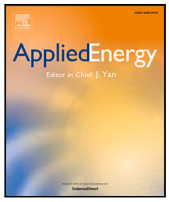
A holistic time series-based energy benchmarking framework for applications in large stocks of buildings

Applied Energy 357 (2024) 122550

Contents lists available at ScienceDirect

Applied Energy

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



A holistic time series-based energy benchmarking framework for applications in large stocks of buildings

Marco Savino Piscitelli *, Rocco Giudice, Alfonso Capozzoli

Department of Energy "Galileo Ferraris", TEBE Research Group, BAEDA Lab, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, 10129, Italy

ARTICLE INFO

Keywords:
External energy benchmarking
Building energy performance
Time series analytics
Peer identification
Key performance indicators

ABSTRACT

With the proliferation of Internet of Things (IoT) sensors and metering infrastructures in buildings, external energy benchmarking, driven by time series analytics, has assumed a pivotal role in supporting different stakeholders (e.g., policymakers, grid operators, and energy managers) who seek rapid and automated insights into building energy performance over time. This study presents a holistic and generalizable methodology to conduct external benchmarking analysis on electrical energy consumption time series of public and commercial buildings. Differently from conventional approaches that merely identify peer buildings based on their Primary Space Usage (PSU) category, this methodology takes into account distinctive features of building electrical energy consumption time series including thermal sensitivity, shape, magnitude, and introduces KPIs encompassing aspects related to the electrical load volatility, the rate of anomalous patterns, and the building operational schedule. Each KPI value is then associated with a performance score to rank the energy performance of a building according to its peers. The proposed methodology is tested using the open dataset Building Data Genome Project 2 (BDGP2) and in particular 622 buildings belonging to Office and Education category. The results highlight that, considering the performance scores built upon the set of proposed KPIs, this innovative approach significantly enhances the accuracy of the benchmarking process when it is compared with a conventional approach only based on the comparison with the buildings belonging to the same PSU. As a matter of fact, an average variation of about 14% for the calculated performance scores is observed for a testing set of buildings.

1. Introduction

The building sector is one of the top primary energy users, contributing to 40% of final energy consumption and 36% of greenhouse gas emissions in Europe [1]. As a consequence, there is a pressing need to enhance building energy performance to achieve the decarbonization targets set for 2050 [2]. Within this context, considering that the operational phase of a building life cycle accounts for at least 80% of total energy consumption [3], and with the proliferation of IoT sensors and metering infrastructure, the research is increasingly focused on the analysis of monitoring data to track and optimize the actual energy performance of buildings. The availability of data on the actual energy performance of buildings makes it possible to address the so-called energy performance gap, which refers to the difference between the intended and actual measured energy consumption, caused by several reasons, such as occupant behavior, micro-climate, and design versus as-built configurations [4,5]. To reduce such a gap, external energy benchmarking plays a key role in the identification of sub-optimal performance of a building by means of a comparison against its peers, such as buildings with the same PSU under the same boundary conditions (i.e., climatic conditions) [6]. A benchmarking process is usually used by regulators or public authorities to push owners to improve the energy efficiency of their buildings identified as poorly performing. In this sense, the identification of reference baseline, that is representative of the current/intended performance of similar buildings is at the basis of an external building energy benchmarking process [7]. In this perspective, data-driven energy benchmarking leverages the measurement of actual energy consumption data to define a reference baseline, by employing statistical approaches [8], data analytics techniques [9], and machine learning models [10].

One of the most common tasks in conventional energy benchmarking systems is the identification of reference target values or baseline models to estimate performance indicators such as the EUI for a specific PSU category as a function of influencing variables [11]. However, this approach has shown limitations, due to the fact that indicators as EUI are not always able to properly describe the causes of energy inefficiencies. The literature also demonstrated that the PSU cannot

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

https://doi.org/10.1016/j.apenergy.2023.122550
Received 4 November 2023; Received in revised form 13 December 2023; Accepted 21 December 2023
Available online 29 December 2023

Acronyms

ANN	Artificial Neural Network
AMI	Advanced Metering Infrastructures
BDGP2	Building Data Genome Project 2
CDD	Cooling Degree Days
EDM	E-Divisive with Medians
EEI	Energy Efficiency Index
EPBD	Energy Performance of Buildings Directive
EPC	Energy Performance Certificate
EUI	Energy Use Intensity
GEPIII	Great Energy Predictors III
HDD	Heating Degree Days
K-NN	K-Nearest Neighbors
KPI	Key Performance Indicator
MLR	Multiple Linear Regression
MOB	MOdel-Based recursive partitioning
MSTL	Multi Seasonal and Trend decomposition
PSU	Primary Space Usage
RLP	Reference Load Pattern
SWH	Service Water Heating
SVR	Support Vector Regression
PS	Performance Score
F	Load shape factor
LV	Load Volatility
AR	Anomaly Rate
AEC	Anomalous Energy Consumption
FR	load shape Frequency Ratio
LP	Load Profile
MAD	Median Absolute Deviation
DB	Davies Bouldin index
DR	Demand Response

be considered as the only driver to normalize the evaluation of a target KPI to benchmark buildings, while there is a need to deeply analyze the operational performance over time and identify consistent peers for extracting credible targets [12,13]. As a consequence, the concept of energy benchmarking in buildings is increasing its complexity becoming more focused on how the energy is consumed over time rather than on how much energy is consumed during a specific reference period. Nowadays, the availability of building energy-related data provides the opportunity to go beyond the use of standard KPIs investigating, by means of more sophisticated analysis, energy use inefficiency and improper or infrequent operational patterns [14,15]. To this aim, the definition of user-friendly KPIs extracted from energy consumption time series can support building owners in identifying buildings with sub-optimal performance in a more straightforward way [14]. As a result, the use of data-driven energy benchmarking systems is increasingly being developed to meet this need [16–18]. Data analytics techniques, both unsupervised and supervised, have experienced a rapid spread in the field of building energy benchmarking especially due to their enhanced capability to handle huge collections of operational data and to support the systematic extraction of helpful knowledge [19–21]. In this context, this paper introduces a novel external energy benchmarking methodology based on data-driven processes that relies on the analysis of electrical energy time series for a stock of buildings, using the open dataset Building Data Genome Project 2 (BDGP2) [22]. The work proposes a robust peer identification framework to properly identify similar buildings for the definition of a reference baseline. Moreover, a number of compact and powerful operational KPIs extracted from building energy consumption time series are introduced.

The paper is organized as follows. Sections 1.1 and 1.2 provide the research context and highlight the contribution of the work accordingly. Section 2 describes the dataset used in this work. Then, Section 3 introduces the proposed methodological framework to perform a data-driven building energy benchmarking analysis. In Section 4, all the implemented KPIs, algorithms, and statistical analyses are presented and discussed. Consequently, Section 5 reports the results of test buildings to prove the robustness of the proposed methodology while Section 6 critically discusses the outcomes and summarizes the results. Eventually, Section 7 provides the conclusions of this work and an overview of the future perspective.

1.1. Related works on data-driven energy benchmarking of buildings

In the last recent years, the topic related to building energy benchmarking has been widely discussed in the literature. Specifically, a great effort has been devoted to data-driven approaches that can be employed to address this task. Typically, two types of data-driven approaches can be found in the literature: statistical-based benchmarking and data analytics-based benchmarking.

Statistical models have been widely employed for the development of energy benchmarking systems in the building sector [19] to extract baseline models from aggregated data such as those collected from monthly bills or referred to the total seasonal or annual energy consumption. For example, the authors in [23] employed the Energy Star score method [24] to develop an energy benchmarking model based on a Multiple Linear Regression (MLR) for Malaysian hospitals, to identify parameters that mainly affect the building energy consumption. To characterize building attributes and energy performance of Brazil's non-residential buildings, the authors in [25] have conducted a top-down analysis of more than 10,000 buildings classified in 12 typologies. The authors employed a statistical analysis to assess the correlation between EUI and influencing variables, and an ANOVA test and a regression analysis to investigate the influence of energy usage indicators for each building typology. Statistical baseline models to estimate EUI have been also employed in [26], where 587 bank buildings in Turkey have been analyzed by means of a MLR model, using as explanatory variables attributes related to climate conditions and building features. In less recent years, research has been focused on the definition of general Energy Efficiency Index (EEI)s to assess in a quantitative way the energy performance at the building system level [27–29]. In fact, with the advent of Advanced Metering Infrastructures (AMI), the increased detail of monitored data allowed energy and facility managers to assess and track the energy performance over time for each energy service and sub-system. As a reference, authors in [30] have introduced a set of system-level KPIs, which cover four major end-use energy services in buildings: lighting, plug load, HVAC, and Service Water Heating (SWH). However, energy benchmarking systems that merely rely on the calculation of compact and aggregated indicators such as EUI, do not take into account changes in the operational behavior of buildings over time. Different researchers faced this issue using time series analytics techniques to extract meaningful KPIs from energy consumption data of buildings. In [31] the authors employed a novel framework based on multiple data-mining techniques on residential building load profiles to characterize occupant behavior, filtering unrelated effects, and ranking buildings in terms of achieved and potential savings thanks to the definition of a performance indicator. Authors in [14] extracted operational KPIs from time series data of office buildings and ranked them on percentile values assessing the potential energy saving for each building. As a reference, in [15] was introduced an innovative KPI that indicates the discrepancy between working hours and facility hours to estimate the impact on the prediction of EUI. Eventually, with the introduction of the recent regulatory framework (e.g., the Energy Performance of Buildings Directive (EPBD) [32]) and concepts

related to grid-interactive efficient buildings, the energy benchmarking is evolving also considering the analysis of a set of KPIs to assess the flexibility potential of buildings by measuring their capability to provide load matching/load shifting services and the reduction of energy demand during demand response events [33–35].

With regard to energy benchmarking systems based on data analysis, two main tasks have been addressed in the literature: the development of energy performance estimation models, and the extraction of reference load patterns through energy load profiling processes.

Several research works focused on the use of machine learning techniques to estimate the expected energy consumption for a building that can be compared against the actual one in order to assess its performance. To this aim supervised learning techniques, in the form of both regression and classification models, have been widely employed as reported in [36,37].

Authors in [38] employed Support Vector Regression (SVR) optimized, with different heuristic methods, to predict heating and cooling loads for both residential and commercial buildings, achieving low estimation error. On the other hand, authors in [39] used a linear regressor, over a transformed feature, space to predict the energy consumption of industrial buildings. Authors in [40] determined the most important features extracted from time series electricity consumption to predict building type, performance class, and operational strategy, by using a random forest model. Galli et al. [6] introduced an explainable framework to predict the performance class of buildings using data from Energy Performance Certificate (EPC) collections. In the field of feature importance assessment, authors in [10] introduced a holistic energy benchmarking system for the city of Singapore to predict total energy usage, finding the air conditioning floor area as the most affecting variable on energy consumption. Authors in [20] extracted quantitative and qualitative features from energy consumption time series of office buildings to analyze energy performance and predict building PSU category, magnitude of energy consumption, and type of operational strategy. A similar approach was adopted in [41], where the authors extracted features from time-series data to train an ensemble classifier that predicts PSU category of buildings. They found that 22.4% of buildings were mislabeled or used for different purposes than those declared. The authors in [42] employed clustering and random forest algorithm to decompose the interaction among the factors that affect building energy performance, identifying 36 principal factors that reliably explain building energy efficiency variations in CBES dataset [43]. On the same data-set, authors in [44], applied a MOdel-Based recursive partitioning (MOB) and identified the most influencing variables on energy consumption in buildings. Quevedo et al. [45] used synthetic data generated from parametric simulation of archetypes of university buildings to develop an energy benchmarking system that classifies buildings as efficient or inefficient, giving insights into the causes of poor performance. Geraldi et al. [46] applied an Artificial Neural Network (ANN) to several building archetype models to reduce the modeling uncertainty and predict the EUI, benchmarking their performances. What emerged as a common finding is the need for effective segmentation of buildings in groups/classes of similar peers to achieve good performance in the estimation without over-fitting problems. This aspect is particularly crucial in the definition of a robust energy benchmarking system given that the comparison of a building against a group of peers that do not share enough similarity could lead to misleading results and to the identification of not credible energy performance targets [47,48]. For what is concerned the definition of load profiling processes for benchmarking purposes, several researchers pushed towards the concept of load similarity focusing on the analysis of energy consumption patterns over a certain period of time to understand how energy is used in buildings [49,50].

Several frameworks have been introduced in the literature to analyze energy consumption data gathered from various buildings, typically with the aim of identifying homogeneous groups of energy customers that share similarities among their typical daily load patterns in terms of shape and/or magnitude [51]. In that way, according to the membership of a building to a specific group of customers, it is possible to assess its performance against its peers [52,53]. To this purpose, usually, an unsupervised clustering technique is employed to identify the most representative groups of daily load profiles among customers, while a supervised classification algorithm is used to estimate the membership of a new customer to one of the pre-identified groups [54]. The use of clustering techniques in load profiling processes was widely explored in recent years by [55–57], finding that K-means algorithm and its variations, coupled with the Euclidean distance as proximity measure, is one of the most employed configurations. Authors in [58] employed clustering to discover misfit buildings in the same PSU category, finding that around 30% of buildings were misclassified according to their load patterns. Load profiling analysis can be developed also by exploiting features extracted from energy consumption time series. As a reference, authors in [59,60] employed clustering algorithms using statistical features, peak and valley information extracted from load profiles to characterize Reference Load Pattern (RLP)s of buildings.

Although the load profiling process showed good performance in the characterization of the building's energy usage patterns for energy benchmarking, some issues emerged from the reference literature. Specifically, in most cases, load profiling is performed only on the whole building's electrical energy consumption time series, de facto making the energy benchmarking task a monovariate problem. As a consequence, the shape similarity between the electrical load profiles of buildings becomes the only driver for conducting a comparison among them. In addition, other aspects such as the load magnitude, the climatic conditions, and the presence of thermal-sensitive electrical sub-loads, were not properly taken into account by the existing literature. For example, if one building is equipped with a gas boiler as a heating system and another similar building with the same PSU is instead equipped with a heat pump, their electrical loads during the winter season may not be comparable, posing a potential risk of labeling the former as more efficient in terms of electrical load density per unit of floor area. In this sense, a good identification of peers still remains the key aspect to develop a robust benchmarking process. According to the reviewed literature, the next section introduces the main contributions of this work.

1.2. Research gap and contribution of this paper

The electrical energy consumption of a building can be challenging to be benchmarked due to differences in building characteristics such as PSU, size, layout, age, and equipment. Furthermore, the presence of thermal-sensitive loads and different occupancy patterns can also significantly affect the way the energy is used over time. Consequently, relying solely on the PSU category as a key driver for peer identification can result in partial or inaccurate benchmarking process [48,49,61]. Data analytics technologies can support the characterization of distinctive features of energy consumption time series and through the definition of KPIs is then possible to benchmark the energy behavior of buildings during operation. Different works have already covered the topic of KPI definition for energy benchmarking purpose [14,15,20,62], but more effort is needed to better generalize some of them and to introduce new ones to enable a novel energy performance comparison among buildings.

From this perspective, the main contributions of this paper can be summarized as follows:

- A clear pipeline to benchmark electrical energy consumption of a building against consistent peers based on time-series analysis is introduced. The benchmarking process automatically analyses yearly electrical energy consumption time series evaluating distinctive features and performing a dynamic and robust identification of the baseline for deploying the benchmarking process. The pipeline was conceived to be easily deployed as an automatic tool.

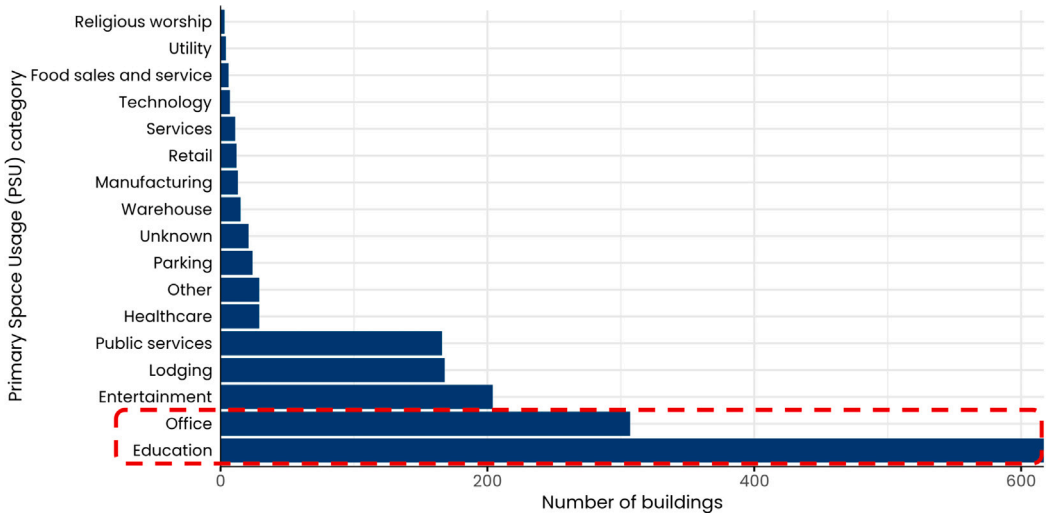


Fig. 1. Number of buildings included in the raw dataset [22,63] with evidence of the two PSU categories considered in this study.

- Development of an automatic and dynamic peer identification process to identify, in an existing stock of buildings, the group against which benchmark the performance of a new (i.e., out-of-sample) building according to its specific features and energy behavior. In the proposed framework, beyond the PSU category (e.g. office, education, residential, etc.), the peer identification process takes into account the magnitude of the electrical load, its shape, its sensitivity to the climatic conditions. The proposed process allows the analyst to identify in relation to the distinctive features of the building to be benchmarked a reference group of consistent similar peers from which extract target values for a set of KPIs.
- Definition of a set of meaningful KPIs based on time-series feature extraction processes to effectively benchmark the energy usage of a building against its peers. Once a reference group of peers is identified, a set of KPIs is introduced to rank the energy performance of a building considering aspects related to energy consumption volatility, operational schedules, and rate of potential energy anomalies associated with its operation.

