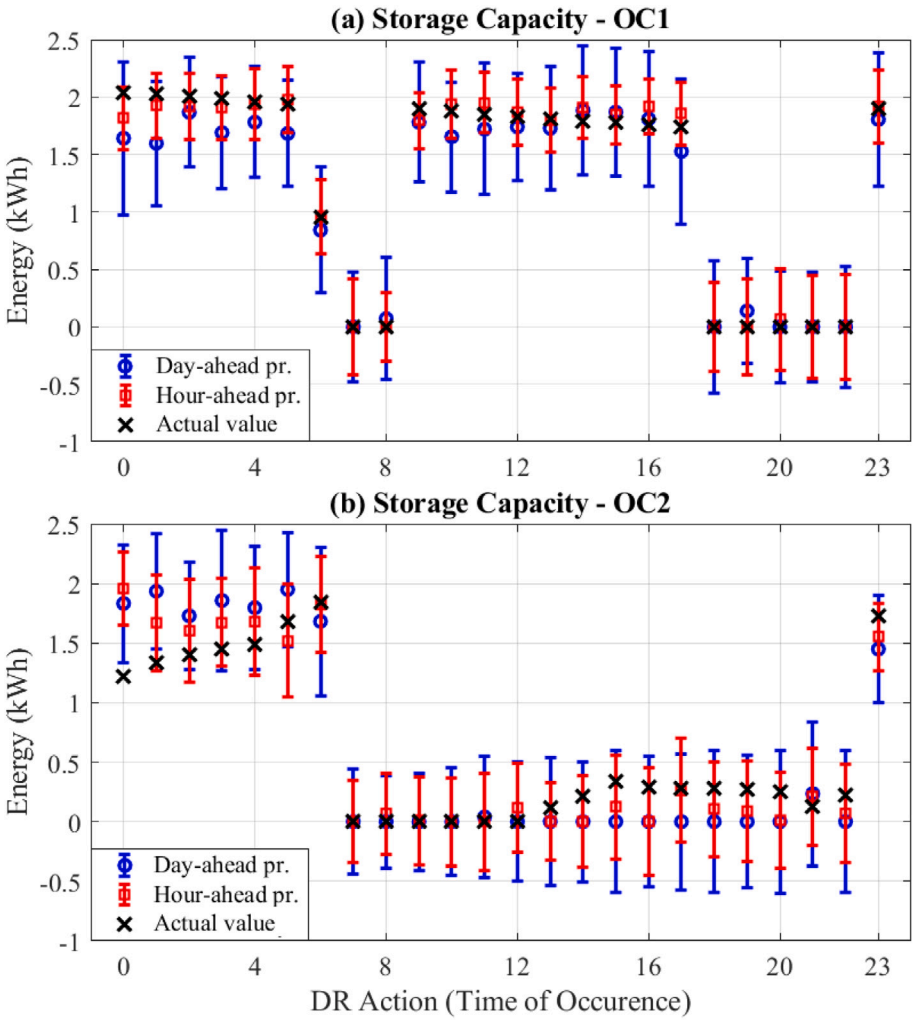


Fig. 18. Hourly uncertainty map of the heat pump storage capacity (hourly up-flex DR actions): (a) OC1, (b) OC2.

variables are not considered as inputs for the developed Bayesian convolutional neural networks, thus not contributing to model uncertainty estimation. To evaluate weather influence on uncertainty, future research will examine buildings with different thermal inertia using short-term weather forecasts.

The synthetic database utilised to build and assess the Bayesian convolutional neural network, as well as to evaluate the energy flexibility of the various energy conversion systems, is generated by using a calibrated physics-based model and average daily occupancy profiles resulting from categorising residential weekday diaries and represent 56% of the survey sample [71]. However, the selected occupancy profiles are indicative and aim to exemplify the performance of the proposed flexibility evaluation framework for various power consumption patterns and assess its effectiveness in capturing the uncertainties related to the energy shifting capability of the various building energy systems and the resulting zone temperature deviations. This methodology is applicable to individual energy systems, buildings, or groups of buildings to assess the predictability and the uncertainty associated with building energy flexibility. However, uncertainty analysis at different levels of aggregation is not part of this study and will be addressed in future research. To further validate the performance of the developed prediction models, future research could examine the impact of occupant behaviour on uncertainty estimates using stochastic occupancy models. It is important to note that this is early-stage research, and the proposed framework has not yet been implemented in a real-life setting.

