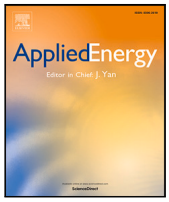
Applied Energy 301 (2021) 117502

Contents lists available at ScienceDirect

Applied Energy

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



A data-driven methodology for enhanced measurement and verification of energy efficiency savings in commercial buildings

Benedetto Grillone [a,*], Gerard Mor [b], Stoyan Danov [a], Jordi Cipriano [b,c], Andreas Sumper [d]

[a] Centre Internacional de Mètodes Numèrics a l'Enginyeria. Building Energy and Environment Group. CIMNE - UPC Campus Terrassa Edifici GAIA, C. Rambla Sant Nebridi 22, 08222 Terrassa, Barcelona, Spain

[b] Centre Internacional de Mètodes Numèrics a l'Enginyeria. Building Energy and Environment Group. CIMNE-Lleida. Pere de Cabrera 16 2. Office G., 25002 Lleida, Spain

[c] Department of Environmental and Soil Sciences, INSPIRES, University of Lleida, Rovira Roure 191, 25198 Lleida, Spain

[d] Centre d'Innovació Tecnològica en Convertidors Estàtics i Accionaments (CITCEA-UPC), Departament d'Enginyeria Elèctrica, Universitat Politècnica de Catalunya, ETS d'Enginyeria Industrial de Barcelona, Av. Diagonal 647, Pl. 2, 08028 Barcelona, Spain

ARTICLE INFO

Keywords:
Building energy retrofit
Measurement and verification
Data driven approach
Generalized additive models
Building energy performance
Energy savings estimation

ABSTRACT

Methods to obtain accurate estimations of the savings generated by building energy efficiency interventions are a topic of great importance, and considered to be one of the keys to increase capital investments in energy conservation strategies worldwide. In this study, a novel data-driven methodology is proposed for the measurement and verification of energy efficiency savings, with special focus on commercial buildings and facilities. The presented approach involves building use characterization by means of a clustering technique that allows to extract typical consumption profile patterns. These are then used, in combination with an innovative technique to evaluate the building's weather dependency, to design a model able to provide accurate dynamic estimations of the achieved energy savings. The method was tested on synthetic datasets generated using the building energy simulation software EnergyPlus, as well as on monitoring data from real-world buildings. The results obtained with the proposed methodology were compared with the ones provided by applying the time-of-week-and-temperature (TOWT) model, showing up to 10% CV(RMSE) improvement, depending on the case in analysis. Furthermore, a comparison with the deterministic results provided by EnergyPlus showed that the median estimated savings error was always lower than 3% of the total reporting period consumption, with similar accuracy retained even when reducing the total training data available.

1. Introduction

In 2018, 36% of final energy use and 39% of energy and process related CO_2 emissions worldwide were attributed to the building and construction sector, a 2019 report from the IEA showed [1]. In the same document, it was reported that the global emissions of the building sector increased 2% in 2018, reaching a record amount of 9.7 gigatonnes of carbon dioxide, and marking a 7% increase from 2010. Different studies have analysed the low turnover rate of the existing built environment and forecasted that up to three quarters of the existing building stock might still be standing in 2050 [2,3]. This means that improving the energy efficiency of the existing building stock is a crucial step to lower the energy demand of the building sector, and to reach the climate goals set in Europe [4] and in the rest of the world [5]. In the last years, there is been many studies that tried to asses the energy impact of entire sectors and possible carbon reduction scenarios [6,7]. On the other hand, it was found that when it comes to the implementation of energy retrofitting measures in individual buildings, one of the greatest challenges is measuring the savings achieved, this process is commonly referred to as the Measurement and Verification process [8].

Measurement and verification (M&V) is defined as the process of using measurements to reliably determine energy savings generated within an individual building or facility by an energy efficiency intervention. Since energy savings cannot be directly measured, as they represent the absence of energy use, they are usually determined by comparing the facility's energy consumption before and after the implementation of a retrofit measure, considering appropriate adjustments for possible changes in conditions. Different methodologies and protocols have been developed for this scope, the main being: the International Performance Measurement and Verification Protocol (IPMVP), and the ASHRAE Guideline 14 [9]. Both these methodologies are based on the use of a baseline energy model to compare the energy

* Corresponding author.
E-mail address: bgrillone@cimne.upc.edu (B. Grillone).

https://doi.org/10.1016/j.apenergy.2021.117502
Received 26 November 2020; Received in revised form 19 July 2021; Accepted 28 July 2021
Available online 6 August 2021

Abbreviations

BIC	Bayesian Information Criterion
COP	Coefficient Of Performance
CV(RMSE)	Coefficient of Variation of the Root Mean Square Error
EEM	Energy Efficiency Measure
GAM	Generalized Additive Model
GHI	Global Horizontal Radiation
GBM	Gradient Boosting Machine
GOF	Goodness of Fit
HVAC	Heating, Ventilating and Air Conditioning
IPMVP	International Performance Measurement and Verification Protocol
M&V	Measurement and Verification

consumption before and after the Energy Efficiency Measure (EEM) implementation. The baseline model can be defined as the energy characterization of the starting situation, and its role is fundamental in the assessment of energy savings. In fact, the baseline model has the task of segregating the effects of a retrofit programme from the effects of other simultaneous changes that can affect the energy consumption, hence improving the accuracy with which energy savings are estimated.

Recent breakthroughs in advanced metering infrastructure technologies and data analytics techniques have initiated a transition of M&V to a new stage, also known to practitioners as *advanced measurement and verification* (or M&V 2.0). The main characteristic of M&V 2.0 is the possibility of performing real-time energy efficiency savings estimations, thanks to the analysis of datasets having high sampling frequency and resolution [10]. This translates into the possibility of providing dynamic energy insights, maximizing the savings estimation accuracy, and obtaining a detailed characterization of the building's energy usage and performance [11]. In the framework of advanced M&V, great effort is being put into the development of advanced baseline models that can reach high estimation accuracy thanks to the application of state-of-the-art statistical and machine learning techniques. In addition to that, many advanced M&V tools are now offering several other energy services, such as consumption profile data mining, equipment fault detection, and building energy benchmarking [12].

Comprehensive reviews regarding the state-of-the-art of data-driven measurement and verification methodologies were carried out in the last years [13,14]. While different methods have been identified to have different strengths and weaknesses, the reviews highlighted the following knowledge gaps:

- none of the reviewed methodologies make use of the typical consumption patterns detected in the analysed facilities as a direct predictor variable,
- all of the reviewed methodologies utilize energy baseline models which are trained only on the data that precedes the EEM implementation. This causes the loss of valuable information regarding how energy consumption fluctuates as a consequence of the variations of outdoor climate variables, which is contained in the data that follows the EEM implementation and which can be extracted (and uncoupled from the effects induced by the measures) by means of statistical methods introduced in the present paper.

As a consequence, the results provided by most of the reviewed methodologies are affected by a rapid degradation when less than a full year of training data is available. At the same time, most of the existing methodologies only provide an estimation of energy efficiency savings, with no additional information about the energy performance and behaviour of the analysed facilities. The novel approach proposed in this publication addresses the presented gaps by developing an enhanced data-driven methodology able to:

- consider typical building usage profiles as predictor variables in the energy baseline model, introduced by means of daily consumption patterns detected with a clustering methodology,
- perform a detailed characterization of the building energy usage, harnessing also data from the post-measure application period,
- provide accurate results even when less than a full year of training data is available,
- provide, together with the savings estimation, a range of actionable insights into the energy performance of the analysed facilities, such as typical load consumption profiles, change-point temperature of the building, and weather dependence analysis.

This represents a significant advance to the state-of-the-art, especially in practical applications, e.g. implementation of Energy Performance Contracts, where accurate short-term monitoring with high granularity could be feasible, but year-round data acquisition is not practical from a business point of view. This feature considerably widens the scope of application of the proposed methodology, also considering that the calculations performed are based on data that can be easily obtained, such as energy consumption and climate data, making the method easy to apply to a wide range of existing buildings.

The viability of the approach was tested in two different case studies: the first was carried out with a set of synthetic data generated using the energy modelling software EnergyPlus [15], while the second was performed using a set of real consumption data from three commercial buildings located in Barcelona (Spain), where different EEMs have been implemented. For case study 1, the operation of three commercial building typologies was simulated in different climate locations, and for each of them two EEMs were introduced, at different points of the time-series, for a total of 54 unique simulation cases. Each of the buildings was simulated with and without the energy efficiency intervention, thereby allowing a comparison between the effective energy savings, simulated with EnergyPlus, and the ones estimated with the data-driven statistical model. Such a comparison made it possible to verify that low statistical model errors are translated into accurate estimations of the energy savings achieved by the EEMs. The proposed approach is also compared to another method based on the time-of-week-and-temperature (TOWT) model [16], a commonly employed method in measurement and verification applications that was the object of different studies to estimate prediction accuracy [13,17,18] and uncertainty [19] in measurement and verification applications. The TOWT model is also the reference model of the CalTRACK protocol [20]. For the second case study, data from three existing buildings that implemented EEMs were analysed. The two case studies have two complementary objectives: with the first it was possible to test the methodology and showcase its features in a controlled environment, where the true savings values are known, since they can be estimated using EnergyPlus deterministic calculations. This provides a concrete value to compare the results to, and a benchmark for model performance. Conversely, the second case study has the goal of demonstrating that the proposed methodology is able to work properly not only with simulated data, but can also provide accurate results with low error when analysing monitoring consumption data from existing buildings.

The present study provides significant research contribution to each of the needs described at the beginning of this section, demonstrating considerable accuracy improvements, compared to the industry benchmark, which are even more evident when the training data used is reduced to less than a full year. At the same time, the results yielded are easily interpretable even for non experts in statistics, providing not only energy savings quantification but also a wide range of actionable insights into the energy performance of the analysed building.