In order to demonstrate the added value of the proposed methodology, the whole process has been tested on hourly electrical energy consumption data available for a stock of buildings as described in the following section.

2. Description of the dataset

The proposed methodology is validated on a subset of the BDGP2 dataset [22], i.e. the 2017 data of all the meters and sites from the Great Energy Predictors III (GEPIII) competition [63]. The employed dataset originally includes, for about 1200 buildings, the energy consumption time series related to (i) electricity, (ii) chilled water, (iii) steam, (iv) and hot water production, together with outdoor air temperature, and metadata such as the PSU, gross floor area and year of construction.

For the purposes of the study, the dataset used in this paper is a subset of the one used in the GEPIII competition which has been filtered considering only buildings with electrical energy measurements, outdoor air temperature time-series and gross floor area.

In this work, only Office and Education PSU categories are considered, which include a substantial number of buildings (i.e., 307 Office buildings and 617 Education buildings) as shown in Fig. 1. These buildings are then further filtered, according to the pre-processing analysis described in Section 3.1, to obtain the final reference dataset. Eventually, from the final reference dataset, 10 Office buildings and 10 Education buildings are randomly selected and used as test cases to validate the methodology developed, as reported in Section 5.

3. Methodology

This section introduces the methodological framework at the basis of the proposed benchmarking process. The framework is based on the analysis of yearly electrical energy consumption time series of buildings by means of data analytics and statistics with the aim to characterize their operational behavior. To this purpose, for each building included in the analyzed dataset, are extracted the following data: one year of hourly electrical energy consumption data, one year of hourly mean outdoor air temperature, the PSU category, and the gross floor area. The outcome of the benchmarking process are performance scores for a number of KPIs, which provide feedback about the energy performance of a building against a set of its similar peers.

As shown in Fig. 2, the methodological framework unfolds over three main steps: preliminary analysis, identification of peers, KPI calculation and evaluation of performance scores.

3.1. Preliminary analysis

The first step of the analysis is devoted to clean and pre-process the data. This process aims to identify and replace outliers, missing values, zeros, and continuous constant values in the time series. In particular, the Multi Seasonal and Trend decomposition (MSTL) method is employed to automatically identify statistical outliers considering the seasonality and trend of the considered time series. Details on the employed algorithm are provided in Section 4.1.1. In addition, during this step of the analysis, the pre-processed yearly energy consumption time series is further analyzed to extract features and statistically identify the so-called “ON-hours” and “OFF-hours” during workdays, without any a-priori knowledge of the actual operational schedule of the building. During a workday, “ON-hours” are those characterized by an energy consumption above a certain statistical threshold [49] during which it is possible to infer that there is activity inside the building (e.g., the building is occupied, energy systems are turned on). Specifically, for each workday included in the yearly energy consumption time series, each hour is then labeled as ON or OFF hour. Details on the employed algorithm for the labeling are provided in Section 4.1.2.

3.2. Identification of peers

The peer identification process is aimed to identify a group of similar buildings against which to compare an out-of-sample new building to be benchmarked. This identification process considers several aspects, such as:

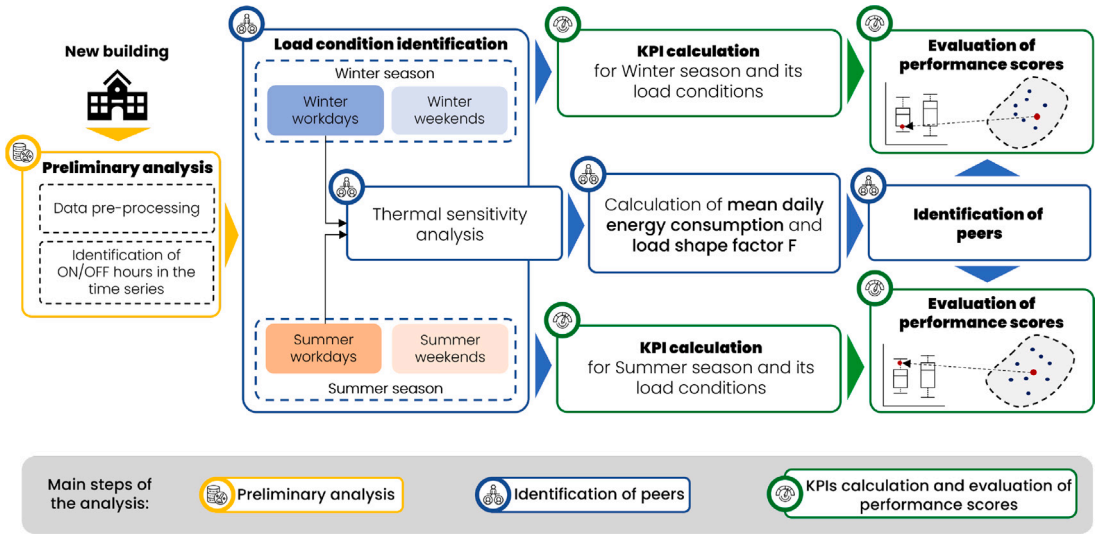


Fig. 2. Methodological framework of the energy benchmarking process.

- The PSU category of the building: when a new building is benchmarked, its peers are searched among buildings with the same PSU category;
- Reference load conditions among the yearly electrical energy consumption time series: when a new building is benchmarked its yearly energy consumption time series is firstly chunked into sub-sequences of daily length and all the obtained daily load profiles are then grouped in "Load conditions" a-priori identified with a domain-expert approach (i.e., winter workdays, winter weekends, summer workdays, summer weekends, non-working days and holidays).
- The sensitivity of the electrical load to the outdoor air temperature: when a new building is benchmarked, for some of the pre-identified load conditions (i.e., winter workdays, summer workdays) the electrical load is classified, by means of a statistical correlation analysis, as thermal sensitive or not.
- The mean daily energy consumption: when a new building is benchmarked, for some of the pre-identified load conditions (i.e., winter workdays, summer workdays) is calculated the mean daily energy consumption.
- A load shape factor F: when a new building is benchmarked, for some of the pre-identified load conditions (i.e., winter workdays, summer workdays) is calculated an indicator that is representative of the shape of daily load profiles.

In that way, a building is benchmarked, individually for each load condition, against a group of peers that share the same PSU category, that exhibit similar sensitivity of the electrical load to outdoor air temperature during winter workdays and summer workdays, and for the same two load conditions have daily load profiles that on average are similar in terms of both magnitude (i.e., mean daily energy consumption) and shape. The process of peer identification is further detailed in Section 4.2.

3.3. KPI calculation and evaluation of performance scores

After the identification of peers, to effectively benchmark the load condition of a new building it is necessary to evaluate KPIs and compare those values with a reference distribution extracted from the considered peers. The KPIs calculated in this step are derived from five different indicator categories:

- Energy Use Intensity: the ratio between the total energy consumption and the floor area. In the case of a load condition that is classified to be sensitive to the outdoor air temperature this KPI is also normalized on degree days.
- Operational schedules: this group of KPIs summarizes the impact of energy usage during weekends and workday OFF-hours against to workday ON-hours.
- Volatility of energy consumption: this KPI summarizes the degree of variability or fluctuations that characterize the daily load profiles of a building in a specific load condition [62].
- Anomalies in energy consumption: this group of KPIs is associated with the rate of anomalous daily load profiles of a building in a specific load condition.
- Load shape pattern frequency: this KPI describes the shapes of the daily load profiles of a building in a specific load condition in order to understand if those shapes are frequent or not, considering all the buildings within the same PSU.

Once the KPIs are calculated for a new building, those values are compared with the statistical distributions extracted from the peers and converted into percentile values. To return to a common convention of a Performance Score (PS) where a score of 0 is representative of low performance and a score of 100 means excellent performance [14], the following calculation is applied when needed:

$$Performance\ score\ (KPI_i) = 100 - pct\ (KPI_i) \quad (1)$$

where pct is the percentile associated to the value of the KPI i. In the case the percentile is already consistent with the 0–100 convention, then it is directly used as a performance score.

4. Methods and algorithms

According to the methodological framework above described, this section introduces the methods and algorithms used in each step of the benchmarking process.

4.1. Preliminary analysis

This subsection describes the steps followed to conduct the preliminary analysis (i.e. pre-processing of data and feature extraction).

4.1.1. Pre-processing of data

The first step of the analysis involves data cleaning and pre-processing. Specifically, extreme values related to very low or very high observations in the energy consumption time series are removed

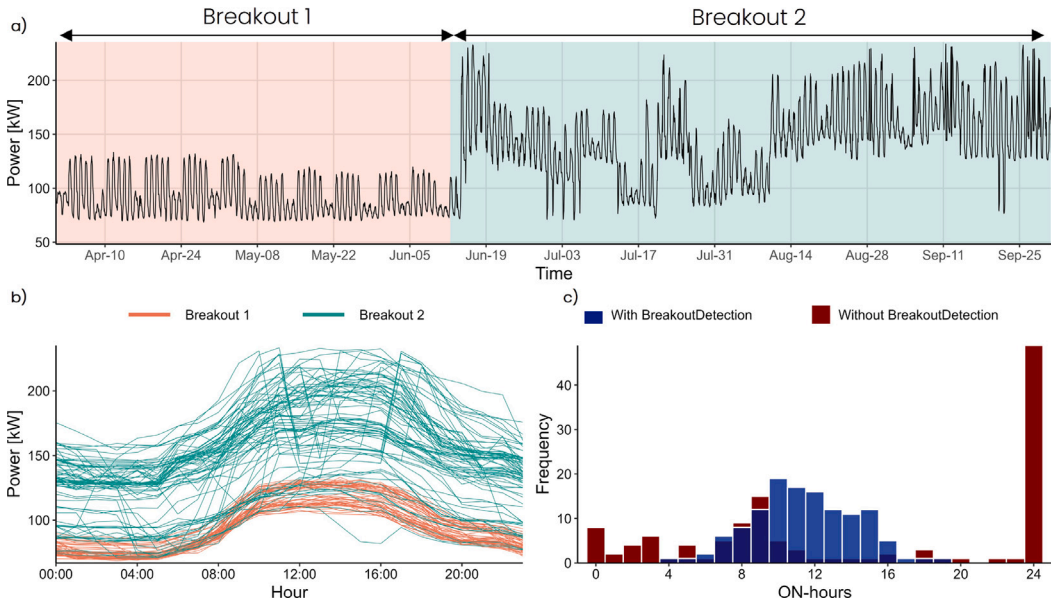


Fig. 3. Effect of breakout identification on the assessment of ON hours for the building encoded as Bear_education_Chun. (a) Identification of breakouts in the summer season. (b) Representation of summer workday load profiles according to the identified breakouts. (c) ON-hours distribution with and without the implementation of the breakout detection process.

using threshold values. The thresholds are identified by using the following Eqs. (2) and (3).

$$LT = \frac{pct_5}{2} \tag{2}$$

$$HT = pct_{95} \cdot 2 \tag{3}$$

where LT is the low value threshold, HT is the high value threshold and pct_n indicates the n-percentile of distribution of the energy consumption time series. As a second step, the MSTL decomposition technique is employed to further pre-process the building energy consumption time series, using the *forecast* package in R [64]. This technique takes into account seasonal and trend-related information to identify outliers through the decomposition of the time series into three components: seasonality, trend, and remainder. The way in which these components are extracted is reported in [64,65]. Specifically, outliers are evaluated on the remainder component of the time series, employing the interquartile range method.

Once the outliers and constant observations have been detected, they are removed from the original dataset and treated as missing values. Eventually, after the implementation of the pre-processing step, the buildings that have more than 10% of time series records corrupted (i.e. outliers, missing values, or continuous constant values) are filtered out from the analyzed dataset. For the remaining buildings, all the missing values are then replaced using linear interpolation if there are less or equal to two consecutive missing values, or with a look-up table for longer gaps in the time series of consecutive missing values. The lookup table is filled with the mean values of the energy consumption calculated per day of the season, week, and hour of the day. After this step of analysis, the total number of buildings analyzed is 622, of which 201 belong to the Office PSU category, and 421 to the Education one.

4.1.2. Feature extraction

The second step of the preliminary analysis consists in the extraction of features from the pre-processed energy consumption time series, in order to label ON and OFF hours in workdays without any a-priori knowledge of the building operational schedules. For this reason, a statistical approach is employed, identifying as ON-hours those characterized by an energy consumption above a certain statistical threshold, with an approach similar to the one introduced in [49].

In order to robustly extract energy consumption thresholds a breakout detection process is carried out, using the *BreakoutDetection* library from Twitter, an open-source R library that makes use of the E-Divisive with Medians (EDM) algorithm [66].

In particular, a breakout is a significant change observed in a time series that can consist of a mean shift or a sudden increase from one steady state to another. An example of breakout detection is shown in Fig. 3. Specifically, Fig. 3(a) graphically identifies two breakouts on a time series while in Fig. 3(b) are reported the corresponding daily load profiles.

Once the breakouts are identified, the minimum load variation threshold ΔL to distinguish ON-hours from OFF-hours is extracted from each of them as follows:

$$\Delta L = 0.25 \cdot (pct_{95} - pct_{15}) \tag{4}$$

where pct_{95} and pct_{15} are the 95^{th} and 15^{th} percentile respectively extracted from the hourly electrical energy consumption data that pertain to the workdays of an identified breakout. Then, for all the workdays belonging to the same breakout, an ON-hour is identified by means of the following Eq. (5).

$$ON - hour\ threshold = pct_{15,day} + \Delta L \tag{5}$$

where $pct_{15,day}$ is the 15^{th} percentile extracted from the hourly electrical energy consumption data of a workday included in the same breakout for which ΔL is calculated. An example of ON/OFF-hour identification is shown for three working days in Fig. 4. In addition, as a further demonstration of the positive effect of breakout analysis in the ON/OFF-hour identification, in Fig. 3(c) the daily ON-hour distribution for the analyzed time series is reported. In particular, without breakout identification (red histogram) there is a significant occurrence of days with 24 or 0 ON-hours per day differently from what happens when the breakout identification is performed. In fact the change in the mean energy consumption between breakout 1 and 2, if not properly considered, could determine the labeling of 24 ON-hours for the working days with an energy consumption above the mean (i.e., in breakout 2) and 0 ON-hours for the days below the mean (i.e., in the breakout 1). However, if the ON-hour threshold is evaluated for each breakout this effect is minimized generating a more compact distribution of daily ON-hours (i.e., blue histogram).

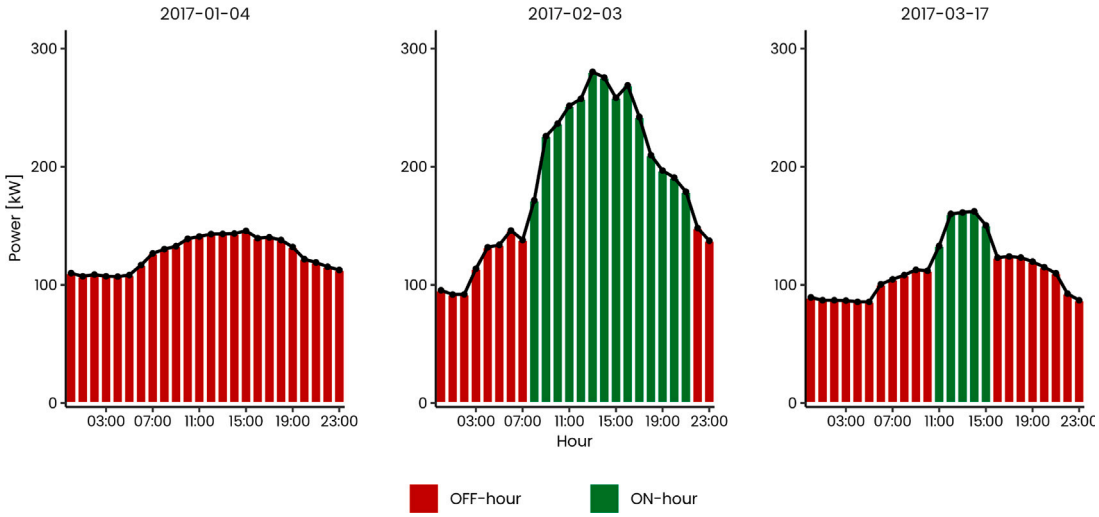


Fig. 4. Identification of ON and OFF hours in three different days for the building encoded as Wolf_education_Ursula.

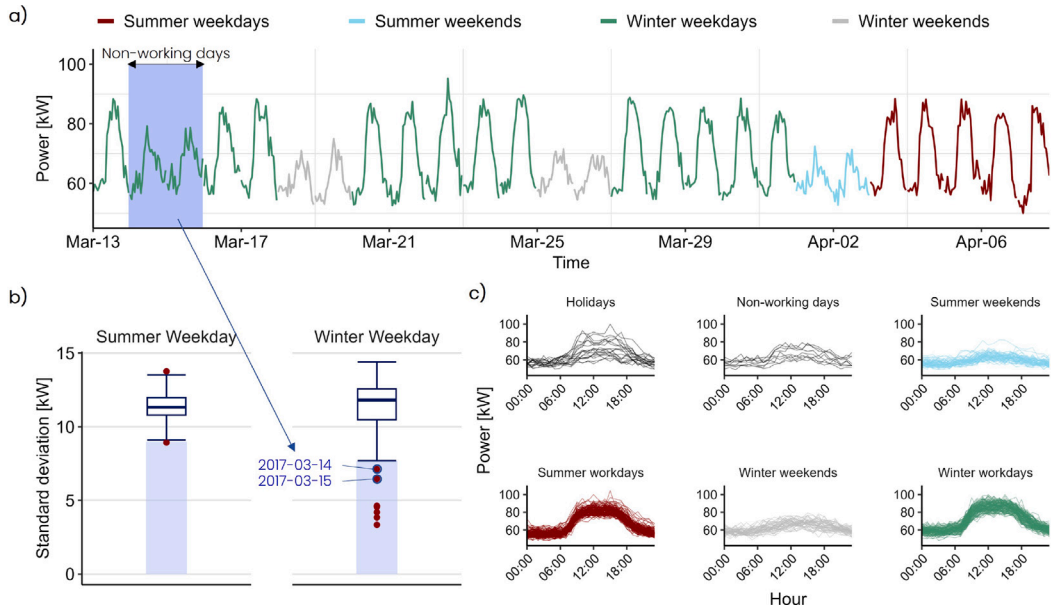


Fig. 5. Identification of load conditions on a part of the energy consumption time series referred to the building encoded as of Cockatoo_education_Brendan. (a) Preliminary identification of load conditions based on expert knowledge. (b) Identification of non-working days in the winter workdays load condition by means of statistical analysis. (c) Final results of the load condition identification process.

