A novel data-analytics based process for load profiling and meter-level anomaly detection in building energy consumption time series

Energy & Buildings 368 (2026) 117829

Contents lists available at ScienceDirect

Energy & Buildings

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

A novel data-analytics based process for load profiling and meter-level anomaly detection in building energy consumption time series

Marco Savino Piscitelli [a,*], Giacomo Buscemi [a], Carlo Roselli [b], Elisa Marrasso [b], Giovanna Pallotta [b], Alfonso Capozzoli [a]

[a] *Politecnico di Torino Department of Energy, TEBE Research Group, BAEDA Lab, Corso duca degli Abruzzi, 24, Turin, 10129, Italy*
[b] *University of Sannio, Department of Engineering, Piazza Roma 28, Benevento, 82100, Italy*

ARTICLE INFO

Keywords:
Load profiling
Anomaly detection
Building energy management
Energy benchmarking
Time series analytics

ABSTRACT

The ongoing digitalization of building operations, combined with growing demands for energy efficiency and sustainability, has intensified the need for advanced methodologies to monitor building energy performance and promptly detect infrequent and unexpected consumption patterns. However, many existing anomaly detection approaches either depend on rigid threshold rules lacking adaptability or employ complex black-box models that offer limited interpretability and require high-resolution monitoring infrastructures, which are often unavailable in many existing buildings.

In response to these challenges, this paper presents a novel data analytics-based methodology for detecting anomalies in building energy consumption time series, addressing the critical need for precise and scalable tools for energy performance monitoring. Tested on a university campus energy consumption time series, the proposed approach explicitly separates and models the shape and magnitude components of daily electrical load profiles, enabling a detailed characterization of typical consumption behaviors. Historical daily electricity data are clustered to identify recurring load shapes, serving as the groundwork for predictive models that estimate both the expected profile shape and total daily energy use based solely on readily available contextual variables, such as weather conditions and calendar information.

By integrating classification and regression techniques with unsupervised learning, the framework generates synthetic daily load profile benchmarks, enabling continuous or recurrent comparisons with actual energy consumption. This supports timely, sub-daily anomaly detection while maintaining low computational cost.

Through validation over a 24-week testing period, the framework successfully identified 17 days exhibiting anomalous consumption patterns, corresponding to a total overconsumption of approximately 1400 kWh. These results confirm the robustness and reliability of the approach in defining statistically informed operational baselines and detecting deviations. The method also provides practical diagnostic insight and supports scalable implementation, particularly in contexts where sensor infrastructure is limited.

1. Introduction

In the context of building energy management, ensuring optimal energy performance is a critical concern for facility managers and stakeholders. Buildings account for nearly 40% of global energy consumption, making their efficient operation essential not only for achieving sustainability goals and reducing greenhouse gas emissions but also for minimizing operational costs and improving system reliability [1,2]. Addressing these challenges requires robust methodologies that enable continuous monitoring, assessment, and improvement of energy consumption [3].

One of the most pressing challenges in building energy management is the ability to track energy performance over time and promptly identify deviations from expected behavior [4–6]. This necessitates the use of tools and methodologies capable of analyzing vast amounts of operational data, uncovering patterns, and isolating anomalies [5,7]. A particularly effective way to address this challenge is through the development of an energy benchmarking process [8–10]. Benchmarking enables comparisons between actual energy consumption and expected values, which are typically derived from historical operational data while considering critical factors such as weather conditions, occupancy levels, and operational schedules [11,12]. In this context, benchmarking stands

* Corresponding author.
E-mail address: marco.piscitelli@polito.it (M.S. Piscitelli).

https://doi.org/10.1016/j.enbuild.2026.117829
Received 2 November 2025; Received in revised form 7 June 2026; Accepted 18 June 2026
Available online 22 June 2026

out as an effective tool, providing a structured framework for anomaly detection at the whole building level [13].

Benchmarking processes focus on establishing a baseline or expected consumption based on historical data. By systematically analyzing past energy consumption patterns, benchmarking determines what is considered "normal" for building energy use under typical operating conditions [14]. This process involves accounting for a variety of influencing factors, such as seasonal variations, occupancy trends, and equipment schedules, to define a robust baseline that represents the expected performance [15]. This reference estimation becomes the starting point for identifying anomalies, as deviations from the expected consumption often indicate inefficiencies, equipment faults, or operational issues [16]. For instance, a sudden change in energy usage trends outside of known influencing factors may suggest a malfunctioning system or suboptimal operational practices. In this sense, the strength of benchmarking lies in its ability to provide a clear and objective measure of normal performance, enabling real-time monitoring and facilitating rapid identification and resolution of energy performance issues [12].

Forecasting, by contrast, focuses on predicting future energy consumption using historical trends, operational data, and external variables [17,18]. This involves the dynamic application of statistical and machine learning models that thanks to autoregressive formulations are able to adapt to changes in building behavior and external conditions [17]. In this perspective, While benchmarking primarily identifies deviations from historical baselines [19], forecasting aims to enhance prediction accuracy, even in scenarios where a building is diverging from expected patterns. This capability makes forecasting models particularly valuable for enable a strategic operation of building energy systems and optimizing their management over time, but at the same time less sensitive to the deviations of building energy consumption from expected patterns [20].

This study aims to establish a scientifically rigorous energy benchmarking framework specifically designed for anomaly detection at the whole building level. It systematically explores methodologies for constructing robust baselines, examines the critical features that explain the occurrence of energy consumption patterns, and drafts the practical applications of benchmarking models in diagnosing inefficiencies and detecting operational anomalies in a consistent way.

2. Literature review

Research on anomaly detection and energy benchmarking in buildings has become increasingly relevant, driven by the demand for enhanced energy efficiency and management of buildings and energy systems during their operation [21]. These efforts address a broad spectrum of issues related to the detection of point, collective, and contextual anomalies in time series [4,22,23]. Point anomalies, which represent singular data points deviating significantly from expected behavior, are often linked to sensor malfunctions or equipment failures [22]. Collective anomalies, in contrast, consider groups of records that, while appearing normal individually, exhibit systemic inconsistency when viewed together [22,23]. Eventually, contextual anomalies are deviations considered abnormal only within specific contexts, such as specific operational schedules or environmental conditions [22,23]. In the context of building energy consumption, anomaly detection processes often prioritize the identification of infrequent or abnormal collective and contextual anomalies on a daily scale [10,12,24]. This focus is due to the quasi-periodicity of energy consumption time series, particularly in commercial, office, and public buildings [25]. These time series typically exhibit regular cycles influenced by predictable factors such as work hours, climate conditions, and occupancy patterns. In this context, anomalies can emerge from disruptions in either the shape of the time series or its magnitude, or both [23]. Magnitude anomalies, can occur when energy consumption levels exceed or fall below expected thresholds, even if the overall shape of the time series matches with expectations. These deviations can be attributed to equipment faults, such as a chiller operating at reduced efficiency, or external factors, such as extreme weather conditions that increase heating or cooling demands. Magnitude anomalies might also indicate energy wastage, such as equipment left running during non-operational hours, or significant underutilization of building resources, such as an HVAC system set to operate at full capacity despite low occupancy. Shape anomalies, in contrast, reflect deviations from the expected pattern or structure of the time series. For instance, in an office building, an unexpected flattening or peaking of the daily energy usage curve might indicate operational issues of the energy systems. These could arise from unexpected scheduling of energy systems or loads. Examples include unplanned overtime activities, maintenance operations conducted outside normal hours, or erroneous system activations, such as lighting systems turning on during unoccupied periods.

To these purposes, unsupervised learning approaches have been widely adopted for identifying patterns and detecting anomalies in building energy consumption data [26]. Lei et al. [27] proposed a dynamic anomaly detection framework combining entropy-weighted clustering and local outlier factors to identify point and collective anomalies. This approach demonstrates the utility of unsupervised learning in extracting patterns from unlabeled data. Similarly, Zangrando et al. [28] compared multiple algorithms for detecting anomalies in quasi-periodic energy consumption data, highlighting the strengths of classical machine learning methods such as Isolation Forest and One-Class SVM in generalizing across different operational contexts. These studies provide valuable insights into pattern recognition but remain focused on retrospective analyses, often lacking predictive capabilities.

Supervised learning techniques have played a significant role in detecting contextual anomalies [29]. Piscitelli et al. [24] developed a regression-based framework integrating decision trees and neural networks to predict daily energy consumption and identify deviations. This method effectively contextualizes anomalies within operational trends, offering precise predictions and identifying deviations that could signal inefficiencies. Lei et al. [30] employed regression combined with residual analysis for anomaly detection in HVAC systems, demonstrating the importance of combining statistical and machine learning techniques to improve fault detection. Despite their accuracy, these supervised approaches heavily depend on annotated datasets, which are challenging to obtain at scale. Song et al. [23] propose an ensemble anomaly detection framework for industrial buildings that combines Local Outlier Factor, Deep Isolation Forest, and Anomaly Transformer to detect subsequence anomalies in power consumption. The model addresses missing data via transformer-based imputation and achieves high accuracy (AUC > 0.96) on real-world cooling energy datasets. To enhance interpretability, Extreme Gradient Boosting links anomalies to key influencing factors such as weekdays and dew point temperatures, supporting data-driven energy management.

Dynamic analysis methods have gained prominence in recent years, emphasizing real-time adaptability. Zhou et al. [31] propose a low-cost anomaly detection method for central air conditioning systems (CACS), framing the problem as a binary classification task without relying on predefined thresholds. By using information entropy as the key feature of daily energy consumption patterns, the method reduces false positives and missed detections typically associated with traditional parameters. The approach also incorporates online updates of normal behavior profiles based on three key influencing factors, improving detection accuracy compared to conventional methods such as standard K-means clustering. Lei et al. [30] further developed a hybrid framework that integrates supervised and unsupervised methods to dynamically identify anomalies, enhancing performance through particle swarm optimization. Tested on real building data, the approach achieved over 80% clustering accuracy, demonstrating its potential for improving building management through more accurate anomaly detection.

Pattern recognition techniques have also been extensively employed to understand typical energy consumption behaviors, serving as a foundation for anomaly detection [32,33]. Capozzoli et al. [10] used symbolic aggregate approximation (SAX) to identify recurring energy patterns, integrating these with decision tree models for anomaly detection. This approach was particularly effective in identifying historical patterns and deviations, though it lacks the ability to predict future behaviors or contextualize anomalies dynamically. Mascali et al. [34] present a semi-supervised anomaly detection framework that combines Symbolic Aggregate approXimation (SAX) encoding with two complementary machine learning models–a Classification and Regression Tree (CART) for interpretability and a Multi-Layer Perceptron (MLP) for accuracy. Aimed at addressing the limitations of purely supervised approaches and the interpretability challenges of many ML techniques, the framework was tested on real-world electricity consumption data and demonstrated improved performance in detecting and contextualizing anomalies. Chiosa et al. [9] present a self-tuning anomaly detection and diagnosis process at the meter level, designed to identify abnormal energy performance patterns and support decision-making in building operations. The methodology integrates supervised and unsupervised techniques with the contextual matrix profile (CMP) algorithm to detect infrequent subsequences in energy consumption time series under specific boundary conditions. Tested on a year-long dataset from a university campus, the process successfully detected and diagnosed 55 anomalies by analyzing sub-load-level information, offering a scalable and context-aware approach to energy anomaly management.

Deep learning methodologies have also emerged as powerful tools for modeling complex relationships in energy consumption data. Azzalini et al. [35] explored the use of deep autoencoders with attention mechanisms for anomaly detection, focusing on addressing the challenges of noisy datasets and the bypassing effect in attention-based models. Similarly Fan et al. [36] investigated the use of autoencoders for unsupervised anomaly detection in building energy data, addressing the challenges of unavailable anomaly labels in real-world operational datasets. They proposed an ensemble autoencoder-based method and provided a comprehensive comparison of different architectures and training schemes. Noorchenarboo et al. [37] extended these efforts by leveraging Shapley Additive Explanations (SHAP) to enhance the interpretability of anomaly detection results, allowing for the identification of contextually relevant features that contribute to energy deviations. These studies highlight the potential of deep learning models in improving detection accuracy but reveal gaps in their application to scalable, real-time frameworks.

Hybrid approaches have shown significant promise in addressing the limitations of individual methodologies. Shi et al. [38] introduced a federated learning model for anomaly detection, employing sparse attentive aggregation to address data privacy and heterogeneity across distributed datasets. This framework demonstrates the scalability of hybrid approaches in complex environments. Similarly, Lei et al. [27] proposed a dynamic anomaly detection framework combining entropy-weighted k-means (EWKM), classification and regression trees (CART), and ensemble outlier detection methods (Isolated Forest and Local Outlier Factor). The method targets both point and collective anomalies in energy consumption data. By integrating dynamic time warping for sequence comparison and enabling continuous dataset updates, the EWKM-CART-IF-LOF approach demonstrated superior detection performance (AUC = 0.90, accuracy = 0.98) in real-world applications, significantly enhancing energy management capabilities in buildings. In this perspective, authors in [39] presented an Asymmetric Hybrid Encoder-Decoder (AHED) model for detecting point and collective anomalies in building energy consumption. By integrating supervised and unsupervised learning with advanced encoding-decoding and cross-correlation techniques, the model transforms temporal multivariate data into feature matrices for precise anomaly detection. Canale et al. [40] propose a model for detecting abnormal thermal energy consumption in residential buildings with limited metering infrastructure. Using data from heat cost allocators, indoor temperature sensors, and outdoor climate conditions, the model applies a hybrid method combining interquartile range-based outlier detection, regression analysis, and rule-based diagnosis. It benchmarks consumption and temperature at dwelling, floor, and building levels to identify anomalies and assign fault labels. Validated on four heating seasons in a case study building, the model successfully identified behavioral faults, underheating, overheating, and envelope inefficiencies, supporting both diagnosis and user awareness.

Other studies have emphasized the importance of normalization and scaling in energy data analysis.

Kong et al. [41] extended this work by incorporating spatio-temporal features using graph convolutional networks, enhancing the detection of anomalies in electricity consumption data while addressing data heterogeneity. These approaches highlight the utility of normalization and pattern recognition but reveal gaps in integrating these elements with predictive models. Similarly, Fan et al. [42] proposed a graph-based methodology for analyzing building operational data, addressing challenges in representing complex, multi-relational datasets and visualizing analysis results. By converting diverse operational data into graph structures, the method enables effective knowledge discovery, including the detection of atypical operations and high-level behavior patterns. Copiaco et al. [43] propose a deep anomaly detection method for building energy data that transforms 1D time series into 2D image representations and leverages convolutional neural networks (CNNs) through transfer learning. Features extracted from pre-trained models (AlexNet and GoogleNet) are used with an SVM classifier, achieving F1-scores up to 99.89% and demonstrating both high accuracy and computational efficiency suitable for edge applications.