2. Methodology

2.1. Methodology overview

The energy savings calculation methodology proposed in this article is based on two main concepts: (i) the baseline energy use of

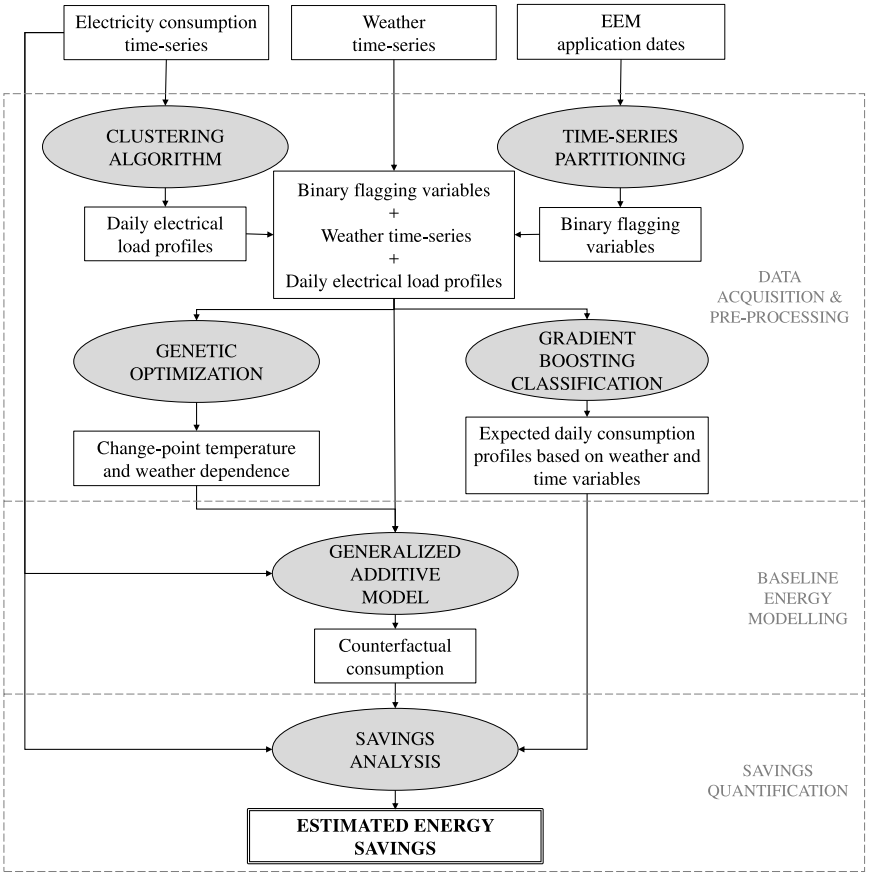


Fig. 1. Methodology flowchart.

a commercial building or facility can be represented by a statistical model; (ii) when an EEM is implemented in that building, the variation in the energy use caused by this EEM can be represented by adding supplementary terms to the initial baseline model. This last concept defines one of the main differences of this paper's approach and the M&V methodologies currently employed in industry: traditional methods are based on the theory that it is possible to detect energy behaviour changes by training a model with historical data of the period that precedes the application of the EEM, and then to predict energy consumption on the post EEM application period. In contrast, in the methodology presented in this article, the detection of energy behaviour changes does not happen thanks to the choice of a specific training dataset, but through the definition and selection of different additive terms in the model itself. There are two main advantages related to this revised paradigm: the first one is that the whole time-series data can be used to train the baseline model, which enables an improved description of the buildings' energy performance; the second one is that, in case of more than one EEM implementation, there is no need to fit a different model for each measure. The technical details of this procedure are further explained in Sections 2.3 and 2.4 .

With regard to the method to evaluate the impact of EEMs, it is supposed that implemented measures can have a two-fold effect on the energy usage of a building: they either cause a change in the daily load profile, or they modify the way the building consumption increases (or decreases) as a consequence of the variations of outdoor climate variables. Both effects are considered in the assessment of the counterfactual consumption, which is the estimated energy consumption following an intervention, as if the intervention had not taken place. To evaluate the first of these two effects, when estimating the daily counterfactual usage, instead of using the detected load profile for that day, the profile used is the one the building would have had if the measure was not applied, calculated using a prediction algorithm based on time of the year and weather variables. On the other hand, to account for weather related changes of the energy consumption, different additive terms are considered in the baseline model, that have the goal of capturing how the building weather dependence changes after the application of a given measure. The technical details of this process are explained in detail in Sections 2.2, 2.3, and 2.4.

In order to estimate the energy savings due to EEMs, the proposed approach requires the following data:

- energy consumption data (with hourly or sub-hourly granularity),
- outdoor temperature,
- global horizontal radiation (GHI),
- wind speed,
- public holidays calendar,
- date of application of the analysed EEMs.

Once these data sets are collected, the procedure to estimate the energy savings can be initiated. The methodology structure can be divided in three main phases: data acquisition and preprocessing, model computation, and energy savings quantification. Fig. 1 illustrates the process flow diagram of the methodology.

2.2. Phase 1: data acquisition and preprocessing

In this first phase of the approach, different techniques are used to extract all the variables necessary for the calculation of the baseline model, including the identification of the load profile patterns, and an analysis of the weather dependence of the building.

2.2.1. Time-series partitioning through binary flagging variables

Binary flagging variables are introduced in this model to mark the periods of application of energy efficiency measures. For every Energy Efficiency Measure, EEM_i, that is eligible for evaluation, the sections of the time-series before and after the measure application are marked with a binary variable, $m_{i,j}$, having 0 values for the days before the measure application date, and values of 1 for the days after the measure application date. Thanks to this binary flagging variable, it is possible to separate the sections of the time series where the EEM is applied and the ones where it is not. The mathematical definition of this binary variable is expressed as:

$$m_{i,j} = \begin{cases} 0, & \text{if measure } i \text{ is not applied on day } j \\ 1, & \text{if measure } i \text{ is applied on day } j \end{cases} \quad (1)$$

In addition to $m_{i,j}$, another variable m is computed and assigned to each day j of the time-series, indicating the total number of EEMs applied in the building:

$$m_j = \sum_{i=1}^{n} m_{i,j} \quad (2)$$

where n is the total number of measures applied on the building. This means that if in a building two different measures were applied, one in January and one in September of the same year, m will have value 0 before January, 1 between January and September, and 2 after September. m and m_i mark which measures are applied (and which are not) on any given day of the time series, these variables allow for an easy identification of the consumption trends during the model training phase, and help in the selection of the additive terms that form the baseline model.

2.2.2. Identification of the daily electrical load profiles

To identify the daily energy usage patterns of the building, a clustering algorithm is applied. The different days of the time-series are separated into clusters of similar daily behaviour, using as input of the algorithm the building's energy consumption values, sampled every hour. Different clustering methods have been tested in literature for this purpose, with k-means clustering [21], self organizing maps (SOM) [22], and Symbolic Aggregate approXimation (SAX) [23], being popular choices. In this research, the clusters are identified using a Gaussian mixture model [24,25], and the optimal number of clusters k is chosen according to the Bayesian Information Criterion (BIC) of the model. The result of the clustering technique is a set of k centroid curves for 24 h electrical load profiles, which define the typical patterns for the analysed building. To avoid including any potential change of profile generated by an EEM application in the clusters, the profile patterns are identified using only consumption data previous to the first EEM application. Adopting the symbols introduced previously, the clustering is run on all the days that have $m = 0$, while for the days having $m > 0$, the profiles are identified by assigning each day to one of the clusters previously detected. This profile classification is achieved by calculating the cross euclidean distance matrix between the centroids of each of the clusters and the consumption profiles of each day.

2.2.3. Prediction of expected daily profiles based on weather and time variables

Once the daily load profiles have been classified in different clusters, an additional model predicts, for each day of the time-series, which should be the expected load profile on that day. The variables on which this model is based are: day of the week, day of the year, outdoor temperature, global horizontal radiation (GHI), and the electrical load profiles of the 7 days previous to the considered day. More specifically, the goal is to predict, based on these variables, which would have been the load profile of any day after the first EEM application, if no EEMs were applied in the building. To achieve this, the time-series is split in different sections, according to the binary flagging variables previously introduced. Then, for every section of the time-series after the first EEM implementation, the expected profile of the day is estimated using a model trained on data from the preceding time-series section. A concrete example follows, to better illustrate this process: in a building with two applied EEMs, the time series will be split in three parts: the first part, identified by $m = 0$, prior to the first measure application, the second part ($m = 1$), that includes all the days between the implementation of the first measure and the implementation of the second, and the third and last section ($m = 2$), formed by all the days from the application of the second measure until the end of the time-series. Once the split is completed, two different models are run: to predict the daily profiles during section $m = 1$, the model is trained with all the data contained in section $m = 0$, while to predict the profiles of the section that follows the application of the second measure ($m = 2$), the model is trained using the data of section $m = 1$.

To improve the accuracy of this classification, the input variables undergo a preprocessing phase: the day of the week and day of the year variables are encoded as cyclical features using Fourier's transformation [26], while the outdoor temperature and GHI are transformed using spline functions. The classification model used for this task is a gradient boosting machine (GBM), a machine learning algorithm that has been previously applied in other research works related to building energy, and has been shown to have strong predictive performance and flexibility [27,28]. To execute the model, the XGBoost [29] library was used, in the R programming environment. The results generated here, are then used in the savings quantification phase to detect if applied EEMs caused a change in the load profile of the building and if this lead to energy efficiency savings.

2.2.4. Weather dependence analysis

In the weather dependence analysis, the temperature dependence of the building is identified, more specifically the change-point temperature of the building is calculated, and therefore the days of the year when the consumption is directly dependent on the outdoor temperature values. In order to perform this analysis, the temperature data is pre-processed by applying a first order low-pass filter [30]. This pre-processing is required to take into account that the energy consumption is not directly dependent on the short-term outdoor temperature variation because of the effect of the building's thermal inertia and the thermal resistance. The low-pass filter allows to ignore the fluctuations and to consider only the longer-term trend. The low-pass filtered outdoor temperature is calculated as follows:

$$T_{lp}(t) = \overline{T}(t-1)\alpha + \overline{T}(t)(1-\alpha) \quad (3)$$

where $\overline{T}(t)$ and $T_{lp}(t)$ are the hourly average outdoor temperature and the low-pass filtered temperature at hour t. $\overline{T}(t-1)$ is the hourly average outdoor temperature at hour $t-1$, and α is the smoothing factor of the filter (in this analysis a value of $\alpha = 0.1$ was chosen). The expression 'hourly average temperature' refers to the case of having more than one temperature measurement per each hour, in which case the average of these measurements is taken to be the reference temperature for hour t.

Once the low-pass filtered temperature has been calculated, a piecewise linear model is fitted with these input data. The most significant points of this model are defined by the heating and cooling change-point temperatures, estimated as a change-point temperature T_{cp} plus–minus a hysteresis h which can take on positive or negative values. These two parameters define the outdoor temperatures from which a significant relationship between the building's energy consumption and the outdoor temperature conditions is detected. T_{cp} and h are calculated using a linear regression model and should be understood as a measure of how the daily aggregated electricity consumption varies with the daily average outdoor temperature values.

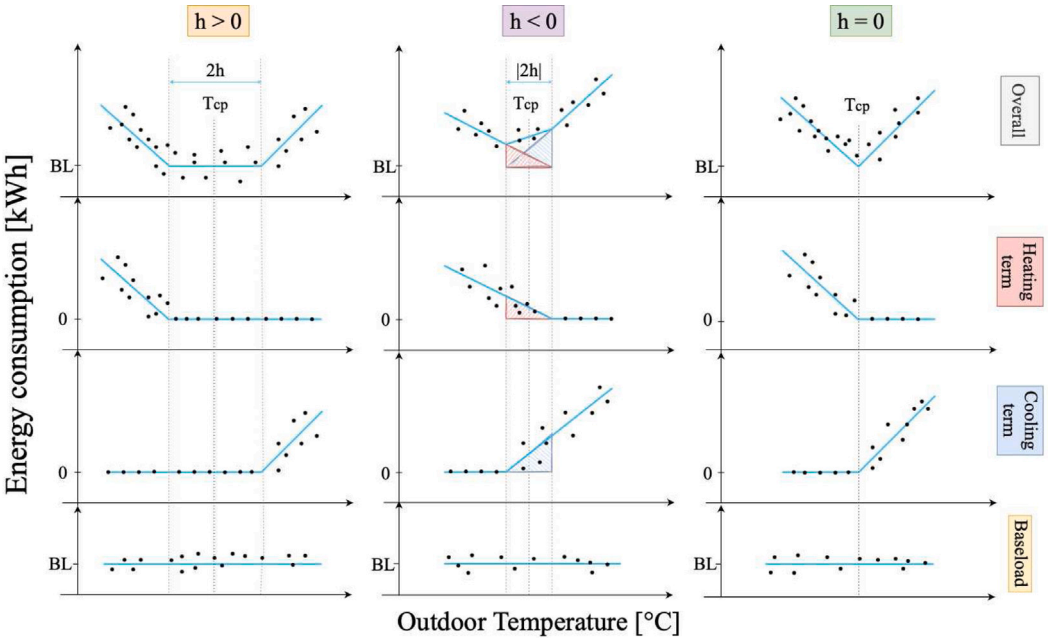


Fig. 2. Possible consumption scenarios depending on T_{cp} and h values.

