ORIGINAL RESEARCH **Open access**

Non-Intrusive Machine Learning for Reconstructing Hourly Residential Load Profiles in Renewable Energy Communities

Fernando Diaz[1], Lucia Morales[1*], Diego Perez[2], Valeria Soto[1]

Abstract

Effectively deploying Renewable Energy Communities (RECs) requires maximizing collective self-consumption, especially within legal frameworks that reward shared energy. Across numerous regions, the lack of measurement infrastructure delivering hourly or sub-hourly data for low-voltage consumers (such as homes and businesses) complicates the planning of novel energy communities. This paper introduces a non-invasive machine learning approach to synthesize hourly residential power usage profiles from monthly billing statistics alone, seeking to evaluate the energy shared within RECs. The introduced technique operates in three stages: Identifying representative home demand configurations via k-Means clustering. Applying a Random Forest model informed by monthly utility invoices to pinpoint typical demand shapes. Rebuilding the hourly power demand profile using an empirical scaling approach. The performance of this strategy was verified using a REC pilot case featuring 37 household consumers supplied by a 70kWP photovoltaic installation. The Normalized Mean Absolute Error (NMAE) and the Normalized Root Mean Squared Error (NRMSE) were calculated across a full annual cycle and specifically during periods of communal energy sharing. The Relative Absolute Error was likewise determined for shared energy calculations on both a monthly (MRAE) and annual (RAE) basis. Contrasting the synthesized REC load curves against actual consumption patterns revealed an overall NMAE of 20.04% and an NRMSE of 26.17%, alongside errors of 18.34% and 23.87% during periods of active energy sharing, respectively. Additionally, the model produced relative absolute errors of 8.31% and 0.12% for monthly and annual shared energy estimations, respectively.

Keywords Renewable energy community, Load profiling, Non-intrusive machine learning, Data-driven models, Data analytics, Shared energy estimation

*Correspondence:
Lucia Morales
lucia.morales@gmail.com
[1] Department of Artificial Intelligence in Electrical Engineering, Faculty of Engineering, University of Buenos Aires, Buenos Aires, Argentina
[2] Department of Smart Energy Analytics, Faculty of Engineering, National University of La Plata, La Plata, Argentina

Introduction

Context and motivation

Renewable Energy Communities (RECs) represent community-led projects investing in green energy to satisfy local demand and ecological targets, thus supporting the expansion of sustainable power sources [1]. RECs are localized structures that unite residents, small and medium-sized enterprises (SMEs), public entities, and diverse end-users to establish initiatives focused on generating, distributing, and locally utilizing clean power. Sustainable power generation

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

within RECs typically uses technologies like solar photovoltaic arrays, wind generators, and hydro systems, spanning applications such as electricity generation, thermal building management, and electric vehicle refueling.

Furthermore, RECs aim to accelerate the integration of green energy assets [2], advance efficiency benchmarks, drive the adoption of eco-friendly alternatives, curb carbon emissions, and encourage the development of resilient local organizations. In addition, they empower individuals to engage directly in power markets as prosumers [3], facilitating internal self-consumption while enabling community members to act as creators and proprietors of green production assets. RECs act as developing participants in power markets, executing diverse operations to provide socio-economic gains to their stakeholders and affiliates. They offer a concrete demonstration of how the shift toward clean energy can become an inclusive, cooperative movement, empowering local actors to oversee regional energy governance and define their future power strategies [4].

The standard lifecycle of a nascent REC initiative encompasses three main steps: design, establishment, and management, as illustrated in **Figure 1.**

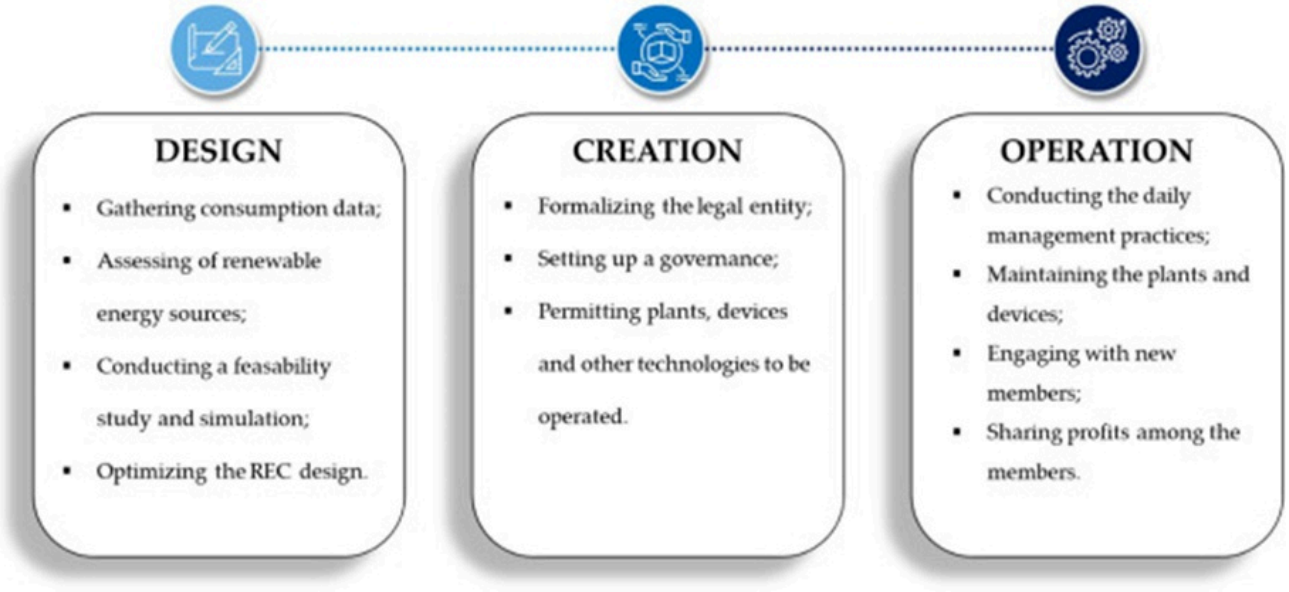


Figure 1. The evolution timeline of an REC.

The planning stage begins when an assembly of residents, companies, institutions, and local stakeholders unite around a shared goal and commit to collaborating on the project. Initially, they map out the baseline project parameters by recruiting potential participants, validating geographic contiguity, compiling demand statistics, and evaluating accessible renewable energy sources (RES) or sites for prospective infrastructure [4]. They then conduct a viability assessment to model power dynamics, refine the REC configuration (specifically to increase communal energy sharing), and outline an initial implementation roadmap. Lastly, they launch a public invitation to residents, backers, and related parties to evaluate and endorse the project's layout and capital commitments.

The establishment stage starts by legally formalizing the corporate structure of the Renewable Energy Community, assigning participant responsibilities, and setting up the regulatory framework for the RES. Thereafter, required funding and regulatory approvals for constructing the new power production sites are secured. This step concludes with assembling the production facilities, instruments, and software required for REC management, along with securing operational authorization for the community.

The operational stage involves the routine supervision of the REC, including tracking energy metrics and system health, servicing equipment, managing financial tracking, communicating with participants, offering support services, and

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

distributing financial returns among members.

A major obstacle encountered during the planning period involves accessing highly granular (hourly or sub-hourly) consumption records from community participants. Access to these granular statistics directly influences the ideal dimensioning of generation systems, which is critical to the financial and operational durability of an REC initiative, particularly in subsidized frameworks where revenue depends heavily on the fiscal monetization of shared power. Extensive academic literature has explored RECs. These investigations cover multiple dimensions, including the influence of RECs within intricate socio-economic frameworks [5-7], the application of digital environments to facilitate their growth [4], regulatory frameworks [8, 9], alongside their design and capacity optimization.

Recent literature has analyzed the consolidation of prosumers in RECs through both financial and thermodynamic viewpoints [10]. These works have integrated community-level storage systems and reversible solid oxide cells, evaluating their effects on local distribution infrastructure [11-13]. Moreover, innovative commercial frameworks and optimization models targeting energy clusters have introduced balanced profit-sharing mechanisms, fair compensation models for aggregators, and termination terms [14, 15]. These publications have advanced cutting-edge analytical tools to identify ideal control strategies for microgrids, even inside policy-supported environments [16, 17]. Additionally, artificial intelligence frameworks have been deployed to assess the deployment and control of multi-carrier energy and storage assets [18-20].

Nonetheless, these contemporary studies depend significantly on granular empirical demand data, which remain exceedingly difficult to source during the initial REC planning stages, particularly within the housing sector [16]. Additionally, the computational frameworks applied in these investigations necessitate parameters equivalent to those presented in **Table 1.**

Table 1. Attributes of datasets used in recent academic publications concerning RECs.

References	Research purpose	REC development phases	Methods	Data type	Data resolution
[10-12]	Simulation and optimization of REC	Design	MILP/Linear programming	Real data from industrial, agricultural, tertiary, grid, household, and commercial activities	Hourly
[14, 15]	Simulation, optimization, and assessment of sharing mechanisms in REC	Design and Operation	Energy- and economic-based simulation	Real and synthetic residential and commercial activity data	Hourly / Sub-hourly
[13, 16-18]	Simulation, optimization, and management of REC	Design and Operation	Energy Hub / Time Delay Neural Network / Energy-based simulation	Real residential and building data	Hourly / Sub-hourly

Evidently, the planning, simulation, sharing framework verification, and management configuration assessments in the cited works were derived via high-frequency telemetry. In contrast, practical applications often struggle to acquire hourly or finer time-step data, which restricts the applicability of conventional methods like MILP, Energy Hub concepts, and Time Delay Neural Networks.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

The accessibility of high-frequency data, critical for REC sizing, depends fundamentally on the market penetration of advanced metering systems. Numerous European nations targeted an 80 % deployment rate for smart meters by 2020 [21], a deadline that was subsequently delayed to 2024 for several regions [22]. Based on a contemporary European assessment, nations such as Austria, Greece, and the United Kingdom missed their projected 2020 large-scale implementation schedules [21]. Furthermore, a recent report [22] indicates that a mere thirteen European Union member states (Sweden, Denmark, Finland, Estonia, Spain, Norway, Luxembourg, Latvia, Italy, France, Malta, Slovenia, and the Netherlands) fulfilled the 2020 benchmark.

Nations like Portugal, Austria, the UK, and Ireland shifted their 80% penetration objectives to 2024, whereas Belgium, Croatia, Poland, Slovakia, Lithuania, and Hungary remain in the early phases of infrastructure expansion. Additionally, Bulgaria, Cyprus, the Czech Republic, Germany, and Greece have deployed minimal or zero advanced meters. Hence, it is clear that advanced meter implementation across over half of the European Union is either delayed or entirely stagnant. This situation underscores the impracticality of extracting fine-grained time-series information across numerous EU territories, particularly given that even nations meeting recent rollout goals need additional time to compile and format actionable datasets for engineering assessments.