While recent advances have significantly improved the accuracy and applicability of data-driven anomaly detection in building energy systems, many methods remain limited by their reliance on either rigid threshold rules or black-box models with limited interpretability. In addition, most approaches focus on detecting anomalies based solely on energy magnitude, often overlooking the operational significance of temporal consumption patterns. The distinction between shape and magnitude as complementary dimensions of abnormal behaviour has been acknowledged in the anomaly detection literature [24], yet existing methods rarely treat them as independently modelled and jointly informative components. Approaches rooted in load disaggregation and pattern recognition [44] have demonstrated the value of characterising typical consumption behaviours, but primarily target retrospective analysis rather than predictive benchmarking. Similarly, baseline construction methods widely adopted in measurement and verification frameworks [12] capture deviations in overall consumption levels without explicitly accounting for the temporal structure of daily load profiles. Hybrid approaches have improved contextualisation of anomalies [9,10], yet they typically entangle shape and magnitude within a single analytics step.

In response to these gaps, this study proposes a multi step methodology that characterizes both the total daily energy consumption and the shape of the daily load profile of a building. The analysis is conducted on a real case study involving a university campus in Italy. The method begins with the normalization of historical daily electricity profiles, followed by the application of clustering algorithm to identify a discrete set of typical load shapes. These clusters capture recurring patterns from historical building operation data and provide the reference basis for detecting shape anomalies.

Once the shape classes are defined, a classification model is trained and compared to predict the most probable load shape class for each day using autogenous input variables such as outdoor air temperature, calendar variables, and other contextual features. In parallel, a regression model estimates the expected total daily energy consumption exploiting the same feature set of the shape classifier. These two predictive models function independently but are combined to reconstruct an expected hourly profile that integrates both the predicted shape and magnitude.

The resulting profile acts as a synthetic benchmark against which the actual daily consumption is evaluated. Distance metrics are employed to quantify deviations in both shape and magnitude, allowing the identification and ranking of anomalies. This analytics framework supports

Fig. 1. University of Sannio.

daily execution of the benchmarking pipeline while enabling sub-daily monitoring of the building performance for operational diagnostics. The novelty of the approach is detailed in the following section.

2.1. Novelty and contribution of the paper

This work introduces a novel methodology for anomaly detection in building energy consumption time series, addressing several critical limitations found in current research. A central innovation is the explicit separation of shape and magnitude analysis in daily load profiles. This allows the method to capture not only deviations in total energy use but also atypical temporal distributions that may indicate operational anomalies, even when daily totals appear consistent.

The approach incorporates contextual variables such as calendar type, weather conditions, and occupancy schedules, enhancing the interpretability of anomalous patterns and reducing false positives by framing energy deviations within their operational context.

Designed for practical deployment, the methodology operates effectively with aggregated or low-resolution data. Its modular structure supports application across diverse building types and usage profiles, ensuring scalability and robustness in real-world scenarios.

Differently from methods that focused solely on retrospective analysis, the proposed framework provides a predictive benchmark for the day ahead by estimating both the expected energy magnitude and load shape. This benchmark enables subdaily anomaly detection through comparison with real-time observations, without requiring continuous baseline computation. As a result, the method maintains low computational demand while supporting timely operational diagnostics.

Finally, the framework supports anomaly ranking by distinguishing between days with both unexpected consumption and shape, and those where only one of the two dimensions deviates from expectations. This differentiation improves the precision of diagnostics and helps prioritize responses based on the nature of the anomaly. The rest of the paper is organized as follows: Section 3 provides an overview of the considered case study. In Sections 4 and 5, the methods and overall methodology employed in this study are detailed. Section 6 presents the results for the analyzed case studies. Eventually, Sections 7 and 8 offer a comprehensive discussion of the findings and outline the next steps in this field of research.

3. Description of the case study

The analyzed case study is the historical building of the University of Sannio, located in Benevento, Southern Italy (Fig. 1). The building accommodates administrative offices, academic staff offices and university classrooms. It falls within Italian climate zone C, with 1316 heating degree days (HDD), as defined by national legislation. The structure has a rectangular layout and comprises three floors, with a conditioned floor area of 1,008.39 m^2 and a gross heated and cooled volume of 5,940.46 m^3. The surface-to-volume ratio is 0.46 m^{-1}.

According to Italian regulations, the standard heating season for Benevento extends from November 15th to March 31st, allowing for a maximum of 10 h of heating per day. However, a ministerial decree issued in October 2022 introduced energy-saving measures that reduced the heating season to the period between November 22nd and March 23rd, and limited daily heating operation to a maximum of 9 h.

The cooling period typically spans from June 1st to September 30th, during which there are no regulatory constraints on the activation schedule or duration of space cooling systems.

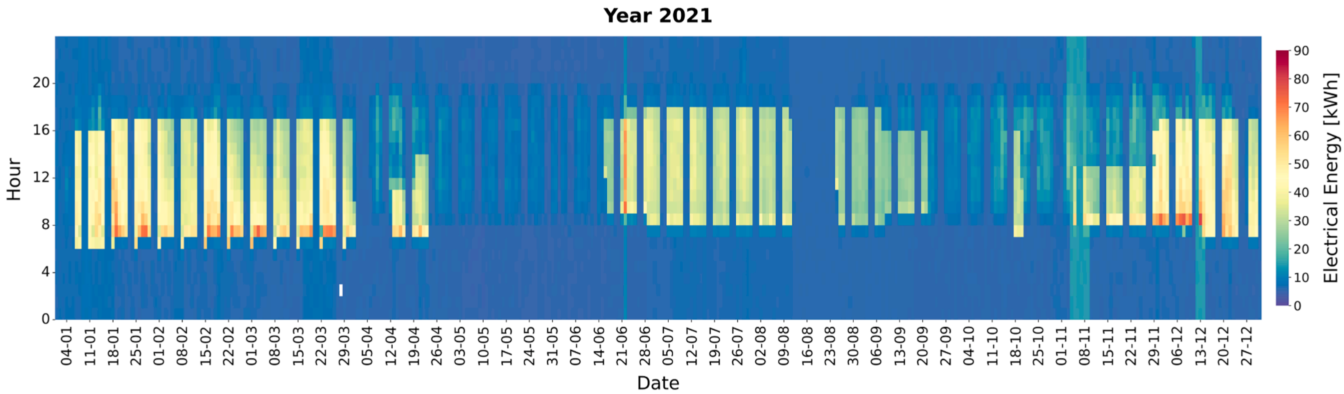
Space heating and cooling are provided by a rooftop air-to-water electric-driven heat pump (EHP) using R410A refrigerant. The nominal performance characteristics of the unit are an Energy Efficiency Ratio (EER) of 2.14 and a Coefficient of Performance (COP) of 3.10. The HVAC system also includes a 750-liter thermal and cooling energy storage unit and three circulating pumps, each serving a separate branch of the distribution circuit. The pumps operate in an on-off mode without inverters and run at nominal power during scheduled HVAC operation. The combined auxiliary electric load of the pumps is constant at 6.96 kW, as determined by current clamp meter measurements conducted over two weeks in both the cooling and heating seasons. The building is occupied on weekdays from 7:45 AM to 8:00 PM. Holidays and scheduled closures are explicitly included in the analysis. Artificial lighting is controlled by occupancy sensors and daylight dimming systems, ensuring no lighting load during unoccupied periods. The HVAC system operates according to a predefined schedule that specifies the heating and cooling seasons and the maximum daily operating hours; this information is incorporated into the analysis as a fixed boundary condition. For the sake of completeness, Fig. 2 report the heatmaps of hourly energy consumption for 2021 and 2022, which were used to develop and test the proposed anomaly detection process, and show the temporal patterns and variability that characterize the analysed dataset.

4. Materials and methods

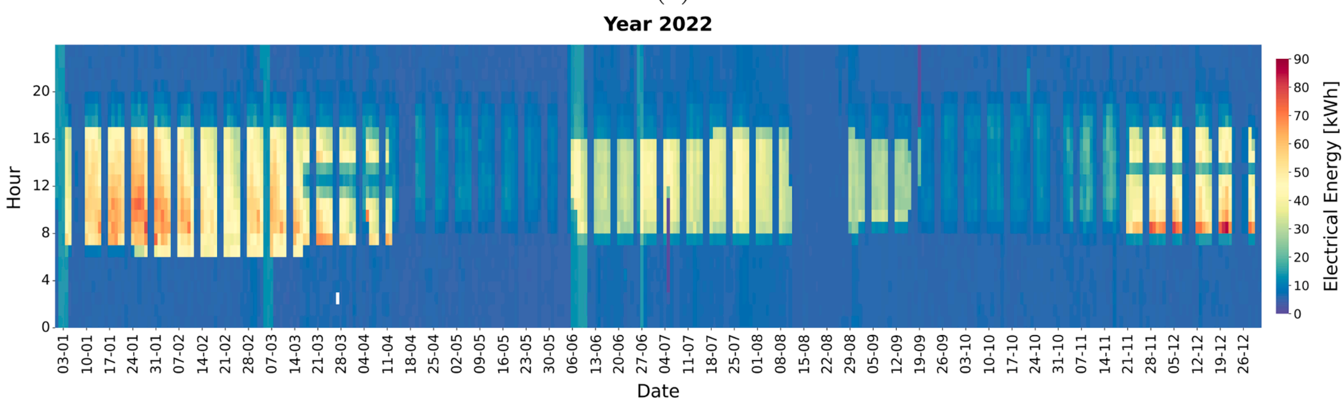
This section provides a concise overview of the data analytics methods underlying the proposed anomaly detection process. Rather than offering an exhaustive description, it highlights the key aspects of the algorithms relevant to this study objectives. Specifically, it presents the fundamental concepts of the supervised and unsupervised algorithms, which were employed to establish the normal operation baseline of the considered building.

4.1. Cluster analysis

Hierarchical clustering is a widely used technique that builds a tree of clusters (dendrogram) without requiring a predefined number of clusters. It can be performed using either an agglomerative (bottom-up)



(a)



(b)

Fig. 2. Heatmaps of the hourly electrical energy consumption for 2021 (a) and 2022 (b).

or divisive (top-down) approach. The agglomerative method starts by treating each data point as its own cluster and successively merges the closest pairs based on a distance matrix, continuing until a single cluster remains or the desired level is reached. The divisive approach, by contrast, begins with all data points in one cluster and recursively splits them into smaller groups. While more computationally demanding, it provides a top-down perspective of the data structure. The final number of clusters can be determined either by setting a threshold or by inspecting the dendrogram. A key advantage of hierarchical clustering is that it reveals the nested structure of the data and does not require prior specification of cluster count. However, it may become computationally expensive for large datasets due to repeated distance calculations.

4.2. Classification models

The following models have been employed in this study to perform the classification task. A brief overview of their main theoretical foundations is provided below.

- Classification and Regression Trees (CART) is a non-parametric decision tree algorithm used for classification and regression. For classification tasks, it builds a binary tree by recursively splitting the input space to increase class homogeneity. The Gini index is used as the splitting criterion, quantifying impurity at each node. The process continues until a stopping condition is met, producing interpretable and efficient decision rules.
- Random Forest is an ensemble method that constructs multiple decision trees from bootstrap samples of the dataset. At each node, a random subset of features is considered, which increases model diversity and reduces overfitting. Predictions are made by averaging outputs in regression or majority voting in classification. Random Forest is robust to noise and effective for high-dimensional, non-linear problems.
- Boosting Trees are sequential ensemble methods that combine weak learners, typically shallow decision trees, to form a strong predictor. Each new tree is trained to reduce the residual errors of the current model. Boosting emphasizes poorly predicted observations and aggregates tree outputs to improve accuracy. It is especially useful in capturing complex, non-linear patterns.
- K-Nearest Neighbors (KNN) is a non-parametric, instance-based algorithm. For classification, it assigns a class based on the majority among the k closest neighbors, typically using Euclidean distance. KNN does not require training and stores the entire dataset, making prediction computationally intensive for large samples. The choice of k significantly influences model behavior.
- Support Vector Machines (SVM) are supervised models that aim to find the optimal hyperplane separating data classes. If linear separation is not possible, kernel functions are used to project the data into a higher-dimensional space. In this study, a polynomial kernel is applied to capture non-linear relationships through polynomial transformations of the input features.

4.3. Regression models

The following models have been employed in this study to perform the regression task. A brief overview of their main theoretical foundations is provided below.

- GAMs Generalized Additive Models (GAMs) are a flexible class of regression models that allow each predictor to influence the response through a smooth, potentially non-linear function. Unlike traditional

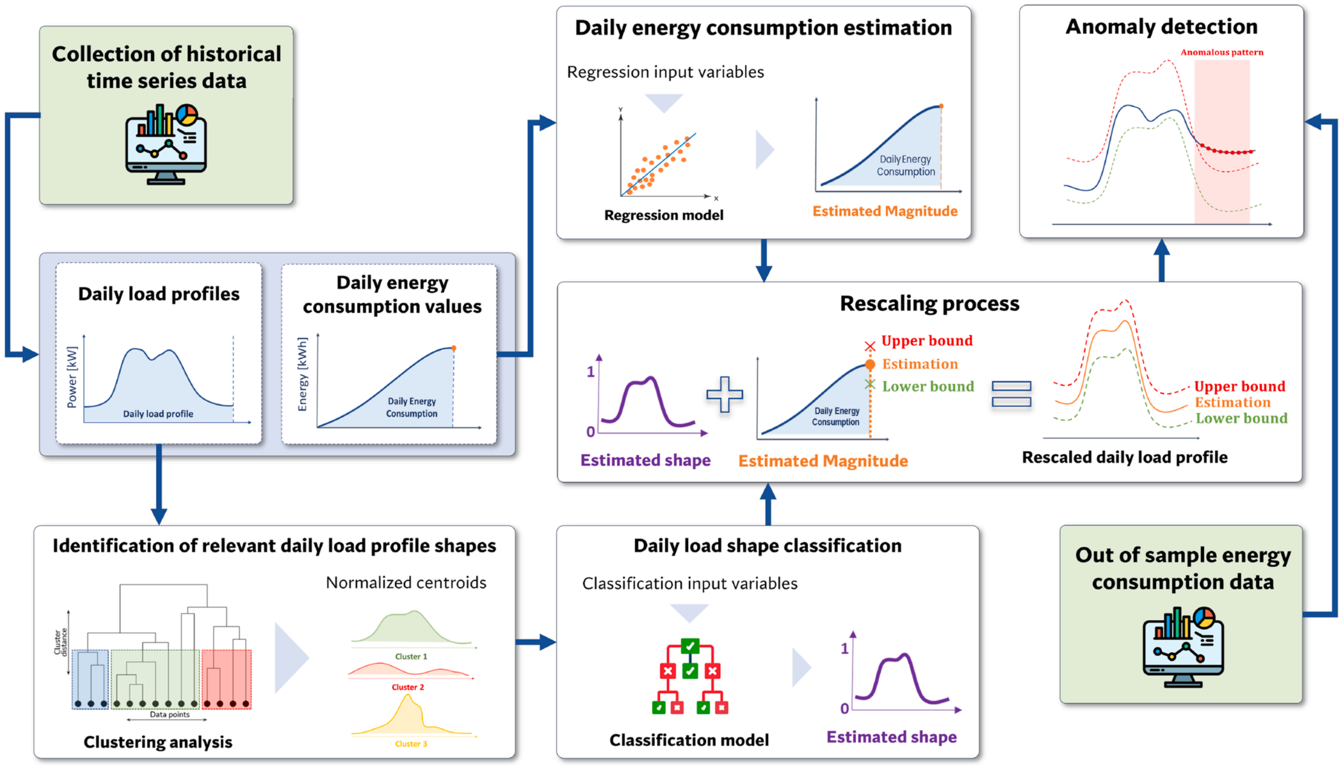


Fig. 3. Methodological framework.

linear models, which assume a fixed linear relationship between inputs and output, GAMs model the contribution of each predictor separately, using data-driven smoothing techniques such as splines. This structure provides a clear interpretation of how each variable affects the outcome while retaining the ability to capture complex trends in the data.