In order to estimate T_{cp} and h, three possible building operation modes are identified, depending on the hourly average outdoor temperature $\overline{T}$:

$$mode = \begin{cases} heating & \text{if } \overline{T} < (T_{cp} - h) \\ cooling & \text{if } \overline{T} > (T_{cp} + h) \\ equilibrium & \text{if } (T_{cp} - h) \leq \overline{T} \leq (T_{cp} + h) \end{cases} \quad (4)$$

From Eq. (4), it is visible that the equilibrium mode is only possible for buildings having $h > 0$, and that for buildings with $h < 0$, the heating and cooling modes can happen at the same time. Fig. 2 shows the different consumption scenarios for a hypothetical building, depending on the possible values of h. Positive h will result in the building being characterized by an equilibrium range of temperatures around T_{cp}, where the energy consumption is not affected by the change of outdoor temperature. Negative h means that there is a range of temperatures around T_{cp}, in which the building can have heating and cooling dependence at the same time. Finally, $h = 0$ means that the building has heating dependence for $\overline{T} < T_{cp}$ and cooling dependence for $\overline{T} > T_{cp}$.

Once the mode for each day of the time series has been identified, the daily mean dependent temperature T^*_{mode} is calculated, defined as:

$$T^*_{mode} = \begin{cases} (T_{cp} - h) - \overline{T} & \text{if mode = heating} \\ \overline{T} - (T_{cp} + h) & \text{if mode = cooling} \\ 0 & \text{if mode = equilibrium} \end{cases} \quad (5)$$

Regardless of the operation mode, a higher T^*_{mode} represents a higher deviation from the change-point temperature of the building, meaning higher energy requirements to keep the desired internal temperature. The T^*_{mode} is then used to fit the step-wise linear model that estimates the temperature dependence of the building energy usage. It is assumed that the ratio by which the building's consumption depends on the temperature T^*_{mode} varies according to the load profile and the operation mode of the day, for this reason the model is constituted of $K \times 3$ different linear terms, where K is the number of clusters identified as described in 2.2.2. The formulation of the step-wise linear temperature dependence model is the following:

$$Y_d = \sum_{mode,k} C_k \alpha_k + C_k \beta_{mode,k} T^*_{mode} \quad (6)$$

where Y_d is the daily electricity consumption of the building, α_k is the base load for cluster k, $\beta_{mode,k}$ are the linear coefficients that mark the temperature dependence according to the load profile and mode of the day (i.e. $\beta_{heating,1}$ is the coefficient that marks the temperature dependence of the building when it is in heating mode and the day's load profile is represented by cluster 1), and C_k is a variable that marks the cluster of the analysed day and is equal to 1 if the cluster of the day is equal to k, or 0 in all the other cases. In this way, only the relevant linear coefficient is used to describe the temperature dependence of the day (i.e. when the load profile of the day corresponds to cluster 1, $C_1 = 1, C_2 = 0, C_3 = 0, \ldots, C_K = 0$). To identify the linear coefficients $\beta_{mode,k}$, the lasso method (least absolute shrinkage and selection operator) is used. The optimization is initiated with an arbitrary T_{cp} and h, that are then optimized with a genetic algorithm that uses as score the R^2 of the model in Eq. (6). Note that a single T_{cp} is evaluated for the whole building, regardless of the different load profiles, meaning that effect of the change-point temperature variation between different clusters on the daily aggregated consumption is considered to be negligible. Once T_{cp} and h are optimized, the mode of the building for each day of the time-series is calculated using Eq. (4). The building mode identifies the days in which the energy usage of the building is weather-dependent, since it is considered that the weather variables will only affect the electricity consumption if the building needs active heating or cooling. Once the optimized building mode is identified, modified temperature, GHI, and wind speed vectors are calculated, having non-zero values only on the days where the building is considered weather dependent, these vectors are marked by the superscript 'dep':

$$T_h^{dep} = \begin{cases} (T_{cp} - h) - \overline{T} & \text{if mode = heating} \\ 0 & \text{if mode = cooling or mode = equilibrium} \end{cases} \quad (7)$$

$$T_c^{dep} = \begin{cases} \overline{T} - (T_{cp} + h) & \text{if mode = cooling} \\ 0 & \text{if mode = heating or mode = equilibrium} \end{cases} \quad (8)$$

$$GHI^{dep} = \begin{cases} \overline{GHI} & \text{if mode = cooling or mode = heating} \\ 0 & \text{if mode = equilibrium} \end{cases} \quad (9)$$

$$Ws^{dep} = \begin{cases} \overline{Ws} & \text{if mode = cooling or mode = heating} \\ 0 & \text{if mode = equilibrium} \end{cases} \quad (10)$$

where $\overline{GHI}$ and $\overline{Ws}$ represent the average global horizontal radiation and average wind speed of the day. By including these modified vectors T_h^{dep}, T_c^{dep}, GHI^{dep} and Ws^{dep} in the baseline estimation model, it is

possible to take the weather variables into account only on the days when the building is directly dependent on them.

Finally, a last set of vectors necessary for baseline estimation is calculated, by multiplying T^{dep}, GHI^{dep}, and Ws^{dep} by the binary flagging variables m_i. The resulting vectors are used to characterize how the weather dependence of the building changed in time, as different EEMs were applied:

$$T^{dep}_{h,m_i} = T^{dep}_{h} * m_i \tag{11}$$

$$T^{dep}_{c,m_i} = T^{dep}_{c} * m_i \tag{12}$$

$$GHI^{dep}_{m_i} = GHI^{dep} * m_i \tag{13}$$

$$Ws^{dep}_{m_i} = Ws^{dep} * m_i \tag{14}$$

These vectors have a crucial role in the methodology, as they contain the information about the effect of the implemented EEMs on the building energy consumption. It is thanks to these vectors that it is possible to use the whole time-series to train the baseline model, while still excluding the energy reductions caused by the measures when estimating the savings. This process is explained in more detail in the next section, which presents the characteristics of the baseline model.

2.3. Phase 2: baseline energy modelling

This phase has the goal of developing a statistical model able to estimate the daily baseline energy demand of the building under analysis. Among the different possible statistical models that could be used for this task, the authors decided to work with generalized additive models (GAM), which is a specific method for supervised learning originally developed by statisticians Trevor Hastie and Robert Tibshirani [31]. GAMs are flexible statistical methods that can be used to identify and characterize nonlinear effects. In the regression setting, a generalized additive model has the form:

$$g(E(Y)) = \beta_0 + f_1(X_1) + f_2(X_2) + \cdots + f_p(X_p) \tag{15}$$

where X_1, X_2,…,X_p represent the predictors, and Y is the outcome. The f_j may be functions with a specified parametric form (polynomial or un-penalized regression spline, for example), or unspecified 'smooth' functions, to be estimated by non-parametric means. This means that the model allows for rather flexible specification of the dependence of the response on the covariates. The GAM developed in the framework of this research aims to represent daily electricity consumption as a function of outdoor temperature, sun altitude, wind speed, daily load profile, and a holiday flagging variable. The model has the following form:

$$\begin{aligned} g(E(E_d)) = & \alpha T^{dep}_{h} + \beta T^{dep}_{c} + f(GHI^{dep}) + f(Ws^{dep}) \\ & + \sum_{i=1}^{n} [\alpha_i T^{dep}_{h,m_i} + \beta_i T^{dep}_{c,m_i} + f(GHI^{dep}_{m_i}) + f(Ws^{dep}_{m_i})] \\ & + \sum_{i=0}^{n} m_i (\delta_{h,i} d_h + \delta_{nh,i} d_{nh} + \sum_{k=1}^{K} \gamma_{i,k} C_k) \end{aligned} \tag{16}$$

where:

- $E(E_d)$ is the estimator of the daily electricity consumption,
- $\alpha T^{dep}_{h} + \beta T^{dep}_{c} + f(GHI^{dep}) + f(Ws^{dep})$ are the linear and smooth terms marking the part of consumption that depends on weather variables, evaluated on both pre and post measure application data,
- $\alpha_i T^{dep}_{h,m_i} + \beta_i T^{dep}_{c,m_i} + f(GHI^{dep}_{m_i}) + f(Ws^{dep}_{m_i})$ also identify climate-affected consumption, but only during specific sections of the time-series, since the vectors with the subscript m_i, defined in 2.2.4 are non-zero only when measure i is applied,
- n is the total number of energy efficiency measures applied in the building during the considered time-frame,
- $\sum_{k=1}^{K} \gamma_{i,k} C_k$ is the term representing the effect of the daily load profile on the consumption, where $\gamma_{i,k}$ is the coefficient that specifies the impact of the profile, C_k is a variable that marks the profile of the analysed day, being equal to 1 if the profile of the day is profile k, 0 in all the other cases, and K is the total number of daily load profiles identified by the clustering algorithm,
- $\delta_{h,i} d_h + \delta_{nh,i} d_{nh}$ is the term that marks the effect of calendar holidays on the consumption: d_h is a coefficient equal to 1 if the day is a calendar holiday, and 0 if it is not, d_{nh} the opposite, while $\delta_{h,i}$ and $\delta_{nh,i}$ are the coefficients that specify the impact of the holiday variable depending on the EEM applied in a certain period of the time-series,
- m_i are the EEM binary flagging variables, equal to 1 when measure i is effective, 0 elsewhere (m_0 is considered 1 when there are no measures applied, 0 elsewhere).

All the model regressors are estimated in one single training phase, but fitting on different periods of data depending on the regressor. For each of the weather variables, two typologies of regressors are defined: baseline regressors, which are estimated using the whole time-series; and measure effect regressors, that represent the variation from the baseline and are estimated using the time periods where the measure is applied. E.g. for the GHI case, there is a baseline smooth term that represents the building dependence on solar radiation throughout the whole time-series: $f(GHI^{dep})$. Then, additional GHI regressors $\sum_{i=0}^{n} f(GHI^{dep}_{m_i})$ allow to estimate the variations from the baseline introduced by the measures. Thanks to this fitting technique, additional data from the post EEM application date can be included in the evaluation of the baseline dependence coefficients, since the variations generated by the EEM are included in the measure effect set of coefficients. This allows an improved accuracy of the model, even when pre measure implementation time-series data available is short. The methodology used to harness this model for the measurement and verification of energy savings is explained in the next section

2.4. Phase 3: Energy savings quantification

In the last phase of the methodology, the baseline model is used to calculate the counterfactual energy consumption, that allows to estimate the energy savings. The accuracy metrics used to evaluate the goodness of fit of the model are also introduced.