Literature review on clustering, load profiling and classification techniques

In this section, we review the primary methodologies and algorithms associated with load profiling, consumer pattern recognition, and classification frameworks. We introduce a methodology that integrates these three components to overcome the limitation of sparse.

In the clustering domain, numerous investigations have shown that k-Means clustering delivers superior performance compared to alternative strategies when extracting load patterns from substantial datasets [22-24]. Specifically, k-Means is frequently paired with validation metrics to identify the ideal number of clusters, as demonstrated by Pérez-Chacon *et al*. [24]. Conversely, Wang *et al*. [25] introduced an alternative implementation of k-Means, presenting a dual-stage framework that pairs k-Means with Dynamic Time Warping Clustering (DTWC).

The objective of this dual-stage strategy was to overcome difficulties in characterizing standard building load curves, proving more effective than standalone k-Means execution. Nevertheless, implementing such a complex clustering process is generally practical only when dealing with a wide variety of distinct load curves. In a similar vein, Fang *et al*. analyzed multiple clustering frameworks for pattern recognition [26], testing k-Means, Hierarchical clustering, Gaussian Mixture Models, C-vine Copula Mixture Models, and R-vine Mixture Models. These were paired with various classification models, such as K-Nearest Neighbors (KNN), Classification and Regression Trees, Naive Bayes (NB), and Random Forest. Their findings indicated that pairing clustering algorithms with a Random Forest Classifier yielded the highest classification accuracy.

Regarding classification frameworks, the research by Pérez-Ortiz *et al*. [27] offered a thorough evaluation of classification algorithms and their utilization across diverse Renewable Energy (RE) applications. Their study concentrated on delivering a comprehensive assessment of classification tools and their underlying learning frameworks. Similarly, Neelamegam and Ramaraj [28] provided an extensive overview of classification methodologies utilized within data mining. Both papers deliver granular assessments of diverse classification methods, outlining their respective benefits, constraints, and sector-specific utility. Furthermore, Neelamegam and Ramaraj [28] analyzed multiple evaluation metrics and validation protocols for measuring classification model success, emphasizing indicators like accuracy, precision, and recall.

Looking at load profiling, Alrawi *et al*. [29] investigated a data-driven framework to analyze consumer electrical behavior. Their core findings indicate that property size, billing responsibility, and the specific cooling systems are the primary drivers of electricity usage. However, acquiring details regarding building infrastructure, household appliances, and

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

behavioral habits is highly challenging during the initial phases of a Renewable Energy Community (REC) initiative; consequently, load profiling methods dependent on such granular metrics are impractical for REC applications.

Similarly, Piscitelli *et al*. [30] examined a non-intrusive approach to manage user classification within the building sector, designing a tool to forecast typical monthly consumption curves for new consumers. This model utilizes a Globally Optimal Decision Tree algorithm trained on predictive features from monthly utility bills alongside supplementary data collected via telephone polling. Additionally, this methodology estimates the scale of typical consumption curves using billing data via distinct scaling factors.

However, as noted in the study by Alrawi *et al.* [29], the extra data gathered via phone surveys—such as occupant schedules and routines—is typically unavailable when modeling and planning an REC initiative. To address this, Lazzeroni *et al*. [31] introduced a data-driven model to forecast consumption curves for distinct end-user types (such as working days, Saturdays, and holidays) utilizing minimal inputs, namely monthly power bills. They leveraged a K-Nearest Neighbors approach to estimate typical consumption curves based on historical billing similarities. They validated their framework against two baselines: one utilizing Standard Load Profiles (SLPs) and another utilizing clustering paired with decision rules to match new users to representative curves based on time-of-use billing. Their outcomes demonstrate that the proposed framework outperformed both benchmarks across several error criteria for nearly every consumer type, establishing itself as a credible alternative to traditional clustering and classification-based profiling.

In summary, the literature surrounding clustering, classification, and load profiling indicates that k-Means stands out as the most practical clustering method for massive datasets. Furthermore, classification frameworks paired with this clustering approach achieve peak performance when utilizing Random Forest techniques. Additionally, the surveyed profiling methods heavily depend on datasets supplemented with extra metadata regarding appliances, user behavior, and occupancy schedules, except Lazzeroni *et al.* [31], which offered a strategy to reconstruct demand profiles using only monthly utility statements. However, this alternative method may not be ideal for synthesizing aggregate electrical demand curves within an REC framework. Ultimately, existing research relies on enriched datasets detailing occupant schedules, routines, and appliances, which are difficult to source during the planning phase of an REC. Furthermore, the reviewed load profiling studies focus on reconstructing individual consumer curves, which diverges from the aggregate-level analysis required for RECs.

Research gap, novelty and paper organization

As demonstrated by prior sections, current literature surrounding REC initiatives relies heavily on high-resolution data streams that include supplemental information regarding occupant behavior, schedules, and specific appliances. A scarcity of this data hinders the planning stage of new REC developments, even within nations boasting widespread smart meter deployment, due to administrative delays in accessing usable datasets. Furthermore, existing load profiling studies in the literature have not explicitly focused on evaluating the fundamental energy metric of an REC: shared energy. Consequently, given the sparse availability of high-resolution residential data, the difficulties in gathering supplementary information like occupant schedules, and the absence of data-driven profiling studies evaluating key REC metrics, this research aims to:

· Propose a data-driven framework, capable of being trained on datasets across diverse geographical regions, to reconstruct aggregate-level hourly power consumption for residential users.

· Reconstruct hourly residential electrical demand using non-intrusive machine learning approaches that function solely on readily available parameters, such as monthly usage data, without requiring extra metadata regarding user schedules, habits, or appliances.

· Apply the introduced framework to assess shared energy within a Renewable Energy Community setup under realistic data scarcity conditions, offering a model that allows potential aggregators to precisely project shared energy

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

within an REC.

Data

This study utilized the open-access dataset from the Low Carbon London project. UK Power Networks led this initiative, which took place in London between November 2011 and February 2014 [32]. The database contains a comprehensive time-series collection of electricity usage for 5567 residential consumers in London during this timeframe. Among these, 1122 participants were placed on an experimental dynamic Time of Use (ToU) pricing plan in 2013. The data is recorded in kWh at 30 min intervals. These readings are highly reliable: a maximum of 3.4% of data is missing for any given half-hour slot, and the average rate of missing data per period was only 0.17% throughout the 2013 trial year [32].

The dataset is split into two distinct consumer categories. The first group includes roughly 1100 users who experienced Dynamic Time of Use (dToU) tariffs during the entire 2013 calendar year. The second group consists of the remaining approximately 4500 consumers whose energy readings were not affected by the dToU pricing. Before its release, the database underwent validation to remove any negative or missing entries.

Initial data exploration revealed that data gathering was not uniform across the project's duration, with varying amounts of data collected each year, as detailed in **Table 2.**

Table 2. Preliminary dataset analysis.

Parameter	2014	2013	2012	2011
Number of participating users	4065	4411	4433	354
Relevant months	2	12	12	2
Percentage of participating users	91.5 %	99.3 %	99.8 %	7.9 %

· 1 Month recording a minimum of one consumption reading per user.

· 2 Proportion of users relative to the overall project participant count.

During 2011, data was successfully gathered from only 7.9% of the total participants across the four-year project, and only two months contained at least one electricity usage reading per user. In contrast, 2012 yielded the most extensive data collection, capturing 99.8 % of all participants and featuring twelve complete months with at least one reading per user. The year 2013 included slightly fewer participants compared to 2012 but also maintained twelve months of at least one usage measurement per user. Lastly, 2014 saw a significant proportion of users participating, yet only two months yielded at least one measurement per individual. Additionally, data for March was largely missing across the entire project duration. Consequently, March was dropped from all subsequent evaluations.

Modeling Framework and Methods

Modeling framework

This paper presents a modeling architecture designed to project the hourly electricity usage of aggregated residential consumers, relying on monthly consumption records and utilizing non-intrusive machine-learning approaches. The

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

primary goal of this framework is to leverage the modeled hourly usage to assess shared energy within a REC project. This enables prospective aggregators to accurately gauge shared energy under realistic conditions of limited data, thereby supporting more robust and reliable economic forecasts and pre-feasibility assessments.

The suggested approach is split into two main phases: model creation and case study application.

· Model creation: This stage focuses on building a tool capable of reconstructing hourly electricity profiles for an aggregate group. This model requires input in the form of real or simulated data with at least hourly resolution for training purposes.

· Case study application: This stage reconstructs the required aggregate hourly electrical load and computes key evaluation metrics for that specific profile. This step necessitates real or simulated monthly electricity usage data for the specific consumers whose aggregate profile needs reconstruction.

The model creation phase typically follows five sequential stages: i) pre-processing, ii) data filtering, iii) load pattern recognition, iv) classification, and v) scaling coefficient evaluation, as illustrated in **Figure 2.**

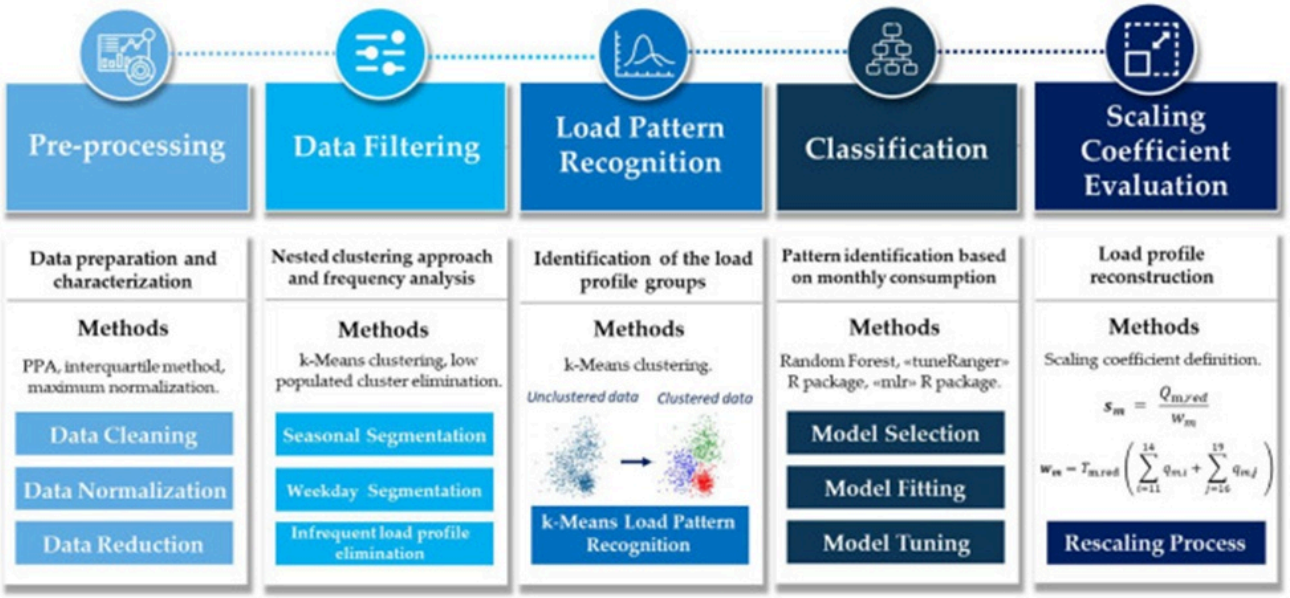


Figure 2. Framework of the model creation phase.