4.2. Identification of peers

The peer identification process is based on different steps of analysis that aim to identify and highlight similarities among buildings as reported in the following subsections.

4.2.1. Load condition identification

In order to effectively compare the energy consumption patterns among buildings, four reference load conditions are identified as shown in Fig. 5.

Specifically, from the energy consumption time series, the daily load profiles pertaining to bank holidays are filtered out according to the geographic location of each building. Then, a domain-expertise approach is used to categorize the remaining daily load profiles in four load conditions i.e., Winter workdays, Winter weekends, Summer workdays, and Summer weekends (Fig. 5(a)), where the winter season is referred to the months between October and March and the summer season to the months between April and September. On the other hand, the workdays are defined according to the conventional working week i.e., from Monday to Friday. As a further step, in the case the holiday filter is not sufficient to remove all the non-working days from the dataset, in the load conditions of Winter workdays and Summer workdays the load profiles with low variability and low energy consumption are identified, and removed following the same process reported in [54], as represented in Fig. 5(b). For the sake of completeness, Fig. 5(c) reports an example of the final result of the load condition identification process by displacing each daily load profile in the pre-determined groups.

4.2.2. Temperature sensitivity analysis

The temperature sensitivity analysis aims to assess the strength of the correlation between the electrical energy consumption of a building and the outdoor air temperature. For each building, this analysis is performed only in two load conditions i.e., Winter workdays and Summer workdays. After carrying out the analysis, those two load conditions can be labeled as “Thermal sensitive” or “Non-thermal sensitive” and as a

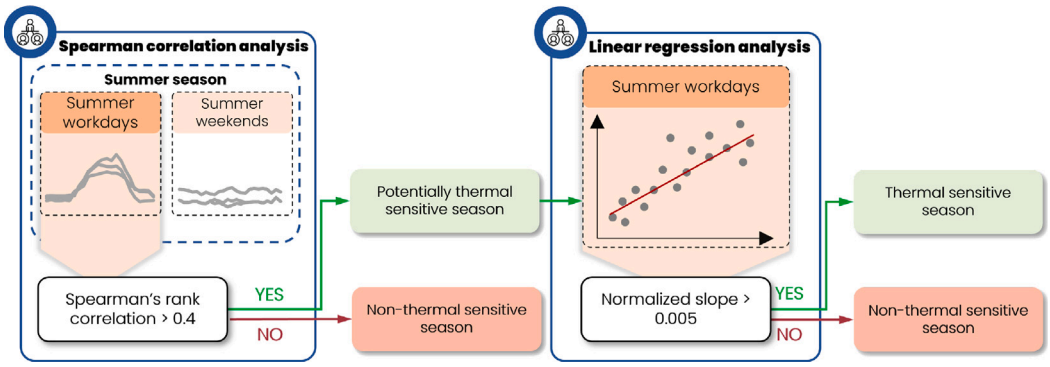


Fig. 6. Methodological framework of temperature sensitivity analysis.

consequence also the corresponding season (i.e., winter and summer season).

The analysis consists of a two-step process as shown in Fig. 6.

Firstly, the Spearman's rank correlation coefficient ρ is estimated, using the top 90 hottest or coldest workdays of the summer and winter season respectively, using the daily electrical energy consumption and the daily average outdoor air temperature. Then all the load conditions with $\rho > 0.4$ are tagged as "Potentially thermal sensitive", whereas for lower values of ρ the load condition is labeled as "Non-thermal sensitive". Secondly, for all the "Potentially thermal sensitive" load conditions, a simple linear regression analysis between the normalized daily electrical energy consumption and the average daily outdoor air temperature is performed. For the normalization of daily electricity consumption, the max value across the load condition is considered, as highlighted in Eq. (6).

$$E_{i,norm} = \frac{E_i\ [\text{kWh}]}{max\ \overrightarrow{E_{lc,n}}\ [\text{kWh}]} \tag{6}$$

where $E_{i,norm}$ is the normalized daily energy consumption of the day i, E_i is the daily energy consumption of the day i, and $\overrightarrow{E_{lc,n}}$ is the vector of the daily energy consumption for the days included in the load condition n.

The slope value (i.e., α) of the fitted regression line is then used to distinguish dubious cases in "Thermal sensitive" and "Non-thermal sensitive" conditions. The threshold value selected, employing a domain-expert approach, is $\alpha = 0.005$, which means that for "Thermal sensitive" load conditions the daily electrical energy consumption varies by at least 0.5% of the maximum daily electrical energy consumption per unit change of the average daily outdoor air temperature.

4.2.3. Identification of similar peers in the building stock

When a new building(i.e., out-of-sample) is benchmarked, its yearly time series of electrical energy consumption is segmented into the above-defined load conditions. In particular, its peers are identified separately for the summer and winter season through the analysis of Summer and Winter workdays (i.e., the peers for the load condition winter weekends are the same as the load condition winter workdays). Then, individually for winter and summer workdays, the building is compared with its peers that share similarities in the energy consumption patterns and belong to the same PSU category. In addition, the peer identification is conducted also considering the membership to the "Thermal sensitive" or "Non-thermal sensitive" class as a constraint. Once the main categories in which search the potential peers are defined, the identification of the most similar peers to the building under analysis is performed considering two different metrics. The first metric is F defined as:

$$F = \frac{E_{night}\ [\text{kWh}]}{E_{day}\ [\text{kWh}]} \tag{7}$$

where E_{night} is the amount of energy consumption in the time interval (20:00–07:00) and E_{day} is the amount of energy consumption in the interval (08:00–19:00) separately evaluated for all the days in a load condition (i.e., Winter workdays or Summer workdays). This indicator is then averaged ($\overline{F}$) among all the days in the same load condition. A value of $\overline{F}$ much lower than 1 is associated with load conditions during which buildings consume energy mainly during the daytime, while an $\overline{F}$ value greater than 1 means that the energy consumption is much more concentrated during night hours. On the other hand, the second metric is the mean daily energy consumption $\overline{E}$ calculated in the load condition analyzed (i.e., Winter workdays or Summer workdays).

In this way, both Winter and Summer workday load conditions of each building included in the analyzed dataset are identified by a tuple $(\overline{F_{lc}}, \overline{E_{lc}})$ and in order to perform the peer identification for a new building all the tuples are scaled through a min–max normalization by extracting the min and max values for both metrics from the entire sample. Then, when a new building is benchmarked, after the identification of the category of potential peers and the computation of the normalized tuple $(\overline{F_{norm,lc}}, \overline{E_{norm,lc}})$, the 30 nearest neighbors are extracted from the dataset using the Euclidean distance as a similarity measure in the geometrical space separately for Winter and Summer workdays. As a result, the set of peers identified for the winter workdays are the same for the entire winter season and could differ from those identified for the summer season, de facto increasing the flexibility of the entire benchmarking process. In particular, the calculated Euclidean distance is differently weighted among the two metrics, giving 70% of the importance to the peer similarity on $\overline{E}$ and the remaining 30% of the importance to the peer similarity on the load shape factor $\overline{F}$.

Eventually, from the identified 30 nearest neighbors is extracted a statistical distribution for a set of KPIs to assess the performance of the building under analysis. The calculated KPIs are in the following described.

4.3. Key performance indicators

In the following subsections, all the implemented KPIs are discussed and explained. In particular, some of those KPIs are evaluated for each load condition, while others are assessed for an entire season (e.g., considering Summer workdays and weekends together).

4.3.1. KPI for energy use intensity

EUI is a metric used to measure the energy efficiency of a building and it is defined as the amount of energy consumed by a building per unit of floor area per year. EUI is useful because it allows building owners, managers, and energy professionals to compare the energy performance of different buildings, immediately identifying those with lower performance. In the proposed benchmarking process the EUI is calculated in the two seasons: Winter and Summer. However, when the considered season is classified as "Thermal sensitive" the EUI is further adjusted considering a weather-related variable, in order to make buildings located in different regions and climates comparable between each other [67,68]. In Eqs. (8) and (9) is summarized the

calculation of EUI for both thermal and non-thermal sensitive seasons of a building.

$$EUI_{non-thermal\ sensitive} = \frac{E_{season}\ [\mathrm{kWh}]}{Floor\ area\ [\mathrm{m^2}]} \tag{8}$$

$$EUI_{thermal\ sensitive} = \frac{E_{season}\ [\mathrm{kWh}]}{Floor\ area\ [\mathrm{m^2}] \cdot DD\ [^{\circ}\mathrm{C}]} \tag{9}$$

where E_{season} is the total electrical energy consumption during the considered season, while DD are the degree days calculated on the same period (Heating Degree Days (HDD) or Cooling Degree Days (CDD) according to the season). Degree days are calculated as the sum of the daily differences between a reference indoor air temperature of 18.3 °C and the average daily outdoor air temperature [68].

4.3.2. KPIs for the characterization of operational schedule

Operational schedule KPIs aim to report useful insights about the use of energy over time during the day. In particular, these KPIs allow to compare the energy use during OFF-hours and weekends against the energy consumption during workdays and workday ON-hours [49]. When the ON/OFF-hours are tagged for each workday based on what is described in Section 4.1.2, two schedule-based KPIs are calculated (i.e., *OFF-impact* and *weekend impact*).

In particular, the *OFF-impact* is calculated for each workday load condition (i.e., Winter workdays and Summer workdays) while the *weekend impact* is assessed for an entire season (e.g. calculated for Winter workdays and Winter weekends) as reported in the following equations:

$$OFF-impact = \frac{E_{OFF-hours}\ [\mathrm{kWh}] - E_{ON-hours}\ [\mathrm{kWh}]}{E_{ON-hours}\ [\mathrm{kWh}]} \cdot 100 \tag{10}$$

$$weekend\ impact = \frac{E_{weekends}\ [\mathrm{kWh}] - E_{ON-hours}\ [\mathrm{kWh}]}{E_{ON-hours}\ [\mathrm{kWh}]} \cdot 100 \tag{11}$$

Specifically, the *OFF-impact* is the ratio between the difference of the energy that is consumed during workday OFF-hours and workday ON-hours with respect to that consumed during workday ON-hours, while the *weekend impact* is the ratio between the difference of energy consumed during weekends and workday ON-hours respect to the amount of energy consumed during the ON-hours of the workdays belonging to the same season. Similar indicators have been already implemented in [14], but with a different approach for the encoding ON/OFF-hours.

These two KPIs are essential to extract insights into the operation of buildings. In fact, both indicators should be negative and have the lowest value as possible, which means that the building consumes a lower amount of energy during periods when it is statistically reasonable that there is no activity inside the building. On the other hand, high *OFF-impact* values would mean that the building is characterized by a relatively high energy consumption during OFF-hours (i.e., high baseload consumption or few hours of operation). A similar reasoning can be then applied to the *weekend impact* KPI which is related to the intensity of load reduction during weekends with respect to workdays.

4.3.3. KPI for volatility of energy consumption

The assessment of volatility of energy consumption is crucial in different tasks related to building energy management such as energy benchmarking and forecasting [62,69]. Determining a KPI for energy consumption volatility for a given temporal period can effectively identify buildings that exhibit non-regular electrical loads with respect to their peers. The volatility of energy consumption generally refers to the concept of variability of the Load Profile (LP) in buildings [62] where a low volatility is associated with load profiles that share both shape and magnitude similarities under the same conditions.

In the proposed benchmarking process, the volatility KPI is evaluated for each load condition with different approaches for those belonging to thermal and non-thermal sensitive seasons.

Specifically, considering a non-thermal sensitive load condition of a building, a K-Nearest Neighbors (K-NN) algorithm is employed on each daily electrical load profile LP_i in order to retrieve its K-nearest neighbors (in terms of Euclidean distance) in the same load condition. In this study, the number of neighbors considered is 10% of the load profiles included in the load condition. Indeed for a load condition with 100 LPs, only the 10 nearest LPs to the day i are considered for the calculation of the Load Volatility (LV) referred to the day i itself (LV_i).

First, the full distance matrix M among daily load profiles LP_i in the same load condition is calculated, obtaining a symmetric matrix with zero diagonal values, where each row/column of the matrix includes the Euclidean distances of a daily electrical load profile LP_i from all the others in the load condition. For each row, only the K-minimum values of the distance are selected, representing the K-nearest profiles N_i, where K is the 10% of the load profiles included in the load condition. Load volatility is then evaluated for each day calculating LV_i, as reported in Eq. (12):

$$LV_i = \frac{mean(\vec{d_i})}{E_i} \cdot 100 \tag{12}$$

where $\vec{d_i}$ is the vector containing the distance values between each k-nearest load profile $LP_{i,k}$, with k = 1, … K, also called N_i, and the load profile of the day LP_i, and E_i is the energy consumption of the day i. To obtain a single value for the entire load condition of the building, the daily values of LV_i are averaged among all the days in the load condition obtaining then $\overline{LV}$.

A sketch of the methodology for the calculation of volatility is presented in Fig. 7.

Regarding the "Thermal sensitive" load conditions the calculation process of $\overline{LV}$ is the same, the only difference pertains to the identification of the nearest neighbors for each day in the load condition. In fact, instead of directly using the electrical daily load profiles to identify the nearest neighbors, in this case, the K-NN is applied to the daily outdoor air temperature profiles. In that way, for each day is identified the set of K days that are similar from the climatic point of view and then the corresponding daily electrical load profiles are considered, as previously explained, to calculate $\overline{LV}$.

As a final remark, the use of a number of nearest neighbors rather than the entire set of available days for calculating $\overline{LV}$, allows to consider the existence of different load patterns in the same load condition. This aspect is crucial given that the identification of the load conditions is defined a-priori and the presence of a single typical pattern is not guaranteed.

Eventually, after the evaluation of $\overline{LV}$, a low value indicates that in a load condition, there is the presence of one or more load patterns, represented by at least the 10% of load profiles that exhibit high similarities in load shape/magnitude or both of them. On the contrary, if $\overline{LV}$ value is high, the nearest neighbor profiles are distant from each other, suggesting that the load condition is characterized by a high sparsity of energy consumption patterns.

4.3.4. KPIs for anomalies in energy consumption

Following the process to quantify the LV, two KPIs are proposed to assess the presence or not of energy anomalies in a specific load condition. Such anomalies may be caused by various factors, including failure to shut down systems overnight, and anomalies in lighting, heating/cooling, or ventilation systems that result in excessive energy consumption during the day.

Once nearest neighbor load profiles are identified following the approach described in Section 4.3.3, the mean distance of each load profile with its k-nearest neighbor load profiles ($mean(\vec{d_i})$) is evaluated and stored in the vector $\vec{d}$. Then the components of the vector are subjected to 3 different statistical outlier detection methods, similarly to [70]. Specifically, inter-quartile method, Z-score method, and Median Absolute Deviation (MAD) are applied to detect upper outliers.

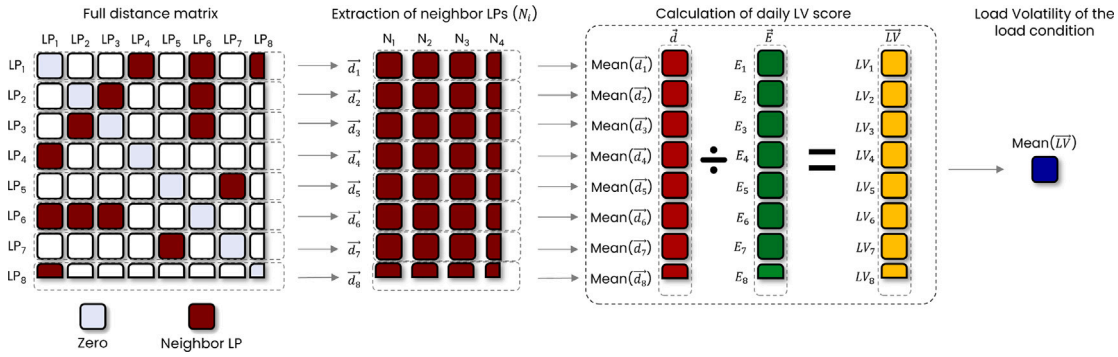


Fig. 7. Process of analysis for the evaluation of the load volatility ($\overline{LV}$) KPI in a load condition of a building.

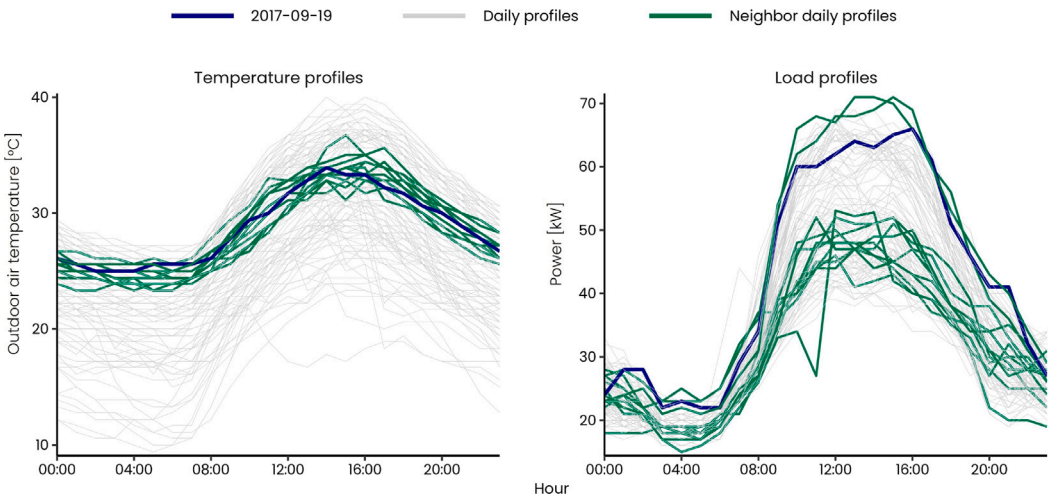


Fig. 8. Identification of neighbor profiles for the day 2017-09-19 in a thermal sensitive load condition of the building encoded as Bull_education_Brady.