6. Conclusions

In this paper, a probabilistic energy flexibility assessment framework has been developed to evaluate the operational energy flexibility of residential buildings considering both aleatoric and epistemic uncertainties associated with building energy use. To this end, periodically updated day-ahead and hour-ahead Bayesian feedforward convolutional neural networks based on Monte Carlo dropout sampling are developed to quantify the uncertainties related to building energy consumption and onsite electricity generation. The developed data-driven models can capture new patterns in data by utilising a sliding window method and by selecting the most recent occupancy patterns. The ensembles of estimates developed by utilising the Monte Carlo dropout sampling can be not only utilised to quantify uncertainty, but they also exhibit increased robustness compared to the constituent models. Subsequently, the probabilistic flexibility quantification framework presented can estimate the energy shifting capability of the energy systems considered (passive TES and electrical energy storage) as well as any zone temperature deviations arising from activating this flexibility. The proposed framework can be used to assess the flexibility of various thermal and electrical systems not only on an integrated common basis, but also by considering the different uncertainties characterising building energy consumption and onsite electricity generation. Moreover, it can be extended for any thermostatically controlled load and battery-based system. The probabilistic data-driven flexibility evaluation framework developed can allow electricity aggregators to evaluate

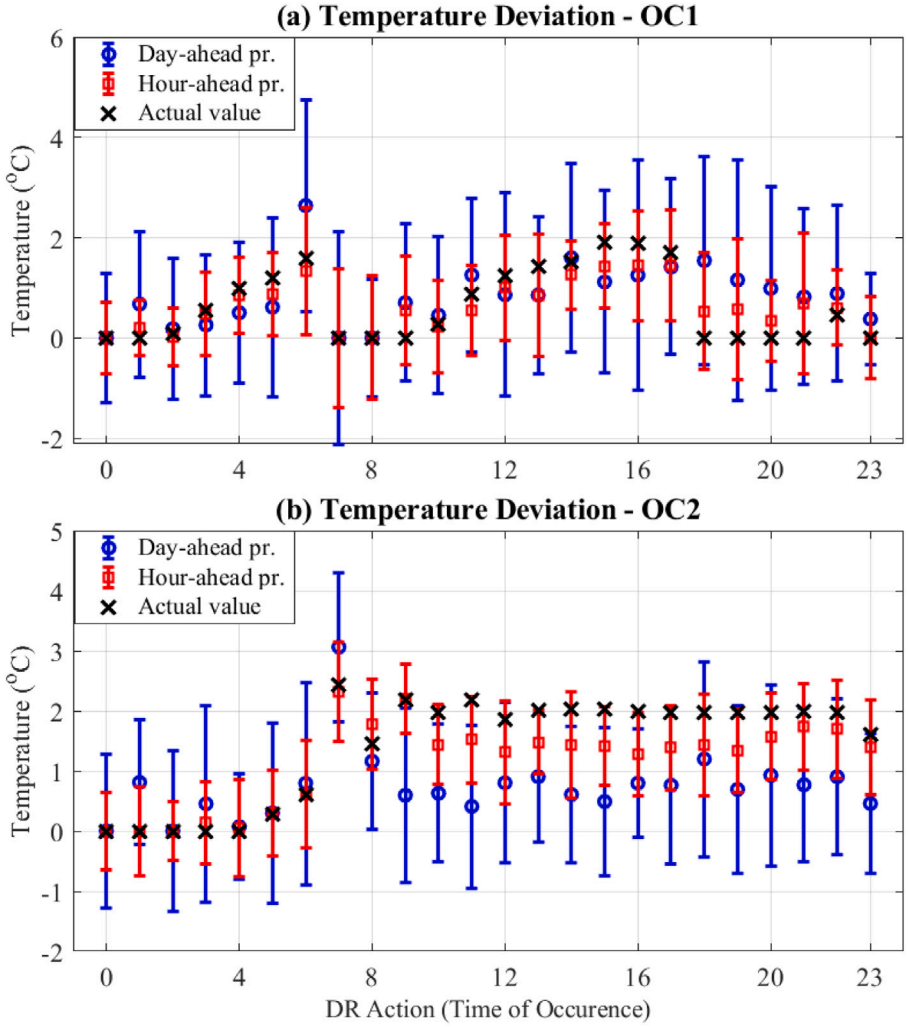


Fig. 19. Hourly uncertainty map of the temperature deviations (hourly up-flex DR actions): (a) OC1, (b) OC2.