The initial stage involves rigorous pre-processing of the raw dataset. This procedure relies on the technique outlined in Schofield *et al*. [32] and seeks to remove dataset anomalies that might degrade the accuracy of subsequent processes and models. This stage is split into two main parts:

1. Data characterization: Aims to deepen the comprehension of the dataset using appropriate visual tools.

2. Data preparation: Focuses on purging the dataset of inconsistencies, outliers, and missing information.

Next, a data filtering step is executed to refine the dataset further by excising rare load profiles. This is accomplished by dividing the dataset using a two-step method: k-Means clustering, followed by evaluating cluster densities to drop the least populated groups.

The load pattern recognition phase follows this. This step's objective is to pinpoint the characteristic electrical load curves of residential consumers via k-Means clustering. Much like the filtering phase, the populations of the resulting clusters are

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

analyzed to discard any load curves that do not accurately represent consumer usage.

After this, the classification process builds a model capable of matching a generic user with one of the previously established typical load curves for every month, based strictly on their monthly electricity usage. In simpler terms, a Random Forest algorithm is deployed, followed by iterative tweaking to locate the model hyperparameters that yield the best performance.

Lastly, a scaling coefficient evaluation calculates the multipliers needed to convert the typical load curves assigned to each user monthly into actual electrical load profiles. This is a critical component, requiring unique scaling coefficients for every individual user and each month. Ultimately, these calculated multipliers are used to rebuild each user's monthly electrical load profile. Thus, by assuming a user's daily electricity usage aligns with these generated monthly profiles, their hourly consumption for the whole year can be accurately reconstructed. The upcoming sections provide a streamlined explanation of this proposed methodology and its results, alongside in-depth breakdowns of the applied models and their core assumptions.

Pre-processing

Data pre-processing is a vital component in executing any data-driven model, acting as a mandatory requirement to guarantee precise outcomes and trustworthy evaluations. It entails formatting and modifying data into a structure primed for the mining procedure. The objectives of this phase include reducing data volume, identifying inherent relationships, normalizing variables, excising outliers, and extracting key features. It employs multiple strategies, such as data cleaning, transformation, and reduction. As depicted in **Figure 3**, this operation is broadly split into two key tasks: data preparation and data characterization. Data characterization is crucial for establishing a foundational understanding of the data, especially given the challenges of clearly visualizing multidimensional datasets. Meanwhile, data preparation is a critical procedure that yields clean, appropriately structured, and usable data.

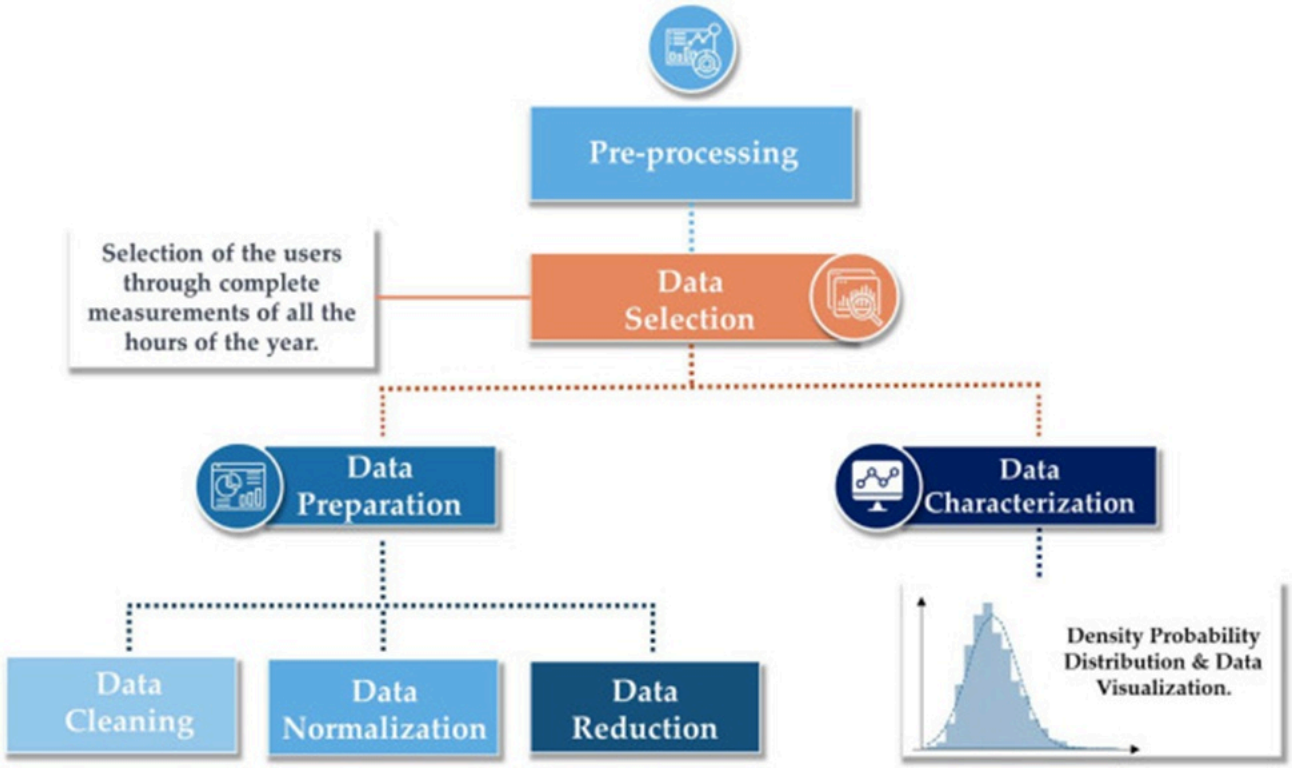


Figure 3. Pre-processing workflow.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

Before initiating the characterization and preparation tasks, an initial data screening is conducted to isolate the portion of the dataset representing the most robust year—measured by the volume of participating users and fully recorded months. The 2012 subset was selected because it proved to be the most heavily populated segment. Furthermore, within this 2012 subset, only users boasting uninterrupted electricity usage records for the entire year were included, narrowing the final pool down to 272 users.

During the data characterization phase, the hourly consumption logs undergo a probability density function analysis. At the same time, the total monthly usage is subjected to a frequency analysis, as displayed in **Figure 4**. This dual approach facilitates a deeper comprehension of the data landscape.

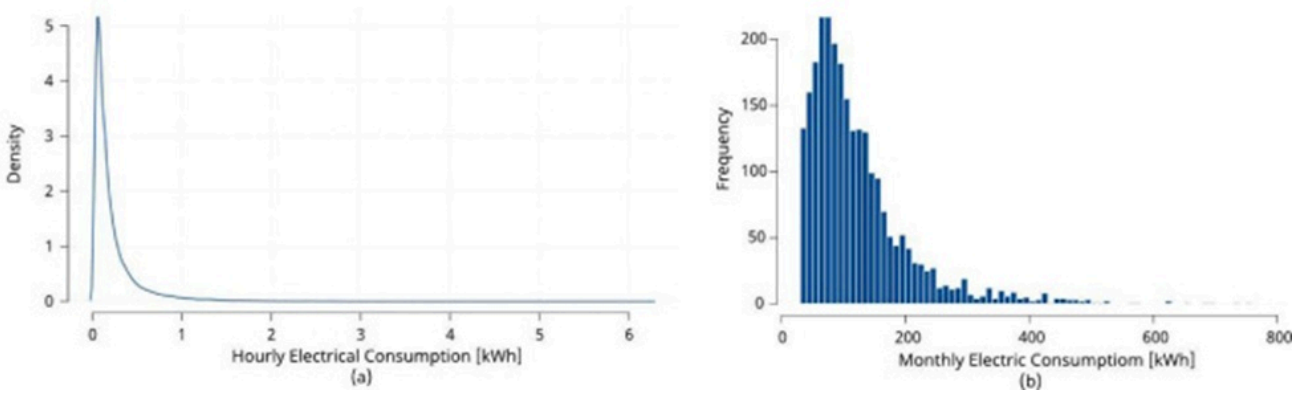


Figure 4. (a) Density probability of the pre-processed data; (b) Monthly consumption frequency analysis of the pre-processed data.

As depicted in the figure, the pre-processed hourly electricity usage metrics overwhelmingly reflect low consumption tiers. Two primary elements drive this trend:

· First, overnight electricity usage levels are notably lower than daytime figures.

· Second, a significant portion of the user base inherently exhibits lower electrical demand, as reinforced by **Figure 4b**, which shows that most data points cluster around medium-low monthly usage.

However, **Figure 4b** also highlights a few exceptionally high monthly consumption spikes, nearing 800 kWh in certain cases. This indicates the existence of energy-heavy consumers during specific months, albeit in smaller quantities, highlighting the varied nature of user profiles within the refined dataset. Wrapping up the pre-processing phase, the data preparation step revealed zero outliers or missing values.

Next, the data preparation procedure is executed, tackling two primary objectives: formatting the data appropriately for the upcoming data mining algorithms, and refining the data to guarantee optimal quality and performance for the models utilized in those operations. The data cleaning addresses missing entries, inconsistencies, and outliers. To resolve missing values, a linear interpolation technique is applied, replacing the gaps with values plotted on a straight line connecting two valid measurements [33]. Because linear interpolation loses efficacy when valid measurements are spaced too far apart, the maximum allowable gap between two valid data points is capped at 3 during the cleaning routine.

Outliers are managed on an hourly scale utilizing the “interquartile method”. As one of the most straightforward and efficient strategies in data mining, it relies on computing three core metrics [34]:

· The median (Q2), representing the midpoint of the sampled data spread.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

· Q1, recognized as the first quartile, marks the threshold below which 25 % of the original data falls.