Each method produces a boolean vector $B_i = \{0, 1\}$, defining whether a load profile is considered an outlier or not in the $\vec{d}$ vector. Summing all the B_i is obtained the severity score vector $S = \sum_{i=1}^{3} B_i$, in which each cell corresponds to a daily load profile and reports a score in the range (0–3), based on the number of methods that tagged that load profile as an outlier. Finally, only scores equal to three (i.e., all the methods suggest that a daily load profile is anomalous) are considered anomalies, in order to avoid false positives and spurious alerts.

Once the anomalous daily load profiles are identified, two KPIs are calculated for the whole load condition considered: the Anomaly Rate (AR) and the Anomalous Energy Consumption (AEC), reported in Eqs. (13) and (14).

$$AR = \frac{n^{\circ}\ anomalies}{N} \cdot 100 \tag{13}$$

$$AEC = \frac{\sum E_{anomaly,i}\ [\mathrm{kWh}]}{\sum E_i\ [\mathrm{kWh}]} \cdot 100 \tag{14}$$

where in the same load condition, N is the total number of daily load profiles, $n^{\circ}\ anomalies$ is the number of anomalous daily load profiles, $\sum E_{anomaly,i}$ is the total amount of energy consumed during anomalous days and $\sum E_i$ is the total amount of energy consumed in the load condition.

Considering the example of a thermal sensitive load condition, presented in Fig. 8, it is possible to notice that the day 2017-09-19 is very close to its nearest outdoor air temperature profiles. On the other hand, the corresponding daily electrical load profiles are much more sparse. As a result, the day 2017-09-19 has an energy consumption pattern that is not consistent with the identified climatic boundary conditions. Specifically, due to the high distance from its nearest neighbours, the considered daily load profile obtained an anomaly score of 3 out of 3, determining its labeling as *anomaly* in the computation process of AR and AEC.

4.3.5. KPI for load shape pattern frequency

The last KPI included in the proposed benchmarking process assesses how many daily load profiles are characterized by frequent shapes in a building load condition considering the energy behavior of the entire building stock within the same PSU category.

To this purpose, all the energy consumption time series of the buildings with the same PSU and pertaining to the four main load conditions (i.e. Winter workdays, Winter weekends, Summer workdays, and Summer weekends) are collected, chunked in daily load profiles, and normalized on their own maximum. Then, a clustering analysis employing the K-means algorithm is applied, in order to find among different buildings the most relevant groups of normalized daily load profiles. The clustering analysis is performed using the Davies Bouldin index (DB) as a quality metric, searching the optimal configuration in the range of 10–20 clusters. When the optimal number of clusters is identified, each group of normalized daily load profiles is tagged as frequent or infrequent, for each load condition. To this purpose, for each cluster, two metrics are calculated: (i) the percentage of buildings that have at least one daily load profile grouped in the considered cluster and (ii) the percentage of daily load profiles in the considered cluster respect to the total number of profiles analyzed. Specifically, a cluster is considered frequent if the percentage of buildings in the cluster is higher than 50% and the percentage of load profiles in the cluster is above the mean value, considering all clusters (i.e., a frequent cluster is representative of a high number of buildings and includes a high number of daily load profiles).

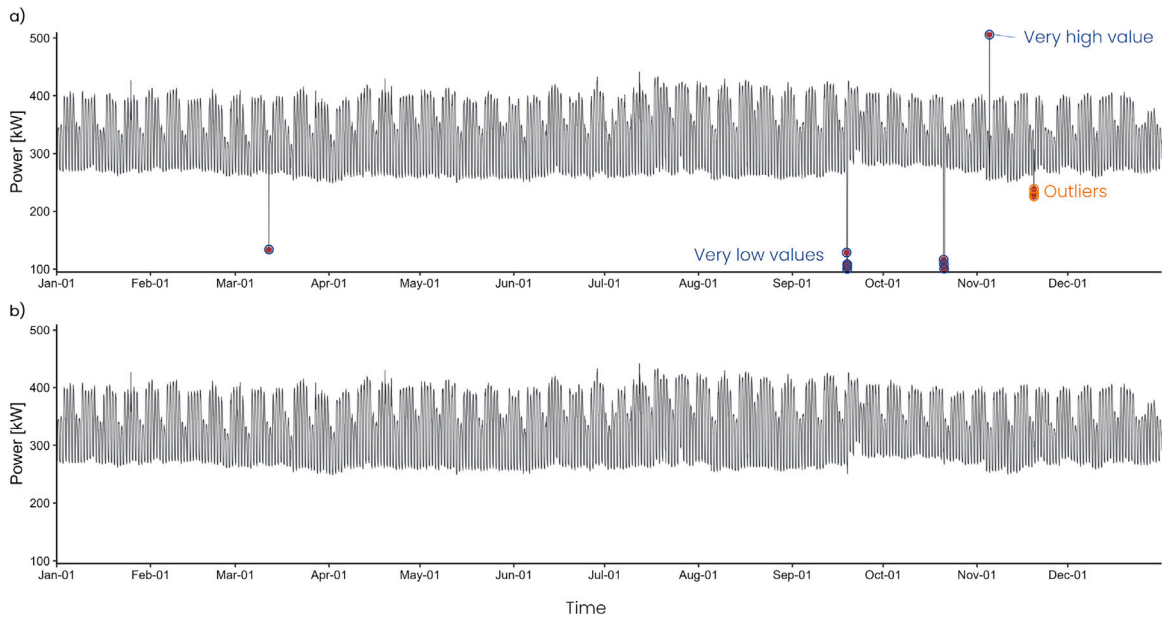


Fig. 9. Outlier detection performed on the energy consumption time series of the building encoded as Hog_office_Eloise. (a) Original time series, with evidence of the detected outliers and high/low values. (b) Cleaned time series.

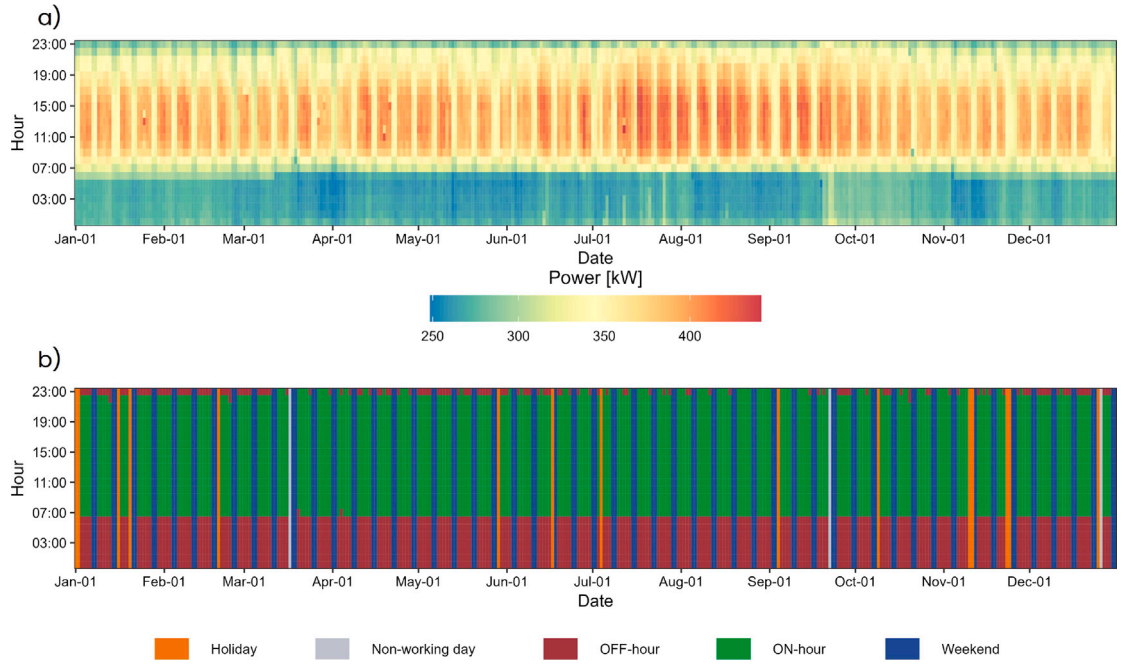


Fig. 10. Operational schedule encoding for the building Hog_office_Eloise. (a) Carpet plot of the energy consumption time series. (b) Carpet plot of the operational schedule obtained by encoding each hour as: ON-hour, OFF-hour, weekend, non-working day, or holiday.

Therefore, to assess the frequency of the load profiles of a new building in one of its load conditions the following steps are followed:

- normalization on max of the daily load profiles of a load condition;
- classification of each normalized daily load profile into the clusters identified for the considered PSU category, using the minimum distance from cluster centroid as a driver for classification;
- assessment of the load shape pattern frequency KPI load shape Frequency Ratio (FR) for the entire load condition.

Specifically, as reported in Eq. (15), FR is calculated as the ratio between the number of daily load profiles classified in frequent clusters and the total number of daily load profiles in the considered load condition.

$$FR = \frac{f}{N} \cdot 100 \tag{15}$$

where N is the number of daily load profiles in a load condition and f the number of daily load profiles classified in frequent clusters. So, if a load condition has a high value of FR means that the building has a frequent behavior (e.g., daily load shape) in that period of time with respect to all the other buildings with the same PSU category.

5. Results

The methodological process described in Section 3 is tested on 10 Office buildings and 10 Education buildings from the BDGP2 dataset

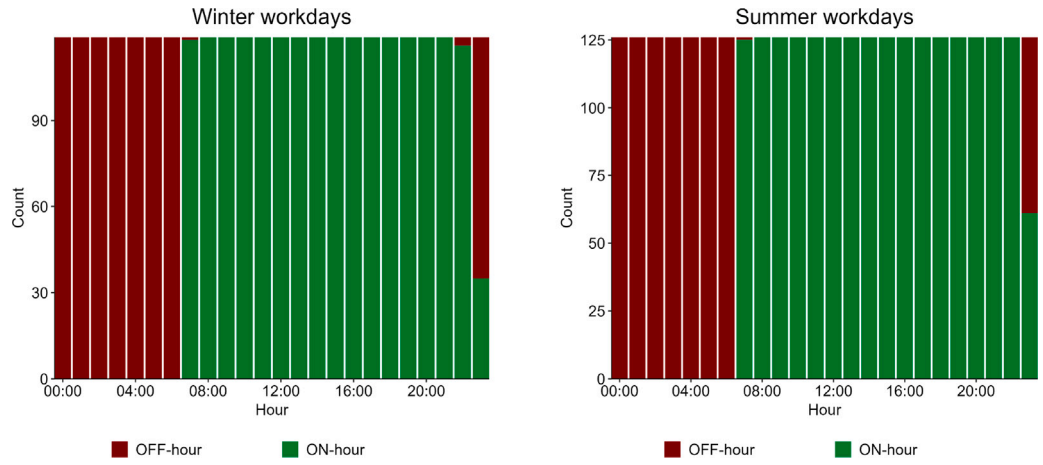


Fig. 11. Operational schedule results for two load conditions of the building encoded as Hog_office_Eloise.

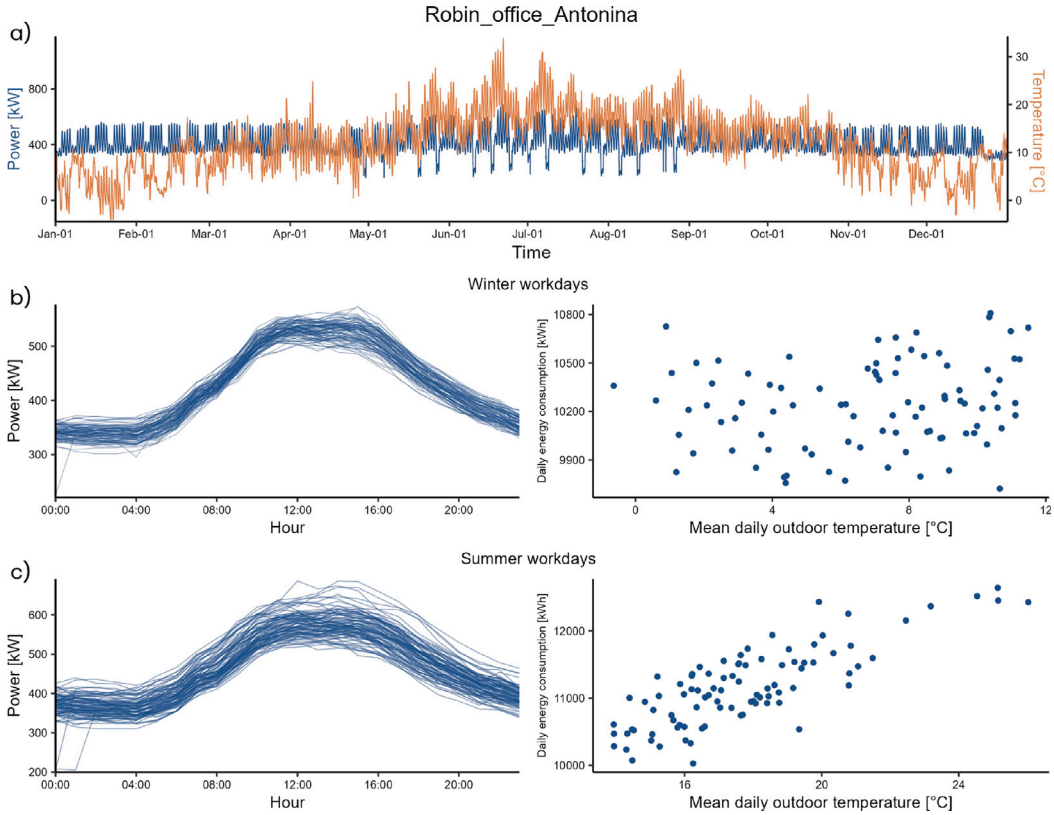


Fig. 12. Results of the thermal sensitivity analysis for the building encoded as Robin_office_Antonina. (a) Outdoor air temperature and energy consumption time series. (b) On the left are represented the daily load profiles for two load conditions, while on the right the scatterplots that put in relation the daily energy consumption with the mean daily outdoor temperature.

[22]. The statistical language R [71] and Python [72] are employed to implement the flow of analysis above described.

5.1. Data pre-processing and feature extraction results

The pre-processing step outlined in Section 3.1 is applied to the buildings tested, and the results for the building encoded as Hog_office_Eloise are shown in Fig. 9. Specifically, 22 outliers (i.e., 0.3% out of the total number of data points included in the time-series) were detected and consistently replaced according to the methodological process previously described.

After the pre-processing step, the feature extraction analysis is performed to automatically label ON/OFF hours in workdays without any a-priori knowledge of the building operational schedules. The results for Hog_office_Eloise are shown in Fig. 10.

In particular, Fig. 10(a) shows the heat map associated with the original electrical power time series while in Fig. 10(b) each hour is encoded as ON-hour, OFF-hour, weekend, or holiday. Comparing

Table 1
Results of the thermal sensitivity analysis for the buildings encoded as Robin_office_Antonina and Panther_education_Vincent.

Building_id	Load condition	Spearman ρ	Slope α	Result
Robin_office_Antonina	Winter workdays	0.17	–	Non-thermal sensitive
Robin_office_Antonina	Summer workdays	0.75	0.011	Thermal sensitive
Panther_education_Vincent	Winter workdays	−0.15	–	Non-thermal sensitive
Panther_education_Vincent	Summer workdays	0.04	–	Non-thermal sensitive

the two figures it emerges how the identified operational schedule is consistent with the actual energy consumption pattern of the analyzed building.

Furthermore, Fig. 11 shows the extraction of the operational schedule of the building employing the feature extraction process detailed in Section 3.1 for Winter workdays and Summer workdays load conditions. Each histogram represents the count of hours of the day in each load condition, filling the bar as OFF-hours (red) and ON-hours (green). As can be observed, in both load conditions, the building is tagged as ON in the time interval (07:00–22:00), while is OFF in (23:00–06:00). Only in summer workdays, almost half of the days are ON also at (23:00–00:00).

5.2. Peer identification results

The peer identification process, as outlined in Section 3.2, aims to identify the most suitable set of buildings, in the same PSU category, that exhibit a high degree of similarity in terms of energy consumption behavior with a new building to be benchmarked.

The peer identification process starts with the identification of the four load conditions (Winter workdays, Winter weekends, Summer workdays, and Summer weekend) using the methodology detailed in Section 4.2.1 and Fig. 5.

After the load condition segmentation, the thermal sensitivity analysis is performed on Winter workdays and Summer workdays to label them (and their corresponding season) as "Thermal sensitive" or "Non-thermal sensitive".

Figs. 12 and 13 show the relation that exists between electrical energy consumption and outdoor air temperature for two different buildings in the testing set. In Figs. 12(a) and 13(a) both variables (i.e., electrical energy consumption and outdoor air temperature) are represented as time series, while in Figs. 12(b)–(c) and 13(b)–(c) are reported the daily load profiles referred to winter and summer workdays together with the scatter plots of the daily electrical energy consumption against the mean daily outdoor air temperature.