- CART (Classification and Regression Trees) in the case of regression, the model recursively splits the data to minimize prediction error within each resulting node. The splitting criterion is based on minimizing the mean squared error, which quantifies the variance of the target variable within each partition.
- Boosting: In the context of regression, boosting follows the same iterative framework used in classification but is adapted to minimize a continuous loss function, typically the mean squared error. At each iteration, a new weak learner is trained to predict the residuals–the differences between the actual and currently predicted values. These residuals represent the unexplained portion of the data, and the learner's goal is to reduce them progressively.
- CatBoost is a gradient boosting algorithm specifically designed to handle categorical variables efficiently and improve model accuracy and robustness. In regression tasks, CatBoost builds an ensemble of decision trees sequentially, where each tree is trained to reduce the residual error of the combined ensemble at that point. A distinctive feature of CatBoost is its native support for categorical features, which are transformed internally using techniques such as target statistics and permutation-driven encoding, eliminating the need for extensive preprocessing.
- Bayesian Additive Regression Trees (BART) is a non-parametric Bayesian approach to regression that models the response variable as the sum of predictions from an ensemble of shallow regression trees. Unlike traditional boosting methods, BART places a prior distribution over the ensemble and uses Bayesian backfitting and Markov Chain Monte Carlo (MCMC) techniques to estimate the posterior distribution of the function being modeled. Each tree in the ensemble contributes a small portion to the overall prediction, and the model as a whole captures complex non-linear relationships with controlled flexibility.

5. Methodology

The methodology implemented in this work aims to define a novel anomaly detection strategy for the implementation energy consumption time series in buildings. The framework unfolds over multiple steps as shown in Fig. 3 and are in the following explained.

5.1. Data preparation

The process begins with the preparation of the energy consumption time series. Due to quality constraints the dataset was limited to the years 2021 and 2022. Raw electrical consumption data were initially acquired at a finer resolution (15-min) and then aggregated to an hourly scale to align with the intended level of analysis. Preprocessing also involved the identification and treatment of missing or anomalous values. Short gaps were filled using linear interpolation to preserve continuity.

5.2. Identification of relevant daily load profile shapes

After data preparation, the energy consumption time series was segmented into daily profiles, each consisting of 24 hourly values. To focus the analysis on consumption patterns rather than absolute energy use, each daily load profile was normalized by its own maximum value. This normalization allowed for the comparison of profile shapes across days with different consumption magnitudes. Clustering analysis was then applied to group daily profiles exhibiting similar shapes, enabling the identification of typical operational patterns. An agglomerative hierarchical clustering algorithm was employed, using Euclidean distance to measure dissimilarity and WARD linkage to ensure compact and well-separated clusters by minimizing within-cluster variance. The optimal number of clusters was determined by evaluating the Silhouette index in the exploration range between 3 and 15 clusters. Once the clustering process was completed, the assigned cluster labels were used as categorical output variables for the classification model. These labels represent typical daily load shapes and serve as target classes to be predicted based on contextual features. For each cluster, the centroid–representing the average normalized daily profile–was also computed and stored. These centroids act as reference profiles against which new or observed daily load patterns can be compared, enabling consistent classification and supporting subsequent anomaly detection based on deviations from expected behavior.

5.3. Daily load shape classification

In this phase, a classification model was developed to estimate the daily electrical load shape of the building based on the cluster labels obtained from the previous clustering step. These labels, representing distinct daily consumption patterns, were used as the categorical output variable in the classification task. To train and evaluate the models, the dataset was split as follows: The fourth week of each month in 2021 and 2022 was included in the test set to evaluate model performance, whereas the remaining weeks were assigned to the training and validation sets to enable model development and tuning. The classification model aims to predict the most likely load shape for a given day using only a limited set of easy-to-collect contextual input variables, including calendar variables, outdoor air temperature, and other operational features. Five classification algorithms were tested and compared: Classification and Regression Trees (CART), Random Forest, Boosting Trees, K-Nearest Neighbors (KNN), Support Vector Machines (SVM) with a polynomial kernel. Model parameters were fine-tuned and the resulting performance metrics were presented and discussed in the Results section in order to select the best performing model. This classification step plays a key role in the overall framework by enabling the prospective estimation of daily load shapes from easily accessible input variables. The predicted class can then be associated with its reference centroid profile derived from the clustering phase.

5.4. Daily energy consumption estimation

In this step, regression models were developed to estimate the total daily energy consumption of the building. For each day in the dataset, the actual energy use was calculated by aggregating the hourly consumption values. The regression task aimed to predict this total based on a limited set of input variables–the same contextual features used in the classification phase. In addition, the predicted cluster label, representing the expected daily load shape, was included as a categorical input to the regression model. This integration allowed the model to account for the influence of typical consumption patterns on total energy use, enhancing its ability to capture shape-dependent variations in daily demand.

To ensure consistency with the classification model development, the regression models were trained and tested on the same days across 2021 and 2022 as those employed in the classification phase.

Five regression algorithms were tested and compared: GAM, CART, Boosting Trees, Catboost, BART. Model parameters were fine-tuned and the resulting performance metrics were presented and discussed in the Results section in order to select the best performing model.

The estimated daily energy consumption was used in the subsequent step to rescale the normalized load shape predicted by the classification model. This rescaling enabled the reconstruction of the expected hourly energy profile for each day, combining predicted shape and magnitude into a complete synthetic benchmark suitable for anomaly detection.

5.5. Rescaling process

The following step of the methodology, integrates the outcomes of the classification and regression processes to reconstruct a complete expected daily energy profile. The classification model assigns each incoming day to the most probable cluster label based on contextual features such as weather and calendar information. Each cluster, previously defined through hierarchical clustering, is associated with a centroid representing the normalized hourly shape of typical energy consumption for that group.

The centroid corresponding to the predicted label is retrieved and used as a reference normalized load profile. To restore the actual magnitude of consumption, the regression model provides an estimate of the total daily electrical energy use. This predicted value is used to scale the normalized profile by applying a rescaling factor, referred to as k. The rescaling factor K is computed as the ratio between the predicted daily energy consumption $\hat{E}$day and the sum of the 24 hourly values of the normalized centroid profile c_h, as shown in Eq. (1):

$$K = \frac{\hat{E}\mathrm{day}}{\sum_{h=1}^{24} c_h} \tag{1}$$

Where $\hat{E}$day is the predicted daily energy consumption provided by the regression model, and c_h is the normalized centroid value at hour h. Each hourly value of the centroid is multiplied by the scaling factor K, producing a rescaled profile that preserves the shape of the typical pattern while aligning it with the expected magnitude of daily energy consumption [45–47]. To incorporate prediction uncertainty, the standard error of the regression model is used to define an acceptability range around this profile. Specifically, two additional scaling factors are computed by replacing the numerator in Eq. (1) with the predicted daily energy consumption increased or decreased by the standard error of the regression model. These factors define the lower and upper bounds of a confidence band around the rescaled profile. The resulting profile, together with its uncertainty bounds, serves as a synthetic benchmark for sub-daily anomaly detection. This enables the identification of deviations not only from the expected trend but also from the predicted range of normal variability, thereby improving diagnostic robustness and supporting more informed operational decisions.

5.6. Anomaly detection

The final step of the methodology focuses on detecting anomalies in the building energy consumption time series at a subdaily scale. This is achieved by comparing the observed hourly load profile of each day against the expected profile reconstructed through the combined classification and regression models, as described in the previous sections.

The reconstructed profile serves as a dynamic benchmark that reflects both the anticipated shape and magnitude of energy use, conditioned on contextual variables. To enhance robustness, an acceptability band is established around the predicted profile using the standard error of the regression model. This band defines upper and lower bounds within which deviations are considered tolerable, accounting for prediction uncertainty.

An anomaly is flagged when the actual measured hourly load falls outside this acceptable range for a significant portion of the day, either due to a deviation in magnitude, shape, or both. This approach allows for the identification of a wide range of operational anomalies, including unexpected peak loads, off-schedule system activations, or periods of underconsumption. The anomaly detection procedure was evaluated on the days belonging to the test set, which was not used during the training of the classification and regression phase. This ensures an unbiased evaluation of the framework performance under real-world conditions.

6. Results

The methodological framework described in Section 5 was tested on the case studies illustrated in Section 3. The considered dataset covered two years 2021–2022 of monitored energy consumption data comprising 730 daily load profiles aggregated at hourly scale. Data corresponding to the fourth week of each month in 2021 and 2022 were included in

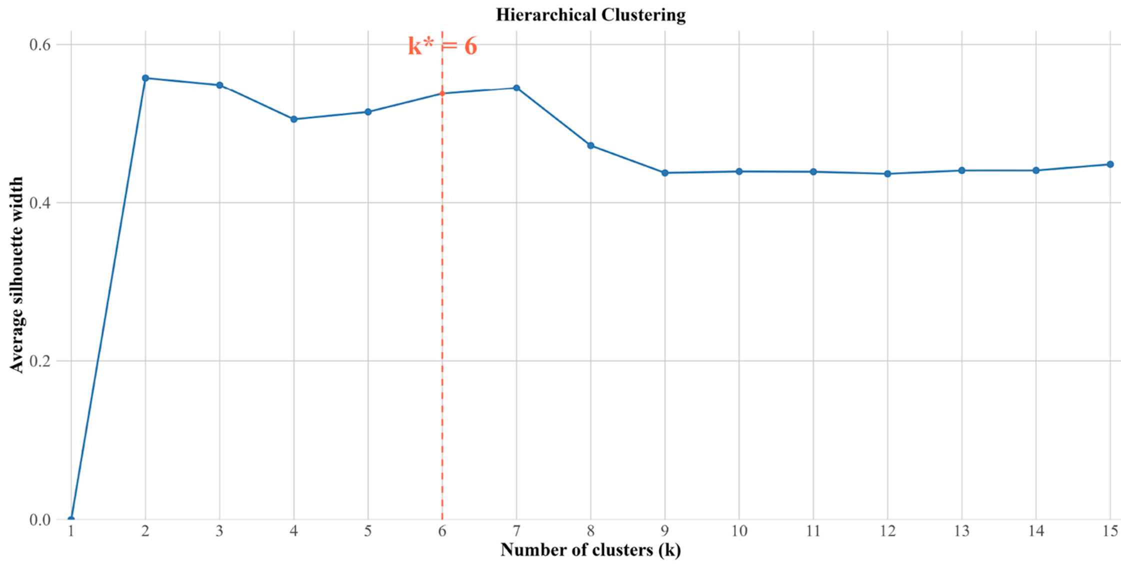


Fig. 4. Average Silhouette value by number of clusters.

the test set to evaluate model performance (i.e., 168 days) while the remaining weeks (i.e., 562 days) were included in the training and validation sets, maintaining a training-to-test ratio of approximately 75-25%. The obtained results were presented and discussed below.

6.1. Clustering results

The clustering analysis focused on hourly daily electrical load profiles of 2021 and 2022. All the daily load profiles were normalized on their own daily max to better highlight shape similarities. The optimal number of clusters within the range of 3 to 15 was explored, employing the Euclidean distance as a measure of similarity between objects. Quality metrics such as the Silhouette index and elbow method were used to determine the final optimal number of clusters. Fig. 4 presents the variation of the Silhouette index as a function of the number of clusters. This visualization supports the identification of the optimal number of clusters by highlighting configurations that maximize intra-cluster similarity, with six and seven clusters emerging as the most promising options.

While the configuration with seven clusters achieved the highest Silhouette index, the six-cluster solution demonstrated a very similar performance in terms of cohesion and separation among groups, while maintaining a lower level of complexity. This distinction is particularly relevant given the downstream use of cluster labels for training a classification model. In such cases, the number of classes directly influences the model's complexity, generalization ability, and overall predictive performance. A higher number of classes can increase the risk of misclassification, especially in the presence of overlapping clusters or limited training data. Therefore, selecting six clusters offered a more favorable balance between representational granularity and classification accuracy. Based on this rationale, the six-cluster configuration was selected as the optimal solution.

As a result, in Fig. 5, the corresponding clusters of normalised daily load profiles are reported.

The figure displays the typical daily electrical load profiles grouped into six clusters, each representing a distinct consumption pattern. The grey lines correspond to individual daily profiles, while the colored lines indicate the cluster centroids, with shaded areas showing the standard deviation around each centroid.

Cluster 1 shows the most stable behavior, characterized by a nearly constant load throughout the day, suggesting baseline or continuous usage. Clusters 2, 3, and 5 capture typical working-day profiles with variations in timing and intensity. Together, these clusters represent the dominant operational patterns observed on active weekdays.

On the other hand, clusters 4 and 6 exhibit more complex and irregular dynamics. with a steep early-morning peak followed by a progressive drop, possibly corresponding to partial operations of the building or of its energy systems.

Fig. 6 shows the clustering results for both 2021 and 2022 using calendar plots. Fig. 6a shows the daily cluster assignments for 2021, with each day color-coded according to its associated cluster, allowing for a clear visual interpretation of temporal patterns and seasonal dynamics in the daily load profiles. Fig. 6b displays the corresponding results for 2022, following the same representation scheme. This comparative visualization facilitates the identification of recurrent behaviors, potential anomalies, and year-to-year variations in the shapes of daily load profiles.

6.2. Classification results

After clustering and labeling the daily load profiles for 2021 and 2022, the cluster labels corresponding to the days in the fourth week of each month were designated as the test set, while those from the remaining days served as the target for training the classification models. The input variables employed in the classification task include key meteorological and temporal features known to influence building energy consumption. Specifically, the temperature-related features were computed on a daily basis and include the average, maximum, minimum, and standard deviation of outdoor air temperature, all derived from hourly measurements. Similarly, for solar radiation, the average, maximum, and standard deviation of the hourly global horizontal irradiance were calculated. Temporal features include the day of the week and the month of the year, which contribute to capture behavioral and seasonal variability in energy consumption. Additionally, a Boolean variable indicating whether a day was a public holiday was included to account for the absence of occupants during non-working days. The complete set of input and output variables, along with their respective types and units of measurement is reported in Table 1.