2.4.1. EEM savings evaluation

The savings estimation process has different substeps, firstly the whole time-series is divided in $n+1$ sections ($P_0, P_1, \ldots, P_n$), where n is the total number of measures applied in the building. In order to calculate the savings provided by an individual EEM during period of time P_i, the model computed in Phase 2 is used to predict what would be the energy consumption of the building during that period of time, if the behaviour of the building would still be the one observed in the previous section of the time series $P_{(i-1)}$ (counterfactual). This prediction is realized by applying the GAM model with the model coefficients fitted for section $P_{(i-1)}$, but with the exogenous variables (outdoor temperature, sun altitude, wind speed, etc.) of period P_i. Additionally, to increase the accuracy of the estimation, the daily load profiles introduced in the model (marked by the term C_k in (16)) are not the ones identified with the classifier described in 2.2.2. Instead, the profiles predicted by the classification model detailed in 2.2.3 are used. Thanks to this, if on a given day, the load profile changed because of the implemented EEM, the baseline consumption for that

day will not be calculated based on the new load pattern, but on the one the building would have had if the measure was not applied. The estimated baseline obtained is then compared with the metered energy consumption during the same period of time P_i. If the measure had a positive effect on the energy usage of the building, the baseline consumption for the considered period will be higher than the metered one, and the energy savings achieved can be obtained by calculating the difference between both time series. The energy savings S_i for period P_i can be described as:

$$S_i = \sum_{j \in P_i} g(E(E_{d_j} | i = i-1, C_k = C_{kpred})) - \sum_{j \in P_i} E_{dm_j} \quad (17)$$

where:

- $\sum_{j \in P_i}$ represents the sum over the j days of period P_i being the baseline model a daily model,
- $g(E(E_{d_j} | i = i-1, C_k = C_{kpred}))$ is the estimated baseline consumption on day j, calculated with the model coefficients fit for period $P_{(i-1)}$, and substituting the load profiles identified by the clustering classifier (C_k) with the load profiles predicted by the gradient boosting machine (C_{kpred}),
- E_{dm_j} represents the metered electricity consumption on day j.

2.4.2. Accuracy metrics

Once the energy savings are estimated, different metrics to evaluate the goodness of fit of the model (GOF) are calculated. The GOF of the baseline model is assessed with the coefficient of variation of the root mean square error (CV(RMSE)). It is a metric frequently employed to evaluate the accuracy of energy baseline models:

$$CV(RMSE) = \frac{\sqrt{\frac{1}{n}\sum_i^n (E_i - \hat{E}_i)^2}}{\bar{E}} \times 100 \quad (18)$$

where n are the total days of the time-series, E_i is the measured energy consumption value on day i, $\hat{E}_i$ is the predicted energy consumption with the baseline model, on the same day, and $\bar{E}$ is the average daily energy consumption across the whole time-series. In Eq. (18), it can be appreciated that the CV(RMSE) is a normalization of the root mean square error by the mean of the measured energy consumption values $\bar{E}$, therefore, it represents the size of the model error in relation to the average energy consumption values. For the scope of this article, two additional accuracy metrics are defined, that can be used to calculate how well different models are performing, compared to the ground truth provided by the EnergyPlus simulations. The first of these measures is the normalized percentage difference between the model estimated savings and the simulated savings calculated with EnergyPlus:

$$S_{diff} = \frac{|S_{Eplus} - S_{model}|}{E_{tot}} \times 100 \quad (19)$$

where S_{Eplus} are the overall savings calculated with EnergyPlus for the time period under analysis, S_{model} are the savings estimated with the data-driven model, for the same period, and E_{tot} is the total monitored energy consumption in that time-frame. S_{diff} can be seen as the deviation of the model estimated savings from the theoretical savings, in terms of percentage of the period total consumption. The other metric introduced is the baseline CV(RMSE), a parameter similar to the one presented in Eq. (18). The difference lays in the fact that the CV(RMSE) of Eq. (18) is generally calculated on data that precedes any EEM implementation, since, in real world use cases, there is no measured data about how the energy consumption of the building would have been if the measure was not implemented. Conversely, being in possession of the ground truth provided by EnergyPlus, in the setting of the present article, it is possible to calculate the difference between the baseline consumption obtained with the data-driven models, and the one calculated using the deterministic approach:

$$CV(RMSE)_{BL}[\%] = \frac{\sqrt{\frac{1}{m}\sum_j^m (B_{E+,j} - \hat{B}_j)^2}}{\bar{B}_{E+}} \times 100 \quad (20)$$

where m are the total days after the measure implementation, $B_{E+,j}$ is the EnergyPlus baseline consumption value on day j, $\hat{B}_j$ is the model predicted baseline energy consumption, on the same day, and $\bar{B}_{E+}$ is the average daily EnergyPlus baseline consumption. S_{diff} is a quick and intuitive metric that can be used to evaluate the accuracy of different data-driven models in the estimation of overall yearly savings, while the $CV(RMSE)_{BL}$ can be used to compare how such models are performing in the prediction of the daily counterfactual.

3. Case study

The enhanced M&V methodology proposed in this article was thoroughly tested on two different case studies: one that uses synthetic data generated with simulation software, and one that uses monitoring data from existing buildings. The two case studies are described in detail in this section.

3.1. Case study 1: synthetic data

For the first case study, synthetic data was generated with the building energy simulation software EnergyPlus. Three different building typologies were designed, using the 3D design software SketchUp [32] and its plugin Euclid [33]. Each of the three building typologies was simulated in three different geographical locations and six different sets of implemented energy efficiency measures were considered, meaning that 54 unique simulations were carried out, in order to test the performance of the proposed methodology. Additional white noise was also added to the final simulation time-series, with the goal of improving their resemblance to the stochastic behaviour of non simulated buildings. For each of the 54 cases, three years of building operation were simulated, with two different EEMs introduced at different dates of the time-series. Three sets of simulations were run for each case: (i) first, the whole three years of operation were simulated without any measure applied; (ii) then, with one EEM applied; (iii) and finally, with both EEMs applied. Thanks to this approach, it was possible to simulate the energy savings associated with each EEM. To estimate the accuracy of the proposed M&V approach, a comparison of the simulated savings and the ones estimated by the data-driven model, was performed. The three years of operation that were simulated are 2016, 2017, and 2018, and the EEMs were considered to be applied each at the end of one year of operation (January 1st 2017, and January 1st 2018). Historical weather data for the selected locations was obtained from the Dark Sky weather web service [34] and used in the simulations. It is important to point out that, while it is true that synthetic data is used in this case study, this is neither an attempt to build a calibrated model, nor to create a simulation-based M&V framework. The methodology presented in this article is purely data-driven and does not use any information derived from the simulations' generation process.

The energy consumption time series of one of the simulated building typologies is shown in Fig. 3. The different colours represent the state of the building in terms of the applied EEMs: for the first year of operation (red) no measure is applied, for the second year (green) the first of the two measures is applied ($m = 1$), while, for the third year (blue), both EEMs are applied ($m = 2$). A quick view of this figure shows an appreciable decreasing trend of the energy consumption, which could be associated with the application of each of the EEMs.

3.1.1. Pilot building typologies

As mentioned before, three different building typologies were analysed in this case study: an office, a primary care centre, and a hospital. The three geographical locations considered were Stockholm (Sweden), Berlin (Germany), and Girona (Spain). The U-values for walls, windows, and roofs were selected using the information contained in TABULA [35,36], a WebTool that comprises information about typical U-values of buildings of different age bands across Europe. With regards to the buildings' construction materials and equipment, the buildings

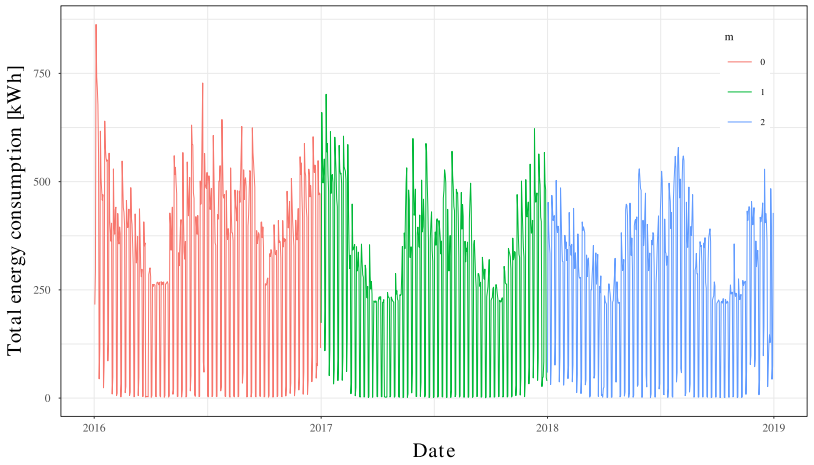
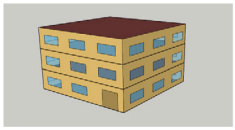
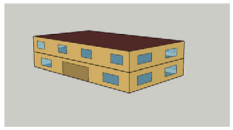
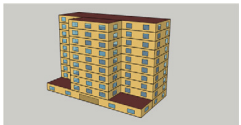


Fig. 3. Consumption time-series for the office building, with colours marking the implemented EEMs.

Table 1
Characteristics of the building typologies used for Case Study 1.

Building block	Pilot 1 - office	Pilot 2 - Primary care	Pilot 3 - Hospital
Building shape			
Total floor area [m^2]	1680	700	9800
Number of floors	3	2	10
Floors to ceiling height [m]	3	3	3
Windows			
U-factor [W/m^2 K]	3	4	3
Solar heat gain coefficient	0.763	0.763	0.763
Visible transmittance	0.812	0.812	0.812
Infiltration	1 ACH	1 ACH	1 ACH
Ventilation (peak) [m^3/(s floor)]	1.32	1.05	2.94
Internal loads (peak)			
Lighting power density [W/m^2]	13	13	8
Plug load power density [W/m^2]	3.55	5	18
Occupancy (peak) [$W/(m^2$ floor)]	4.8	7.2	14.4

were supposed to be built between the 1950s and the 1980s. In all the three building typologies, heat pumps are used for both heating and cooling, with a variable COP ranging between 2.7 and 3.5 depending on the outdoor temperature values. The construction details of the three pilot buildings are presented in Table 1.

3.1.2. EEMs details

Four different EEMs were applied in the pilot building typologies. Their details are outlined here:

- Electric equipment efficiency improvement: the peak plug load power density was reduced by an amount ranging between 2 and 6 W/m^2, depending on the building,
- Lighting efficiency improvement: the peak lighting power density was reduced by 4 W/m^2,
- HVAC set-points shift: the heating temperature set-point was shifted from 21 °C to 19 °C, and the cooling one from 23 °C to 25 °C °, additionally, the functioning hours were changed in order to reduce the HVAC overall usage,
- Building envelope improvement: the total air infiltration was changed from 1 ACH to 0.3 ACH, as well as the UA values for windows, walls and roof. The values depend on the specific buildings and geographic locations, and were chosen taking into account the research carried out within the TABULA project.

These four measures were then combined into six possible combinations of two measures: one at the end of the first year, and one at the end of the second year of simulation. The six combinations are shown in Table 2. Each building typology was simulated in the three previously mentioned geographical locations and with each of the six measure combinations, generating in this way 54 unique simulations.

3.2. Case study 2: monitoring data

The second case study is based on real data from two offices and a cultural building belonging to the regional authority of Catalonia (Spain), and situated in the province of Barcelona. The goal of this case study is to demonstrate that the proposed methodology works well not only on synthetic data coming from a controlled environment simulation, but also on monitoring consumption data coming from real-world buildings. In two of the buildings (that will be named here Building 1, Building 2, and Building 3 due to the confidentiality of the data analysed), a substitution of all the final elements of the illumination system was realized. While for the third, a pack of different retrofit measures was realized, including the replacement of all the building lighting devices, the installation of an intelligent building energy management system, and the installation of solar screens. A summary of the information available for the three buildings, and the respective implemented measures, is available in Table 3.