As previously discussed the first step is to assess the thermal sensitivity of the electrical load by calculating the Spearman's correlation coefficient ρ between daily energy consumption and mean daily outdoor temperature in both Winter and Summer workdays. As reported in Table 1, for Panther_education_Vincent both load conditions are labeled as "Non-thermal sensitive" (as can be easily observed from Fig. 13(b)–(c)) given that the correlation coefficient ρ does not exceed the threshold fixed at $|0.4|$. On the contrary for Robin_office_Antonina, the Summer workday load condition exhibits a value of ρ higher than the threshold, meaning that it may be potentially labeled as "Thermal sensitive". Then a linear regression model is fitted using the normalized daily energy consumption as the output variable and the mean daily outdoor air temperature as the input variable. Extrapolating the slope value α, it results to be greater than the fixed threshold of 0.005, meaning that the load condition (and then the season which includes it) can be eventually labeled as "Thermal sensitive". Indeed for the considered buildings, all the analyzed seasons are characterized by electrical energy consumption that can be considered independent from the weather conditions with the exception of the summer season of Robin_office_Antonina, during which the electrical load varies according to the outdoor air temperature. As a reference, this can be related to the presence of chiller systems in the building that may have a significant impact on its total electrical demand during summer.

The importance of differentiating "Thermal sensitive" from "Non-thermal sensitive" buildings is demonstrated by the obtained results from the analysis of the considered dataset. In fact, by following the thermal sensitivity analysis, in the Education and Office PSU categories, it was found that 43% and 51% of buildings respectively have at least one load condition whose electrical energy consumption can be considered as significantly affected by variations of outdoor air temperature.

After this step of analysis, peers are searched considering the same PSU category and thermal sensitivity by using as drivers two different metrics: the mean daily energy consumption and the load shape factor F, both separately calculated for Winter and Summer workdays, as already detailed in Section 4.2.3.

Specifically the load shape factor F, is useful to distinguish buildings that may use a similar amount of energy but with a different daily shape pattern. As a reference, Fig. 14 shows the normalized daily load profiles of two buildings referred to the Winter workdays load condition. For both buildings is highlighted the average load profile (i.e., solid red line) and the average F. It can be observed, that the building with a higher load shape factor $F = 1,34$ (i.e., Fox Education Melvin) mainly consumes energy during night hours, and given that it is labeled as an education building, it is maybe used as a dormitory. On the other hand, the building with a lower $F = 0,55$ (i.e., Cockatoo Education Arlen) is characterized by a more conventional daily energy pattern due to daytime activities. It is then clear how important is to consider such differences in the energy benchmarking process, given that each building has its own features that could make some comparisons not so consistent if not well reflected in the group of peers.

The identification of peers, for the two buildings analyzed, is then shown in Figs. 15 and 16. Each blue point in the plot corresponds to a building in a specific PSU category, characterized by the same thermal sensitivity, and represented by its own values of mean daily energy consumption and shape factor F evaluated in a load condition. The orange dots are the buildings to be benchmarked while the green points are their 30-nearest peers, identified by computing a weighted Euclidean distance as specified in Section 4.2. In this way, the following evaluation of KPIs can be considered robust and consistent with building PSU category, load condition, thermal sensitivity, magnitude and shape of energy consumption patterns.

5.3. Key performance indicators results

In this section are presented and discussed the results related to the evaluation of different KPIs defined as in Section 4.3. In particular, after the identification of peers, it is possible to extract reference distributions of KPIs that are representative of a group of buildings that share similarities with the one that is the subject of the benchmarking process. Eventually, by means of statistical analysis (i.e., percentile value) it is possible to assess for each KPI, a final performance score for the building of interest. In the following, the results pertaining some test buildings are discussed for each proposed KPI.

5.3.1. Energy use intensity

Fig. 17 shows the results pertaining to the benchmarking of EUI for the test building labeled as Bull_education_Roseann. It is possible to observe that the EUI distributions of peers (i.e., green areas) and the distributions of the whole PSU category (i.e., Education buildings)

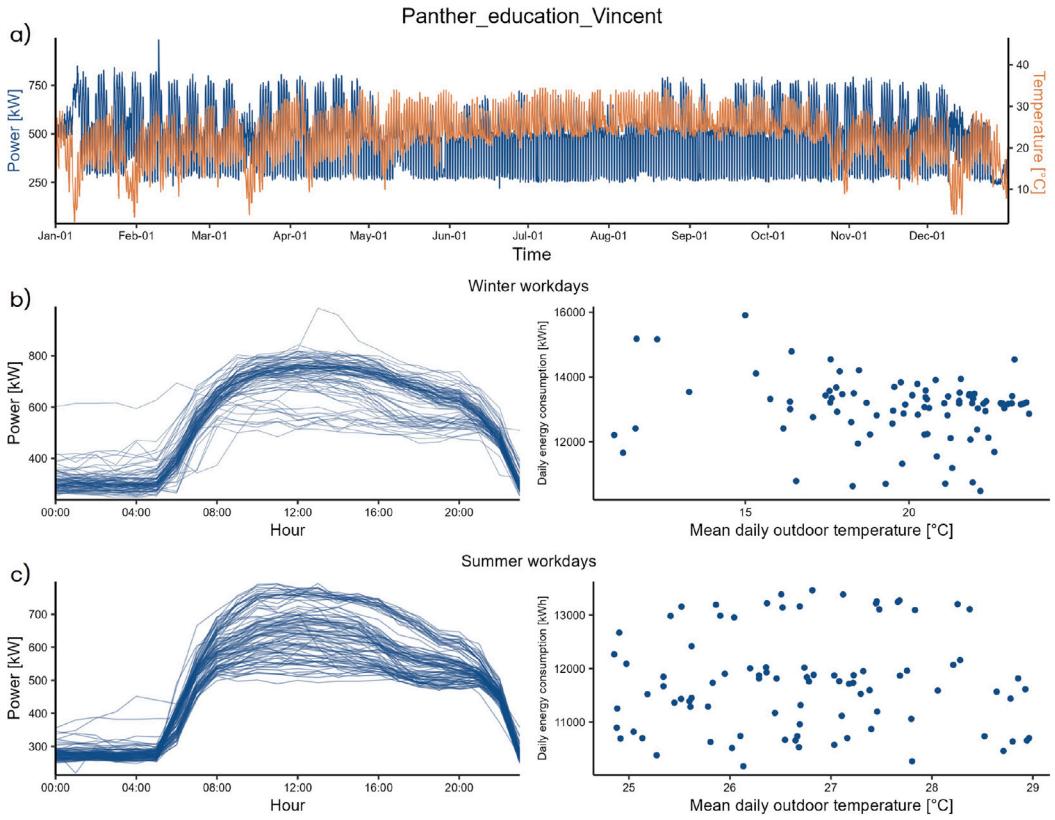


Fig. 13. Results of the thermal sensitivity analysis for the building encoded as Panther_education_Vincent. (a) Outdoor air temperature and energy consumption time series. (b) On the left are represented the daily load profiles for two load conditions, while on the right the scatterplots that put in relation the daily energy consumption with the mean daily outdoor temperature.

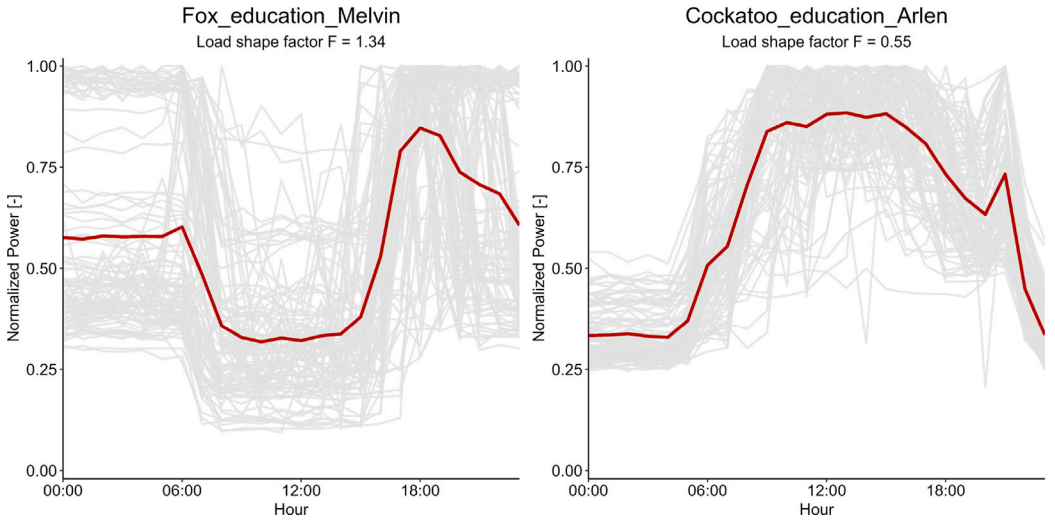


Fig. 14. Normalized daily load profiles referred to the load condition Winter Workdays for two buildings which have different F. In red is highlighted the average profile.

show significant differences in both summer and winter workday load conditions. As a reference, Bull_education_Roseann is characterized by an EUI equal to 111 [kWh/m^2] during summer, which is included between the median and third quartile values considering the peers distribution. It means that the considered building has an EUI value higher than 50% of its peers but it is not too critical. However, if the distribution of the whole PSU category is considered, the benchmarked building performs worse than about the 75% of the entire group, determining a final score on this KPI of about 25 out of 100. The opportunity to rank the building against its peers allowed then to better take into account all the distinctive features that characterize its energy consumption and assess its potential uniqueness in the reference set of buildings. In the previously reported case, it is possible to say that despite Bull_education_Roseann has a high EUI value, it is consistent with the group of its peers.

5.3.2. *Operational schedules*

The second group of KPIs analyzed, pertain with the building operational schedules. Specifically, as outlined in Section 4.3, during the preliminary analysis of the energy consumption time series, ON/OFF hours are automatically evaluated for winter and summer

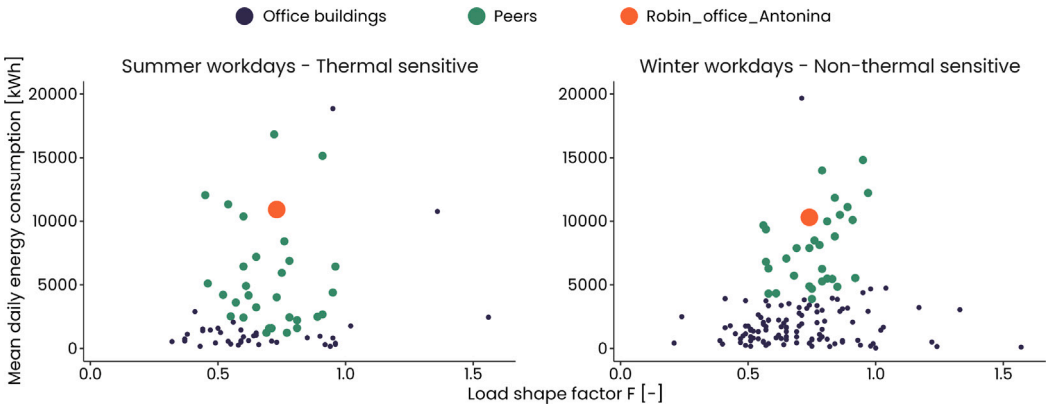


Fig. 15. Identification of peers for the building encoded as Robin_office_Antonina.

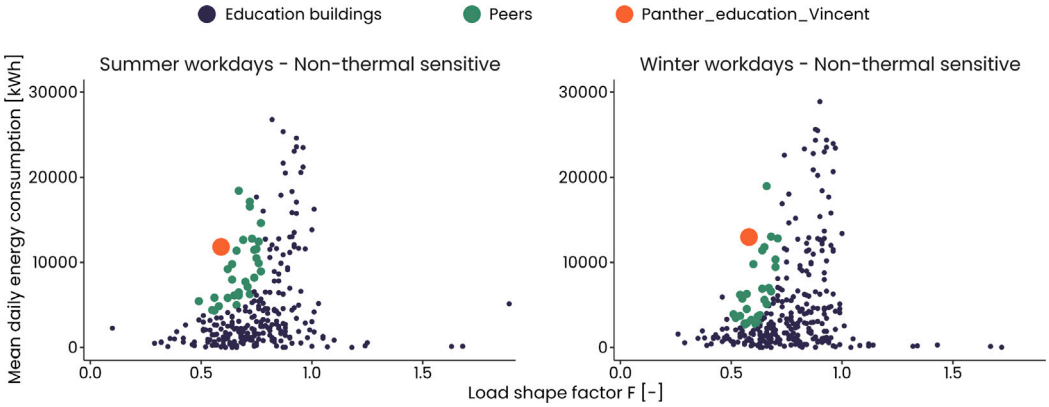


Fig. 16. Identification of peers for the building encoded as Panther_education_Vincent.

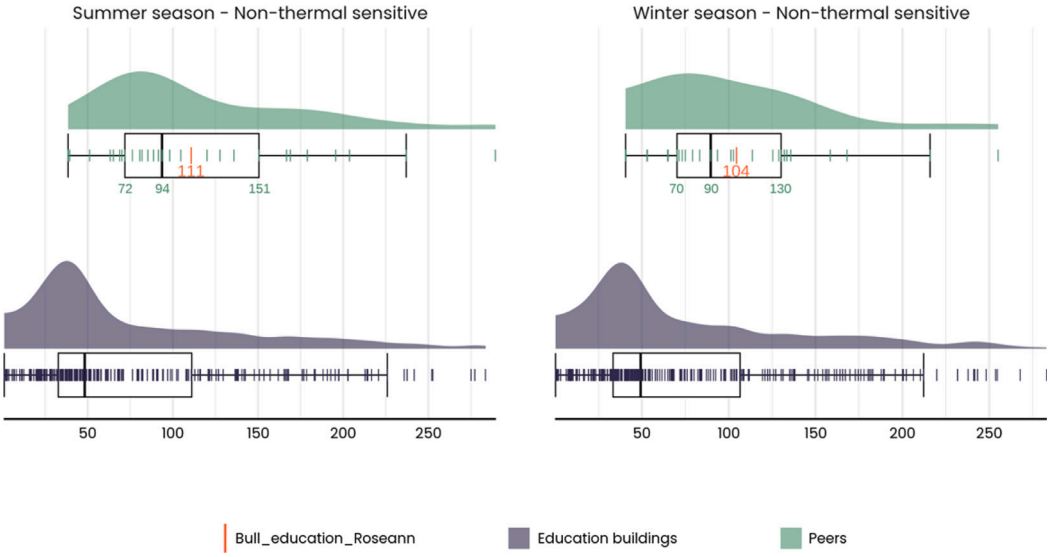


Fig. 17. Benchmarking results for the EUI of the building encoded as Bull_education_Roseann (orange line) in both summer and winter season. In green is shown the EUI distribution of the peers identified for Bull_education_Roseann, while in blue is shown the EUI distribution of the entire education PSU category.

workdays. Then for both the summer and winter season, the OFF-impact and weekend impact indicators are calculated. Fig. 18 shows the ON/OFF/weekend-hours encoding for a winter, (Fig. 18(a)) and a summer week, (Fig. 18(b)), for the building encoded as Fox_office_Sheila. It can be observed that in the summer week analyzed, the energy used during ON-hours is more than that used during the winter week, due to the higher peak of the energy consumption. Also the summer weekend analyzed has an energy consumption higher than the winter weekend. However, in relative terms, the ratio between the energy used during OFF-hours and that used during ON-hours is higher in the winter week, contributing to achieving an higher value for the OFF-impact KPI.

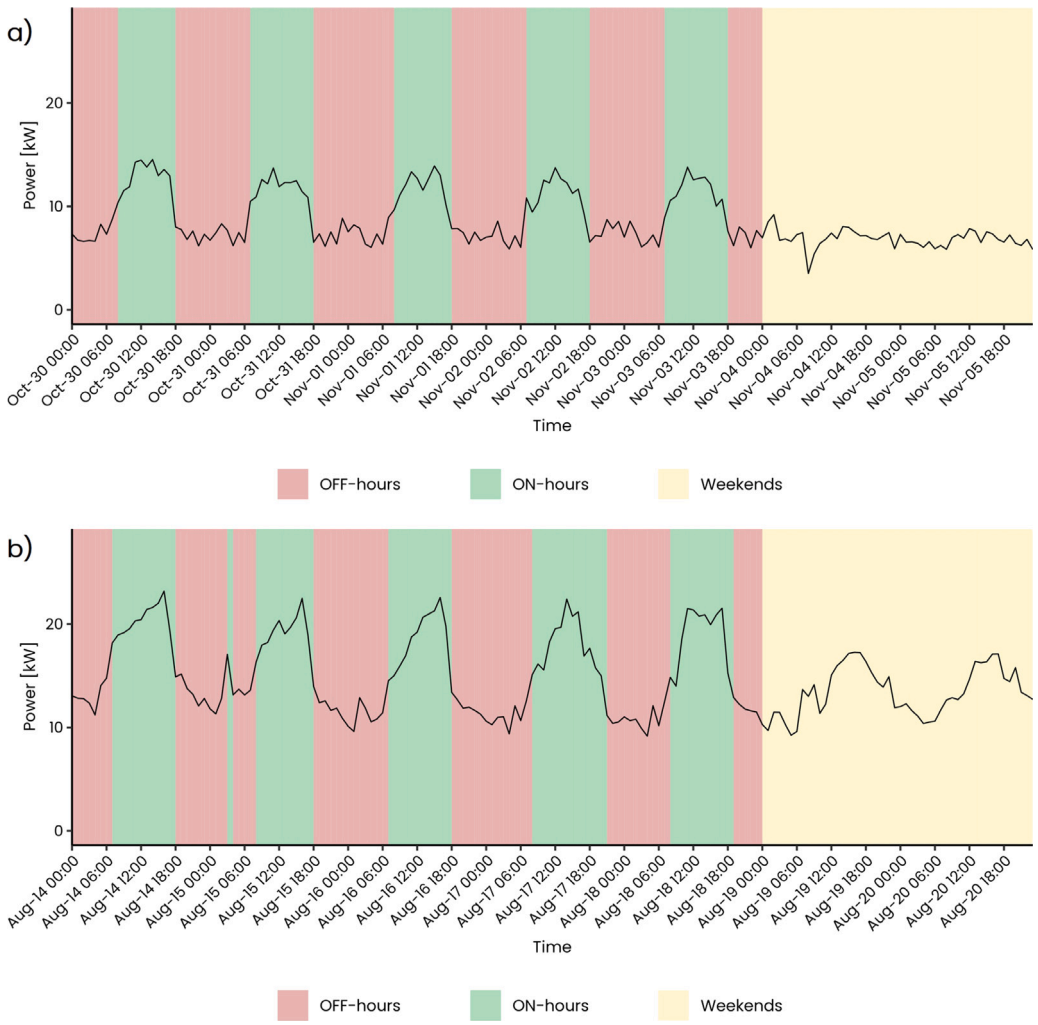


Fig. 18. Operational schedule extraction from the energy consumption time series of the building encoded as Fox_office_Sheila in a summer and winter week. (a) ON and OFF hour encoding for the week 2017/10/30–2017/11/05. (b) ON and OFF hour encoding for the week 2017/08/14–2017/08/20.