The classification task was designed to predict, for each day, the most probable cluster label given the aforementioned boundary conditions. This classification allowed the extraction of the corresponding normalized centroid, which was used as the reference shape for the expected daily load profile. The target variable included six classes (equal to the number of identified clusters). Five classification models were trained

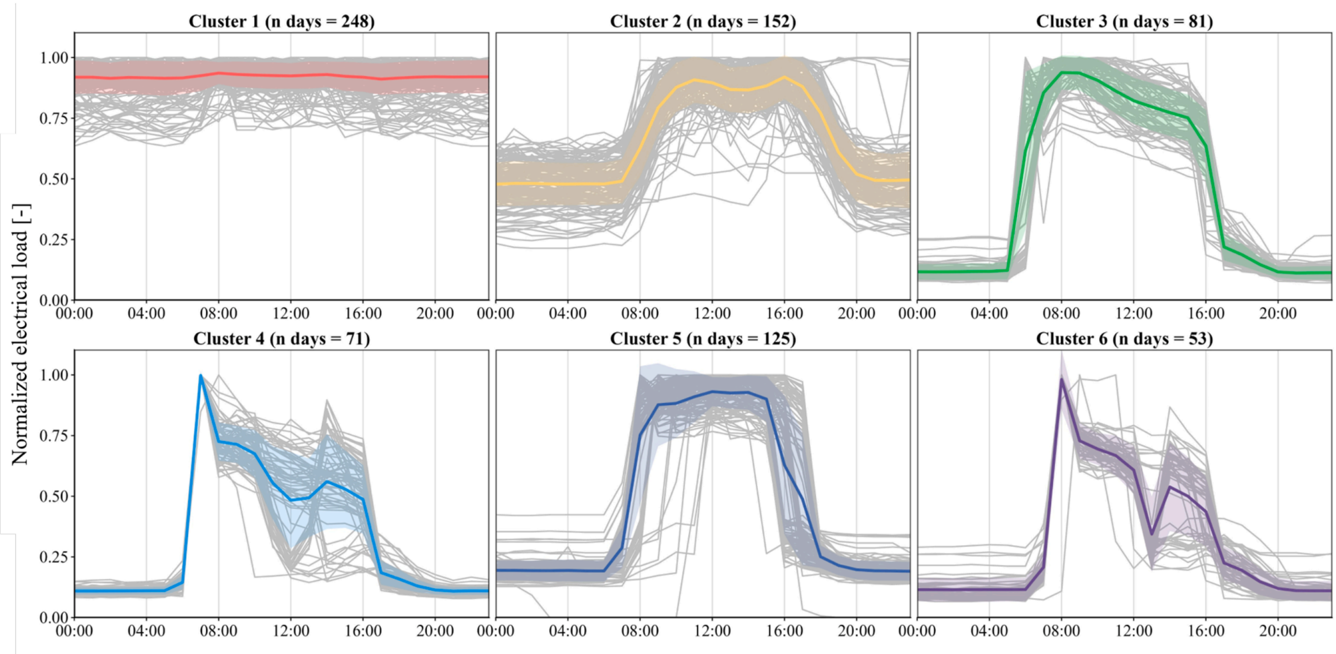


Fig. 5. Clusters of normalised daily load profiles with evidence of their own centroid.

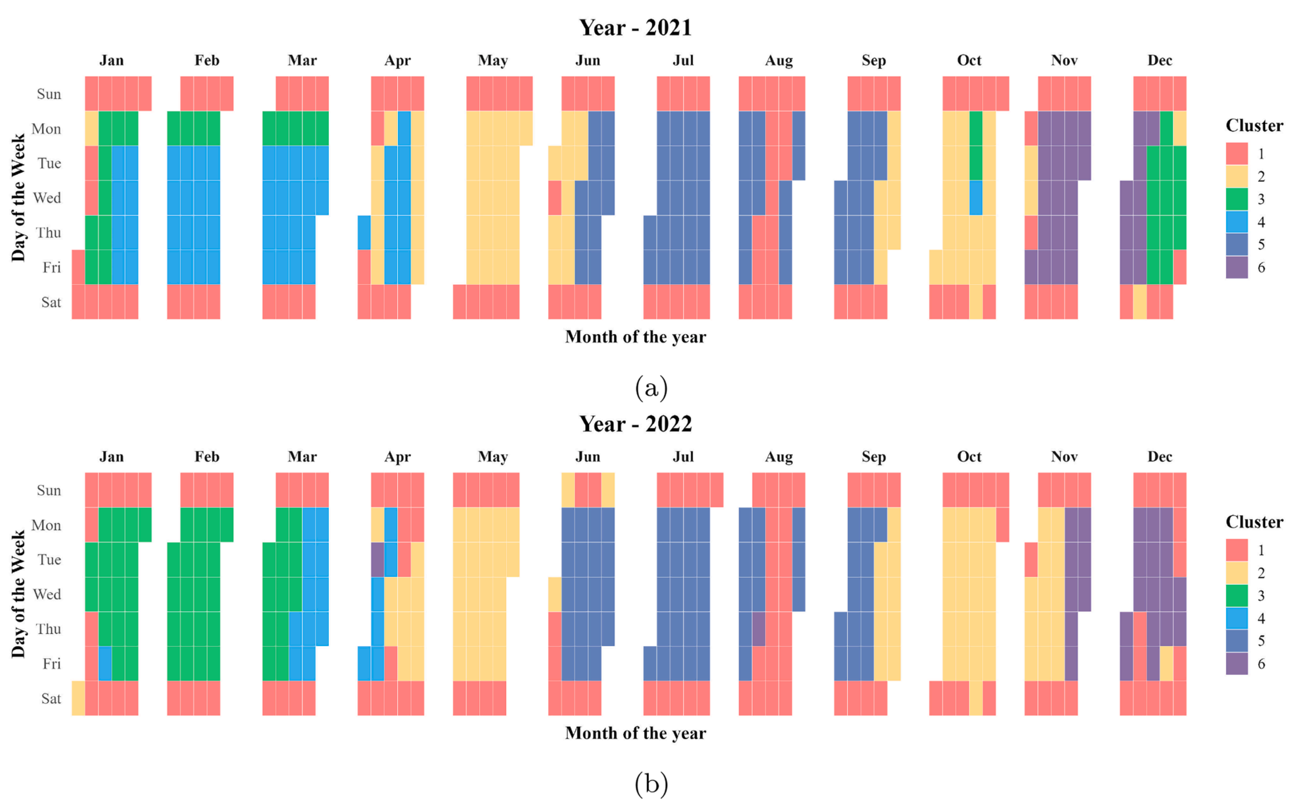


Fig. 6. Daily clustering results for the years 2021 (a) and 2022 (b).

and their performance was evaluated on the defined out-of-sample testing dataset. All models (with the exception of the k Nearest Neighbors model) were trained and validated using k-fold cross validation (with K = 10). The performance of each model on the testing set was assessed in terms of average precision, average recall and accuracy. Table 2 presents a summary of these performance metrics for each developed classification model.

The performance metrics indicate that the Boosting tree model achieved the highest accuracy (85.70%) among all tested algorithms, together with superior precision (83.11%) and recall (81.69%). These results reflect a balanced and robust classification capability, supporting the choice of Boosting as the preferred model. In contrast, k-NN showed the lowest performance across all metrics, whereas the SVM Polynomial model yielded results comparable to Boosting but with a slightly lower

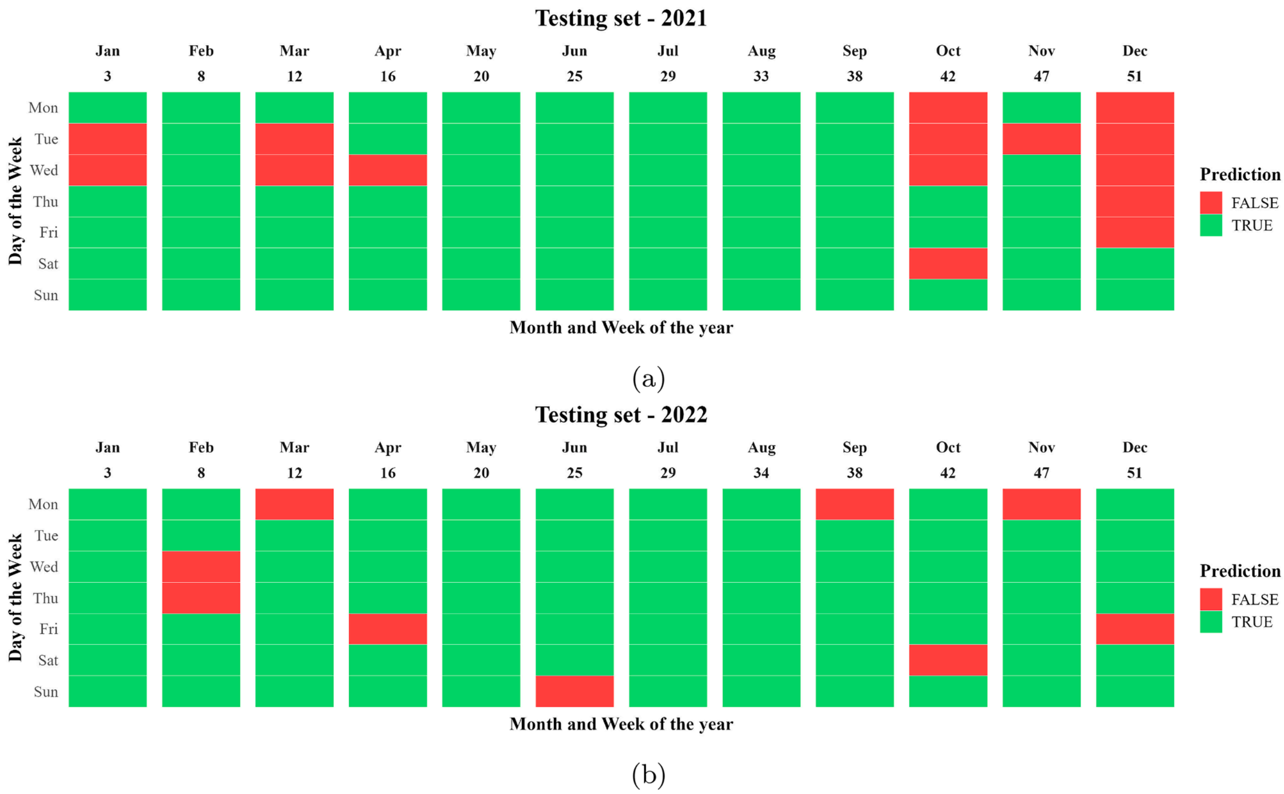


Fig. 7. Calendar plot of misclassification errors for the test set 2021 (a) and 2022 (b).

Table 1
Set of input and output variables of the classification models, along with their respective types and units of measurement.

Variable	Variable Type	Unit
Input Variables		
Daily average outdoor air temperature	Numeric	[°C]
Daily maximum outdoor air temperature	Numeric	[°C]
Daily minimum outdoor air temperature	Numeric	[°C]
Daily standard deviation of outdoor air temperature	Numeric	[°C]
Daily average global horizontal irradiance	Numeric	[W/m^2]
Daily maximum global horizontal irradiance	Numeric	[W/m^2]
Daily standard deviation of global horizontal irradiance	Numeric	[W/m^2]
Day of week	Categorical	[–]
Month of the year	Categorical	[–]
Holiday	Boolean	[–]
Target Variable		
Cluster label	Categorical	[–]

Table 2
Performance metrics assessed for the developed classification models.

Model	Precision	Recall	Accuracy
CART	81.25%	79.98%	83.99%
Random Forest	78.79%	78.23%	82.74%
Boosting	83.11%	81.69%	85.70%
k-NN	74.66%	70.49%	76.26%
SVM Polynomial	82.30%	81.65%	85.14%

accuracy of 85.14%. Both CART and Random Forest displayed competitive, but slightly lower, performance levels, further suggesting the selection of the Boosting tree as the classifier to be used for the following steps of analysis.

Fig. 7 illustrates the daily prediction accuracy of cluster assignments across the testing set. Each cell in the calendar plot represents a day in the fourth week of each month of 2021 and 2022, color-coded to indicate whether the predicted cluster label matched the actual label. Green cells correspond to correct predictions (TRUE), while red cells indicate mismatches (FALSE). From the Fig. 7 it is possible to infer that misclassifications show no discernible pattern, indicating that model errors are likely due to data noise or class overlap rather than systematic bias of the classifier.

6.3. Regression results

In this step, five regression models were developed to estimate the building's total daily energy consumption, using data from 2021 and 2022 organized into training and testing sets, following the same partitioning used in the classification phase. The goal of the regression task was to estimate the daily electrical energy consumption using a limited set of contextual input features i.e., the same variables used in the classification phase. To further improve predictive accuracy, the cluster label was included as an additional categorical input. This integration enabled the regression models to account for the influence of typical consumption patterns, thereby enhancing their ability to capture shape-dependent variations in daily energy demand. A complete overview of the input and output variables used in the regression models is then provided in Table 3.

The performance of the regression models on the test set is summarized in Table 4, using standard evaluation metrics including R^2, RMSE, MAE, MAPE. To ensure an unbiased assessment of each model's predictive capability, the actual cluster labels were used as inputs during this phase. This approach isolates the regression performance from potential classification errors (in subsequent evaluations, the full pipeline performance is assessed using the predicted cluster labels from the classifier).

Among the tested models, BART (Bayesian Additive Regression Trees) achieves the best overall performance, with the highest R^2 value

Fig. 8. Residuals between the estimated and actual daily electrical energy consumption values for 2021 (a) and 2022 (b) using the BART model.

Table 3
Set of input and output variables of the regression models, along with their respective types and units of measurement.

Variable	Variable Type	Unit
Input Variables		
Daily average outdoor air temperature	Numeric	[°C]
Daily maximum outdoor air temperature	Numeric	[°C]
Daily minimum outdoor air temperature	Numeric	[°C]
Daily standard deviation of outdoor air temperature	Numeric	[°C]
Daily average global horizontal irradiance	Numeric	[W/m^2]
Daily maximum global horizontal irradiance	Numeric	[W/m^2]
Daily standard deviation of global horizontal irradiance	Numeric	[W/m^2]
Day of week	Categorical	[–]
Month of the year	Categorical	[–]
Holiday	Boolean	[–]
Cluster label	Categorical	[–]
Target Variable		
Daily electrical energy consumption	Numeric	[kWh]

Table 4
Performance metrics assessed for the developed regression models.