The electricity consumption time-series of Building 3 is presented in Fig. 4, with the red and blue sections representing the consumption

Table 2
Details of the EEM combinations applied in Case Study 1.

Combination	First year measure	Second year measure
1	Equipment efficiency improvement	HVAC rescheduling and setpoints shift
2	HVAC rescheduling and setpoints shift	Envelope improvement
3	HVAC rescheduling and setpoints shift	Lighting efficiency improvement
4	Lighting efficiency improvement	Equipment efficiency improvement
5	Envelope improvement	Lighting efficiency improvement
6	Equipment efficiency improvement	Envelope improvement

Table 3
Characteristics of the three pilot buildings analysed in Case Study 2.

Building	Building 1 - office	Building 2 - office	Building 3 - cultural
General information			
Total floor area [m^2]	4000+	5000+	15 000+
Date start monitoring	01/08/2018	01/01/2019	01/12/2017
Date end monitoring	31/12/2019	31/12/2019	31/12/2019
EEMs applied			
Measure details	Lighting efficiency improvement	Lighting efficiency improvement	Lighting efficiency + BEMS + solar screens
EEM implementation date	01/06/2019	03/08/2019	01/11/2018

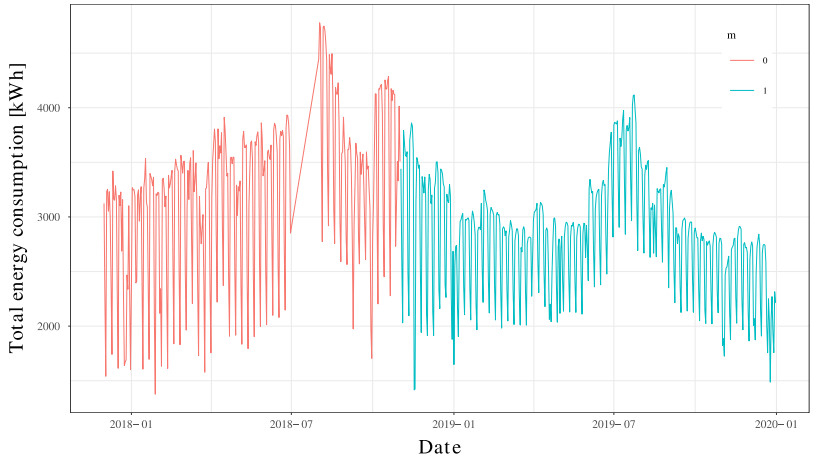


Fig. 4. Consumption time-series for Building 3 of Case Study 2, with colours marking the implemented EEMs.

before and after the implementation of the measures. It can also be seen that one whole month of training data is missing from the consumption time-series, due to a technical malfunctioning of the monitoring system.

4. Results

This section is dedicated to the presentation of the results obtained in the two case studies previously introduced.

4.1. Case study 1

The results for the first case study are divided into two sections: first a single simulated building is analysed in detail, with a comprehensive analysis of all the intermediate results provided by the methodology. Then, the aggregated results obtained for all the 54 analysed buildings are presented and discussed.

4.1.1. Single building analysis

This section has the goal of showcasing the details of the different steps of the methodology, and is related to the analysis of the office building, simulated with Berlin weather data, and having equipment efficiency improvement as first implemented measure and HVAC rescheduling and set-points shift as second implemented measure. This building will be referred to as the *reference building* for the remainder of this section.

The first step of the methodology is the identification of the typical daily consumption profiles. In Fig. 5, the 8 profiles identified for the first year of operation of the office building (before any measure implementation) are displayed: the red lines represent the profile centroids, while the black lines represent the load profiles of days belonging to that cluster.

Once the clusters have been identified, the temperature dependence of the building is estimated for each cluster of days, after obtaining the optimized change-point temperature and hysteresis (for the reference building T_{cp} = 15.2 °C, H = 2.9 °C). Fig. 6 shows the temperature dependence detected for each of the clusters, as the relationship between the daily energy consumption and the T^*, introduced in Eq. (5). Fig. 7 illustrates the time distribution of the different load profiles for year 2016. An analysis of Fig. 6 shows that most of the profiles have heating dependence, except for clusters 5 and 6, representing typical weekday summer profiles. Cluster 8 exemplifies the typical unoccupied profile (weekends and holidays), having very low consumption and slight temperature dependence for both heating and cooling. Distinction between profiles with heating dependence and profiles with cooling dependence can also be drawn by analysing the consumption peaks in Fig. 5: winter profiles have peaks in the early morning, while summer profiles have peaks in the hours of highest solar radiation.

The peak consumption values shown in Figs. 5 and 6 reveal how the heating load for this building is considerably higher than the cooling one. This seems realistic, since the pilot energy use was simulated using weather data from Berlin, where the climate is characterized by cold

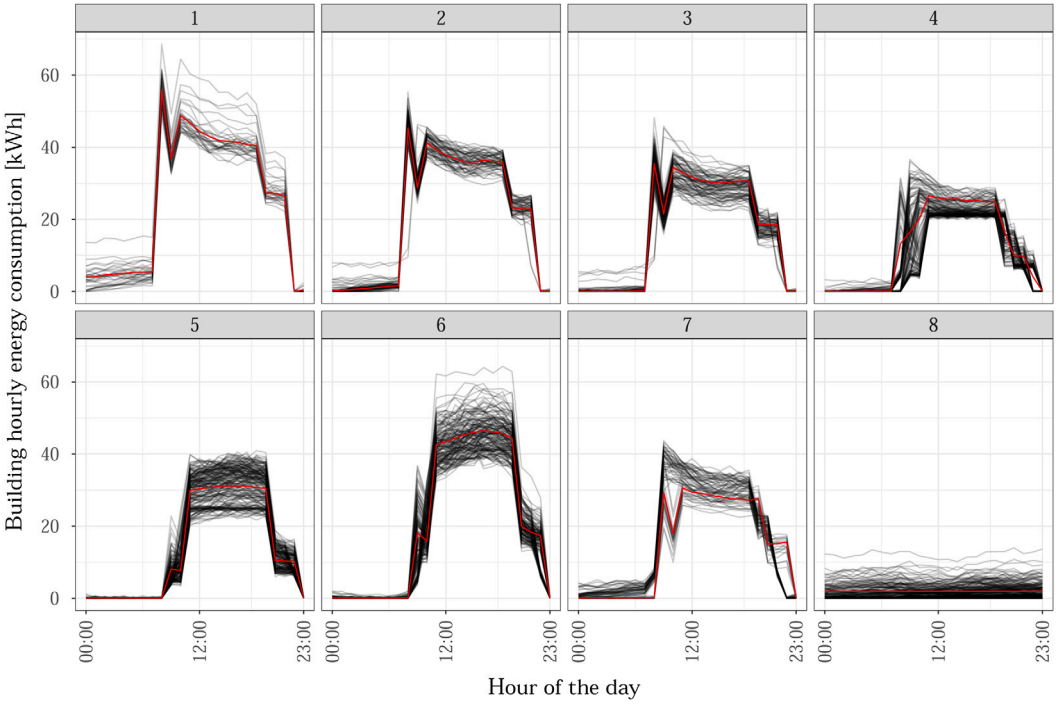


Fig. 5. Load profiles identified for the reference building.

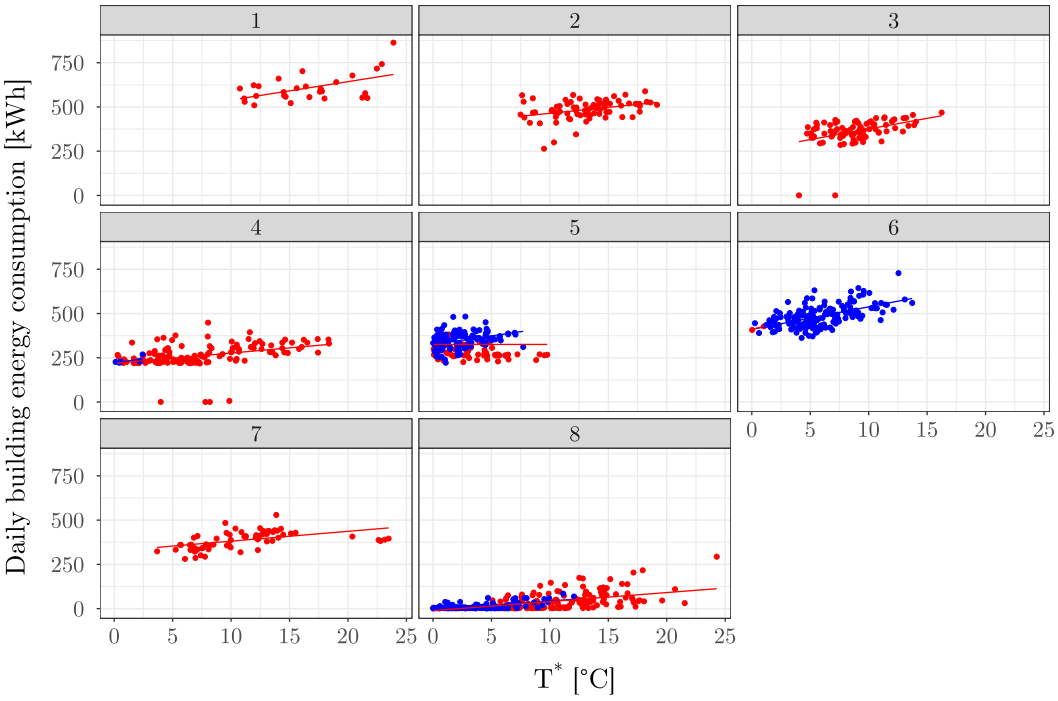


Fig. 6. Heating (red) and cooling (blue) temperature dependence by cluster identified for the reference building (T_{cp} = 15.2 °C h = 2.9 °C).

winters and moderately warm summers. A similar conclusion can be reached by studying Fig. 8, where $T^*_{cooling}$ and $T^*_{heating}$ are shown, with $T^*_{heating}$ reaching higher values compared to $T^*_{cooling}$.