In particular, the benchmarking results for Hog_office_Sheila are shown in Fig. 19. It can be observed that the building performs poorly considering both KPIs and for both seasons compared to its peers.

5.3.3. Volatility of energy consumption

The load volatility LV gives a compact information about the variability of the daily energy consumption patterns in the same load condition. If a load condition is characterized by high volatility means that, on average, each load profile, included in it, is far away from the set of its closest neighbors. This information is then useful to understand how much repetitive and frequent are the daily energy patterns of a building in terms of both shape and magnitude during specific periods of the year. This feature has been discussed in the literature as an important driver for the definition of load prediction and anomaly detection processes in buildings [62,69]. As a reference, Fig. 20 shows the results associated with the calculation of the LV for the test building labeled as Bear_education_Paola. Specifically, in Fig. 20(a) are displayed the daily load profiles for the load condition winter workdays. At first glance, the daily profiles seem to be widely spread, however, such potential high variability needs to be better investigated considering that the load condition was labeled as "Thermal sensitive". In this case, high volatility exists if compared to the one of a "not-thermal sensitive" building, but it is of interest to understand if, when a certain outdoor air temperature pattern occurs, the corresponding electrical load profiles are or not close to each other. For this purpose, in Fig. 20(b) is displayed the distance matrix reporting the euclidean distances between the daily load profiles included in the load condition and their closest neighbors (i.e., which number is fixed to the 10% of the load profiles). In the analyzed case, the closest neighbors are identified according to the outdoor air temperature daily profiles (due to thermal sensitivity) while the distances are a-posteriori computed on the electrical load values. Then, for each day is obtained the load volatility LV_i reported on the bar plot, and by averaging all the LV_i the $\overline{LV}$ is assessed for the whole load condition (red dashed line). For the load condition Winter workdays of the building Bear_education_Paola, $\overline{LV}$ is equal to 3.82%. It means that on average the daily load profiles of this building are far from their closest neighbors for an amount of energy of about 4% of their daily energy consumption. Considering the $\overline{LV}$ distribution of the peers and whole PSU category Fig. 20(c), it is possible to observe that the building Bear_education_Paola is characterized by low volatility in its energy consumption during the load condition, and more specifically that the volatility is fully consistent with the weather conditions (i.e., low sparsity of daily load profiles during similar boundary conditions).

5.3.4. Anomalies in energy consumption

Following the same procedure implemented for assessing load volatility in a load condition, it is also possible to identify a number

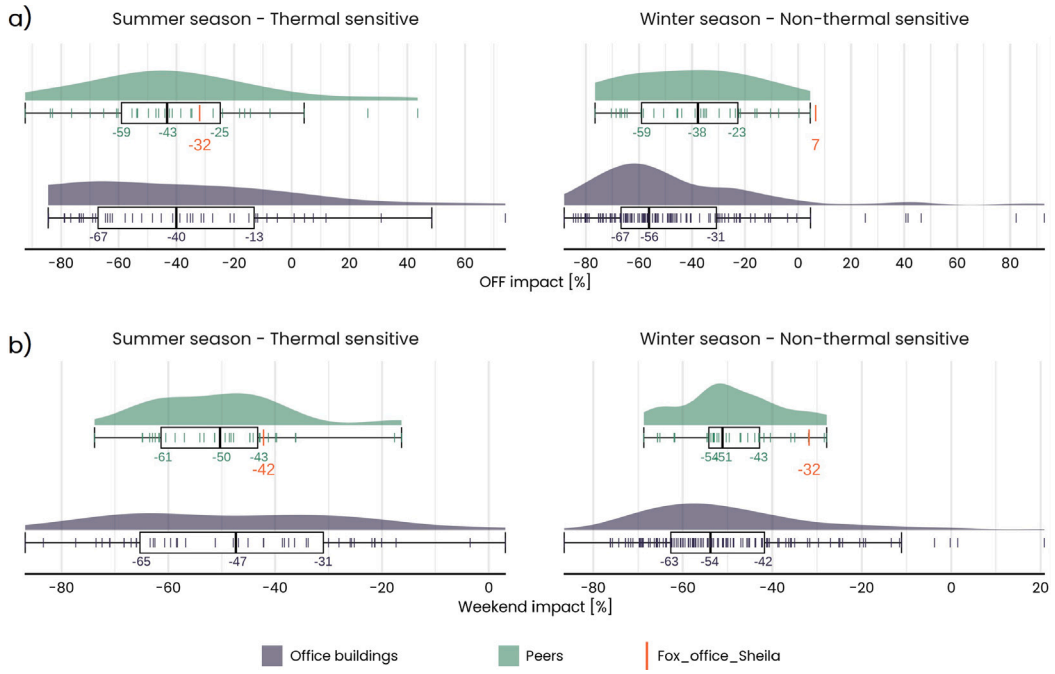


Fig. 19. Benchmarking results for the OFF-impact KPI (a) and the Weekend-impact KPI of the building encoded as Fox_office_Sheila.

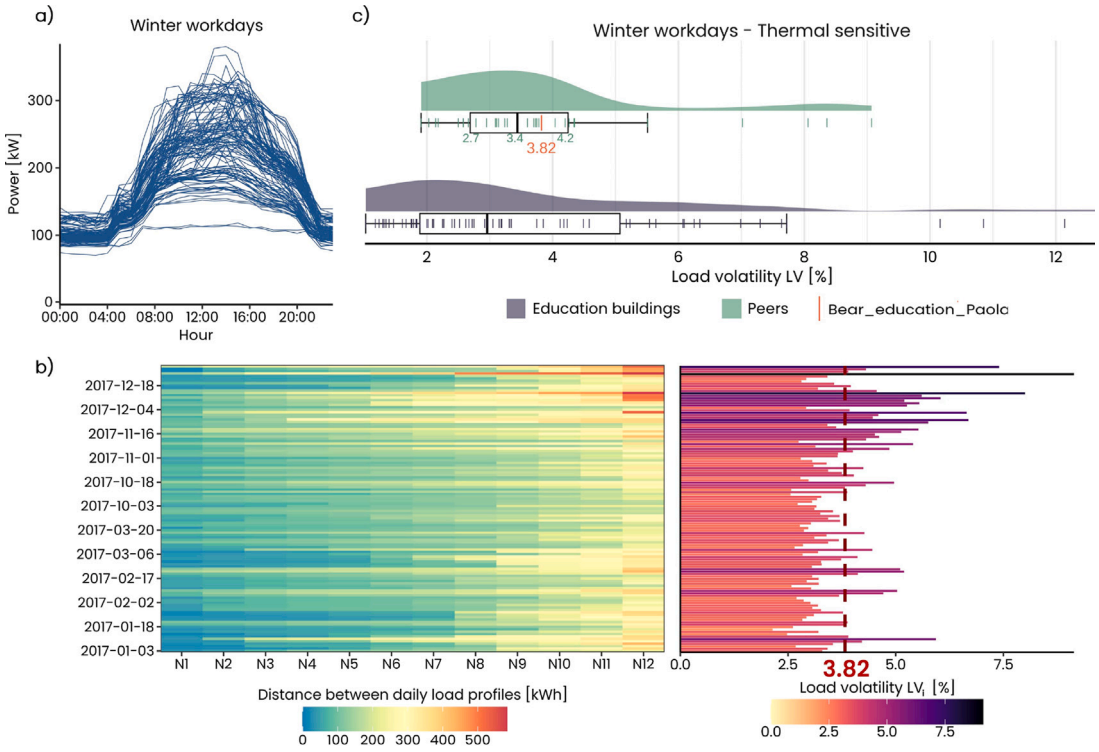


Fig. 20. Load volatility assessment for the building Bear_education_Paola in winter workdays load condition. (a) Daily load profiles referred to the winter workday load condition. (b) Calculation of the load volatility from the distance matrix. (c) Benchmark results for the load volatility KPI considering the distribution of both peers and entire education PSU category.

of potential anomalous load profiles that are characterized by a high Euclidean distance from their closest neighbors. Specifically, before averaging all the LV_i to calculate the total volatility of the load condition, three statistical methods (i.e., inter-quartile method, Z-score method, and MAD) are implemented to detect upper outliers on the distance vector $\vec{d}$. As a reference, Fig. 21(a) shows in red the daily load profiles detected as anomalous (i.e., with a value of d_i that is out of range) for the building labeled as Fox_education_Shirley in the load condition Winter workdays that is tagged as Thermal sensitive. These profiles exhibit significant dissimilarity in relation to other daily load profiles associated with the most similar conditions in terms of outdoor air temperature patterns. In particular for the day 2017-12-28 Fig. 21(b) shows in red its daily outdoor air temperature profile and in blue its closest neighbors. On the other hand in Fig. 21(c) are

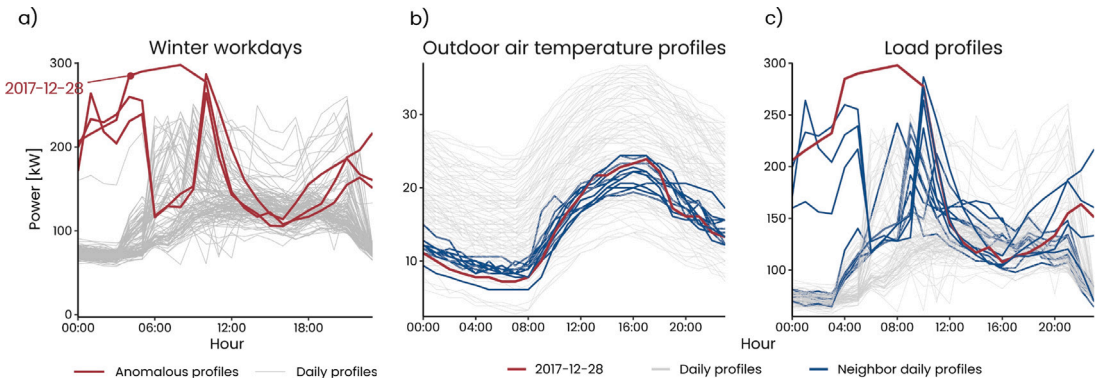


Fig. 21. Identification of anomalies in the winter workday load condition for the building encoded as Fox_education_Shirley. (a) Representation of load profiles in the load condition with evidence of those encoded as anomalous (solid red lines). (b) Representation of the outdoor air temperature profile for the day 2017-12-28 (solid red line) and its neighbor profiles (solid blue lines). (c) Representation of load profiles corresponding to the identified outdoor air temperature profiles.

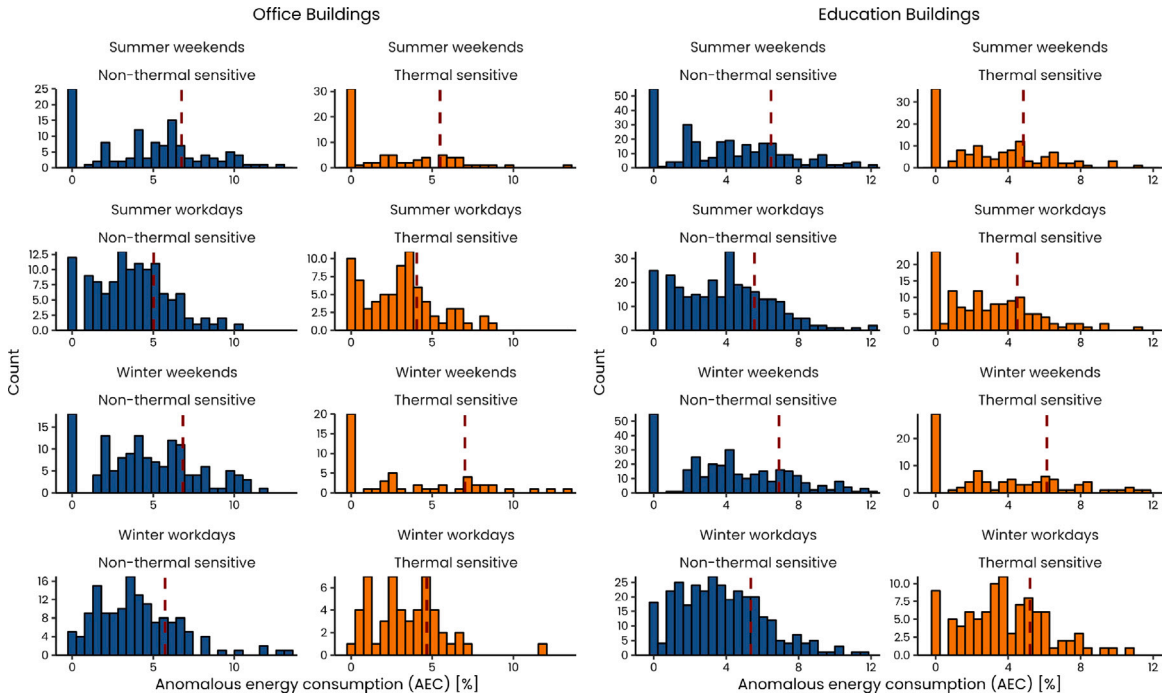


Fig. 22. Distribution of the AEC among all the load conditions for the buildings included in the Education and Office PSU category. For each distribution, with a dashed red line is highlighted the 75^{th} percentile.

reported the corresponding electrical load profiles, which show a great variability despite they are subjected to the same boundary conditions (i.e., similar outdoor air temperature profiles).

Once the potential anomalies are detected, for each load condition, the two KPIs AR and AEC are evaluated. While the first KPI reports the percentage of anomalous daily load profiles in the load condition, the second one assesses the corresponding amount of energy consumed during an anomalous day out the total energy consumption of the load condition as a percentage.

Fig. 22 presents the results obtained for the AEC considering all the buildings in both PSU categories i.e., Office and Education buildings. The results are reported in terms of frequency distributions, separately for each thermal and non-thermal load condition. It is possible to observe that, in most of the cases, there is a high frequency of buildings characterized by a value of AEC close to 0%. It means that no anomalous daily load profile is detected under that load condition. Conversely, a number of buildings are characterized by a higher value of the AEC close to 8%–10%, meaning that the energy consumption related to the detected anomalous daily load profiles is significant with respect to the total electricity consumption of the load condition. In Table 2 the 75^{th} percentile of each distribution included in Fig. 22 is reported and it can be observed that in each load condition, this value falls within the range from 4% to 7%.

5.3.5. Load shape pattern frequency

The last KPI included in the energy benchmarking process aims at understanding how frequent are the shapes of the daily load profiles in a load condition with respect to the entire PSU category.

The results of the clustering analysis and of the frequency of load profile shape analysis are reported in Fig. 23 for the Office buildings.

For what concerns the number of clusters identified through the DB index, the optimal solution is, respectively, 13 for the Office category (Fig. 23(a)) and 10 for the Education one. After the identification of clusters, they are labeled as frequent or infrequent. Specifically, for each cluster is assessed the percentage of included daily load profiles out the total number of profiles in the PSU category, and the corresponding percentage of buildings that have at least one load profile grouped within the cluster itself, as shown in Fig. 23(b). In

Table 2
75^{th} percentile of the Anomalous Energy Consumption KPI extracted from each load condition and for both education and office buildings.

	Winter workdays	Winter weekends	Summer workdays	Summer weekends
AEC 75^{th} percentile - Education buildings				
Non-thermal sensitive	5.4%	6.9%	5.6%	6.5%
Thermal sensitive	5.2%	6.1%	4.5%	4.8%
AEC 75^{th} percentile - Office buildings				
Non-thermal sensitive	5.7%	6.8%	5.0%	6.8%
Thermal sensitive	4.7%	7.0%	4.0%	5.5%

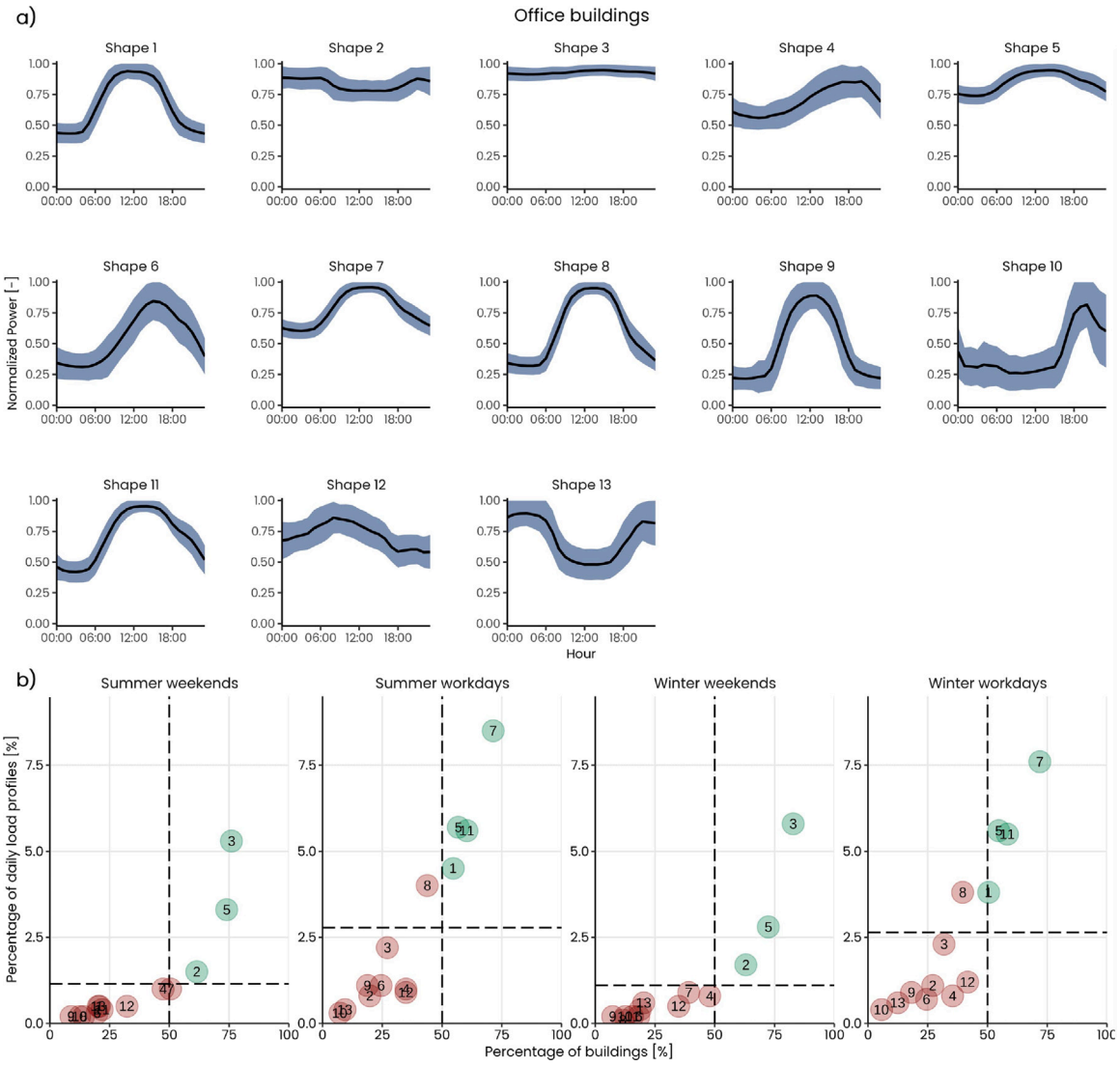


Fig. 23. Clustering results for office PSU category. (a) The solid black lines are the centroids obtained from the cluster analysis, while the blue areas are the one-standard deviation intervals around the centroids. (b) Frequency analysis results in each load condition to label clusters as frequent (green bubbles) or infrequent (red bubbles).

this way, it was possible to label as frequent only the clusters that have both percentage values above the mean, for each load condition.