or optimise building portfolios within prediction intervals and assess the flexibility potential of various combination measures in a scalable and user-tailored manner. Building flexibility evaluation in a context of uncertainty can provide confidence bounds for decision making, thus facilitating optimal bidding in electricity markets.

Simulation results suggest that shorter prediction horizons are more suitable to forecast any thermal comfort deviations resulting from the activation of the passive thermal energy storage. Moreover, the storage capacity resulting from harnessing the heat pump downward flexibility for both day-ahead and hour-ahead prediction models shows very good performance for both occupancy profiles, attaining coefficients of determination between 0.93 and 0.99. Finally, building flexibility potential and the associated predictability depend on weather conditions and/or occupant behaviour.

CRediT authorship contribution statement

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

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.

Data availability

Data will be made available on request.

Acknowledgements

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, Ireland. 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.

Appendix A

See Tables A.1 and A.2

Table A.1
Overview of Bayesian deep-learning methodologies.

Reference	Case study	ML model	Uncertainty quantification	Prediction horizon
[25]	Various building energy performance outputs	Bayesian neural network and Gaussian process	Only epistemic	Annual
[26]	residential air conditioning system	Bayesian convolutional neural network	Aleatoric	6 hour-ahead
[27]	Building electricity consumption	Hybrid network based on CNN and GRU	Aleatoric	4 hour-ahead
[14]	Regional electric load	Bayesian mixture density network	Aleatoric and epistemic	Day-ahead
[28]	Regional electric load	Deep residual network	Aleatoric and epistemic	Day-ahead
[29]	National electric load	Concrete dropouts, deep ensembles, Bayesian neural networks, deep Gaussian processes, and functional neural processes	Aleatoric and epistemic	Hour-ahead and day-ahead
[12]	Net load prediction	Bayesian deep LSTM network	Aleatoric and epistemic	Day-ahead
[30]	Building electricity consumption	Recurrent neural network, LSTM, and GRU	Aleatoric and epistemic	Day-ahead
[31]	Building electricity consumption	Multitask Bayesian neural network	Aleatoric and epistemic	Hour-ahead

Table A.2
Overview of energy flexibility quantification studies focusing on individual residential buildings.

Reference	Case study energy system	Flexibility indicator	Model type	Deterministic/ Probabilistic
[36]	Passive TES	Storage capacity and efficiency	Physics-based	Deterministic
[37]	Passive TES	Storage capacity, energy, power deviation, rebound energy	Physics-based	Deterministic
[38]	Passive TES	Flexible energy, rebound energy, flexible energy efficiency, maximum flexible power	Gray-box	Deterministic
[39]	Active TES	Storage capacity, storage efficiency, power shifting capability, flexibility factor	Physics-based	Deterministic
[40]	Passive and active TES, batteries	Forced and delayed flexibility factor	Physics-based	Deterministic
[41]	Passive and active TES	Storage capacity, storage efficiency	Data-driven	Deterministic
[44]	Wet appliances, EV	Net energy consumption, net delivered energy	Data-driven	Deterministic
[43]	Passive TES	Flexibility factor, supply cover factor, load cover factor	Data-driven	Deterministic
[38]	Passive and active TES	Achievable energy flexibility	Data-driven	Deterministic
[45]	Passive and active TES, PV	PV self-consumption, PV self-sufficiency, heat pump solar contribution	Physics-based	Deterministic
[46]	Passive and active TES	Flexibility index	Gray-box	Deterministic
[6]	Passive TES	Power reduction percentage	Gray box	Probabilistic
[49]	Passive TES	Storage capacity, storage efficiency	Gray-box	Probabilistic
[48]	Passive TES	Average shiftable power	Data-driven	Probabilistic