After completing the cluster and weather analyses, the GAM model is fitted to the input data. In Fig. 9, it is possible to see the consumption predictions obtained for the office building in Phase 2 of the methodology, that is before applying the savings evaluation technique. The baseline estimated consumption (in red) is compared to the simulated consumption obtained using EnergyPlus (in black). Fig. 9 shows a high GOF of the model, which appears to be able to capture the energy dynamics of the building, and to accurately predict its electricity consumption for the three years analysed. Note that the statistical model developed in this phase has the goal of simulating the normal operation of the building: no savings are estimated yet at this stage. For the case study proposed, a train/test split was performed in the GAM fitting phase: the model was trained on 80% of the available time-series data, and then tested on the remaining 20%. This is a common practice when dealing with predictive techniques and has the goal of making sure that

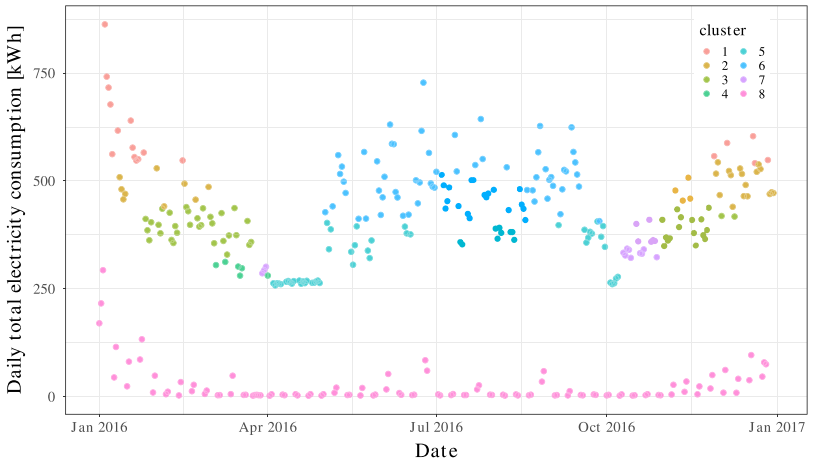


Fig. 7. Cluster distribution identified for the reference building in the first year of operation.

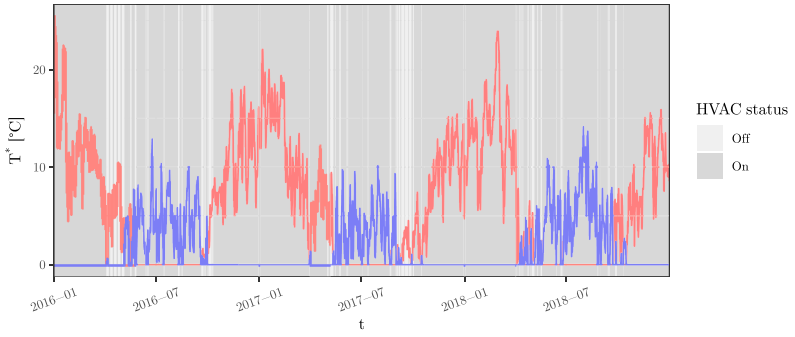


Fig. 8. $T^*_{cooling}$ (blue) and $T^*_{heating}$ (red) profiles identified for the reference building ($T_{cp} = 15.2\,°C$, $H = 2.9\,°C$).

the model is not over-fitting the training data and is able to provide good predictions on new observations. Ultimately, once the GAM coefficients are obtained, the savings for the implemented measures are calculated following the procedure described in Section 2.4.

4.1.2. Overall performance analysis

This part of case study 1 is dedicated to the analysis of the model fit and savings estimation accuracy obtained for the 54 buildings analysed. In order to compare the GAM based methodology and the TOWT model, three different accuracy metrics were calculated:

- the CV(RMSE), calculated on a random holdout set of the training data,
- S_{diff}, the normalized percentage difference between the model estimated savings and the simulated savings calculated with EnergyPlus,
- the daily baseline $CV(RMSE)_{BL}$.

Each of these three metrics has a specific objective: the CV(RMSE) calculated on a holdout set of the training data is a commonly employed metric in the M&V setting to evaluate the goodness of fit of the model and its ability to generalize to unseen data. The normalized difference between the aggregated savings calculated with EnergyPlus and the ones obtained using the data-driven models can be used for a quick estimation of how well the models perform at estimating overall savings across a whole year of operation. While the savings difference is really intuitive in its interpretation, it does not provide information about the accuracy of savings estimations on the daily scale. For this reason, the baseline CV(RMSE) is also included in the analysis. These indicators are described in detail in Section 2.4.2.

Figs. 10, 11, and 12 show the $CV(RMSE)$, S_{diff} and $CV(RMSE)_{BL}$ values obtained for the 54 buildings analysed, where each box represents the distribution of the results for the 18 different simulations belonging to the noted building use (office, primary care centre, hospital). The graphs contain data for the GAM based methodology, and for the CalTRACK implementation of the TOWT model, first using a full year of training data, and then reducing the training data to 9 months. The training data reduction was performed in order to test the hypothesis that one of the main advantages of the methodology proposed in this article is that, thanks to its ability of harnessing data from both the periods before and after an EEM is applied, it is able to accurately predict energy savings even when less than one whole year of historical data, before an EEM is applied, is available.

To estimate the CV(RMSE) for the scenarios with less training data, a 10-fold cross validation approach was used. It is a technique commonly employed within the machine learning community and was introduced in the M&V environment by Touzani et al. [19]. The section of the time-series allocated for training was divided into 10 different sub-samples, called folds. In the first iteration, the baseline model was calculated using the first 9 folds for training, while the validation error was calculated on both the held out fold of the training data *and* the rest of the time-series which was not used for the training. The procedure was then repeated 10 times, holding out a different fold every time, and the final CV(RMSE) was calculated as the mean of the results obtained in each iteration. It was decided to use this technique because it allows the CV(RMSE) to better represent how well the model is able to generalize to a whole year of conditions, when only part of that year is available as training data.

4.2. Case study 2

The results presented for the second case study are more focused on the performance of the prediction algorithms, rather than on the

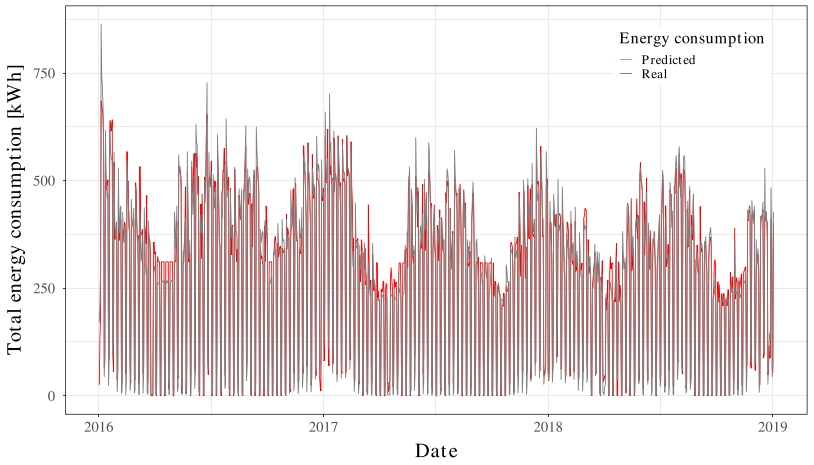


Fig. 9. Predicted (baseline estimated) and real (simulated) consumption time-series for the reference building.

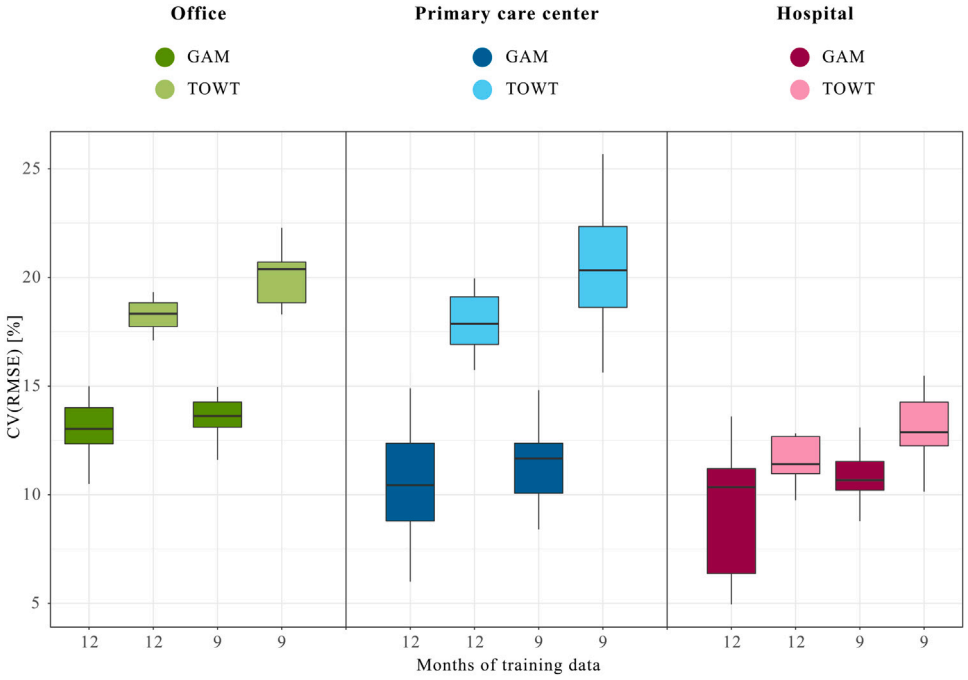


Fig. 10. CV(RMSE) distribution for the 54 buildings analysed in Case Study 1.

intermediate results, such as profile pattern detection and temperature dependence, which have been discussed in detail for case study 1. While the first case study has the goal of showing that the proposed methodology is able to provide actionable insights and accurately estimate energy efficiency savings, the objective of case study 2 is to prove that the methodology works well not only in simulated cases, but also with monitoring time-series data coming from real-world buildings. Synthetic time-series data generated with simulation engines usually does not include stochastic variability observed in the actual energy use of existing occupied buildings, therefore it is important to test the methodology for real-world use cases and ensure that it still provides reasonable results with high goodness of fit.

For the three buildings and measures described in Section 3.2, the methodology was applied and the results obtained were compared with the ones provided by the TOWT model, implemented according to the CalTRACK methodology. The TOWT results were calculated using the Python library *eemeter* [37]. Table 4 summarizes the results obtained for the three test cases. Similarly to case study 1, the table includes: the optimal number of clusters representing the load profiles, the change-point temperature and hysteresis of the building, the CV(RMSE) of the model, calculated on the test set, and the estimated savings for each measure, both in kWh and as percentage of the total reporting period consumption:

The proposed methodology provides 2%–5% lower CV(RMSE) than the TOWT case for all the three analysed buildings. In Fig. 13, the overall fit of the model for Building 1 is also presented, showing high GOF, similar to the one seen for synthetic data in case study 1.

For Case Study 2, the limited availability of monitoring data (less than 1 year for training) did not allow to perform an analysis with reduced training data. Also, the lack of simulated data for this case prevents the comparison of the accuracy of the final savings estimations for the two tested models.

5. Discussion

The two analysed case studies provide important results that showcase the proposed methodology's strengths. The detailed single building analysis of the first case study presents the additional insights that the methodology provides, such as the consumption profiles clustering and

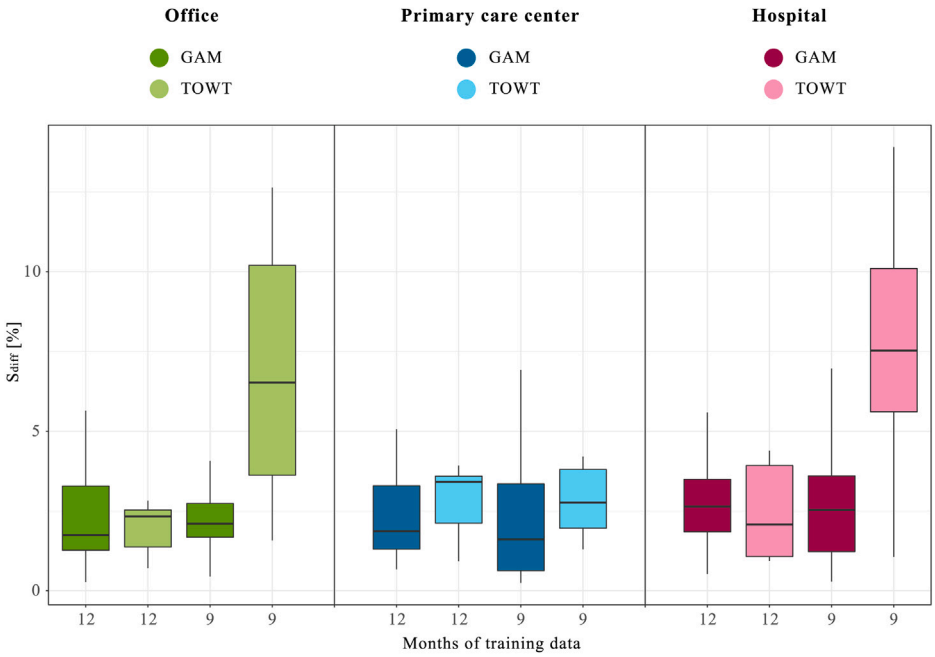


Fig. 11. S_{diff} distribution for the 54 buildings analysed in Case Study 1.