· Q3, recognized as the third quartile, marks the threshold below which 75 % of the original data falls.

The gap between Q3 and Q1 is termed the interquartile range (IQR). This span is then used to establish a new boundary, called the decision range, and any data points falling outside this boundary are flagged as outliers.

Eqs define the Lower and Upper Bounds. **Eqs. 1 and 2.**

$$Lower\ Bound = Q_1 - 1.5*IQR \quad (1)$$

$$Upper\ Bound = Q_3 + 1.5*IQR \quad (2)$$

Subsequently, data transformation takes place through the normalization of the raw data, which is an absolute requirement for specific machine learning algorithms [30]. Specifically, the data undergoes normalization via maximum scaling, as defined in **Eq. 3.**

$$\hat{x}_{i,m} = \frac{x_{i,m}}{\max\left(x_{i,m}\right)} \quad (3)$$

Where:

· $x_{i,m}$ denotes the monthly load profile for the i-th user during the m-th month;

· $\max(x_{i,m})$ signifies the peak value within the load profile of the i-th user during the m-th month.

Data filtering

This section outlines a targeted procedure designed to purge the pre-processed dataset of elements that might degrade the accuracy of the subsequent machine learning and data-driven algorithms. This operation is split into two distinct stages aimed at pinpointing daily load profiles that are highly unusual in either shape or intensity.

During the initial stage, daily profiles belonging to months with exceptionally low electricity usage are removed for every user in the pre-processed dataset. This is executed using a threshold limit defined via **Eq. 4**, which provides an estimated monthly electricity baseline for English residential consumers.

$$TV_{months} = 0.10*\frac{Q_{UK,avg}}{12} \quad (4)$$

Where:

· $Q_{UK,avg}$ represents the mean annual electricity usage for households in England, established at 3800 kWh for the year 2014 [35].

Next, a tiered methodology is applied to spot daily load profiles with unusual shapes. Initially, the pre-processed data is divided by season **(Table 3)** and subsequently by the day of the week. Following this segmentation, k-Means clustering is executed on each distinct subset, a process illustrated in **Figure 5.**

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

Table 3. Seasonal dataset segmentation.

Season	Months
Summer	June, July, August
Winter	December, January, February
Spring	March, April, May
Autumn	September, October, November

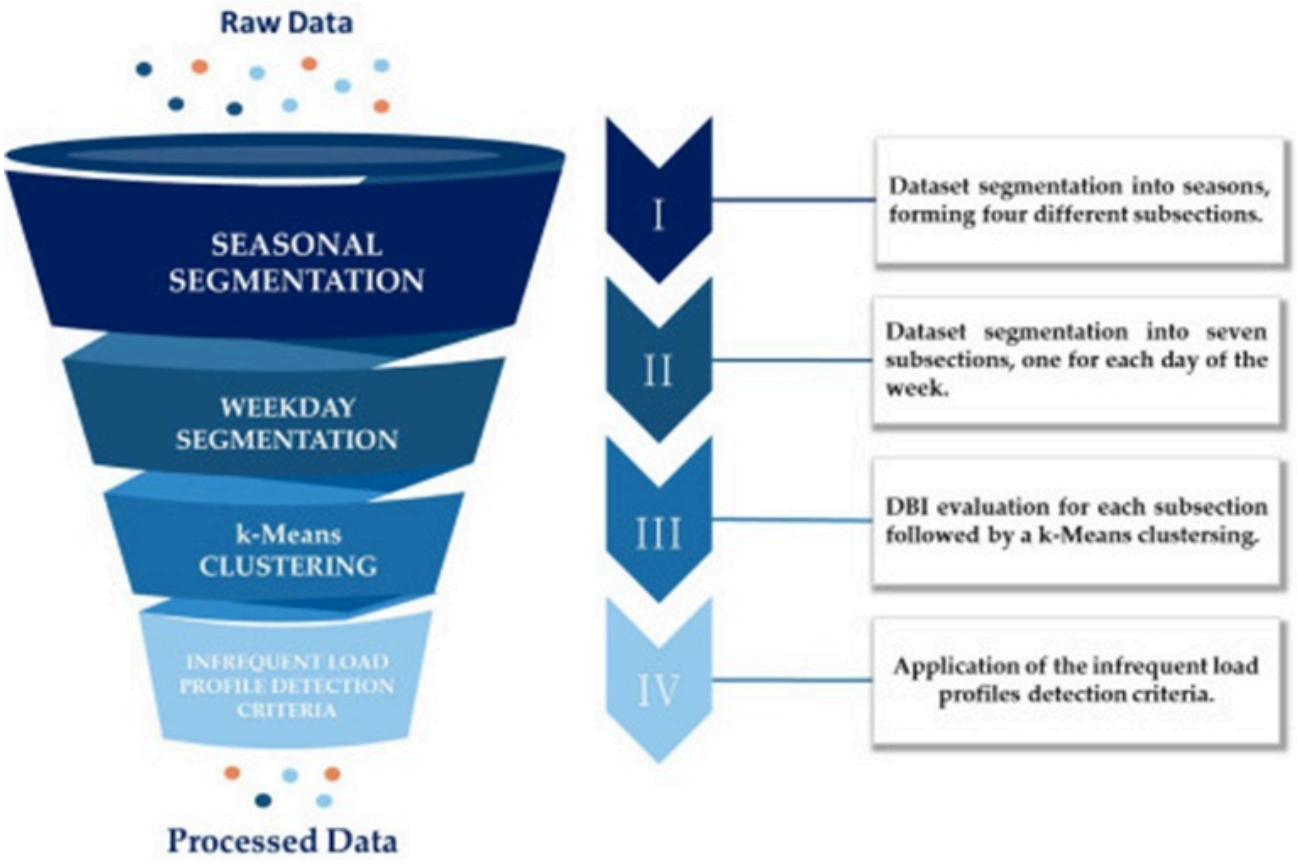


Figure 5. Workflow of the infrequent daily load profile detection.

By utilizing the thresholds defined in **Eqs. 5 and 6**, this strategy makes it possible to flag subgroups within the subsets that house rare load patterns or patterns strictly tied to a narrow fraction of users.

$$TV_u = 0.05 * N_{max,u} \quad (5)$$

$$TV_p = 0.05 * N_{max,p} \quad (6)$$

Where:

· TV_U denotes the threshold limit for the number of users in a given cluster.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

- TV_u denotes the threshold limit for the number of profiles in a given cluster.
- $N_{max,u}$ signifies the absolute maximum user count within a subset.
- $N_{max,p}$ signifies the absolute maximum profile count within a subset.

As demonstrated in **Figure 6,** clusters with the lowest populations are isolated based on the computed thresholds for each data segment. The daily load curves contained within these sparse clusters are then permanently dropped from future analysis.

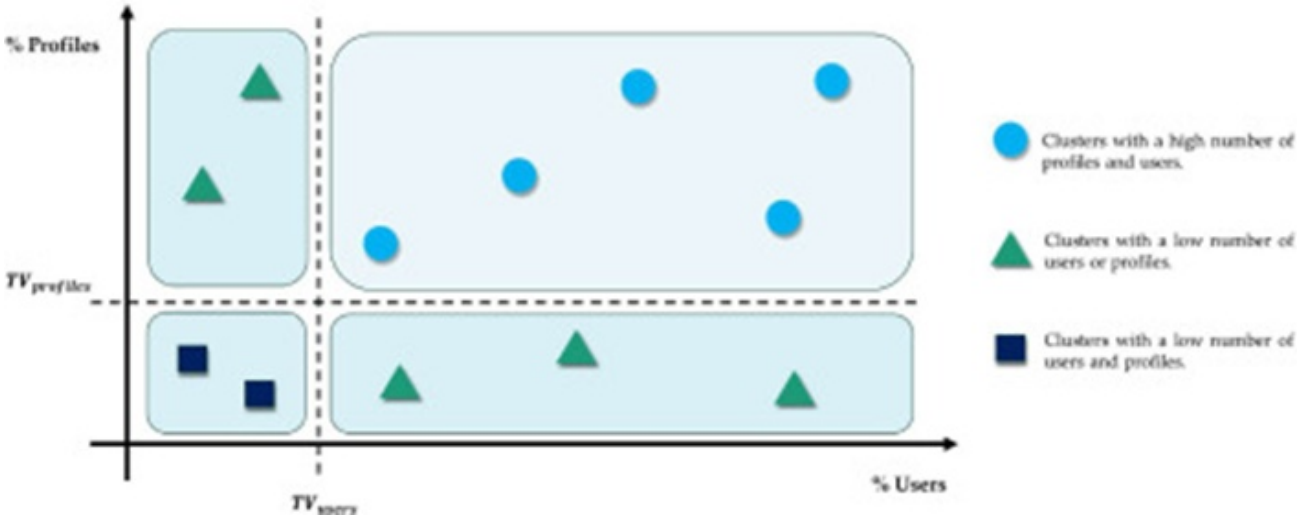


Figure 6. Infrequent load profile detection criteria.

To determine the ideal number of clusters during the partitioning of each subset, the Davies–Bouldin Index (DBI) is applied. The DBI [36] serves as a cluster evaluation metric built on the principle that optimal partitioning requires both maximized distance between different clusters and maximized tightness within individual clusters. **Eq. 7** is used to calculate the DBI.

$$DBI = \frac{1}{K}\sum_{k=1}^{K} \max_{k \neq i}\left(\frac{\delta_k - \delta_i}{d_{k,i}}\right) \quad (7)$$

Where:

- k indicates the total cluster count.
- $d_{k,i}$ represents the Euclidean distance separating the centroids of cluster Ck and cluster Ci.
- δ_k, δ_i signify the standard deviations of the distances connecting data points within clusters Ck and Ci, respectively.

In essence, the DBI gauges how similar clusters are by analyzing their internal spread and how far apart they sit. A low DBI score reflects a strong, logical division. Conversely, a high score suggests a flawed division where clusters blur together and contain highly mismatched elements.