Model	R^2 [–]	RMSE [kWh]	MAE [kWh]	MAPE [%]
GAM	0.90	55.59	31.36	10.44%
CART	0.91	54.45	34.12	10.81%
Boosting	0.90	54.49	35.97	13.05%
Catboost	0.92	48.57	29.99	11.12%
BART	0.93	45.80	29.67	11.00%

of 0.93, indicating excellent explanatory power. Furthermore, it records the lowest RMSE (45.80 kWh) MAE (29.67 kWh) and MAPE (11.00%), confirming its superior accuracy in estimation. Given these results, the developed BART model was selected for being used in the proposed framework of the analysis.

Fig. 8 shows the residuals between the estimated and actual daily electrical energy consumption for 2021 and 2022, as calculated using the BART model. Each cell corresponds to a day within the testing weeks of both years, with the color intensity representing the magnitude of residuals in kWh.

Overall, the residuals remain relatively low throughout most of the year, demonstrating the model's strong predictive capability and its effectiveness in capturing general consumption patterns across both years. Importantly, the BART model successfully reproduces intra-week variability as well as longer-term seasonal trends, underscoring its robustness and reliability.

The calendar plots distinctly highlight specific weeks and weekdays where deviations are more evident, suggesting periods influenced by atypical conditions or unique operational patterns. Such insights are valuable, as they help identify potential anomalies or systematic discrepancies in model performance, thereby supporting further refinement and targeted improvements in future analyses.

Beyond predictive performance, a significant advantage of BART is its Bayesian framework, which inherently provides confidence intervals around predictions. This means that, differently from purely point-estimate models, BART can quantify the uncertainty associated with its predictions.

Technically, BART models the target variable as the sum of many regression trees, each contributing a small part to the final prediction. Through Bayesian inference and a markov chain monte carlo process, it samples many possible configurations of trees from the posterior distribution. This ensemble of sampled trees allows BART to produce not only a mean prediction but also an estimate of the entire distribution of possible outcomes, from which confidence intervals are derived. Such confidence intervals were then evaluated for each estimation in the testing set and used to define an acceptability range around the rescaled daily load profile as further detailed in the next section.

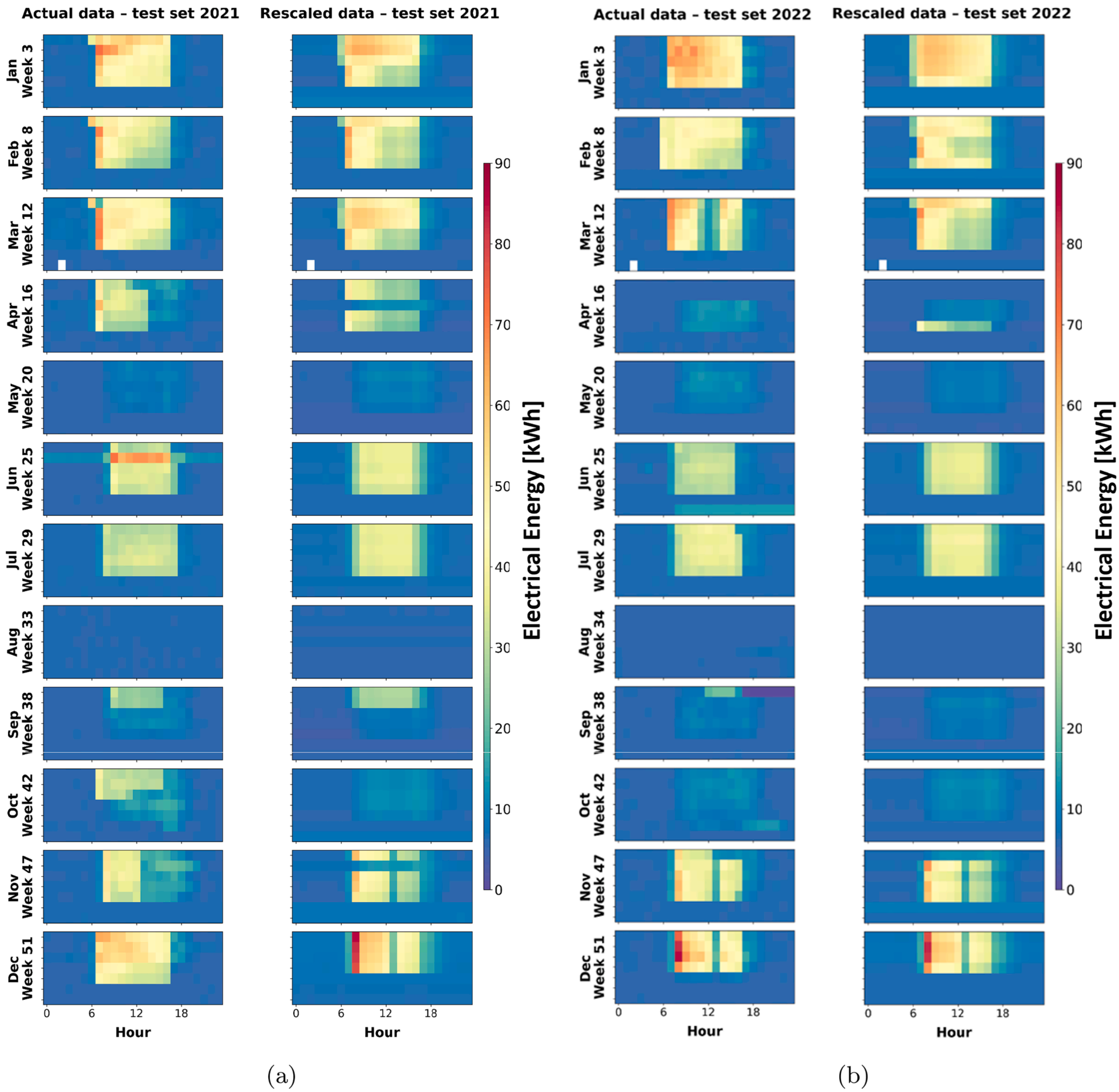


Fig. 9. Actual and rescaled hourly electrical energy consumption values for the test weeks of 2021 (a) and 2022 (b).

6.4. Magnitude rescaling

The final step of the methodological workflow consists in rescaling the estimated normalized daily load profile shapes based on the predicted total daily electrical energy consumption. This was achieved by multiplying the 24 hourly values of the normalized cluster centroid by the corresponding daily scaling factor K, as described in Section 5. The procedure was applied across the entire test set, generating rescaled hourly energy consumption profiles. These were then compared against the actual measured values to assess the overall performance and accuracy of the proposed framework.

Fig. 9 shows a comparative heatmap of the actual (Fig. 11a) and rescaled (Fig. 11b) hourly electrical energy consumption values for the entire test set dividing the test weeks of 2021 from 2022.

The rescaled values, obtained through the full methodological framework (i.e., by sequencing the classification and regression tasks) closely replicate the actual hourly profiles. The process successfully captures both shape and magnitude seasonal variations and intra-day dynamics, particularly the winter and early spring peaks during morning hours and the more uniform patterns observed in summer.

From the Fig. 9 it is possible to infer that the rescaled values generally follow the actual ones suggesting the absence of systematic bias in the process. Nevertheless, some discrepancies appear during early morning and late afternoon hours in specific periods, particularly in late winter and early summer, where peak intensities were occasionally under- or overestimated. These deviations may stem from local irregularities in usage patterns not fully captured by the normalized load profiles. Despite these minor variations, the overall consistency and the

Table 5
Performance metrics assessed for different tests.

Test	R^2 [–]	RMSE [kWh]	MAE [kWh]	MAPE [%]
Classification + Regression	0.86	5.12	2.30	19.42%
Classification + Actual Data	0.90	4.25	1.82	15.10%
Actual Data + Regression	0.91	3.87	2.00	17.53%
LSTM	0.70	7.34	3.64	32.09%
TCN	0.82	5.45	2.61	20.93%
TFT	0.86	4.87	2.33	20.61%

absence of persistent error patterns confirm the reliability and generalization capability of the proposed methodological framework. To assess the accuracy of the rescaling process at the hourly scale MAPE, RMSE, MAE, and R^2 metrics were calculated and are reported in Table 5. These metrics were also evaluated across three additional comparative experiments designed to determine which step (i.e., shape classification or daily energy consumption estimation) has the most significant influence on the hourly results. Additionally, an experiment was conducted to benchmark the performance of the proposed framework against a set of deep learning approaches formulated as a sequence-to-sequence model to predict the entire 24 h load profile for the following day. The three models developed for conducting the comparisons were: a Long Short-Term Memory (LSTM) neural network representing the state of the art, a Temporal Convolutional Network (TCN), representing the convolutional paradigm, and a Temporal Fusion Transformer (TFT), representing the attention-based hybrid paradigm [48].

The first test, named "Classification + Regression" reports the performance achieved by sequentially applying the classification and regression models, where the regression model is fed with the cluster label predicted by the classifier, thus emulating the real error propagation of the pipeline in estimating the expected load profile for the day ahead (hourly results are reported in Fig. 9. The second test, "Classification + Actual Data" evaluates the rescaling performance when the shape predicted by the classification model is combined with the actual observed daily energy consumption, instead of the value estimated by the regression model. Conversely, the third test, "Actual + Regression," uses the actual cluster label to select the load shape while relying on the regression model to estimate daily energy consumption.

The results indicate that the accuracy of the regression model has the greatest impact on the final outcomes. Notably, when the predicted load shape is combined with the actual daily energy consumption value to reconstruct the expected daily load profile, the MAPE decreases by approximately 2% respect to the "Classification + Regression" test. However, this is not necessarily a drawback, since the purpose of the regression model is not to minimize as much as possible the forecast error on future values but rather to serve as a benchmarking tool, estimating the most likely expected energy consumption (i.e., normal behaviour) under specific exogenous conditions instead of predicting its exact temporal evolution. In fact, while rolling autoregressive approaches can achieve higher accuracy in capturing evolving trends in energy consumption, they may also reflect anomalies as part of their forecasts rather than serving as a stable reference for detecting deviations.

Finally, the tests labeled "LSTM", "TFT" and "TCN" in Table 5 reports the performance achieved by directly employing a single regression model (i.e., Long short term memory neural network, Temporal Convolutional Network and Temporal Fusion Transformer) rather than independently modeling shape and magnitude. In these configurations, the models were implemented as a sequence-to-sequence models, with the objective of predicting the 24-hourly energy consumption profile for the day ahead. The models use the preceding 24 hourly energy consumption values, along with seven additional explanatory variables, each lagged both backward and forward by 24 h. These variables include external temperature, global solar radiation, and cyclical encodings (i.e., sine and cosine) of the month and hour, as well as the day of the week. As a result, the input for each day is structured as a matrix of dimensions 24 × 15, where each of the 24 hourly time steps is represented by 15 features (accounting for the backward and forward lags of the seven variables and the backward lag of the energy consumption). The corresponding target output is a vector containing the 24 hourly energy consumption values for the subsequent day. According to the results reported in Table 5, the proposed process achieved the lowest MAPE (19.42%) and MAE (2.30 kWh), outperforming not only the LSTM but also the TCN and TFT, while the TFT achieves a lower RMSE. Figs. 10 and 11 display comparative line plots for selected weeks in the testing sets of 2021 and 2022, where actual electrical power consumption is shown in yellow, the proposed rescaling process in blue, and the LSTM predictions in green. The LSTM was selected as the visual benchmark as it represents the most widely recognised baseline for sequential energy consumption prediction in the building energy literature, providing the most meaningful reference for contextualising the performance of the proposed framework.

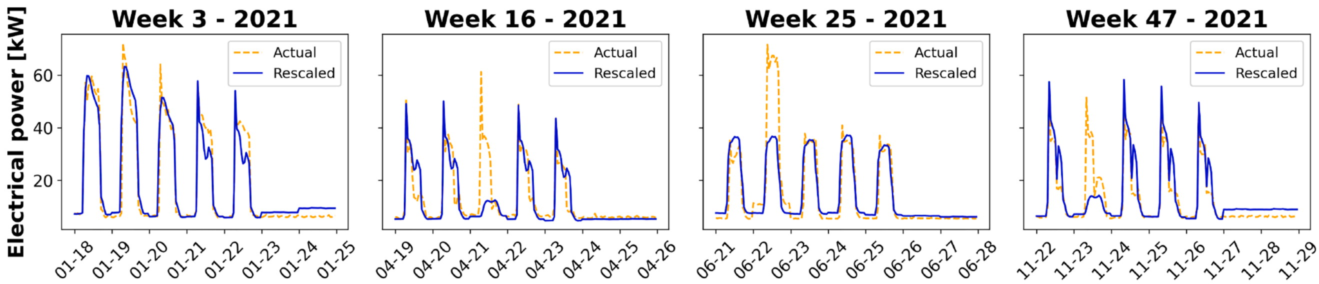
A visual comparison of these figures reveals important insights into model performance. The plots consistently demonstrate that the proposed rescaling process achieves closer alignment with the actual measurements across the weeks analyzed, accurately reproducing both the amplitude and timing of daily peaks and troughs in electrical demand. Notably, the rescaled estimates effectively capture sharp fluctuations and daily operational patterns, indicating the model's capacity to reflect dynamic system behavior under varying conditions (week 16 and 47 of 2021 and week 47 of 2022).

In contrast, the LSTM model often underestimates or smooths out key variations, particularly during periods of pronounced consumption peaks or abrupt changes. This is evident in several weeks, such as Weeks 16 and 47 of 2021 and 2022, where the LSTM predictions exhibit reduced magnitude and fail to track the rapid transitions observed in the actual data. Moreover, the proposed rescaling process maintains strong predictive performance across different seasons and load conditions, as seen from the consistent tracking of the actual profiles in both low- and high-consumption periods. This highlights its robustness and adaptability to seasonal trends, operational schedules, and potential anomalies. However, for certain days, the residuals exhibit consistently high absolute values across multiple hours. These deviations are of particular interest, as they may indicate the presence of anomalous consumption patterns not captured by the representative daily load profiles. Such persistent discrepancies suggest that the actual energy usage on those days deviated substantially from expected behavior, potentially due to atypical operational conditions, equipment faults, or external influences. Identifying and analyzing these patterns is therefore of critical interest, as they may reveal opportunities for anomaly detection, further model refinement, or targeted investigation into building system performance.