Appendix B

Nomenclature

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)
$PICP_{ep}$	PICP related to epistemic uncertainty
$PICP_{al}$	PICP related to aleatoric uncertainty
$PICP_{tot}$	PICP related total uncertainty
$PIMW_{ep}/PINMW_{ep}$	PIMW/PINMW related to epistemic uncertainty
$PIMW_{al}/PINMW_{al}$	PIMW/PINMW related to aleatoric uncertainty
$PIMW_{tot}/PINMW_{tot}$	PIMW/PINMW related to total uncertainty
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
PICP	Prediction interval coverage probability
PIMW	Prediction interval mean width
PINMW	Prediction interval normalized mean width
GSHP	Ground Source Hear Pump
ML	Machine Learning
BDL	Bayesian deep learning
BCNN	Bayesian Convolutional Neural Network

References

[1] Chase A, Gross R, Heptonstall P, Jansen M, Kenefick M, Parrish B, et al. Realising the potential of demand-side response to 2025. Department for Business, Energy, and Industrial Strategy; 2017.

[2] International Renewable Energy Agency. Demand side flexibility for power sector transformation. 2019.

[3] Eissa MM. Developing incentive demand response with commercial energy management system (CEMS) based on diffusion model, smart meters and new communication protocol. Appl Energy 2019;236:273–92. http://dx.doi.org/10.1016/j.apenergy.2018.11.083.

[4] International Energy Agency and the United Nations Environment Programme. Global status report for buildings and construction: Towards a zero–emission, efficient and resilient buildings and construction sector. 2021.

[5] 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.

[6] Hu M, Xiao F. Quantifying uncertainty in the aggregate energy flexibility of high-rise residential building clusters considering stochastic occupancy and occupant behavior. Energy 2020;194. http://dx.doi.org/10.1016/j.energy.2019.116838.

[7] Tian W, Heo Y, de Wilde P, Li Z, Yan D, Park CS, et al. A review of uncertainty analysis in building energy assessment. Renew Sustain Energy Rev 2018;93:285–301. http://dx.doi.org/10.1016/j.rser.2018.05.029.

[8] Shamsi MH, Ali U, Mangina E, O'Donnell J. A framework for uncertainty quantification in building heat demand simulations using reduced-order grey-box energy models. Appl Energy 2020;275. http://dx.doi.org/10.1016/j.apenergy.2020.115141.

[9] Hayes KR. Uncertainty and uncertainty analysis methods: Issues in quantitative and qualitative risk modeling with application to import risk assessment. Commonwealth Scientific and Industrial Research Organisation; 2011.

[10] 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.

[11] Martina-Perez S, Simpson MJ, Baker RE. Bayesian uncertainty quantification for data-driven equation learning. Proc R Soc A: Math, Phys Eng Sci 2021. http://dx.doi.org/10.48550/arXiv.2102.11629.

[12] 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.

[13] Abdar M, Pourpanah F, Hussain S, Rezazadegan D. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Inf Fusion 2021;76:243–97. http://dx.doi.org/10.1016/j.inffus.2021.05.008.

[14] Brusaferri A, Matteucci M, Spinelli S, Vitali A. Probabilistic electric load forecasting through bayesian mixture density networks. Appl Energy 2022;309. http://dx.doi.org/10.1016/j.apenergy.2021.118341.

[15] 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.

[16] 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.

[17] 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.

[18] Li Z, Huang G. Re-evaluation of building cooling load prediction models for use in humid subtropical area. Energy Build 2013;62:442–9. http://dx.doi.org/10.1016/j.enbuild.2013.03.035.

[19] Gao D, Sun Y, Lu Y. A robust demand response control of commercial buildings for smart grid under load prediction uncertainty. Energy 2015;93:275–83. http://dx.doi.org/10.1016/j.energy.2015.09.062.

[20] Mahadevan S, Sarkar S. Uncertainty analysis methods. Issues Quant Qual Risk Model Appl Import Risk Assess ACERA Proj 2009;93:1–26.

[21] 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.

[22] Rätz M, Javadi AP, Baranski M, Finkbeiner K, Müller D. Automated data-driven modeling of building energy systems via machine learning algorithms. Energy Build 2019;2019:109384. http://dx.doi.org/10.1016/j.enbuild.2019.109384.