Figure 7 provides a practical example of this anomaly detection logic, successfully isolating rare daily curves for autumn Sundays and spring Thursdays.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

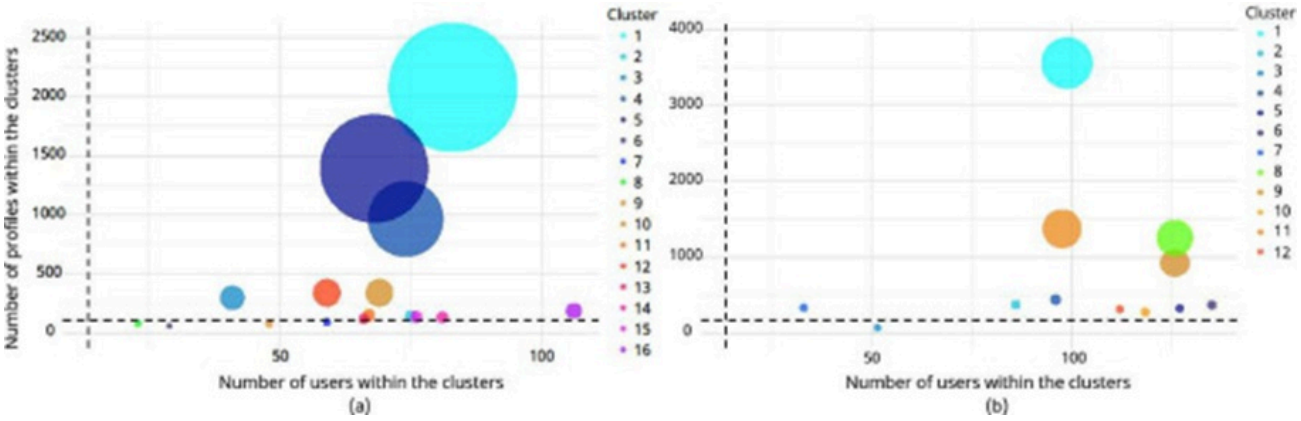


Figure 7. (a) Detection of the daily-scale outliers for Sundays in the autumn months. Clusters within the threshold values were considered infrequent or non-representative and excluded from the analysis; (b) Detection of the daily-scale outliers for Thursdays in the spring months. Clusters within the threshold values were considered infrequent or non-representative and excluded from the analysis. The size of the circles in both figures represents the number of profiles and users within a cluster.

In **Figure 7a**, the DBI pinpointed 16 as the optimal cluster count. The threshold limits for users (TVu) and profiles (TVp) were calculated at 8 and 120, respectively. As a result, clusters 6, 7, 8, and 9 were dropped from the dataset because they held too few profiles, though none violated the user threshold. Likewise, for **Figure 7b**, the DBI established 12 optimal clusters, with user and profile thresholds sitting at 7 and 220. Cluster 3 was the sole group to fall under these limits in this subset, leading to the removal of its profiles.

Clearly, the chosen total number of clusters heavily dictates how many groups fall beneath the threshold lines. The DBI is instrumental here, helping to select a balanced number of clusters that avoids both over-segmentation and overly broad groupings that could blur profile distinctions. Ultimately, as these examples highlight, TVp is the most dominant factor when removing anomalies, responsible for discarding 87 % of the clusters across the operation. The remainder is triggered by either the user threshold alone or a combination of the two limits.

Upon concluding the filtering phase, exactly 7223 daily load profiles were flagged as atypical. **Table 4** breaks down the count of these rare profiles by season, alongside their relative percentages.

Table 4. Daily-scale outliers identified for each season.

Season	Profile percentages[1]	Infrequent daily load profiles
Summer	39.5 %	2850
Winter	10.4 %	750
Spring	27.4 %	1981
Autumn	22.7 %	1642

1 Percentages are related to the maximum number of daily load profiles for each season.

The data shows that summer generated the vast majority of rare profiles, hitting 2850. In contrast, winter produced the fewest, with only 750 atypical profiles. This divergence is likely linked to consumer electricity habits fluctuating more wildly during the summer—driven by varied appliance routines and air conditioning usage—whereas winter routines tend to be

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

steadier. However, a deep dive into the correlation between seasonal shifts and electricity habits falls outside the current study's parameters.

Overall, the data filtering operation stripped away 12.8% of the original dataset. This total consists of 8.0% removed as rare daily profiles and 4.8% discarded due to abnormally low monthly usage, as summarized in **Table 5.**

Table 5. Results obtained from the data filtering process.

Dataset Subsection	Percentage
Valid data	87.2%
Excluded data	12.8%
Infrequent daily load profile data	8.0%
Low monthly consumption data	4.8%

Load pattern recognition

Following pre-processing and data filtering, a k-Means clustering routine is applied to the monthly electricity usage records. As in previous steps, the DBI evaluates the optimal number of clusters. Applying the same logic used during the filtering phase, clusters hosting minimal users or profiles are discarded. For this phase, the threshold cutoff is set at 10% of the total profiles and users. A key difference in this step, however, is the implementation of the Hartigan and Wong algorithm rather than the Lloyd algorithm used earlier. While the Hartigan and Wong method demands more computing power, it delivers vastly more stable and resilient clustering. This is especially true when handling a large volume of clusters or when the starting centroids are placed sub-optimally [37, 38].

The DBI identified 5 as the ideal cluster count for this phase (analyzed across a range of 4 to 20 clusters), as plotted in **Figure 8.** The resulting typical load curves mapped out by the k-Means clustering are illustrated in **Figure 9**, displaying five distinct consumption shapes.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

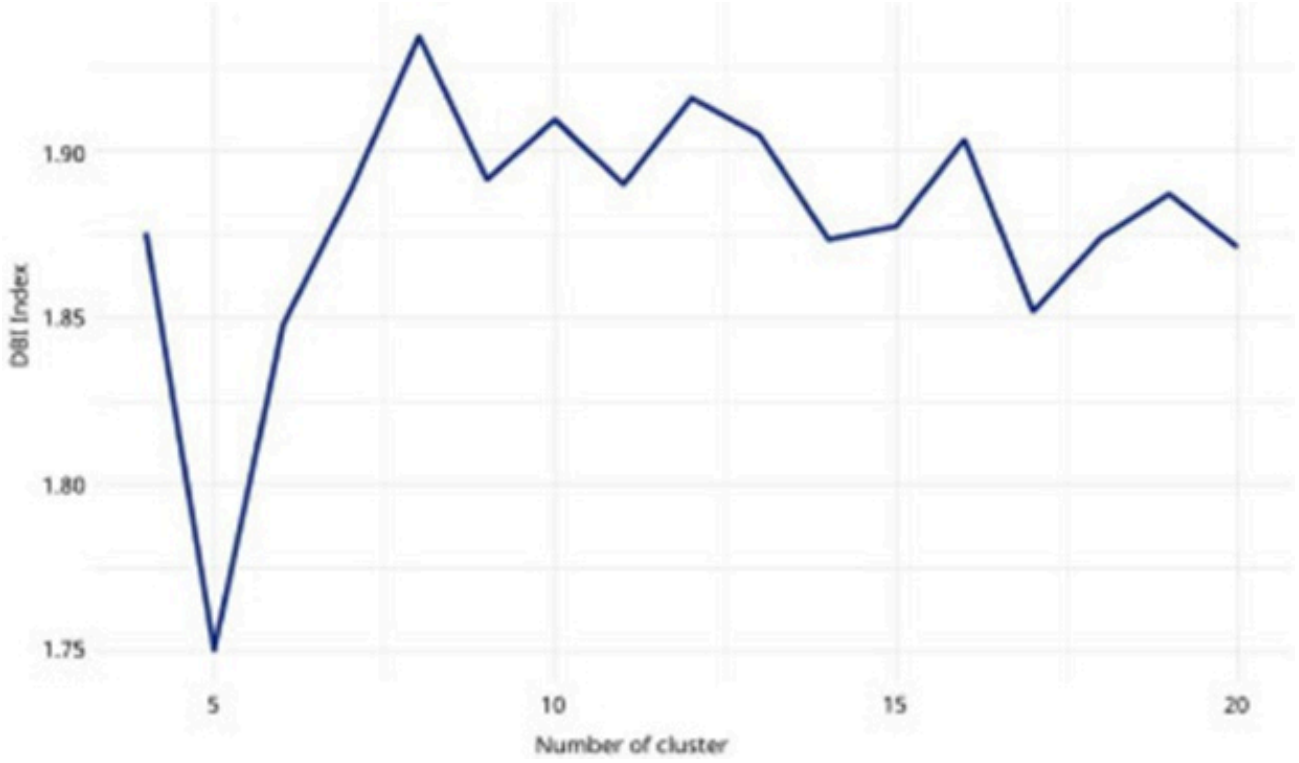


Figure 8. Evaluation of the optimal number of clusters using the DBI index for the load pattern recognition process.

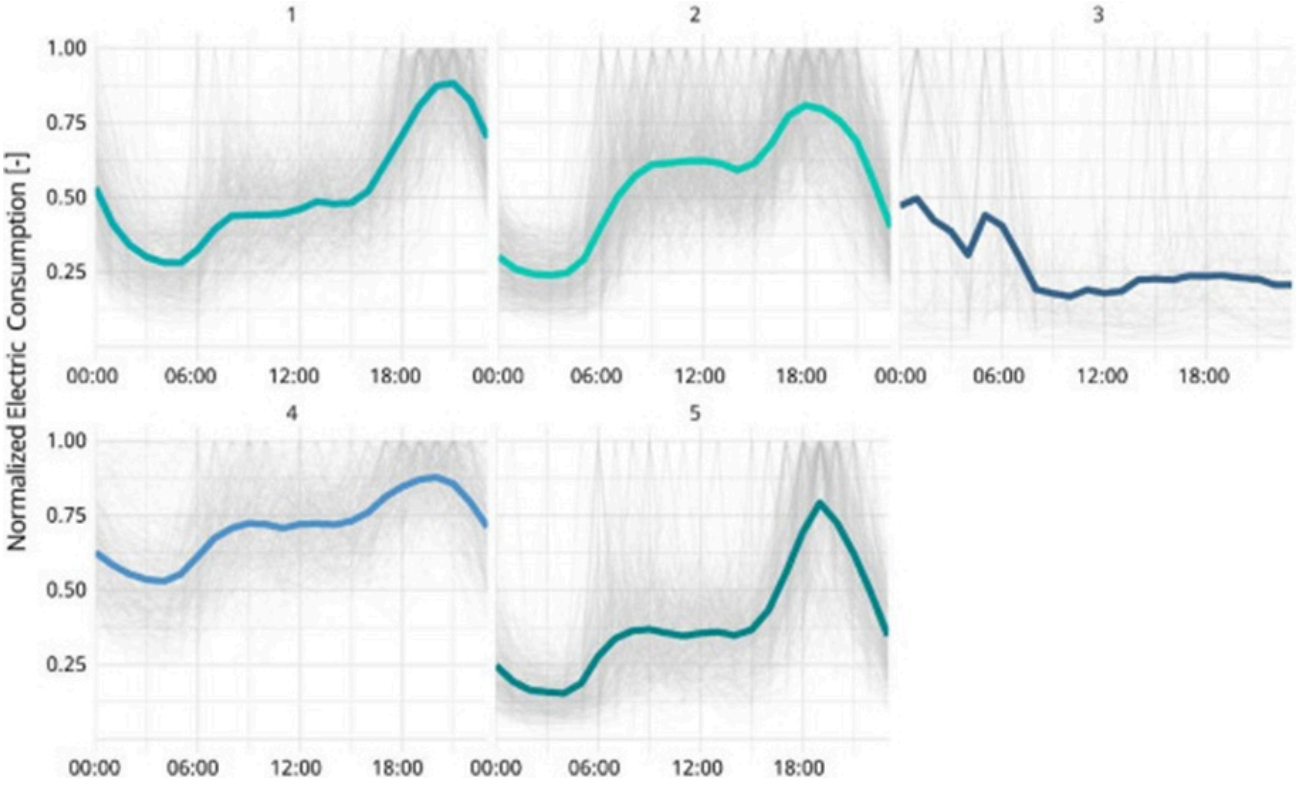


Figure 9. Typical electrical load patterns obtained through the load pattern recognition process.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

· Profile 1: Features a steep drop from 00:00 until roughly 05:00, followed by a steady climb into a flat, sustained load through the middle of the day. A sharp spike begins around 15:00, maxing out near 19:00, before dropping off again.