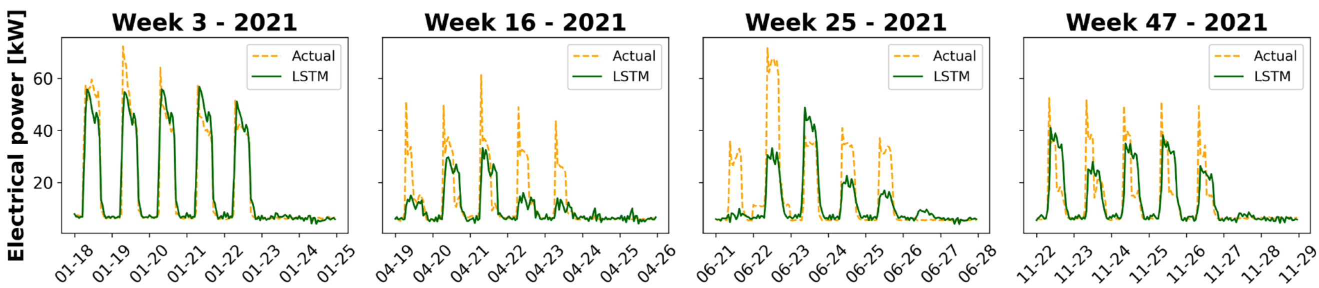
6.5. Anomaly detection results

This section presents the results of the anomaly detection process conducted for the considered case study. As previously explained in Section 5, an anomaly in daily energy consumption is flagged when the actual measured hourly load deviates beyond an acceptable range for a significant portion of the day. Such deviations can manifest as differences in magnitude, changes in the characteristic shape of the consumption profile, or a combination of both. The acceptability range is derived from the uncertainty of the regression model used to estimate the daily energy consumption. In fact, A key advantage of the BART model is that it produces not only point estimates for daily energy consumption but also quantifies the associated uncertainty through a confidence interval. This interval provides a probabilistic measure of the model uncertainty under the specific operating conditions observed.

To establish the boundaries for acceptable daily energy use, a 95% confidence interval around the predicted daily energy consumption was calculated to define upper and lower limits. These bounds provide statistically grounded thresholds that account for both the variability in building operations and the uncertainty in the model predictions.

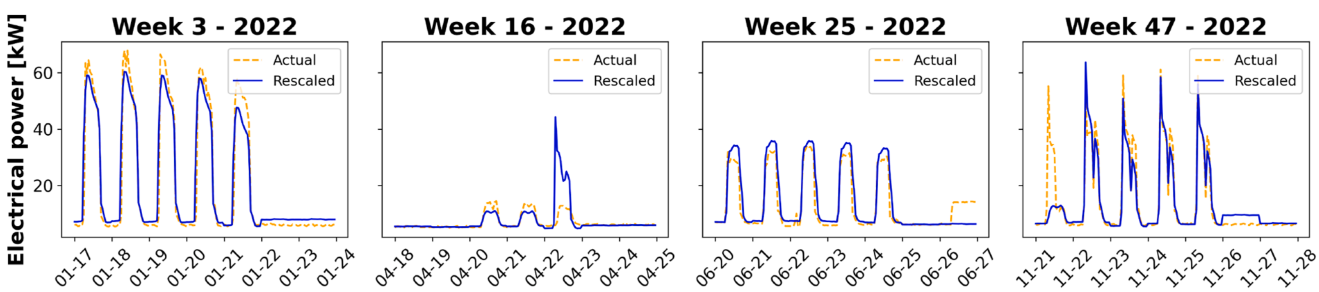


(a)

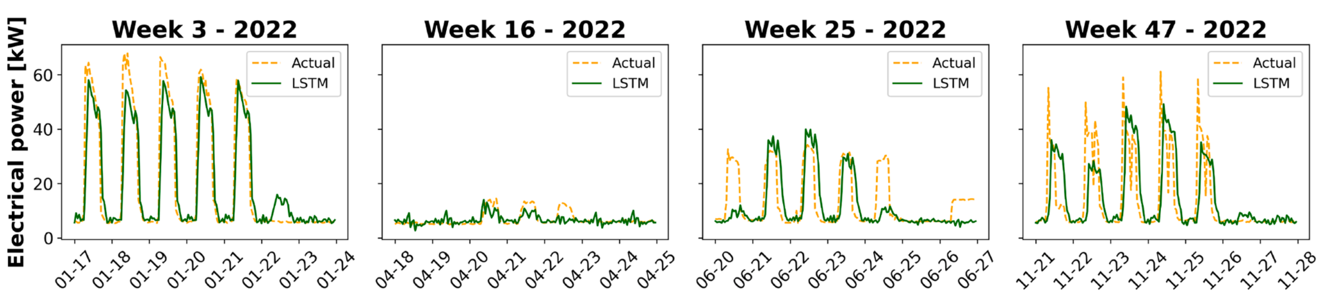


(b)

Fig. 10. Comparison of actual consumption with the rescaled values by means of the proposed methodology (a) and against of LSTM predictions, for a sample of testing weeks in 2021.



(a)

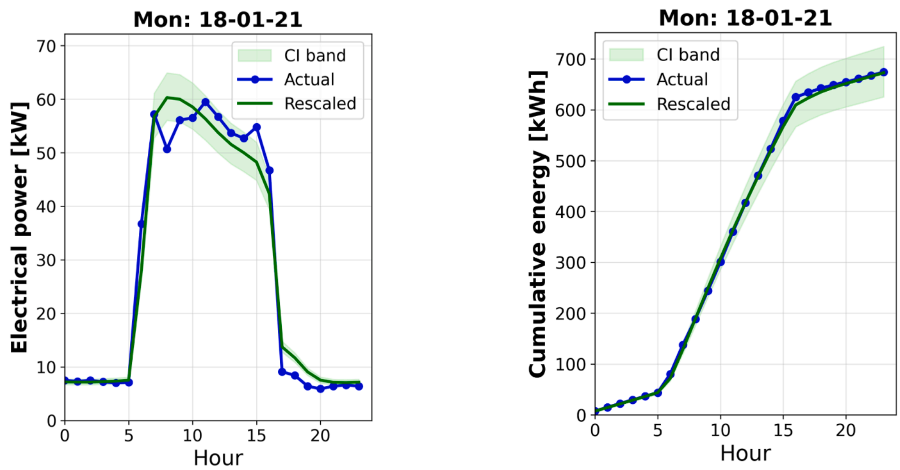


(b)

Fig. 11. Comparison of actual consumption with the rescaled values by means of the proposed methodology (a) and against of LSTM predictions, for a sample of testing weeks in 2022.

The computed upper and lower limits were then added to or subtracted from the estimated value to rescale the normalized daily load profile derived from the classification phase, as outlined in Eq. (1). In this approach, the classification model predicts the normalized shape of the daily load profile for the day ahead, while the regression model provides three rescaling values: the estimated daily consumption, as well as the upper and lower limits corresponding to the 95% confidence interval.

For reference, Table 6 presents all input and output features for 18-01-2021, which were used to define the corresponding reference daily load profile for benchmarking and anomaly detection purposes. Alongside the input variables for both the classification and regression models,



(a) (b)

Fig. 12. Expected and actual daily load profiles on hourly scale with evidence of the acceptability range. (a) Cumulated expected and actual hourly energy consumption over 24 h with evidence of the acceptability range (b).

Table 6
Input and output features for the day 18-01-2021 used for define its reference rescaled daily load profile.

Features for the day 18-01-2021	Value	Unit
Daily average outdoor air temperature	2.73	[°C]
Daily maximum outdoor air temperature	7.03	[°C]
Daily minimum outdoor air temperature	−1.49	[°C]
Daily standard deviation of outdoor air temperature	2.78	[°C]
Daily average global horizontal irradiance	166.70	[W/m^2]
Daily maximum global horizontal irradiance	329.70	[W/m^2]
Daily standard deviation of global horizontal irradiance	111.80	[W/m^2]
Day of week	Monday	[–]
Month of the year	January	[–]
Holiday	NO	[–]
Predicted cluster label	Cluster 3	[–]
Sum of the 24 hourly values of cluster 3 centroid	10.60	[–]
Predicted daily energy consumption	672.91	[kWh]
Actual daily energy consumption	659.30	[kWh]
Upper bound confidence interval	724.47	[kWh]
Lower bound confidence interval	625.47	[kWh]
Scaling factor K	63.43	[–]
Scaling factor K_{upper_bound}	68.28	[–]
Scaling factor K_{lower_bound}	58.96	[–]

the table reports that, for this particular day, the predicted shape of the daily load profile corresponds to the centroid of cluster 3, as depicted in Fig. 5. The sum of the 24 hourly values of the normalized centroid profile is given by $\sum_{h=1}^{24} c_h = 10.60$.

Furthermore, the regression model estimated a daily energy consumption of 672 kWh, with the upper and lower bounds of the confidence interval at 724 kWh and 625 kWh, respectively. By applying Eq. (1), three different scaling factors were calculated by dividing the regression model predictions (including the confidence interval bounds) by $\sum_{h=1}^{24} c_h$, resulting in scaling factors of $K = 63.43$, $k_{upper\_bound} = 68.28$, and $k_{lower\_bound} = 58.96$.

By multiplying all normalized values of the predicted cluster centroid by these scaling factors, it becomes possible to rescale the daily load profile to its expected magnitude, along with defining its acceptable variability range.

Fig. 12, reports then the rescaling results for the same day (i.e., 18-01-2021).

Starting from the info in Table 6 the Fig. 12 integrates two fundamental representations. The first in Fig. 12a depicts the rescaled load profile for the day (green line) overlaid with acceptability bands derived from the confidence interval of the regressor prediction (green area). These bands quantify the expected range of normal behavior, thus enabling detection of deviations in the temporal distribution of energy use throughout the day. On the other hand the blue line represents the actual consumption. For the considered day the building use energy as expected with small fluctuations around the rescaled values.

The second visualization is reported in Fig. 12b where the expected load profile (together with the upper an lower bound) and the actual profile reported in Fig. 12a were cumulated across the 24 h period.

This cumulative curves illustrates the expected progression of energy consumption from midnight, providing hour-by-hour insights into how much energy should have been consumed up to any given time. From this plot it can be seen that despite fluctuations occurred around the expected values, the total energy consumed at the end of the day (659 kWh) is included in the confidence interval (i.e., between 625 and 724 kWh) suggesting a normal operation pattern. The combined application of these two analyses offers significant advantages. The instantaneous load profile allows precise identification of the timing and nature of specific anomalies, such as shifts in demand peaks, unexpected load ramp-ups, or drops in consumption. This facilitates pinpointing the operational context of anomalies within the daily schedule. Conversely, the cumulative profile serves to assess the overall impact of any detected anomalies on total daily consumption, revealing whether deviations observed in specific hours translate into significant overconsumption or underconsumption when considered across the entire day. Together, these perspectives provide a robust framework for anomaly detection, enabling both fine-grained diagnostic analysis and broader evaluation of building energy performance.

In order to better demonstrate the tool usability, results were reported for an entire week in the testing set in Fig. 14 and Fig. 13.

From the weekly visualizations, it is clear that model performance exhibits variations across different days. During week 25 of 2021, with the exception of Tuesday, the actual cumulative consumption generally follows the benchmark profile and remains within acceptable limits, reflecting standard operational conditions as shown in Fig. 13.

The hourly profiles presented in Fig. 14 offer further depth by capturing intraday fluctuations and enabling a more precise identification of deviations within specific periods of the day.

Tuesday, however, stands out as an evident anomaly in both figures, as the cumulative curve starts to diverge significantly from the benchmark in the early afternoon, signaling a sustained and substantial in-

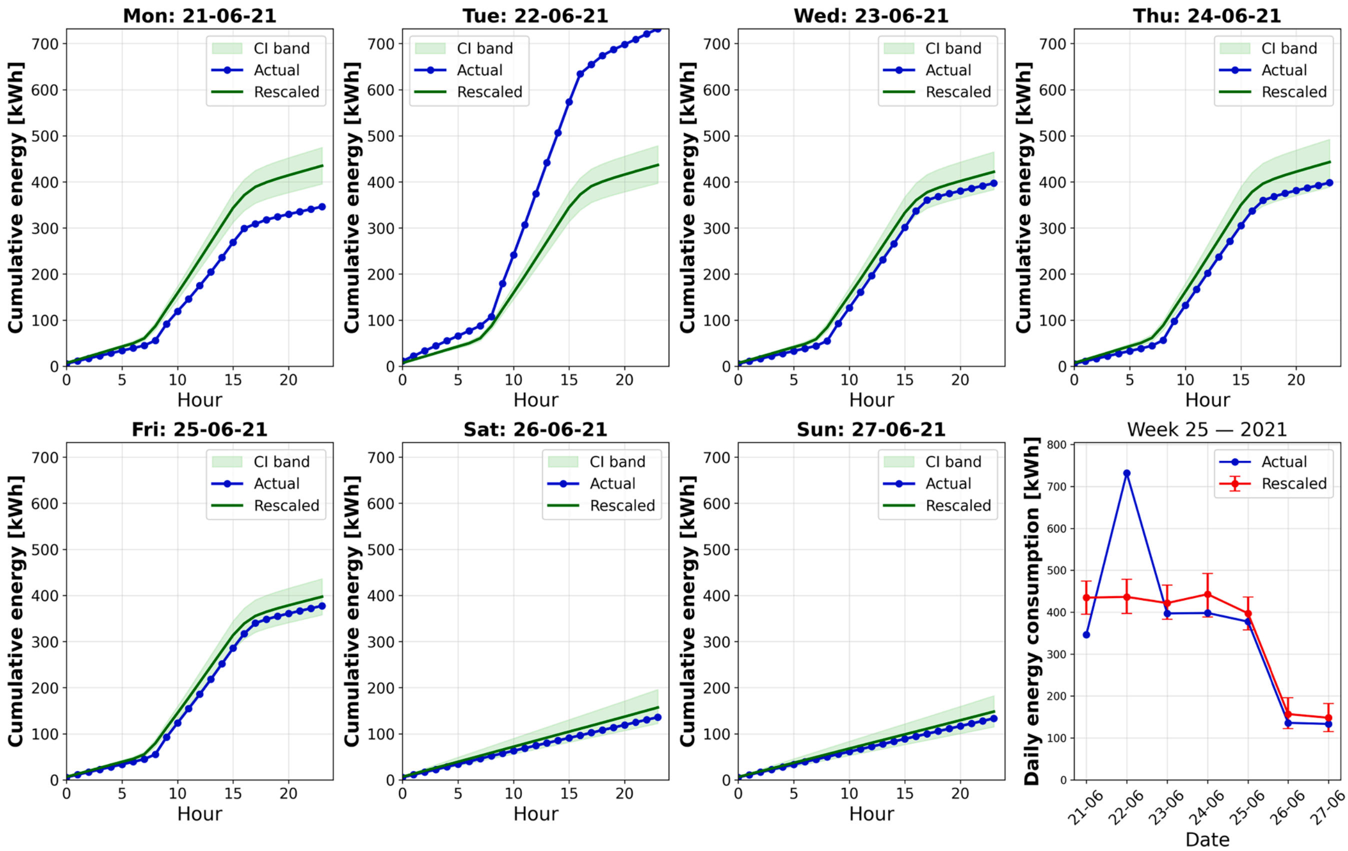


Fig. 13. Cumulated expected and actual daily energy consumption over a week in the test set with evidence of the acceptability range.