[23] Wang H, Yeung DY. Towards Bayesian deep learning: A framework and some existing methods. IEEE Trans Knowl Data Eng 2016;28(12):3395–408. http://dx.doi.org/10.48550/arXiv.1608.06884.

[24] Kronheim BS, Kuchera MP, Prosper HB. TensorBNN: Bayesian inference for neural networks using tensorflow. Comput Phys Comm 2022;270. http://dx.doi.org/10.1016/j.cpc.2021.108168.

[25] Westermann P, Evins R. Using Bayesian deep learning approaches for uncertainty-aware building energy surrogate models. Energy AI 2021;3. http://dx.doi.org/10.1016/j.egyai.2020.100039.

[26] Lork C, Li WT, Qin Y, Zhou Y, Yuen C, Tushar W, et al. An uncertainty-aware deep reinforcement learning framework for residential air conditioning energy management. Appl Energy 2020;276. http://dx.doi.org/10.1016/j.apenergy.2020.115426.

[27] Xuan W, Shouxiang W, Qianyu Z, Shaomin W, Liwei F. A multi-energy load prediction model based on deep multi-task learning and ensemble approach for regional integrated energy systems. Int J Electr Power Energy Syst 2021;126. http://dx.doi.org/10.1016/j.ijepes.2020.106583.

[28] Chen K, Chen K, Wang Q, He Z, Hu J, He J. Short-term load forecasting with deep residual networks. IEEE Trans Smart Grid 2018;10(4):3943–52. http://dx.doi.org/10.48550/arXiv.1805.11956.

[29] Al-Gabalawy M, Hosny NS, Adly AR. Probabilistic forecasting for energy time series considering uncertainties based on deep learning algorithms. Electr Power Syst Res 2021;196. http://dx.doi.org/10.1016/j.epsr.2021.107216.

[30] Xu L, Hu M, Fan C. Probabilistic electrical load forecasting for buildings using Bayesian deep neural networks. J Build Eng 2022;46. http://dx.doi.org/10.1016/j.jobe.2021.103853.

[31] Yang Y, Li W, Gulliver TA, Li S. Bayesian deep learning-based probabilistic load forecasting in smart grids. IEEE Trans Ind Inf 2020;16(7):4703–13. http://dx.doi.org/10.1016/j.jobe.2021.103853.

[32] Kannan T, Lork C, Tushar W, Yuen C, Wong NC, Tai S, et al. Energy management strategy for zone cooling load demand reduction in commercial buildings: A data-driven approach. IEEE Trans Ind Appl 2019;5(6):7281–99. http://dx.doi.org/10.1109/TIA.2019.2930599.

[33] Li WT, Gubba SM, Tushar W, Yuen C, Hassan NU, Poor HV, et al. Data driven electricity management for residential air conditioning systems: An experimental approach. IEEE Trans Emerg Top Comput 2019;7(3):380–91. http://dx.doi.org/10.1109/TETC.2017.2655362.

[34] Li WT, Thirugnanam K, Tushar W, Yuen C, Chew KT, Tai S. Improving the operation of solar water heating systems in green buildings via optimized control strategies. IEEE Trans Emerg Top Comput 2018;14(4):1646–55. http://dx.doi.org/10.1109/TII.2018.27970182.

[35] 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.

[36] Kathirgamanathan A, Péan T, Zhange K, De Rosa M, 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.

[37] 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.

[38] Zhang K, Kummert M. Energy flexible buildings: An evaluation of definitions and quantification methodologies applied to thermal storage. Build Simul 2021;14:1439–52. http://dx.doi.org/10.1007/s12273-020-0751-x.

[39] 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.

[40] 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 Networks 2020;21. http://dx.doi.org/10.1016/j.segan.2020.100304.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] Ren H, Sun Y, Albdoor AK, Tyagi V, Pandey A, Ma Z. Improving energy flexibility of a net-zero energy house using a solar-assisted air conditioning system with thermal energy storage and demand-side management. Appl Energy 2021;285. http://dx.doi.org/10.1016/j.apenergy.2021.116433.