· Profile 2: Shows a swift load increase kicking off at 05:30, hitting an initial crest at 09:00. Usage plateaus until approximately 14:30, dips slightly, and then rockets to a secondary peak at 18:00. The load then falls off heavily overnight.

· Profile 3: Stands out as highly unusual compared to the rest. It carries a much lower overall average load, characterized by two tight peaks during the night. A rapid decline follows at 07:30, after which it flatlines at a remarkably low level compared to its nighttime usage.

· Profile 4: Generally sustains a higher baseline load than the other profiles. It shows a minor overnight drop, slowly builds to a flat midday plateau, forms a shallow peak near 19:30, and then steadily declines.

· Profile 5: Mirrors the first profile, but with shifted timeframes. Its nighttime drop is shallower, and its flat midday plateau rests lower than Profile 1. Additionally, its concentrated evening peak hits slightly earlier, at roughly 18:30.

After isolating these typical curves, a secondary screening was deployed to strip out sparsely populated clusters, perfectly mirroring the logic used to drop rare daily profiles during the data filtering step. Specifically, the thresholds for user and profile limits were calculated via **Eqs. 5 and 6.** The cutoffs were set at 10% of the total users and profiles to flag weak clusters, as visualized in **Figure 10.**

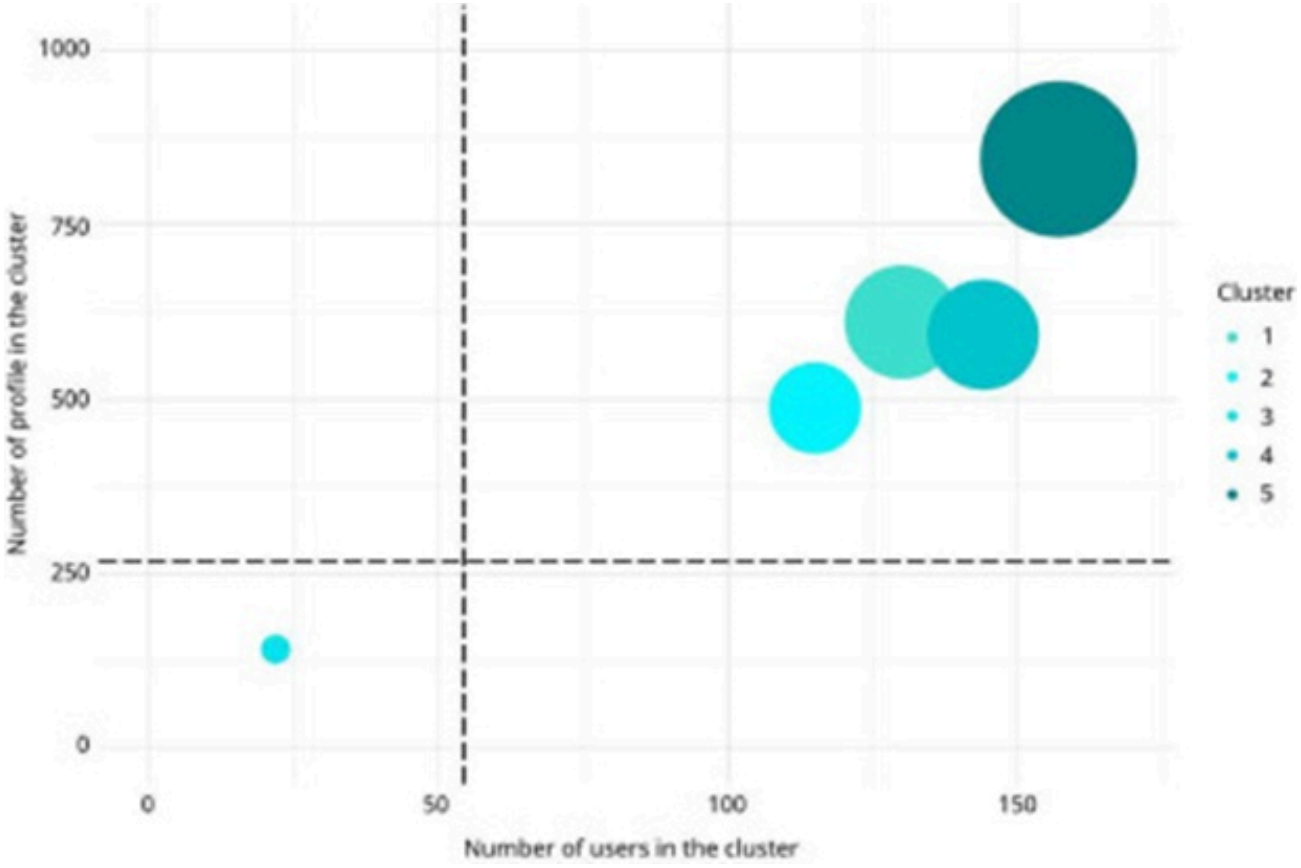


Figure 10. Elimination of the low-populated clusters for the load pattern recognition process.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

For this screening, the limits for user count and profile count were locked at 60 and 262, respectively. The chart clearly indicates that cluster 3 is the weakest group. As noted earlier, this aligns precisely with the highly anomalous third typological load profile.

To wrap up the load pattern recognition phase, typological profiles serve as the structural centroids for the remaining 4 clusters. These finalized patterns are then utilized to train a supervised classification model, empowering it to map specific consumption curves to users by analyzing their monthly usage figures.

Classification

The central aim of the classification phase is to construct a model that can reliably deduce a consumer's month-by-month electrical load curve for the entire year, using their monthly consumption metrics solely, as shown in **Figure 11**. Time of Use (ToU) blocks are structured around the London Power (LOND) pricing tiers, detailed in **Table 6**. These designated time bands remain identical every single day of the year.

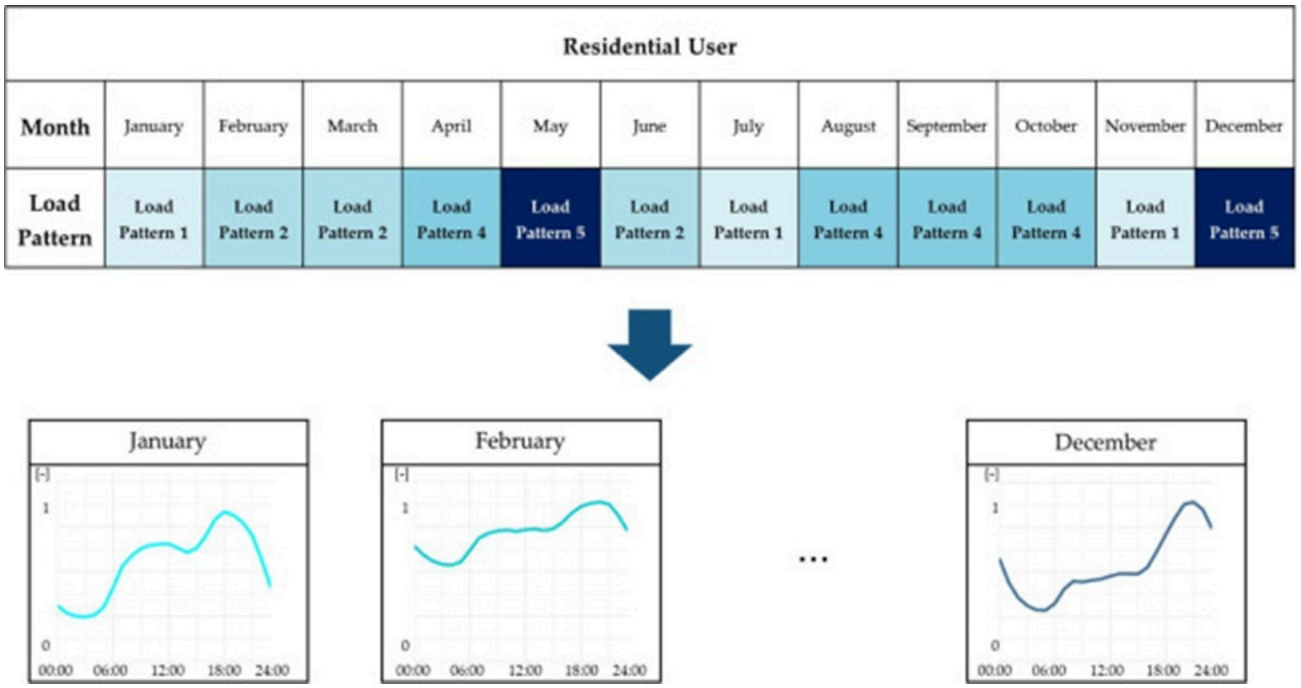


Figure 11. Classification task carried out for a generic residential user.

Table 6. The London Power company's ToU.

Time of use	Hours
Red	11:00–14:00 16:00–19:00
Amber	07:00–11:00 14:00–16:00 19:00–23:00
Green	00:00–07:00 23:00–24:00

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

Building this classification framework requires progressing through five distinct stages: i) choosing the classifier; ii) determining the training inputs; iii) establishing the testing and training data pools; iv) initializing the model; and v) optimizing the model parameters. The initial selection of the classifier is arguably the most vital step. Research by Fang *et al*. [26] highlights that Random Forest frameworks pair exceptionally well with k-Means clustering in classification setups. Consequently, a Random Forest algorithm was built employing the "RandomForest" package within the R coding environment.