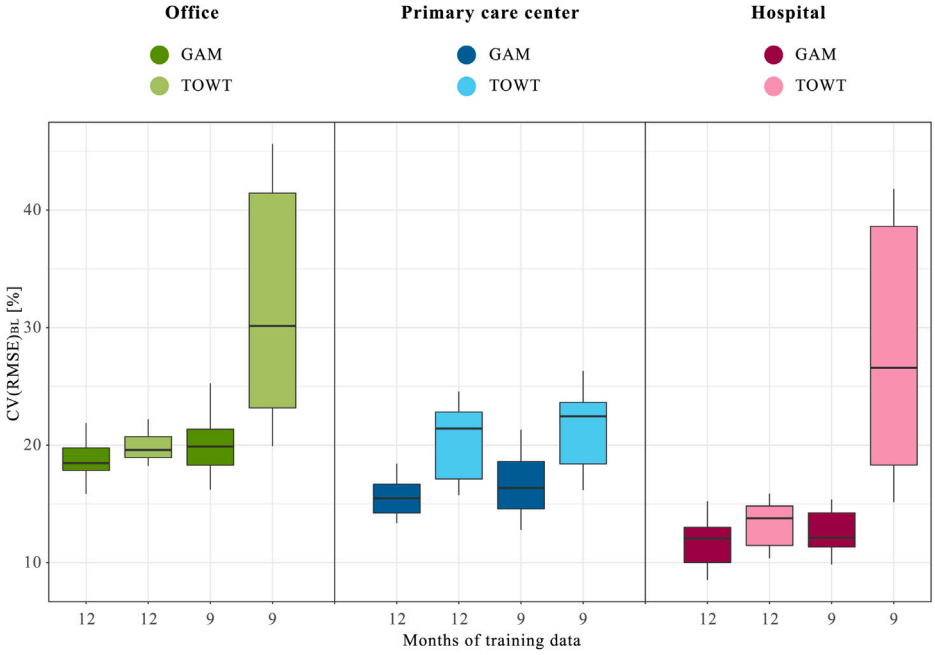


Fig. 12. $CV(RMSE)_{BL}$ distribution for the 54 buildings analysed in Case Study 1.

Table 4
Results obtained for the three buildings in Case Study 2.

Building block	Building 1	Building 2	Building 3
Building information			
Number of clusters	7	7	7
Change-point temperature [°C]	20.8	21.7	21.2
Hysteresis [°C]	3.2	3.8	3
Model metrics			
CV(RMSE)	13%	9.1%	7.8%
CV(RMSE) TOWT	18.1%	11.1%	10%
Savings			
Model estimated savings [kWh]	18 592 (8.6%)	16 644 (10.9%)	108 619 (8.3%)
Savings estimated with TOWT [kWh]	12 979 (6%)	26 313 (17.2%)	122 280 (9,3%)

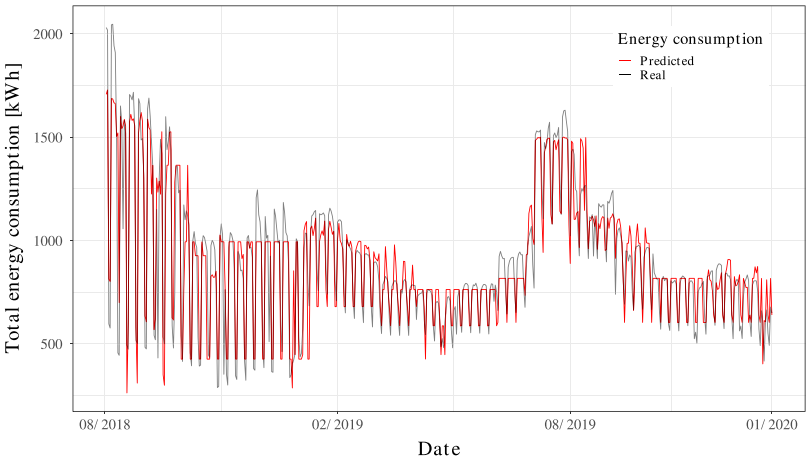


Fig. 13. Predicted (baseline estimated) and real (metered) consumption time-series for Case Study 2, Building 1.

weather dependence analysis. The patterns and dependence detected are coherent with the prior knowledge about the building and the implemented measures, and represent actionable insights that energy building managers can use to improve the operational efficiency of the building. The CV(RMSE) values represented in Fig. 10 show that the GAM based methodology represents a solid GOF improvement compared to the TOWT model, with lower median CV(RMSE) for each of the analysed cases. It is also possible to see that the training reduction from 12 to 9 months marks a definite increase in the model error for the TOWT model, while in the GAM case the median error increase between the 12 and 9 months case is almost negligible. Only in the case of the hospital building, some isolated cases are performing better in the TOWT case than in the GAM, but in terms of median error and distribution, the GAM is still performing better. Fig. 11 shows the savings difference S_{diff} (defined in Section 2.4.2): while the 12 month results are generally comparable between the two models, in two of the three building typologies the training data reduction results into a steep increase in the savings estimation error, with errors higher than 10% of the total building consumption in some cases. While the primary care centre seems to provide more robust results in this sense, likely thanks to a reduced variance in the energy consumption values, which makes it less prone to over-fitting when removing training data, for the other two pilot buildings the TOWT model is characterized by lack of robustness in the estimations, when working with less than a full-year of training data. It is also important to note that the S_{diff} parameter should not be blindly trusted, since good predictions of aggregated yearly savings do not necessarily imply accurate daily baseline estimations. This is evident since a model that underestimates the consumption during certain days of the time-series, and overestimates it during others, will have the same aggregated result as one that provides accurate daily baseline estimation for all of the analysed days. In order to gain a deeper insight into the estimation capabilities of the proposed methodology, and taking into account the mentioned issues of the S_{diff} parameter, an additional metric, $CV(RMSE)_{BL}$ was calculated for all of the 54 analysed cases. The $CV(RMSE)_{BL}$ distribution is shown in Fig. 12, and illustrates again the two trends seen in Figs. 10 and 11: the median of the GAM model is always lower than its TOWT counterpart and, for two of the three buildings, the training data reduction marks a steep increase in the model error. While in terms of overall aggregated yearly savings, the two analysed models are comparable, with each seemingly performing better in certain cases, the $CV(RMSE)_{BL}$ distributions show that in terms of daily savings estimations the GAM model is more accurate than the TOWT model in the majority of the cases, with the TOWT model reaching very high error values when the training dataset is reduced.

The results of case study 2 show that the proposed methodology retains high accuracy also when used to analyse monitoring data from real-world buildings. The CV(RMSE) of the model was below 10% for two of the three buildings of case study 2, and below 15% for the third one. Even though for this case study there is no simulated ground truth to compare the estimated savings, it is still possible to see that even with missing data, and the inherent stochastic variability of real-world buildings, the proposed methodology is able to provide accurate results with high goodness of fit.

The results from both case studies show an overall superior accuracy of the proposed GAM-based methodology on the CalTRACK implementation of the TOWT model, which was taken as the industry benchmark. There might be different factors driving this performance difference: firstly, the proposed methodology implements different techniques to accurately describe the energy usage of the building before modelling the energy consumption and counterfactual. Secondly, the GAM-based methodology is able to harness part of the post EEM application data to increase the accuracy of the energy baseline model. Moreover, the improved performance of the proposed model is also achieved thanks to its ability to harness additional variables which are ignored by the TOWT model, such as global horizontal radiation and wind speed, and to aggregate and transform known variables into information valuable for the prediction, such as typical consumption profile patterns and weather dependence. Another feature that grants the proposed model increased accuracy is the preprocessing that some of the explanatory variables undergo, before being employed in the model. Specifically, time features such as day of the week and time of the year are transformed using Fourier decomposition, and temperature is treated with a low pass filter. Additionally, thanks to the weather dependence analysis module, the weather variables are only included in the model for those days when they are supposed to be affecting the building consumption. Finally, the methods described in this article are specifically aimed at estimating savings on the daily timescale, while the TOWT provides hourly savings which can then be aggregated on a daily level. This means that the TOWT model may have wider applicability, but lower accuracy on the daily timescale.

On a more general level, the results show that with only the use of consumption monitoring and weather data, the proposed methodology is able to achieve low model errors, with CV(RMSE) lower than 15% for all the buildings in both case studies. Regarding the availability of the necessary weather data, accurate historical GHI data can be obtained for European locations through the Copernicus Atmosphere Monitoring Service (CAMS) [38]. Other weather services and APIs are also available both in Europe and globally, providing historical wind speed, GHI, and outdoor temperature data free of charge, some of them are: DarkSky [34], Meteonorm [39], OpenWeatherMap [40]. Given the relative ease of acquisition of the necessary weather data through different weather services and APIs, the proposed methodology is able to accurately detect savings, having as only requirement the availability of

hourly electricity monitoring data. Including in the models additional variables such as occupancy levels or technical details of the different systems in use in the building is likely to increase the accuracy of the savings estimation, but at the same time would highly reduce the applicability of the methodology.

6. Conclusions and future work

In this article, an innovative methodology for the measurement and verification process of energy efficiency savings in commercial buildings was presented. The proposed approach has three main phases: first a data-preprocessing phase, where different algorithms are used to characterize the building and extract valuable information such as typical electricity load profiles, change-point temperature, and climate dependence. In the second phase, a generalized additive model is fit, using daily consumption and climate data, as well as the variables obtained in the previous phase, to estimate the baseline energy usage of the analysed building. Finally, in the last phase, for each energy efficiency measure applied, the savings are estimated as the difference between the metered energy consumption and the estimated baseline consumption for the same period. The presented approach was tested to detect savings in two separate case-studies. The first case study analyses three different building typologies over a three-year period. The buildings' electricity consumption time-series were simulated using the energy modelling software EnergyPlus. Each pilot had a different use (office, primary care centre, hospital), and they were simulated with different climate conditions (oceanic, Mediterranean, continental), and implemented EEM combinations, for a total of 54 unique simulated cases. The second case study involves the analysis of monitoring data from three real-world buildings located in Barcelona (Spain). In both case studies, the model proved to be able to capture the dynamics of the buildings, providing CV(RMSE) below 15%. For case study 1, the median difference between the savings estimated with the proposed methodology, and the ones obtained with the deterministic approach never exceeded 3% of the total reporting period consumption. The results were also compared to the ones obtained by applying the time-of-week and temperature (TOWT) model, following the CalTRACK methodology, with the proposed model showing overall superiority, with up to 10% lower CV(RMSE) than TOWT in both case studies.