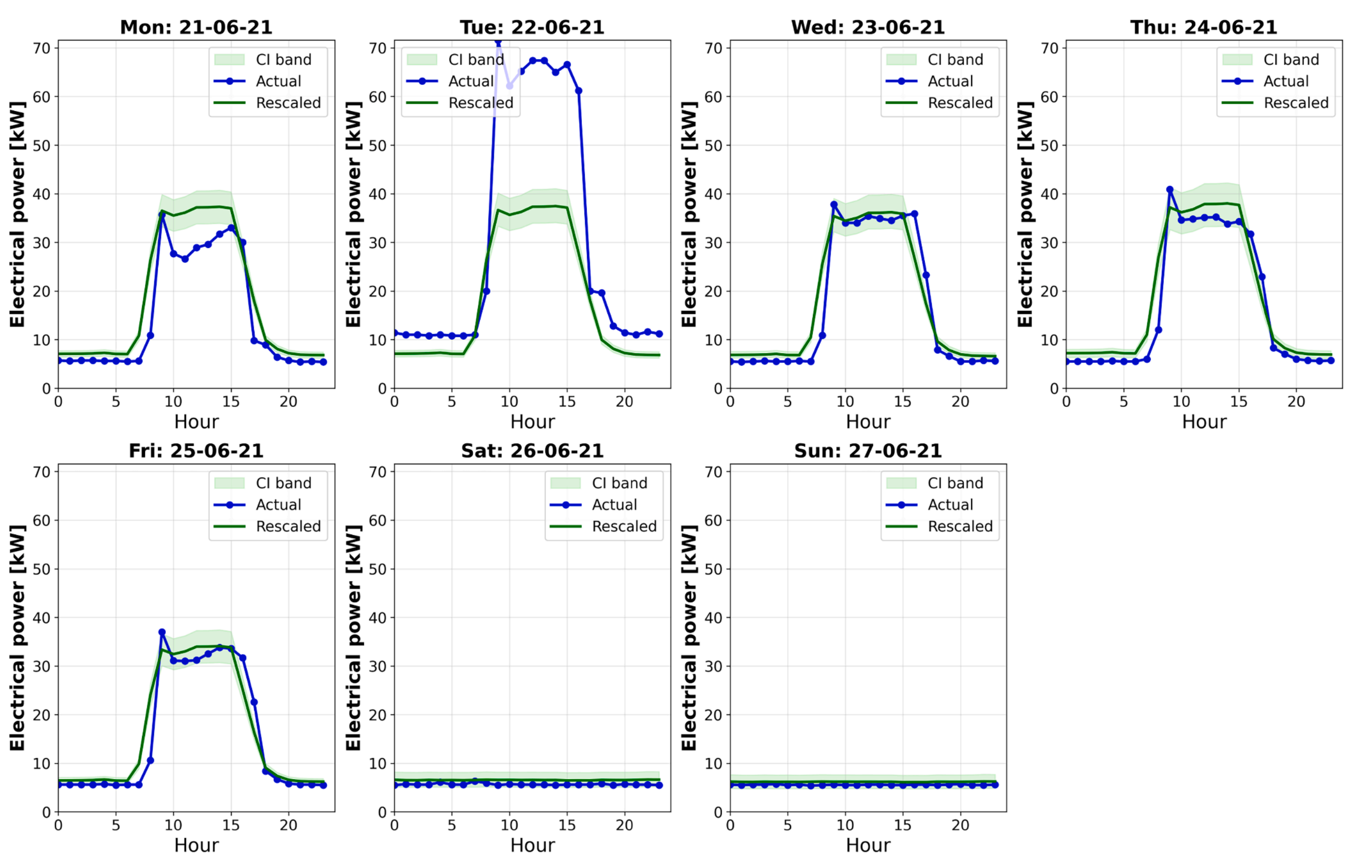


Fig. 14. Expected and actual hourly energy consumption over a week in the test set with evidence of the acceptability range.

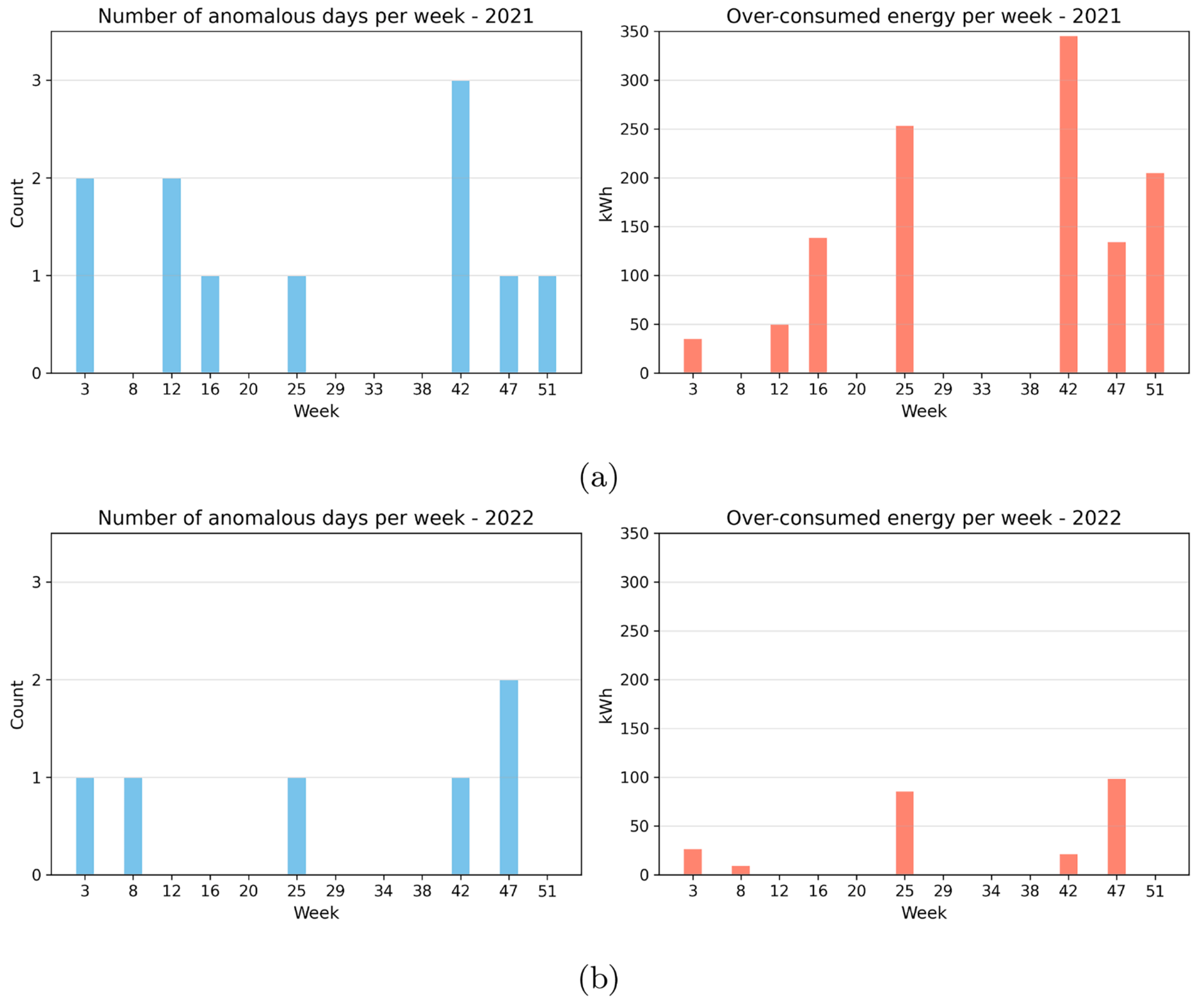


Fig. 15. Detected number of anomalous days and over-consumed energy per week in the testing set of 2021 (a) and 2022 (b).

crease in energy use that continues for the rest of the day rather than a brief, isolated spike.

This observation is supported by the hourly profile, which highlights an evident rise in consumption levels during the same afternoon hours. In contrast to the patterns seen on preceding and subsequent days, the hourly values on Tuesday not only reach higher peaks but also remain consistently elevated across several hours, suggesting a persistent anomaly rather than a temporary fluctuation.

Taken together, the cumulative and hourly analyses provide a comprehensive understanding of the deviation observed for Tuesday, indicating that an unusual operational event or condition led to higher then expected energy demand, and underscoring the importance of examining both aggregate and high-resolution data to accurately detect and interpret anomalies.

Summary results for all weeks in the testing set are presented in Fig. 15, offering a comprehensive evaluation of the potential operational impact of the developed anomaly detection process.

These figures report, for each week of the 2021 and 2022 test sets, two key indicators: (i) the number of days on which actual daily energy consumption exceeded the upper bound of the acceptability range defined by the 95% confidence interval of the model, and (ii) the total amount of energy overconsumed during each week.

Fig. 15a reveals that in 2021, several weeks experienced multiple anomalous days, with peaks of up to three anomalies in a single week. The right panel also shows significant weekly overconsumption, with weeks exceeding 150 kWh above the predicted upper limit. Such deviations suggest periods of substantial deviation from expected operational patterns, potentially caused by unforeseen operational patterns or external boundary conditions.

Conversely, Fig. 15b shows that 2022 exhibited fewer anomalous days and generally lower levels of overconsumption. While some weeks, notably weeks 25 and 47, still recorded an overconsumption close to 100 kWh, the overall magnitude was generally lower than in 2021 suggesting an enhanced operational stability in 2022.

7. Discussions

The present work introduced a data-analytics process for anomaly detection in building energy consumption time series, which relies on comparing measured energy consumption profiles with expected patterns derived from historical data. The process includes classification of daily load profile shapes and estimation of expected energy magnitudes, combining these elements to identify significant deviations that may indicate anomalies. Over the 24-week testing period, the proposed methodology identified 17 days in which the actual daily energy consumption was higher than the upper limit of the acceptability range defined by a 95% confidence interval. These deviations correspond to a cumulative amount of overconsumed energy of approximately 1400 kWh, indicating instances of potential inefficiencies or operational anomalies. The proposed approach addresses both methodological and practical gaps in the state of the art.

A defining feature of this methodology is the adoption of a context-aware benchmarking approach rather than a purely autoregressive forecasting paradigm for anomaly detection. Benchmarking establishes a reference baseline extracted from historical data, characterized by both the shape and magnitude of daily load profiles. This strategy supports robust quantification of deviations as indicators of anomalies. While forecasting methods excel in anticipating future values, they can become less sensitive to certain deviations because they continuously adapt to trends, even those associated with faults or inefficiencies. Benchmarking, conversely, allows to better contextualize the normal and frequent operational patterns maintaining sensitivity to the occurrence of unexpected and infrequent patterns that might otherwise be masked by faster and adaptive learning capabilities of forecast models. However, this rigidity also poses challenges, as infrequent but legitimate operational changes may be flagged as anomalies, increasing the risk of false positives.

The innovation of the proposed framework is the architectural separation between the classification of daily load shapes and the regression of daily energy magnitudes. Rather than treating the load profile as a single entity to be predicted by a unified model, the decoupled structure recognises that shape and magnitude are governed by partially distinct drivers and may evolve independently over time. Shifts in operational schedules or occupancy patterns, for instance, can alter the temporal distribution of energy use without necessarily affecting total daily consumption, while changes in weather conditions or seasonal factors may influence overall magnitude without modifying the characteristic shape of the profile. By modelling these two dimensions separately, the framework better reflects the underlying physical and operational mechanisms driving energy consumption. This separation also has important practical implications for the long-term maintainability of the benchmarking process. When building operations change, only the affected component needs to be retrained or updated, reducing computational overhead and the risk of unnecessary model drift. It also allows each component to be independently evaluated, diagnosed, and replaced with alternative algorithms as new methods become available, without requiring a full redesign of the pipeline. Nevertheless, the coherent integration of the two components remains an aspect requiring careful attention. Since the predicted load shape informs the regression model as a categorical input, classification errors can propagate into magnitude estimation and ultimately affect the quality of the reconstructed benchmark profile. Addressing this error propagation through more robust coupling strategies represents a promising direction for future refinement.

Another significant strength of the proposed methodology is its adaptability to buildings with limited metering infrastructure that represents a crucial consideration given the extensive presence of legacy building stock globally. Instead of depending on high-frequency or detailed sub-metering, the approach effectively exploits meter-level time series data combined with easy-to-collect contextual information. In addition, the process was designed to run once a day allowing its application in monitoring systems characterized by data acquisition with high latency. Daily-level processing mitigates issues related to missing sub-daily values, ensuring continuity of performance monitoring and avoiding gaps that could be problematic in rolling window prediction methods.

Despite its daily focus, the tool is capable of dynamically assessing energy consumption evolution throughout the day with hourly resolution, rather than limiting itself to end-of-day anomaly reporting. By verifying consumption trends against expected daily trajectories (either cumulative or non-cumulative) the process supports near real-time performance monitoring enabling facility managers to promptly detect and address emerging deviations that could drive energy consumption beyond acceptable thresholds. Importantly, it also allows the system to disregard transient local deviations unlikely to have substantial impacts on daily consumption, thus preventing unnecessary interventions. Furthermore, the tool explicitly quantifies uncertainty by calculating confidence intervals around magnitude predictions and reflecting these intervals on rescaled daily load profiles. This probabilistic approach avoids rigid binary judgments, allowing for more tolerant management of potential anomalies and a more effective handling of false alarms. The ultimate objective is to avoid overwhelming users with excessive feedback while providing information of high confidence and relevance. Alerts were designed to be issued only when deviations are likely to have a meaningful impact on daily energy consumption patterns, ensuring that operational interventions from the energy management staff were both justified and timely.

8. Conclusions

This work introduced a novel data analytics-based methodology for anomaly detection in building energy consumption time series, explicitly separating the prediction of daily load profile shapes from the estimation of daily energy magnitudes. Tested on a university campus, the method demonstrated strong predictive accuracy and effectively identified anomalies, confirming its robustness and practical applicability. Compared to deep learning approaches such as LSTM, TCN and TFT, the proposed framework offers superior/comparable performance with lower computational cost, making it particularly suitable for real-world deployment in buildings with limited metering infrastructure and data acquisition characterized by high latency.

Several directions are envisioned to further develop and extend the proposed framework. In the short term, the integration of online learning techniques would enhance the adaptability of the models, allowing them to dynamically adjust to evolving building usage patterns without requiring full retraining. Incorporating richer contextual information such as operational schedules, equipment logs, and external events could further improve detection precision and reduce the occurrence of false positives, which represent one of the main practical limitations of benchmarking-based approaches. From a methodological standpoint, extending the framework to handle multivariate input data, including thermal loads, indoor environmental conditions, and detailed occupancy information, would enable a more comprehensive characterization of building performance and support the detection of anomalies that manifest across multiple energy vectors simultaneously. In the longer term, as monitoring infrastructure evolves and more granular data become available, the framework could be complemented with a dedicated diagnostic layer capable of pinpointing the specific subsystems responsible for anomalies initially detected at the whole-building level. The integration of submetering data from systems such as HVAC, lighting, plug loads, or individual equipment would allow the framework to advance beyond anomaly detection toward precise fault localization, enabling facility managers not only to detect that an anomaly has occurred but also to identify its root cause. Such developments would transform the framework from a monitoring tool into a proactive diagnostic system, increasing its value for preventive maintenance, fault management, and effective energy optimization.