When a new out-of-sample building is benchmarked, all the daily load profiles of a load condition are compared against to the identified cluster centroids, and each of them is classified in the cluster of its closest centroid. As a result, for the analyzed building it is possible to calculate the FR indicator, intended as the ratio between the number of daily load profiles classified in frequent clusters and the total number of daily load profiles in the considered load condition. In Fig. 24 are reported the load shape pattern frequency results for the building encoded as Wolf_office_Elisabeth considering the load condition pertaining to winter workdays. Specifically, all the daily load profiles are displaced in their closest clusters and colored in green or red respectively for frequent and infrequent shapes. As a reference, for the analyzed building, the FR is 82.8% meaning that the majority of the load profiles included in the analyzed load condition are characterized by a shape that frequently occurs in the reference PSU category, such as Shape 3 and Shape 7. On the other hand, since on working days the flat pattern or the morning-peak pattern is infrequent (i.e., Shapes 11

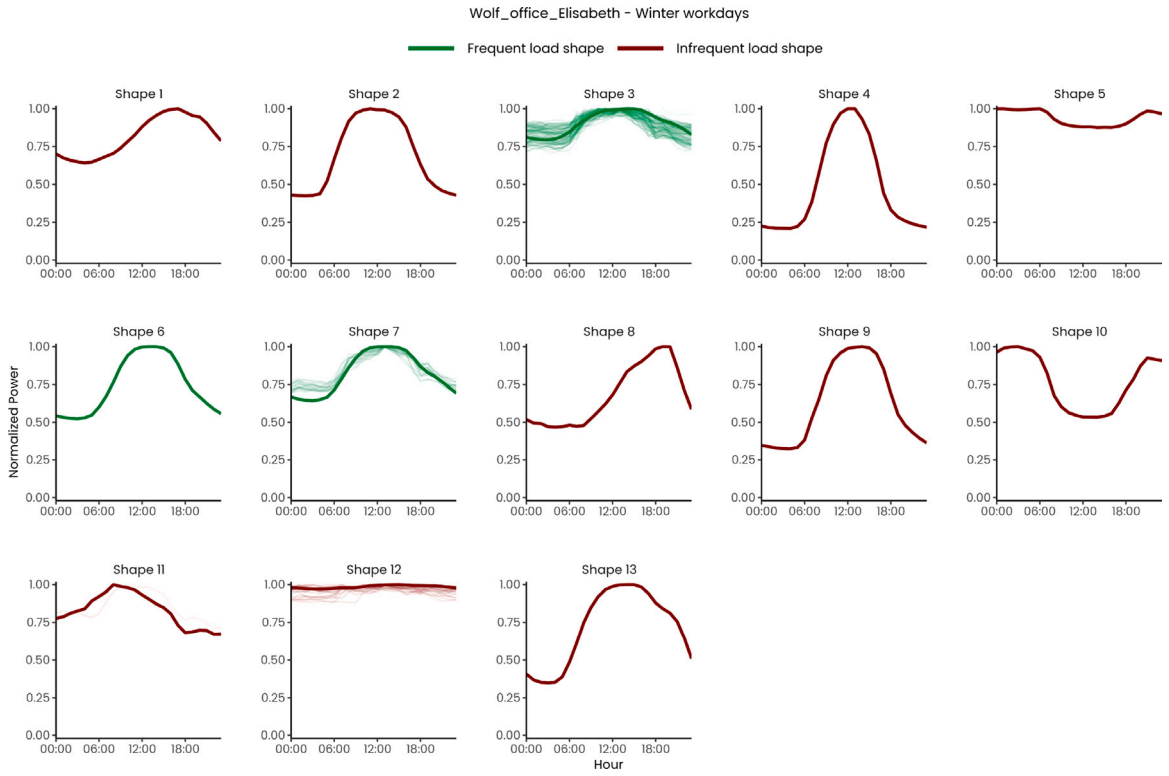


Fig. 24. Classification of the normalized daily load profiles of the building encoded as Wolf_office_Elisabeth for the winter workdays load condition. The solid green and red thick lines represent the centroid of frequent and infrequent clusters respectively, while the thin lines represent the normalized load profiles of the analyzed building.

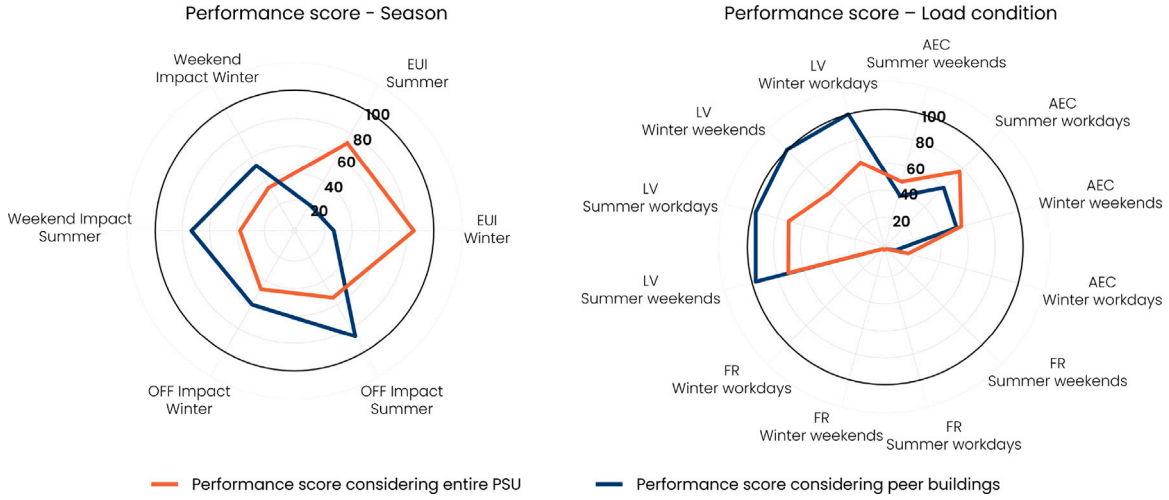


Fig. 25. Radar plot of the performance scores associated to each KPI for the building encoded as Bear_education_Derek. With blue lines are represented the performance scores obtained using the proposed benchmarking process based on a dynamic identification of peers while through the orange lines the performance scores obtained using the entire education PSU category as the reference baseline.

and 12), the shapes of the load profiles that fall into these clusters are considered as infrequent.

5.3.6. Performance score results

As previously discussed in Section 3, the last step of the benchmarking process is to transform the values of each calculated KPI into a performance score. To this purpose, the percentiles are used to report the obtained results in a normalized range between 0 and 100. Where the score of 100 is associated with the building that performs the best, for a specific KPI, considering the performance of its identified peers. This representation is advantageous to inform the final user about the final results in a simple and straightforward way.

A possible visualization of the output of the proposed benchmarking process is reported in Fig. 25, where through radar plots it is possible to easily detect the most critical aspects related to the energy consumption patterns of a building. In Fig. 25, for the building encoded as Bear_education_Derek, the performance scores obtained with the proposed benchmarking process are reported and compared against those obtained using, as a reference set of buildings, the entire PSU category. The performance scores are also reported according to the reference period used for the evaluation of each KPI (i.e., the entire season or a load condition). In general, the building performs better than half of its peers (blue line in the radar plot) in terms of Load Volatility (LV), Anomaly Energy Consumption (AEC) – with exception

Table 3
Summary of KPI values, 1st and 3rd quartiles and performance scores for the building encoded as Bear_education_Derek.

KPI	KPI unit	Value	1st quartile	3rd quartile	Performance score
Winter season					
EUI	kWh/(m^2 DD)	0.12	0.04	0.12	27
OFF-impact	[%]	−39.17	−44.23	14.47	60
Weekend impact	[%]	−38.40	−46.86	−25.25	53
Summer season					
EUI	kWh/(m^2 DD)	1.29	0.05	0.9	20
OFF-impact	[%]	−39.77	−25.74	2.88	87
Weekend impact	[%]	−38.12	−38.08	−26.47	73
Winter workdays					
AEC	[%]	6.02	2.24	4.98	7
LV	[%]	1.23	2.3	4.22	100
FR	[%]	0	1.87	14.64	0
Summer workdays					
AEC	[%]	1.52	0.8	4.5	60
LV	[%]	1.06	1.84	3.66	97
FR	[%]	0	0	6.15	0
Winter weekends					
AEC	[%]	3.61	0.27	4.67	53
LV	[%]	1.35	2.11	4.67	100
FR	[%]	0	0.49	30	0
Summer weekends					
AEC	[%]	3.93	0	4.29	37
LV	[%]	0.03	1.7	3.77	97
FR	[%]	0	0.47	30.63	0

for workdays – and OFF-impact. It is worth to note that, despite some performance scores assume very high values, the daily load profiles are characterized by infrequent shapes (i.e., Performance scores associated to FR values are equal to zero). Additionally, if the entire education PSU category is considered, as the reference baseline for benchmarking the same building, the performance scores for each KPI would be visibly different (i.e., orange line in the radar plot), with a mean absolute variation of about 27%. The same analysis was also performed for the entire test set of buildings leading to an average variation of about 14%. For the sake of completeness, for the building Bear_education_Derek, in Table 3 are reported the following information: calculated values for each KPI, 1st and 3rd quartile values of KPI distributions considering its set of peers, performance scores for each KPI value.

6. Discussions

This work introduces a holistic approach towards energy benchmarking in buildings. The main contribution is related to the high flexibility of the entire approach, overcoming the concept of static comparisons between a building and a reference building stock that only shares the same PSU category. The proposed methodology allows to properly select the best set of peers to conduct a robust energy benchmarking of a building, taking into account a number of distinctive features from its energy consumption time-series (i.e., thermal sensitivity, shape and magnitude of energy consumption, load conditions, and PSU category). Moreover, novel KPIs, such as Load Volatility (LV) and Anomaly Rate (AR) are introduced, while others, such as EUI, have been redefined, allowing them to be more contextualized and informative considering the existence of different boundary conditions. As a key result, the proposed approach leads to an increased accuracy of the energy benchmarking process mainly due to the implemented peer identification process. This demonstrates how impactful is this step of analysis for the definition of a robust benchmarking system and how important is to consider features extracted from time-series to cope with the concept of pattern similarity in energy consumption. Based on what was observed in this study, some aspects also emerged as crucial points to further expand the capabilities of the proposed benchmarking process. The first barrier towards the fully generalizability of the methodology is related to the number of buildings and their PSU categories. Currently, the analysis has been performed on a portion of the open dataset BDGP2 [22], which is one of the few attempts worldwide to make available to the scientific community a large set of building-monitored data. However, the project has some limitations concerning the diversity of data types, lack of user contributions, and missing data [73]. In this context, the collection of extensive datasets will greatly accelerate research in the field of energy and buildings, spanning various domains including building energy management, energy performance assessment, grid management, and socioeconomic analysis. Together with the volume, variety and geographic representativeness of the available open datasets, another barrier is associated with the lack of a unified semantic data representation of building-related data. The use of a formal taxonomy for monitored data and metadata, a unique representation of the relations that exist among energy system features and components, may enable a more comprehensive understanding of the building configuration and support the definition of data-driven processes such as energy benchmarking. Specifically, the use of semantic data representations can be particularly useful in supervising the peer identification process that may leverage the concept of building similarity considering both monitored energy consumption data and building/system metadata (e.g., energy system configuration). In this sense, also the thermal sensitivity analysis may be preliminary driven by the knowledge of the energy services provided to the building and the kind of systems (e.g., presence of heat pumps for space heating) that are installed for that purpose. The knowledge of metadata and contextual information has then a twofold advantage: on one hand it allows to supervise some analyses (e.g., peer identification, thermal sensitivity analysis), while on the other hand to refine the results of the entire benchmarking process. As a reference, the availability of a detailed occupancy schedule and of the real holiday calendar allows a more precise evaluation of load conditions (e.g., workdays) and KPIs such as OFF-impact, weekend impact, Load Volatility (LV) and Anomaly Rate (AR). In fact, in this study those pieces of information were indirectly retrieved by means of

a statistical analysis of the energy consumption time series, without the knowledge of any ground truth. Another aspect that is worth to mention is related to the opportunity to exploit the proposed benchmarking process as a preliminary step of analysis before focusing on specific energy management-related tasks. For example, the evaluation of the AR KPI can be considered as a useful support to estimate the potential impact of an anomaly detection and diagnosis tool to be deployed in the building under investigation, highlighting the load conditions that are of higher interest. Similarly, the assessment of Load Volatility (LV) may be advantageous to preliminary understand some key features of the building under analysis. Firstly, if a building is characterized by a low load volatility, it means that its energy consumption is stable over time or its variation can be well explained by observing the variation of boundary conditions (e.g., outdoor air temperature). In this condition, it has been demonstrated in the literature, that the building energy consumption can be easily predicted over time with high accuracy, de facto enabling the definition of predictive strategies for the optimal management of building energy systems. On the other hand, high values of load volatility are associated with buildings that do not follow a specific pattern in their energy consumption. This aspect may be associated with a high level of inefficiency but in some cases may reveal the high capability of a building to alter its loads more significantly for example in response to a Demand Response (DR) event such as a dynamic pricing plan [74]. In this sense, the knowledge of this KPI can drive the enrollment process of energy customers into demand response programs considering the potential flexibility resource that they would supply to the grid. In this perspective, a comprehensive building dataset which includes information related to the participation of a building in a DR program together with the occurrence and duration ofDR events, would enable the opportunity to evaluate additional flexibility KPIs to assess and track over time the effectiveness of a building in reacting to grid signals [33–35].

7. Conclusions

This paper introduces a novel data-driven external energy benchmarking methodology that extends beyond traditional assessments of Energy Use Intensity (EUI) and comparisons within PSU categories. Instead of merely clustering buildings based on PSU categories, this methodology delves into the details of monitored consumption data to extract valuable insights in the form of KPIs. The methodology is applied to the open-source BDGP2 dataset for Education and Office building categories. In terms of future work, for the sake of generalization and scalability it is necessary to apply this methodology to other categories respect to those analyzed in this paper. Furthermore, additional variables and semantic data representations of the buildings may enable a general refinement and extension of the benchmarking results considering the opportunity to calculate new families of KPIs and to supervise analysis pertaining to the load condition identification, peers identification and thermal sensitivity assessment. In this perspective, another point to be further explored refers to the deployment strategy of the proposed energy benchmarking system. In its current form, the process has been conceived as an offline pipeline of analysis of energy consumption data monitored during a specific year in the past (i.e., 2017). However, the dataset needs to be regularly updated (e.g., yearly) in order to be representative of the energy performance of the building stock over time. Similarly, the benchmarking process can be run, for a specific building, year by year to understand how the KPIs and their associated performance scores evolved and internally compare them with past values. The first steps towards this implementation can be summarized as follows: (i) make the tool publicly available (e.g., by sharing code through an open GitHub project), (ii) make the tool easily integrable with open datasets (e.g., BDGP2 and its future evolutions) (iii) and allow the final user to upload data of a new building and to benchmark it against the available reference stock.

CRediT authorship contribution statement

Marco Savino Piscitelli: Writing – review & editing, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Rocco Giudice:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Alfonso Capozzoli:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

The methodology was developed in the framework of the research collaboration between Politecnico di Torino and Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA) 2022–2024 Three-Year Implementation Plan for “Ricerca di Sistema Elettrico” - project “1.7 Tecnologie per la penetrazione del vettore elettrico negli usi finali” funded by the Italian Ministry for the Environment and Energy Security (MASE).

The work of Marco Savino Piscitelli was carried out within the DM 1062/2021 and received funding from the Italian Ministry of University and Research (MUR) in the framework of the project FSE REACT-EU - PON Ricerca e Innovazione 2014–2020.