To conclude, four main strengths were identified for the proposed methodology, linked with the ability of this approach to provide:

i. high accuracy in the estimation of savings associated with the predictions;
ii. a robust savings quantification that, for the pilot buildings analysed in case study 1, does not to reduce drastically when less than a full year of training data is available;
iii. wide applicability, since the data required is limited to hourly metered electricity consumption and weather data, gradually more available for most commercial buildings;
iv. additional actionable information to stakeholders, such as: typical load consumption profiles, change-point temperature of the building, and weather dependence evaluation.

Although the preliminary results obtained in this study seem promising, it is clear that more work is needed in order to further validate the approach, and accurately define its strengths and weaknesses. Future work might include the use of automated feature selection techniques to include or omit certain weather variables depending on the specific building analysed. An additional level of insight could also be added by estimating an independent change-point temperature for each detected load profile cluster in the building. At the same time, it appears clear that all the listed benefits come at the expense of adding more complexity to other commonly used approaches. Although the proposed methodology provided high accuracy of estimation in both the synthetic and real data use-cases, it would be optimal to test the presented techniques on a large cluster of real-world buildings, in order to assess how well the model generalizes to new buildings, and if the predictive accuracy is similar to the one obtained with simulated data in this study.

CRediT authorship contribution statement

Benedetto Grillone: Conceptualization, Methodology, Software, Writing – original draft. **Gerard Mor:** Methodology, Software, Resources. **Stoyan Danov:** Conceptualization, Writing - review & editing. **Jordi Cipriano:** Validation, Writing - review & editing. **Andreas Sumper:** 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.

Acknowledgements

This work was supported by the European Commission through the H2020 project SENSEI [grant number 847066]. The authors thank the Catalan Institute of Energy (ICAEN) for providing the monitoring and EEM data that was analysed in the case study.

References

[1] International Energy Agency. 2019 global status report for buildings and construction. Tech. rep., 2019, p. 41.
[2] Wang Na, Phelan Patrick E, Gonzalez Jorge, Harris Chioke, Henze Gregor P, Hutchinson Robert, et al. Ten questions concerning future buildings beyond zero energy and carbon neutrality. Build Environ 2017;119:169–82, URL: https://linkinghub.elsevier.com/retrieve/pii/S0360132317301579 (visited on 05/14/2020).
[3] International Energy Agency, ed. Transition to sustainable buildings: strategies and opportunities to 2050. Paris: IEA Publ; 2013.
[4] European Commission-DG Ener. Clean energy for all Europeans. Publications Office of the European Union; 2019, URL: https://publications.europa.eu/en/publication-detail/-/publication/b4e46873-7528-11e9-9f05-01aa75ed71a1/language-en.
[5] United Nations Framework Climate Convention on Climate Change. Paris agreement. 2016, URL: https://unfccc.int/sites/default/files/english_paris_agreement.pdf (visited on 06/03/2020).
[6] Ma Minda, Cai Wei, Cai Weiguang. Carbon abatement in China's commercial building sector: A bottom-up measurement model based on Kaya-LMDI methods. Energy 2018;165:350–68, URL: https://linkinghub.elsevier.com/retrieve/pii/S0360544218318346 (visited on 03/09/2021).
[7] Ma Minda, Ma Xin, Cai Wei, Cai Weiguang. Low carbon roadmap of residential building sector in China: Historical mitigation and prospective peak. Appl Energy 2020;273:115247, URL: https://linkinghub.elsevier.com/retrieve/pii/S0306261920307595 (visited on 03/09/2021).
[8] EVO. International performance measurement and verification protocol. Tech. rep., 2017.
[9] ASHRAE. ASHRAE guideline 14–2014, measurement of energy, demand, and water savings. ASHRAE Atlanta; 2014.
[10] Franconi Ellen, Gee Matt, Goldberg Miriam, Granderson Jessica, Guiterman Tim, Li Michael, et al. The status and promise of advanced M&V: An overview of "M&V 2.0" methods, tools, and applications. Tech. rep. LBNL–1007125, 1350974, 2017, URL: http://www.osti.gov/servlets/purl/1350974/ (visited on 10/28/2019).
[11] Gallagher Colm V, Leahy Kevin, O'Donovan Peter, Bruton Ken, O'Sullivan Dominic TJ. Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0. Energy Build 2018;167:8–22, URL: https://linkinghub.elsevier.com/retrieve/pii/S0378778817336630 (visited on 06/26/2019).
[12] Granderson Jessica, Fernandes Samuel. The state of advanced measurement and verification technology and industry application. Electr J 2017;30(8):8–16, URL: https://linkinghub.elsevier.com/retrieve/pii/S1040619017301999 (visited on 10/25/2019).
[13] Granderson Jessica, Touzani Samir, Custodio Claudine, Sohn Michael D, Jump David, Fernandes Samuel. Accuracy of automated measurement and verification (M&V) techniques for energy savings in commercial buildings. Appl Energy 2016;173:296–308, URL: https://linkinghub.elsevier.com/retrieve/pii/S0306261916305050 (visited on 06/26/2019).

[14] Grillone Benedetto, Danov Stoyan, Sumper Andreas, Cipriano Jordi, Mor Gerard. A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. Renew Sustain Energy Rev 2020;131:110027, URL: https://linkinghub.elsevier.com/retrieve/pii/S136403212030318X (visited on 08/10/2020).
[15] EnergyPlus Development Team. Energyplus engineering reference: the reference to energyplus calculations. 2019.
[16] Mathieu Johanna L, Price Phillip N, Kiliccote Sila, Piette Mary Ann. Quantifying changes in building electricity use, with application to demand response. IEEE Trans Smart Grid 2011;2(3):507–18, URL: http://ieeexplore.ieee.org/document/5772947/ (visited on 10/24/2019).
[17] Granderson Jessica, Touzani Samir, Fernandes Samuel, Taylor Cody. Application of automated measurement and verification to utility energy efficiency program data. Energy Build 2017;142:191–9, URL: https://linkinghub.elsevier.com/retrieve/pii/S0378778817300294 (visited on 06/26/2019).
[18] Miller Clayton, Meggers Forrest. Mining electrical meter data to predict principal building use, performance class, and operations strategy for hundreds of non-residential buildings. Energy Build 2017;156:360–73, URL: https://linkinghub.elsevier.com/retrieve/pii/S037877881732488X (visited on 10/04/2019).
[19] Touzani Samir, Granderson Jessica, Jump David, Rebello Derrick. Evaluation of methods to assess the uncertainty in estimated energy savings. Energy Build 2019;193:216–25, URL: https://linkinghub.elsevier.com/retrieve/pii/S0378778818337940 (visited on 10/17/2019).
[20] CalTRACK Initiative. CalTRACK. Publication Title: CalTRACK. URL: https://www.caltrack.org/ (visited on 10/29/2019).
[21] McLoughlin Fintan, Duffy Aidan, Conlon Michael. A clustering approach to domestic electricity load profile characterisation using smart metering data. Appl Energy 2015;141:190–9, URL: https://linkinghub.elsevier.com/retrieve/pii/S0306261914012963 (visited on 05/15/2020).
[22] Panapakidis Ioannis P, Papadopoulos Theofilos A, Christoforidis Georgios C, Papagiannis Grigoris K. Pattern recognition algorithms for electricity load curve analysis of buildings. Energy Build 2014;73:137–45, URL: https://linkinghub.elsevier.com/retrieve/pii/S0378778814000231 (visited on 05/15/2020).
[23] Capozzoli Alfonso, Piscitelli Marco Savino, Brandi Silvio, Grassi Daniele, Chicco Gianfranco. Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings. Energy 2018;157:336–52, URL: https://linkinghub.elsevier.com/retrieve/pii/S0360544218309617 (visited on 05/15/2020).
[24] Melzi Fateh, Same Allou, Zayani Mohamed, Oukhellou Latifa. A dedicated mixture model for clustering smart meter data: Identification and analysis of electricity consumption behaviors. Energies 2017;10(10):1446, URL: http://www.mdpi.com/1996-1073/10/10/1446 (visited on 05/15/2020).
[25] Li Kehua, Ma Zhenjun, Robinson Duane, Ma Jun. Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering. Appl Energy 2018;231:331–42, URL: https://linkinghub.elsevier.com/retrieve/pii/S0306261918313606 (visited on 05/15/2020).
[26] Bacher Peder, Madsen Henrik. Load forecasting for supermarket refrigeration. In: Applied energy. 163:32–40. URL: https://doi.org/10.1016/j.apenergy.2015.10.046.
[27] Robinson Caleb, Dilkina Bistra, Hubbs Jeffrey, Zhang Wenwen, Guhathakurta Subhrajit, Brown Marilyn A, et al. Machine learning approaches for estimating commercial building energy consumption. Appl Energy 2017;208:889–904, URL: https://linkinghub.elsevier.com/retrieve/pii/S0306261917313429 (visited on 05/15/2020).
[28] Touzani Samir, Granderson Jessica, Fernandes Samuel. Gradient boosting machine for modeling the energy consumption of commercial buildings. Energy Build 2018;158:1533–43, URL: https://linkinghub.elsevier.com/retrieve/pii/S0378778817320844 (visited on 01/14/2020).
[29] Chen Tianqi, Guestrin Carlos. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining - KDD '16. 2016, p. 785–94, URL: http://arxiv.org/abs/1603.02754 (visited on 01/14/2020).
[30] Bacher Peder, Madsen Henrik, Nielsen Henrik Aalborg, Perers Bengt. Short-term heat load forecasting for single family houses. Energy Build 2013;65:101–12, URL: https://linkinghub.elsevier.com/retrieve/pii/S0378778813002752 (visited on 01/14/2020).
[31] Hastie Trevor, Tibshirani Robert. Generalized additive models. Chapman & Hall/CRC; 1990.
[32] Trimble Inc. SketchUp. 2019, URL: https://www.sketchup.com.
[33] The Big Ladder Software. Euclid. 2017, URL: https://bigladdersoftware.com/projects/euclid/.
[34] Powered by Dark Sky. Dark sky API. 2019, URL: https://darksky.net/poweredby/.
[35] Loga Tobias, Calisti Jens. TABULA webtool. IWU – Institut Wohnen und Umwelt GmbH; 2017, URL: http://webtool.building-typology.eu/#bd.
[36] Loga Tobias, Stein Britta, Diefenbach Nikolaus. TABULA building typologies in 20 European countries—Making energy-related features of residential building stocks comparable. Energy Build 2016;132:4–12, URL: https://linkinghub.elsevier.com/retrieve/pii/S0378778816305837 (visited on 01/22/2020).
[37] Ngo Phil. OpenEEmeter documentation. OpenEEmeter Documentation. URL: http://openee-main.webflow.io/docs/documentation (visited on 10/23/2020).
[38] The Copernicus Programme. Copernicus Atmoshpere Monitoring Service (CAMS). URL: https://atmosphere.copernicus.eu/.
[39] Meteotest. Meteonorm. URL: https://meteonorm.com/.
[40] OpenWeather. OpenWeather. URL: https://openweathermap.org/.