Classifiers in this category are ensemble-based machine learning tools that leverage multiple decision trees (DTs). DTs function by categorizing data based on a rigid set of rules. They mimic a branching structure where raw data attributes serve as internal nodes, the logic rules form the branches, and the final aggregated verdicts act as leaf nodes. Random Forest models train these various trees by feeding them randomized slices of the dataset that match the total volume of the original training set. To classify a new object, the model aggregates the final predictions from all its decision trees. The specific variant utilized in this study is the "Ranger Forest," a highly accelerated Random Forest implementation engineered specifically to handle high-dimensional datasets.

Once the algorithm is chosen, the training inputs must be defined. Because the ultimate goal is to assign a specific load shape based on monthly records, the chosen training variables include raw monthly consumption figures, total monthly usage, and the proportion of electricity consumed within the distinct ToU windows, as laid out in **Table 7.**

Table 7. Training variables used to train the classification model.

Training variable	Description	Unit
Amber/Total	Share of monthly electrical consumption during the "Amber" ToU timeframe	[-]
Red/Total	Share of monthly electrical consumption during the "Red" ToU timeframe	[-]
Green/Total	Share of monthly electrical consumption during the "Green" ToU timeframe	[-]
Amber/Red	Electrical consumption in the "Amber" ToU timeframe relative to the "Red" ToU consumption	[-]
Green/Amber	Electrical consumption in the "Green" ToU timeframe relative to the "Amber" ToU consumption	[-]
Green/Red	Electrical consumption in the "Green" ToU timeframe relative to the "Red" ToU consumption	[-]
Green	Electrical consumption during the "Green" ToU timeframe	[kWh]
Amber	Electrical consumption during the "Amber" ToU timeframe	[kWh]
Red	Electrical consumption during the "Red" ToU timeframe	[kWh]
Total	Total monthly electrical consumption	[kWh]

To boost model accuracy during the training phase, additional variables were introduced. These new inputs match those listed in **Table 7**. Still, they are adjusted using a correction factor based on **Eq. 8.** The only exceptions to this adjustment are the total monthly electricity usage and the raw consumption metrics within the "Red", "Amber", and "Green" ToU brackets.

$$f_{m,y(j)} = \frac{T_{m,y}}{T_{m,j}} \quad (8)$$

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

Where:

· $T_{m,t}$ stands for the total hours of the t-th ToU tier during the m-th month.

· $T_{m,j}$ stands for the total hours of the j-th ToU tier during the m-th month.

The correction factor, f, accounts for the temporal proportions of the ToU periods within a given month. Specifically, it conveys the ratio of hours between two distinct ToU blocks—a nuance that is traditionally difficult to feed into a model relying solely on monthly consumption totals.

Next, the datasets for training and testing are established. The specific characteristics of these data collections are outlined in **Table 8.**

Table 8. Training and testing characteristics of the datasets.

Dataset category	Share of the complete dataset	Description
Training set	70%	A randomly selected 70% portion of the complete dataset used for training the classification model.
Testing set	30%	A randomly selected 30% portion of the complete dataset is employed to evaluate the performance of the classification model.

Once the algorithm is chosen and both the datasets and training variables are designated, the hyperparameters that yield the optimum classification performance must be determined. Random Forest classifiers rely on three primary hyperparameters: the total volume of globally generated trees, the number of variables evaluated at each split, and the minimum elements permitted in the leaf nodes.

The total volume of generated trees is a highly critical parameter. It is tightly linked to the database's scale and the balance between processing costs and overall accuracy. While the tree count can theoretically be set to any arbitrary figure, setting it anywhere from a few hundred to a few thousand typically yields solid results across most scenarios. Furthermore, model accuracy generally scales upward with the tree count, though these gains plateau after reaching a certain threshold. The variables per split dictate how many training features are evaluated to formulate a splitting rule at every step. Adjusting this number can drastically shift the model's effectiveness. Lastly, the minimum elements in the leaf nodes establish the lowest allowable data threshold for a terminal node, which directly governs the physical scale of the resulting tree. Setting a low minimum produces massive trees laden with split nodes. Conversely, a higher minimum yields compact trees featuring far fewer splits. For this study, the "tuneRanger" and "mlf" packages in the R programming environment were used to pinpoint the optimal blend of hyperparameters and classification performance. A total of 200 iterations were executed to test various hyperparameter combinations, while the generated tree count was held steady at 4000.

Additional parameters necessary for configuring the model and the optimization cycle include the split rule, importance, and measure. The importance metric dictates how the value of each training variable is scored, configured here to minimize node impurity using the Gini Index across observation groups. Similarly, the split rule relies on the Gini Index to pick the most effective classification variables at any given node. The measure parameter establishes the Key Performance Indicators (KPIs) used to flag the top-performing models throughout the optimization loop. For this task, overall model accuracy was selected as the prime KPI to assess classification success.

In summary, a repetitive optimization routine was executed to secure the most accurate model by finding the ideal hyperparameter mix, holding the tree count constant at 4000. The findings are detailed in **Table 9.**

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

Table 9. Hyperparameters of the random forest classifier obtained through an iterative optimization process.

Hyperparameter	Value	Description
Number of features considered at each split	2	Number of variables required to establish a decision rule
Minimum samples in leaf nodes	16	Number of samples contained in a terminal node
Additional random forest settings		
Feature importance criterion	Gini Index	Metric used to assess variable importance
Splitting criterion	Gini Index	Criterion applied to determine node splits
Evaluation metric	Accuracy	Performance measure used for model assessment

The optimal Random Forest model utilized 2 variables per split. The minimum number of elements in the leaf nodes—a pivotal factor balancing processing cost and predictive capability—was set at 16. This represents under 1 % of the total elements evaluated by the comprehensive model. The classifier's performance, defined as its success rate in making correct categorizations relative to total attempts, achieved accuracies of 66% during the training phase and 63% during the testing phase.

This relatively modest accuracy stems primarily from the fact that the training variables failed to perfectly encapsulate the typological load profiles identified during the pattern recognition stage. Consequently, the algorithm struggles to assign exact load types relying strictly on these inputs. Enhancing the model's predictive power would necessitate extra behavioral or demographic data, such as household occupancy, daily habits, or employment types. However, because this research explicitly operates under conditions of data scarcity, those details were omitted.

A hierarchy of the most critical predictive variables, ranked by the model using the Gini Index, is displayed in **Figure 12.**

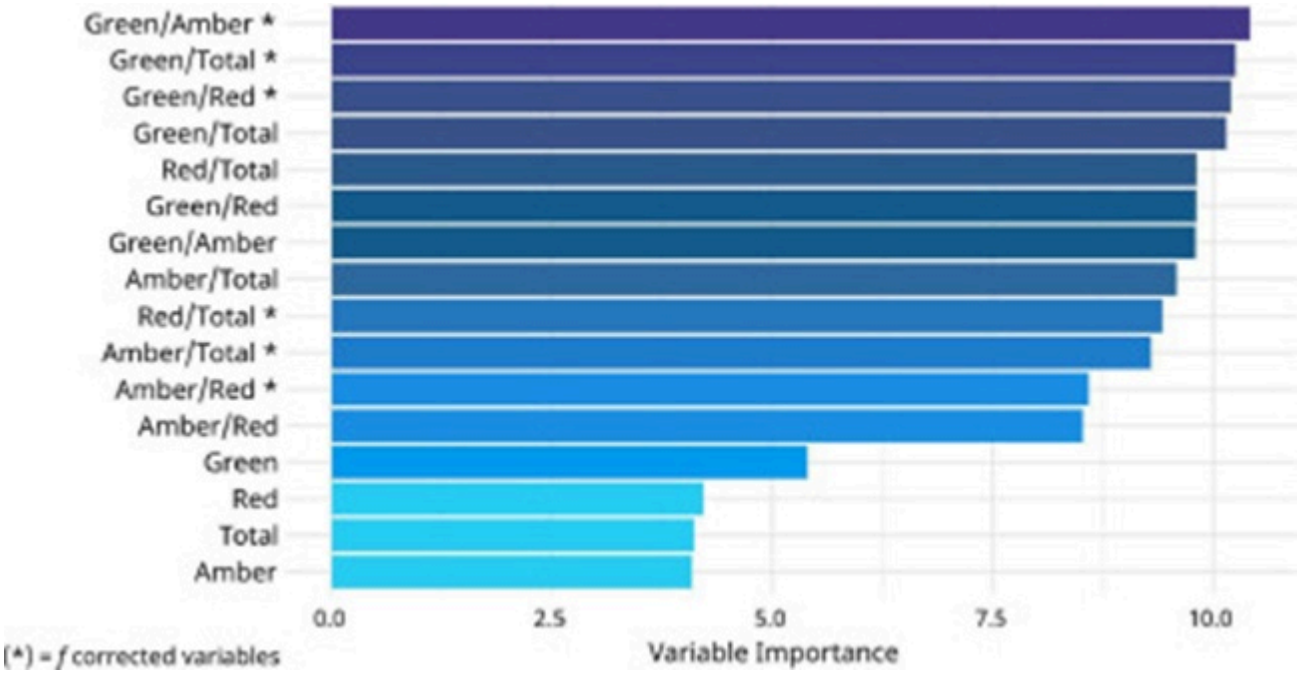


Figure 12. Importance of the variables for the classification process. (*) variables corrected with the correction factor "f".

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

Figure 12 ranks the analyzed variables by their descending impact. Notably, the top 3 most crucial variables are those modified by the correction coefficient "f". This highlights how valuable this supplementary data is for empowering the Random Forest algorithm. Additionally, the chart reveals that absolute monthly electricity totals and raw consumption within specific ToU bands hold drastically less weight than the other metrics. This indicates that relative usage ratios across different ToU periods are significantly more critical than raw consumption totals when attempting to deduce a consumer's load profile shape.