[46] Date J, Candanedo JA, Athienitis AK. A methodology for the enhancement of the energy flexibility and contingency response of a building through predictive control of passive and active storage. Energies 2021;14. http://dx.doi.org/10.3390/en14051387.

[47] Oldewurtel F, Parisio A, Jones CN, Gyalistras D, Gwerder M, Stauch V, et al. Use of model predictive control and weather forecasts for energy efficient building climate control. Energy Build 2012;45:15–27. http://dx.doi.org/10.1016/j.enbuild.2011.09.022.

[48] Martinez S, Vellei M, Dréau JL. Demand-side flexibility in a residential district: What are the main sources of uncertainty? Energy Build 2022;255. http://dx.doi.org/10.1016/j.enbuild.2021.111595.

[49] Amadeh A, Lee ZE, M.Zhang K. Quantifying demand flexibility of building energy systems under uncertainty. Energy 2022;246:15–27. http://dx.doi.org/10.1016/j.energy.2022.123291.

[50] Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: Analysis, applications, and prospects. IEEE Trans Neural Netw Learn Syst 2021;33(12):6999–7019. http://dx.doi.org/10.1109/TNNLS.2021.3084827.

[51] Huang G, Liu Z, Maaten LVD, Wein KQ. Densely connected convolutional networks. In: IEEE conference on computer vision and pattern recognition. CVPR, Honolulu, HI, USA, 2017, http://dx.doi.org/10.1109/CVPR.2017.243.

[52] Gal Y. Uncertainty in deep learning. Doctor of Philosophy, Department of Engineering, University of Cambridge; 2016.

[53] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: A system for Large-Scale machine learning. In: 12th USENIX Symposium on operating systems design and implementation. Savannah, GA: USENIX Association; 2016, p. 265–83, URL https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi.

[54] Chong A, Augenbroe G, Yan D. Occupancy data at different spatial resolutions: building energy performance and model calibration. Appl Energy 2021;286. http://dx.doi.org/10.1016/j.apenergy.2021.116492.

[55] 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.

[56] 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.

[57] Gibbons J, Chakraborti S. Nonparametric Statistical Inference. Routledge, Taylor and Francis; 2020.

[58] Evans JD. Straightforward statistics for the behavioral sciences. Brooks/Cole Publishing Company; 1996.

[59] Zhu Y, Zabaras N. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. J Comput Phys 2018;366:415–47. http://dx.doi.org/10.1016/j.jcp.2018.04.018.

[60] 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.

[61] 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.

[62] Kendall A, Gal Y. What uncertainties do we need in bayesian deep learning for computer vision? Adv Neural Inf Process Syst 2017;30. http://dx.doi.org/10.48550/arXiv.1703.04977.

[63] ASHRAE. Measurement of energy, demand, and water savings. ASHRAE Guideline 14-2014.

[64] 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.

[65] ASHRAE. Thermal environmental conditions for human occupancy. 2020, ANSI/ASHRAE Standard 55-2020.

[66] 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.

[67] EnergyPlus. Energyplus version 9.1. 2022, https://bigladdersoftware.com/epx/docs/9-1/.

[68] Razghandi M, Zhou H, Erol-Kantarci M, Turgut D. Variational autoencoder generative adversarial network for synthetic data generation in smart home. In: IEEE international conference on communications, Seoul, South Korea. 2022, http://dx.doi.org/10.48550/arXiv.2201.07387.

[69] Bolón-Canedo V, Sánchez-Maroño N, Alonso-Betanzos A. A review of feature selection methods on synthetic data. Knowl Inf Syst 2013;34:483–519. http://dx.doi.org/10.48550/arXiv.2201.07387.

[70] 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.

[71] 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.

[72] Krishnadas G. Data-driven modelling for demand response from large consumer energy assets (Ph.D. thesis), University of Edinburgh; 2018.

[73] Pedregosa, et al. Scikit-learn: Machine learning in Python. JMLR 2011;12:2825–30.

[74] Bigladder Software. EnergyPlus weather file (EPW) data dictionary. 2022, URL https://bigladdersoftware.com/epx/docs/9-3/auxiliary-programs/energyplus-weather-file-epw-data-dictionary.html#energyplus-weather-file-epw-data-dictionary.