The work of Rocco Giudice was made in the framework of the project PNRR-NGEU which has received funding from the Italian Ministry of University and Research (MUR) – DM 351/2022.

References

[1] Commission E, Directorate-General for Energy. Clean energy for all Europeans. Publications Office; 2019, http://dx.doi.org/10.2833/9937.

[2] Commission E, Directorate-General for Communication. European green deal : delivering on our targets. Publications Office of the European Union; 2021, http://dx.doi.org/10.2775/373022.

[3] Ramesh T, Prakash R, Shukla K. Life cycle energy analysis of buildings: An overview. Energy Build 2010;42(10):1592–600. http://dx.doi.org/10.1016/j.enbuild.2010.05.007.

[4] de Wilde P. The gap between predicted and measured energy performance of buildings: A framework for investigation. Autom Constr 2014;41:40–9. http://dx.doi.org/10.1016/j.autcon.2014.02.009.

[5] Shi X, Si B, Zhao J, Tian Z, Wang C, Jin X, Zhou X. Magnitude, causes, and solutions of the performance gap of buildings: A review. Sustainability 2019;11(3). URL https://www.mdpi.com/2071-1050/11/3/937.

[6] Galli A, Piscitelli MS, Moscato V, Capozzoli A. Bridging the gap between complexity and interpretability of a data analytics-based process for benchmarking energy performance of buildings. Expert Syst Appl 2022;206:117649. http://dx.doi.org/10.1016/j.eswa.2022.117649.

[7] Chung W. Review of building energy-use performance benchmarking methodologies. Appl Energy 2011;88(5):1470–9. http://dx.doi.org/10.1016/j.apenergy.2010.11.022.

[8] Ding Y, Liu X. A comparative analysis of data-driven methods in building energy benchmarking. Energy Build 2020;209:109711. http://dx.doi.org/10.1016/j.enbuild.2019.109711.

[9] Capozzoli A, Piscitelli MS, Neri F, Grassi D, Serale G. A novel methodology for energy performance benchmarking of buildings by means of linear mixed effect model: The case of space and DHW heating of out-patient healthcare centres. Appl Energy 2016;171:592–607. http://dx.doi.org/10.1016/j.apenergy.2016.03.083.

[10] Arjunan P, Poolla K, Miller C. BEEM: Data-driven building energy benchmarking for Singapore. Energy Build 2022;260:111869. http://dx.doi.org/10.1016/j.enbuild.2022.111869.

[11] Lee W-S. Benchmarking the energy efficiency of government buildings with data envelopment analysis. Energy Build 2008;40(5):891–5. http://dx.doi.org/10.1016/j.enbuild.2007.07.001.

[12] Gao X, Malkawi A. A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm. Energy Build 2014;84:607–16. http://dx.doi.org/10.1016/j.enbuild.2014.08.030.
[13] Zhan S, Liu Z, Chong A, Yan D. Building categorization revisited: A clustering-based approach to using smart meter data for building energy benchmarking. Appl Energy 2020;269:114920. http://dx.doi.org/10.1016/j.apenergy.2020.114920.
[14] Tian Z, Shi X. Proposing energy performance indicators to identify energy-wasting operations on big time-series data. Energy Build 2022;269:112244. http://dx.doi.org/10.1016/j.enbuild.2022.112244.
[15] Tian Z, Zhang X, Shi X, Han Y. Mining operation hours on time-series energy data to identify unnecessary building energy consumption. J Build Eng 2023;63:105509. http://dx.doi.org/10.1016/j.jobe.2022.105509.
[16] Yang Z, Roth J, Jain RK. DUE-B: Data-driven urban energy benchmarking of buildings using recursive partitioning and stochastic frontier analysis. Energy Build 2018;163:58–69. http://dx.doi.org/10.1016/j.enbuild.2017.12.040.
[17] Roth J, Lim B, Jain RK, Grueneich D. Examining the feasibility of using open data to benchmark building energy usage in cities: A data science and policy perspective. Energy Policy 2020;139:111327. http://dx.doi.org/10.1016/j.enpol.2020.111327.
[18] Papadopoulos S, Kontokosta CE. Grading buildings on energy performance using city benchmarking data. Appl Energy 2019;233–234:244–53. http://dx.doi.org/10.1016/j.apenergy.2018.10.053.
[19] Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods 2018;15:233–4. http://dx.doi.org/10.1038/nmeth.4642.
[20] Miller C, Meggers F. Mining electrical meter data to predict principal building use, performance class, and operations strategy for hundreds of non-residential buildings. Energy Build 2017;156:360–73. http://dx.doi.org/10.1016/j.enbuild.2017.09.056.
[21] Miller C, Nagy Z, Schlueter A. A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings. Renew Sustain Energy Rev 2018;81:1365–77. http://dx.doi.org/10.1016/j.rser.2017.05.124.
[22] Miller C, Kathirgamanathan A, Picchetti B, Arjunan P, Park JY, Nagy Z, Raftery P, Hobson BW, Shi Z, Meggers F. The building data genome project 2, energy meter data from the ASHRAE great energy predictor III competition. Sci Data 2020;7:368. http://dx.doi.org/10.1038/s41597-020-00712-x.
[23] Dahlan NY, Mohamed H, Kamaluddin KA, Abd Rahman NM, Reimann G, Chia J, Ilham NI. Energy star based benchmarking model for Malaysian government hospitals - A qualitative and quantitative approach to assess energy performances. J Build Eng 2022;45:103460. http://dx.doi.org/10.1016/j.jobe.2021.103460.
[24] Environmental Protection Agency U, Department of Energy U. Energy star. 2023, http://www.energystar.gov, Accessed: 2023-05-09.
[25] Soares Geraldi M, Melo AP, Lamberts R, Borgstein E, Yujhi Gomes Yukizaki A, Braga Maia AC, Borghetti Soares J, dos Santos Junior A. Assessment of the energy consumption in non-residential building sector in Brazil. Energy Build 2022;273:112371. http://dx.doi.org/10.1016/j.enbuild.2022.112371.
[26] Kükrer E, Aker T, Eskin N. Data-driven building energy benchmark modeling for bank branches under different climate conditions. J Build Eng 2023;66:105915. http://dx.doi.org/10.1016/j.jobe.2023.105915.
[27] González ABR, Díaz JJV, Caamaño AJ, Wilby MR. Towards a universal energy efficiency index for buildings. Energy Build 2011;43(4):980–7. http://dx.doi.org/10.1016/j.enbuild.2010.12.023.
[28] Escrivá-Escrivá G, Álvarez-Bel C, Peñalvo-López E. New indices to assess building energy efficiency at the use stage. Energy Build 2011;43(2):476–84. http://dx.doi.org/10.1016/j.enbuild.2010.10.012.
[29] Abu Bakar NN, Hassan MY, Abdullah H, Rahman HA, Abdullah MP, Hussin F, Bandi M. Energy efficiency index as an indicator for measuring building energy performance: A review. Renew Sustain Energy Rev 2015;44:1–11. http://dx.doi.org/10.1016/j.rser.2014.12.018.
[30] Li H, Hong T, Lee SH, Sofos M. System-level key performance indicators for building performance evaluation. Energy Build 2020;209:109703. http://dx.doi.org/10.1016/j.enbuild.2019.109703.
[31] Ashouri M, Haghighat F, Fung BC, Yoshino H. Development of a ranking procedure for energy performance evaluation of buildings based on occupant behavior. Energy Build 2019;183:659–71. http://dx.doi.org/10.1016/j.enbuild.2018.11.050.
[32] Council of European Union. Directive 2010/31/EU of the European parliament and of the council of 19 may 2010 on the energy performance of buildings. 2021, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02010L0031-20210101.
[33] Al Dakheel J, Del Pero C, Aste N, Leonforte F. Smart buildings features and key performance indicators: A review. Sustainable Cities Soc 2020;61:102328. http://dx.doi.org/10.1016/j.scs.2020.102328.
[34] Li H, Johra H, de Andrade Pereira F, Hong T, Le Dréau J, Maturo A, Wei M, Liu Y, Saberi-Derakhtenjani A, Nagy Z, Marszal-Pomianowska A, Finn D, Miyata S, Kaspar K, Nweye K, O'Neill Z, Pallonetto F, Dong B. Data-driven key performance indicators and datasets for building energy flexibility: A review and perspectives. Appl Energy 2023;343:121217. http://dx.doi.org/10.1016/j.apenergy.2023.121217.
[35] Airò Farulla G, Tumminia G, Sergi F, Aloisio D, Cellura M, Antonucci V, Ferraro M. A review of key performance indicators for building flexibility quantification to support the clean energy transition. Energies 2021;14(18). http://dx.doi.org/10.3390/en14185676.
[36] Olu-Ajayi R, Alaka H, Sulaimon I, Balogun H, Wusu G, Yusuf W, Adegoke M. Building energy performance prediction: A reliability analysis and evaluation of feature selection methods. Expert Syst Appl 2023;225:120109. http://dx.doi.org/10.1016/j.eswa.2023.120109.
[37] Zhang Y, Teoh BK, Wu M, Chen J, Zhang L. Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence. Energy 2023;262:125468. http://dx.doi.org/10.1016/j.energy.2022.125468.
[38] Cai W, Wen X, Li C, Shao J, Xu J. Predicting the energy consumption in buildings using the optimized support vector regression model. Energy 2023;273:127188. http://dx.doi.org/10.1016/j.energy.2023.127188.
[39] Kapp S, Choi J-K, Hong T. Predicting industrial building energy consumption with statistical and machine-learning models informed by physical system parameters. Renew Sustain Energy Rev 2023;172:113045. http://dx.doi.org/10.1016/j.rser.2022.113045.
[40] Najafi B, Depalo M, Rinaldi F, Arghandeh R. Building characterization through smart meter data analytics: Determination of the most influential temporal and importance-in-prediction based features. Energy Build 2021;234:110671. http://dx.doi.org/10.1016/j.enbuild.2020.110671.
[41] Xiao T, Xu P, Ding R, Chen Z. An interpretable method for identifying mislabeled commercial building based on temporal feature extraction and ensemble classifier. Sustainable Cities Soc 2022;78:103635. http://dx.doi.org/10.1016/j.scs.2021.103635.
[42] Wang E. Decomposing core energy factor structure of U.S. commercial buildings through clustering around latent variables with random forest on large-scale mixed data. Energy Convers Manage 2017;153:346–61. http://dx.doi.org/10.1016/j.enconman.2017.10.020.
[43] Energy Information Administration US. Commercial buildings energy consumption survey (CBECS). 2018, https://www.eia.gov/consumption/commercial.
[44] Choi D, Kim C. Diagnosis of building energy consumption in the 2012 CBECS data using heterogeneous effect of energy variables: A recursive partitioning approach. Build Simul 2021;14(6):1737–55. http://dx.doi.org/10.1007/s12273-021-0777-8.
[45] Quevedo T, Geraldi M, Melo A. Applying machine learning to develop energy benchmarking for university buildings in Brazil. J Build Eng 2023;63:105468. http://dx.doi.org/10.1016/j.jobe.2022.105468.
[46] Geraldi MS, Ghisi E. Data-driven framework towards realistic bottom-up energy benchmarking using an artificial neural network. Appl Energy 2022;306:117960. http://dx.doi.org/10.1016/j.apenergy.2021.117960.
[47] Park JY, Yang X, Miller C, Arjunan P, Nagy Z. Apples or oranges? Identification of fundamental load shape profiles for benchmarking buildings using a large and diverse dataset. Appl Energy 2019;236:1280–95. http://dx.doi.org/10.1016/j.apenergy.2018.12.025.
[48] Komatsu H, Kimura O. Customer segmentation based on smart meter data analytics: Behavioral similarities with manual categorization for building types. Energy Build 2023;283:112831. http://dx.doi.org/10.1016/j.enbuild.2023.112831.
[49] Luo X, Hong T, Chen Y, Piette MA. Electric load shape benchmarking for small- and medium-sized commercial buildings. Appl Energy 2017;204:715–25. http://dx.doi.org/10.1016/j.apenergy.2017.07.108.
[50] Zakovorotnyi A, Seerig A. Building energy data analysis by clustering measured daily profiles. Energy Procedia 2017;122:583–8. http://dx.doi.org/10.1016/j.egypro.2017.07.353, CISBAT 2017 International ConferenceFuture Buildings and Districts – Energy Efficiency from Nano to Urban Scale.
[51] Jiang Z, Lin R, Yang F. A hybrid machine learning model for electricity consumer categorization using smart meter data. Energies 2018;11(9). http://dx.doi.org/10.3390/en11092235.
[52] Tureczek AM, Nielsen PS. Structured literature review of electricity consumption classification using smart meter data. Energies 2017;10(5):584. http://dx.doi.org/10.3390/en10050584.
[53] Capozzoli A, Piscitelli MS, Brandi S. Mining typical load profiles in buildings to support energy management in the smart city context. Energy Procedia 2017;134:865–74. http://dx.doi.org/10.1016/j.egypro.2017.09.545, Sustainability in Energy and Buildings 2017: Proceedings of the Ninth KES International Conference, Chania, Greece, 5-7 July 2017.
[54] Piscitelli MS, Brandi S, Capozzoli A. Recognition and classification of typical load profiles in buildings with non-intrusive learning approach. Appl Energy 2019;255:113727. http://dx.doi.org/10.1016/j.apenergy.2019.113727.
[55] Rajabi A, Eskandari M, Ghadi MJ, Li L, Zhang J, Siano P. A comparative study of clustering techniques for electrical load pattern segmentation. Renew Sustain Energy Rev 2020;120:109628. http://dx.doi.org/10.1016/j.rser.2019.109628.
[56] Bourdeau M, Basset P, Beauchêne S, Da Silva D, Guiot T, Werner D, Nefzaoui E. Classification of daily electric load profiles of non-residential buildings. Energy Build 2021;233:110670. http://dx.doi.org/10.1016/j.enbuild.2020.110670.
[57] Ruiz L, Pegalajar M, Arcucci R, Molina-Solana M. A time-series clustering methodology for knowledge extraction in energy consumption data. Expert Syst Appl 2020;160:113731. http://dx.doi.org/10.1016/j.eswa.2020.113731.

[58] Quintana M, Arjunan P, Miller C. Islands of misfit buildings: Detecting uncharacteristic electricity use behavior using load shape clustering. Build Simul 2020;14(1):119–30. http://dx.doi.org/10.1007/s12273-020-0626-1.

[59] Choksi KA, Jain S, Pindoriya NM. Feature based clustering technique for investigation of domestic load profiles and probabilistic variation assessment: Smart meter dataset. Sustain Energy Grids Netw 2020;22:100346. http://dx.doi.org/10.1016/j.segan.2020.100346.

[60] Liu X, Ding Y, Tang H, Xiao F. A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data. Energy Build 2021;231:110601. http://dx.doi.org/10.1016/j.enbuild.2020.110601.

[61] Nichiforov C, Stamatescu G, Stamatescu I, Făgărăşan I. Learning dominant usage from anomaly patterns in building energy traces. In: 2020 IEEE 16th international conference on automation science and engineering (CASE). 2020, p. 548–53. http://dx.doi.org/10.1109/CASE48305.2020.9216794.

[62] Hu M, Stephen B, Browell J, Haben S, Wallom DC. Impacts of building load dispersion level on its load forecasting accuracy: Data or algorithms? Importance of reliability and interpretability in machine learning. Energy Build 2023;285:112896. http://dx.doi.org/10.1016/j.enbuild.2023.112896.

[63] Miller C, Arjunan P, Kathirgamanathan A, Fu C, Roth J, Park JY, Balbach C, Gowri K, Nagy Z, Fontanini AD, Haberl J. The ASHRAE great energy predictor III competition: Overview and results. Sci Technol Built Environ 2020;26(10):1427–47. http://dx.doi.org/10.1080/23744731.2020.1795514, arXiv:https://doi.org/10.1080/23744731.2020.1795514.

[64] Hyndman RJ, Khandakar Y. Automatic time series forecasting: the forecast package for R. J Stat Softw 2008;26(3):1–22. http://dx.doi.org/10.18637/jss.v027.i03.

[65] Cleveland RB, Cleveland WS, McRae JE, Terpenning I. STL: A seasonal-trend decomposition procedure based on loess. J Off Stat 1990;6:3–73.

[66] James NA, Kejariwal A, Matteson DS. Leveraging cloud data to mitigate user experience from 'breaking bad'. In: 2016 IEEE international conference on big data (Big Data). 2016, p. 3499–508. http://dx.doi.org/10.1109/BigData.2016.7841013.

[67] Haas R. Energy efficiency indicators in the residential sector: What do we know and what has to be ensured? Energy Policy 1997;25(7):789–802. http://dx.doi.org/10.1016/S0301-4215(97)00069-4, Cross-country comparisons of indicators of energy use, energy efficiency and CO2 emissions.

[68] Chung W, Hui Y, Lam YM. Benchmarking the energy efficiency of commercial buildings. Appl Energy 2006;83(1):1–14. http://dx.doi.org/10.1016/j.apenergy.2004.11.003.

[69] Coughlin K, Piette MA, Goldman C, Kiliccote S. Statistical analysis of baseline load models for non-residential buildings. Energy Build 2009;41(4):374–81. http://dx.doi.org/10.1016/j.enbuild.2008.11.002.

[70] Chiosa R, Piscitelli MS, Fan C, Capozzoli A. 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 2022;270:112302. http://dx.doi.org/10.1016/j.enbuild.2022.112302.

[71] Core Team R. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022, URL https://www.R-project.org/.

[72] Van Rossum G, Drake Jr FL. Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam; 1995.

[73] Fu C, Quintana M, Nagy Z, Miller C. Filling time-series gaps using image techniques: Multidimensional context autoencoder approach for building energy data imputation. Appl Therm Eng 2024;236:121545. http://dx.doi.org/10.1016/j.applthermaleng.2023.121545.

[74] Jang D, Eom J, Jae Park M, Jeung Rho J. Variability of electricity load patterns and its effect on demand response: A critical peak pricing experiment on Korean commercial and industrial customers. Energy Policy 2016;88:11–26. http://dx.doi.org/10.1016/j.enpol.2015.09.029.