CRediT authorship contribution statement

Marco Savino Piscitelli: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Conceptualization; **Giacomo Buscemi:** Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization; **Carlo Roselli:** Writing – review & editing, Validation, Investigation, Data curation; **Elisa Marrasso:** Writing – review & editing, Validation, Investigation, Data curation; **Giovanna Pallotta:** Writing – review & editing, Validation, Investigation, Data curation; **Alfonso Capozzoli:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: co-author Marco Savino Piscitelli is an editorial board member in the Journal. Given his role, Marco Savino Piscitelli had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor. If there are other authors, they 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

The work of Giacomo Buscemi, Carlo Roselli and Alfonso Capozzoli was carried out in the framework of the project PRIN 2020: OPTIMISM - Optimal refurbishment design and management of small energy micro-grids, funded by the Italian Ministry of University and Research (MUR). The work of Marco Savino Piscitelli was carried out within the Ministerial Decree no. 1062/2021 and received funding from the FSE REACT-EU - PON Ricerca e Innovazione 2014–2020. This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Data availability

Data will be made available on request.

References

[1] H. Li, H. Johra, F. de Andrade Pereira, T. Hong, J. Le Dréau, A. Maturo, M. Wei, Y. Liu, A. Saberi-Derakhtenjani, Z. Nagy, A. Marszal-Pomianowska, D. Finn, S. Miyata, K. Kaspar, K. Nweye, Z. O'Neill, F. Pallonetto, B. Dong, Data-driven key performance indicators and datasets for building energy flexibility: a review and perspectives, Appl. Energy 343 (2023) 121217. https://doi.org/10.1016/j.apenergy.2023.121217

[2] X. Cao, X. Dai, J. Liu, Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade, Energy Build. 128 (2016) 198–213. https://doi.org/10.1016/j.enbuild.2016.06.089

[3] K. Lawal, H.N. Rafsanjani, Trends, benefits, risks, and challenges of IoT implementation in residential and commercial buildings, Energy Built Environ. 3 (3) (2022) 251–266. https://doi.org/10.1016/j.enbenv.2021.01.009

[4] Y. Himeur, K. Ghanem, A. Alsalemi, F. Bensaali, A. Amira, Artificial intelligence based anomaly detection of energy consumption in buildings: a review, current trends and new perspectives, Appl. Energy 287 (2021) 116601. https://doi.org/10.1016/j.apenergy.2021.116601

[5] O. Linda, D. Wijayasekara, M. Manic, C. Rieger, Computational intelligence based anomaly detection for building energy management systems, in: 2012 5th International Symposium on Resilient Control Systems, 2012, pp. 77–82. https://doi.org/10.1109/ISRCS.2012.6309297

[6] M.D. Elu Beily, I. Darwish, A. Mohamed, K. Sawon, M. Shafikuzzaman, A deep-learning approach for detecting anomalies in commercial building load profiles, in: 2024 IEEE 19th Conference on Industrial Electronics and Applications (ICIEA), 2024, pp. 1–6. https://doi.org/10.1109/ICIEA61579.2024.10665197

[7] M. Molina-Solana, M. Ros, M.D. Ruiz, J. Gómez-Romero, M.J. Martin-Bautista, Data science for building energy management: a review, Renew. Sustain. Energy Rev. 70 (2017) 598–609. https://doi.org/10.1016/j.rser.2016.11.132

[8] R. Chiosa, M.S. Piscitelli, A. Capozzoli, A data analytics-based energy information system (EIS) tool to perform meter-level anomaly detection and diagnosis in buildings, Energies 14 (1) (2021). https://www.mdpi.com/1996-1073/14/1/237.

[9] R. Chiosa, M.S. Piscitelli, C. Fan, A. Capozzoli, Towards a self-tuned data analytics-based process for an automatic context-aware detection and diagnosis of anomalies in building energy consumption timeseries, Energy Build. 270 (2022) 112302. https://doi.org/10.1016/j.enbuild.2022.112302

[10] A. Capozzoli, M.S. Piscitelli, S. Brandi, D. Grassi, G. Chicco, Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings, Energy 157 (2018) 336–352. https://doi.org/10.1016/j.energy.2018.05.127

[11] B. Ayegba, M. Abtahi, A.K. Athienitis, H. Frank Nouanegue, Benchmarking building energy flexibility using monitored data, in: 2024 3rd International Conference on Energy Transition in the Mediterranean Area (SyNERGY MED), 2024, pp. 1–5. https://doi.org/10.1109/SyNERGYMED62435.2024.10799373

[12] M.S. Piscitelli, R. Giudice, A. Capozzoli, A holistic time series-based energy benchmarking framework for applications in large stocks of buildings, Appl. Energy 357 (2024) 122550. https://doi.org/10.1016/j.apenergy.2023.122550

[13] S. Maitra, A data mining-based dynamical anomaly detection method for integrating with an advance metering system, 2024,. arxiv preprint arXiv:2405.02574 [cs.LG]

[14] K. Grolinger, H.F. ElYamany, W.A. Higashino, M.A.M. Capretz, L. Seewald, Energy slices: benchmarking with time slicing, Energy Effic. 11 (2) (2018) 521–538. https://doi.org/10.1007/s12053-017-9582-8

[15] X. Luo, T. Hong, Y. Chen, M.A. Piette, Electric load shape benchmarking for small- and medium-sized commercial buildings, Appl. Energy 204 (2017) 715–725. https://doi.org/10.1016/j.apenergy.2017.07.108

[16] J. Zhu, Y. Shen, Z. Song, D. Zhou, Z. Zhang, A. Kusiak, Data-driven building load profiling and energy management, Sustain. Cities Soc. 49 (2019) 101587. https://doi.org/10.1016/j.scs.2019.101587

[17] M. Bourdeau, X.q. Zhai, E. Nefzaoui, X. Guo, P. Chatellier, Modeling and forecasting building energy consumption: a review of data-driven techniques, Sustain. Cities Soc. 48 (2019) 101533. https://doi.org/10.1016/j.scs.2019.101533

[18] J. Amantegui, H. Morais, L. Pereira, Benchmark of electricity consumption forecasting methodologies applied to industrial kitchens, Buildings 12 (12) (2022). https://doi.org/10.3390/buildings12122231

[19] P. Arjunan, K. Poolla, C. Miller, EnergyStar++: towards more accurate and explanatory building energy benchmarking, Appl. Energy 276 (2020) 115413. https://doi.org/10.1016/j.apenergy.2020.115413

[20] M. Neshat, M. Thilakaratne, M. El-Abd, S. Mirjalili, A.H. Gandomi, J. Boland, Smart buildings energy consumption forecasting using adaptive evolutionary bagging extra tree learning models, Energy 333 (2025) 137130. https://doi.org/10.1016/j.energy.2025.137130

[21] J.-S. Chou, A.S. Telaga, W.K. Chong, G.E. Gibson, Early-warning application for real-time detection of energy consumption anomalies in buildings, J. Clean. Prod. 149 (2017) 711–722. https://doi.org/10.1016/j.jclepro.2017.02.028

[22] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, ACM Comput. Surv. 41 (3) (2009). https://doi.org/10.1145/1541880.1541882

[23] Y. Song, S. Kuang, J. Huang, D. Zhang, Unsupervised anomaly detection of industrial building energy consumption, Energy Built Environ. (2024). https://doi.org/10.1016/j.enbenv.2024.12.005

[24] M.S. Piscitelli, S. Brandi, A. Capozzoli, F. Xiao, A data analytics-based tool for the detection and diagnosis of anomalous daily energy patterns in buildings, Build. Simul. 14 (1) (2021) 131–147. https://doi.org/10.1007/s12273-020-0650-1

[25] N. Zangrando, P. Fraternali, M. Petri, N.O. Pinciroli Vago, S.L. Herrera González, Anomaly detection in quasi-periodic energy consumption data series: a comparison of algorithms, Energy Inform. 5 (4) (2022) 62. https://doi.org/10.1186/s42162-022-00230-7

[26] C. Zhang, Y. Zhao, Y. Zhou, X. Zhang, T. Li, A real-time abnormal operation pattern detection method for building energy systems based on association rule bases, Build. Simul. 15 (1) (2022) 69–81. https://doi.org/10.1007/s12273-021-0791-x

[27] L. Lei, X. Guo, R. Zheng, An innovative dynamic anomaly detection method based on hybrid data mining technology for building energy consumption, Energy Build. 319 (2024) 114559. https://doi.org/10.1016/j.enbuild.2024.114559

[28] N. Zangrando, P. Fraternali, P.M. Petri, V. Pinciroli, O. Nicolò, S.L.H. González, Anomaly detection in quasi-periodic energy consumption data series: a comparison of algorithms, Energy Inform. 5 (2022) 62. https://doi.org/10.1186/s42162-022-00230-7

[29] B. Gao, X. Kong, S. Li, Y. Chen, X. Zhang, Z. Liu, W. Lv, Enhancing anomaly detection accuracy and interpretability in low-quality and class imbalanced data: a comprehensive approach, Appl. Energy 353 (2024) 122157. https://doi.org/10.1016/j.apenergy.2023.122157

[30] L. Lei, B. Wu, X. Fang, L. Chen, H. Wu, W. Liu, A dynamic anomaly detection method of building energy consumption based on data mining technology, Energy 263 (2023) 125575. https://doi.org/10.1016/j.energy.2022.125575

[31] X. Zhou, T. Yang, L. Liang, X. Zi, J. Yan, D. Pan, Anomaly detection method of daily energy consumption patterns for central air conditioning systems, J. Build. Eng. 38 (2021) 102179. https://doi.org/10.1016/j.jobe.2021.102179

[32] N.A. Funde, M.M. Dhabu, A. Paramasivam, P.S. Deshpande, Motif-based association rule mining and clustering technique for determining energy usage patterns for smart meter data, Sustain. Cities Soc. 46 (2019) 101415. https://doi.org/10.1016/j.scs.2018.12.043

[33] M. Quintana, T. Stoeckmann, J.Y. Park, M. Turowski, V. Hagenmeyer, C. Miller, ALDI++: automatic and parameter-less discord and outlier detection for building energy load profiles, Energy Build. 265 (2022) 112096. https://doi.org/10.1016/j.enbuild.2022.112096

[34] L. Mascali, D.S. Schiera, S. Eiraudo, L. Barbierato, R. Giannantonio, E. Patti, L. Bottaccioli, A. Lanzini, A machine learning-based anomaly detection framework for building electricity consumption data, Sustain. Energy Grids Netw. 36 (2023) 101194. https://doi.org/10.1016/j.segan.2023.101194

[35] D. Azzalini, B. Flammini, C.A. Emanuele, A. Guadagno, E. Ragaini, F. Amigoni, An empirical evaluation of deep autoencoders for anomaly detection in the electricity consumption of buildings, Energy Build. 327 (2025) 115069. https://doi.org/10.1016/j.enbuild.2024.115069

[36] C. Fan, F. Xiao, Y. Zhao, J. Wang, Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data, Appl. Energy 211 (2018) 1123–1135. https://doi.org/10.1016/j.apenergy.2017.12.005

[37] M. Noorchenarboo, K. Grolinger, Explaining deep learning-based anomaly detection in energy consumption data by focusing on contextually relevant data, Energy Build. 328 (2025) 115177. https://doi.org/10.1016/j.enbuild.2024.115177

[38] Y. Shi, B. Gao, F. Xu, X. Xu, Y. Jiang, Federated anomaly detection with sparse attentive aggregation for energy consumption in buildings, IEEE Internet Things J. 11 (7) (2024) 11737–11749. https://doi.org/10.1109/JIOT.2023.3331351

[39] L. Zhang, J. Guo, P. Lin, R.L.K. Tiong, Detecting energy consumption anomalies with dynamic adaptive encoder-decoder deep learning networks, Renew. Sustain. Energy Rev. 207 (2025) 114975. https://doi.org/10.1016/j.rser.2024.114975
[40] L. Canale, G. Ficco, B. Di Pietra, M. Dell'Isola, A novel model to detect abnormal thermal energy consumption in existing buildings with limited metering infrastructure, Energy Build. 323 (2024) 114803. https://doi.org/10.1016/j.enbuild.2024.114803
[41] J. Kong, W. Jiang, Q. Tian, M. Jiang, T. Liu, Anomaly detection based on joint spatio-temporal learning for building electricity consumption, Appl. Energy 334 (2023) 120635. https://doi.org/10.1016/j.apenergy.2022.120635
[42] C. Fan, F. Xiao, M. Song, J. Wang, A graph mining-based methodology for discovering and visualizing high-level knowledge for building energy management, Appl. Energy 251 (2019) 113395. https://doi.org/10.1016/j.apenergy.2019.113395
[43] A. Copiaco, Y. Himeur, A. Amira, W. Mansoor, F. Fadli, S. Atalla, S.S. Sohail, An innovative deep anomaly detection of building energy consumption using energy time-series images, Eng. Appl. Artif. Intell. 119 (2023) 105775. https://doi.org/10.1016/j.engappai.2022.105775
[44] E. Marrasso, C. Roselli, G. Pallotta, A. Capozzoli, G. Buscemi, M.S. Piscitelli, Novel load-breakdown methodology for a university office building, Energy Build. 351 (2026). https://doi.org/10.1016/j.enbuild.2025.116746
[45] M.S. Piscitelli, S. Brandi, A. Capozzoli, Recognition and classification of typical load profiles in buildings with non-intrusive learning approach, Appl. Energy 255 (2019) 113727. https://doi.org/10.1016/j.apenergy.2019.113727
[46] L. Giannuzzo, F.D. Minuto, D.S. Schiera, A. Lanzini, Reconstructing hourly residential electrical load profiles for renewable energy communities using non-intrusive machine learning techniques, Energy AI 15 (2024) 100329. https://doi.org/10.1016/j.egyai.2023.100329
[47] P. Lazzeroni, G. Lorenti, M. Repetto, A data-driven approach to predict hourly load profiles from time-of-use electricity bills, IEEE Access 11 (2023) 60501–60515. https://doi.org/10.1109/ACCESS.2023.3286020
[48] Y. Eren, I. Kucukdemiral, A comprehensive review on deep learning approaches for short-term load forecasting, Renew. Sustain. Energy Rev. 189 (PB) (2024) None. https://doi.org/10.1016/j.rser.2023.114031