Scaling coefficient evaluation and case study application phase

This section details the essential steps required to compute the scaling coefficients used to transition from the standard normalized load profiles—established during the load pattern recognition stage and allocated to individual users every month during classification—to actual load profiles. This represents the concluding step of the model development phase and links directly with the case study application stage, as illustrated in **Figure 13.**

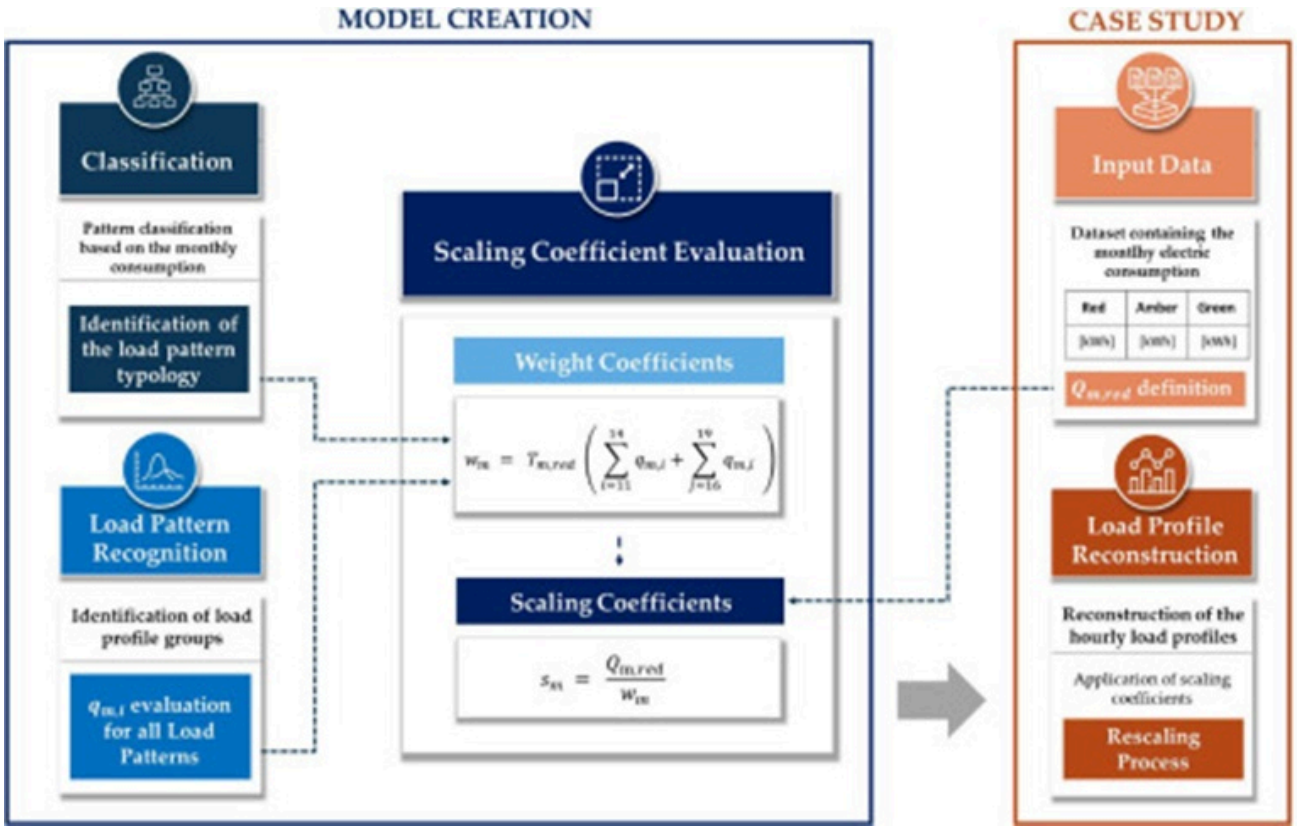


Figure 13. Sequence of steps for evaluating the scaling coefficients.

The factors employed during this evaluation to proportion the monthly consumption volumes for the "Red" Time of Use (ToU) tariff are determined using the normalized standard profile values derived from the pattern recognition stage. These are subsequently assigned to each customer for every month during the classification procedure. The coefficients applied to scale the monthly consumption quantities for the "Red" ToU period are computed based on **Eq. 9.**

$$w_m = T_{m,red}\left(\sum_{i=11}^{14} q_{m,\,i} + \sum_{j=16}^{19} q_{m,j}\right) \quad (9)$$

Where:

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

· $T_{m,red}$ denotes the total hours of the "Red" ToU tariff for the m-th month.

· $q_{m,\,i}$ represents the value of the normalized standard load profiles at the i-th hour within the timeframe specified by the "Red" ToU (specifically, between 11:00 and 14:00, and from 16:00 to 19:00, for the m-th month), acting as the weight for the monthly consumption magnitude.

As demonstrated in **Eq. 9** and **Figure 14**, to establish the weighting coefficients, the data points falling inside the "Red" ToU window were utilized for each designated typological load pattern.

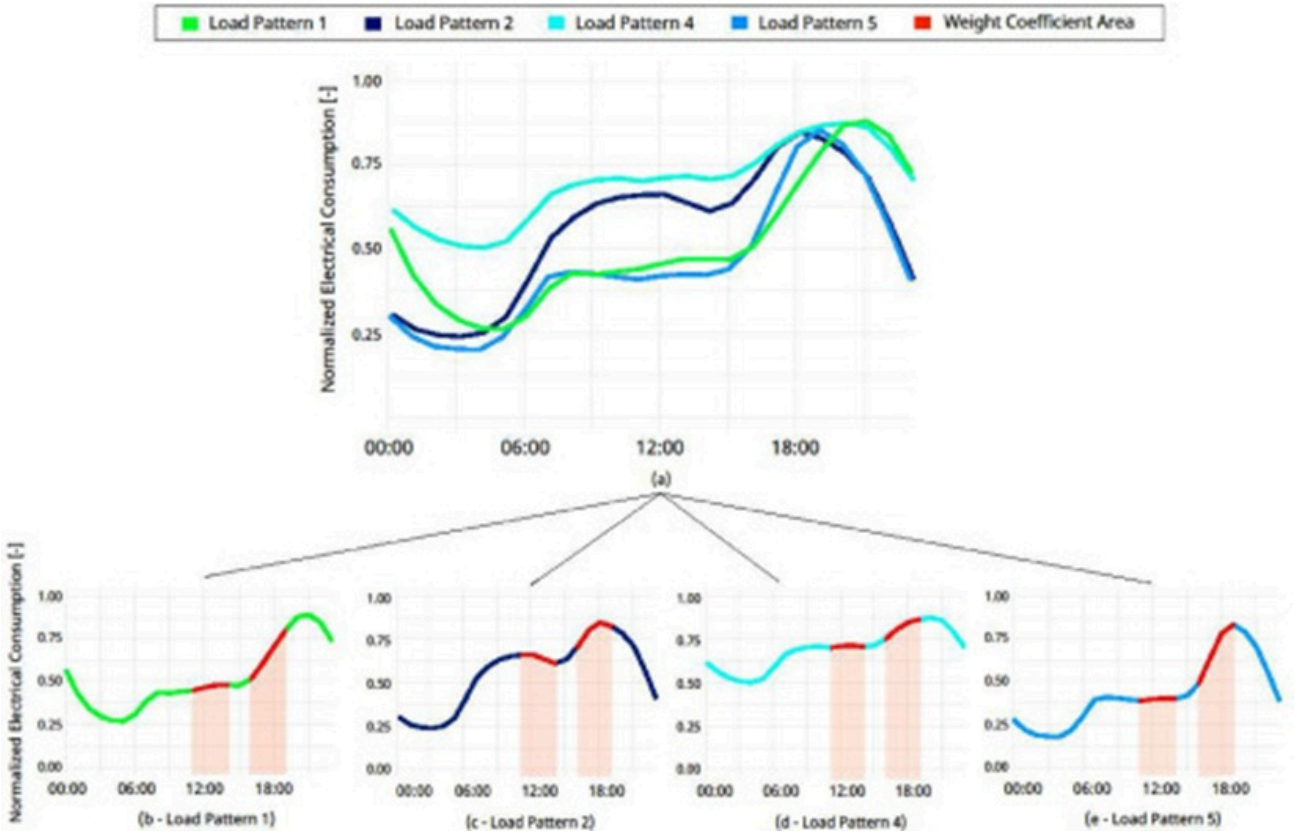


Figure 14. (a) Contrasting standard load patterns; (b) standard load profile figures within the "Red" ToU window for load pattern 1; (c) standard load profile figures within the "Red" ToU window for load pattern 2; (d) standard load profile figures within the "Red" ToU energy consumption interval for load pattern 4; (e) standard load profile figures within the "Red" ToU window for load pattern 5.

An additional comparative analysis of the four distinct typological load profiles generated after the pattern recognition stage is presented in **Figure 14a**. This highlights their variations, which dictate the different coefficients required to translate standard profiles into simulated ones. Meanwhile, **Figures 14b–14e** display visual plots of the typological pattern segments used to determine the scaling coefficients that weigh the monthly consumption amounts. The charts reveal that Load Pattern 4 stands out as the most unique typological profile due to its maximum load intensity. The other patterns share similarities, especially Load Pattern 1 and Load Pattern 5, whose main differences lie in their nighttime curve shapes and the specific timing of the evening peak (which occurs at a slightly different time in Load Pattern 1 compared to Load Pattern 5). Furthermore, while Load Pattern 2 closely mirrors Load Pattern 5 at night, it diverges noticeably during the middle of the day by exhibiting a stronger load intensity.

Once the weight coefficients are acquired, the scaling multipliers needed to shift from the standard normalized patterns to actual load profiles are calculated via **Eq. 10.**

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

$$s_m = \frac{Q_{m,\ red}}{W_m} \quad (10)$$

Where:

$Q_{m,\ red}$ stands for the monthly usage during the “Red” ToU period for the m-th month.

The consumption amount within the “Red” ToU window was selected as the reference point for the scaling factor because it typically encompasses the highest consumption intervals for residential consumers. It also spans the majority of the active consumption period, representing roughly 50 % of the hours where residential users show medium-to-high electricity usage, namely between 8:00 and 20:00. These varying contours of the typological profiles resulted in distinct scaling coefficients, as detailed in **Table 10.**

Table 10. Values incorporated to compute the scaling multipliers applied to weight monthly usage figures.

Load pattern	Weight coefficient (s_m)	18:00	17:00	16:00	13:00	12:00	11:00
1	18.48	0.69	0.58	0.49	0.45	0.44	0.43
2	24.66	0.82	0.76	0.69	0.59	0.62	0.63
4	27.30	0.84	0.81	0.76	0.72	0.71	0.71
5	11.04	0.73	0.58	0.45	0.36	0.37	0.36

The most substantial weighting coefficients correspond to typological load patterns with higher intensity, specifically Load Patterns 2 and 4. As a result, these patterns will exert a stronger impact during the scaling procedure. In contrast, the effect will be more subdued for months categorized by the classifier under Load Patterns 1 and 5.

Ultimately, simulated electricity consumption profiles for each month are generated by multiplying the s_m factor by the normalized load profiles acquired earlier, utilizing **Eq. 11.**

$$\hat{Q}_{m,h} = q_{m,h} * s_m \forall i \in A \quad (11)$$

Where:

· A = {January, February, …., December};

- $_{m,h}$ denotes the modeled consumption amount for the m-th month at hour h;
- $q_{m,h}$ signifies the normalized consumption curve value for the m-th month at hour h;
- s_m acts as the scaling multiplier for the m-th month.

As noted earlier, the culmination of this stage yields a monthly electrical load profile for every user, as illustrated in **Figure 15.**

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

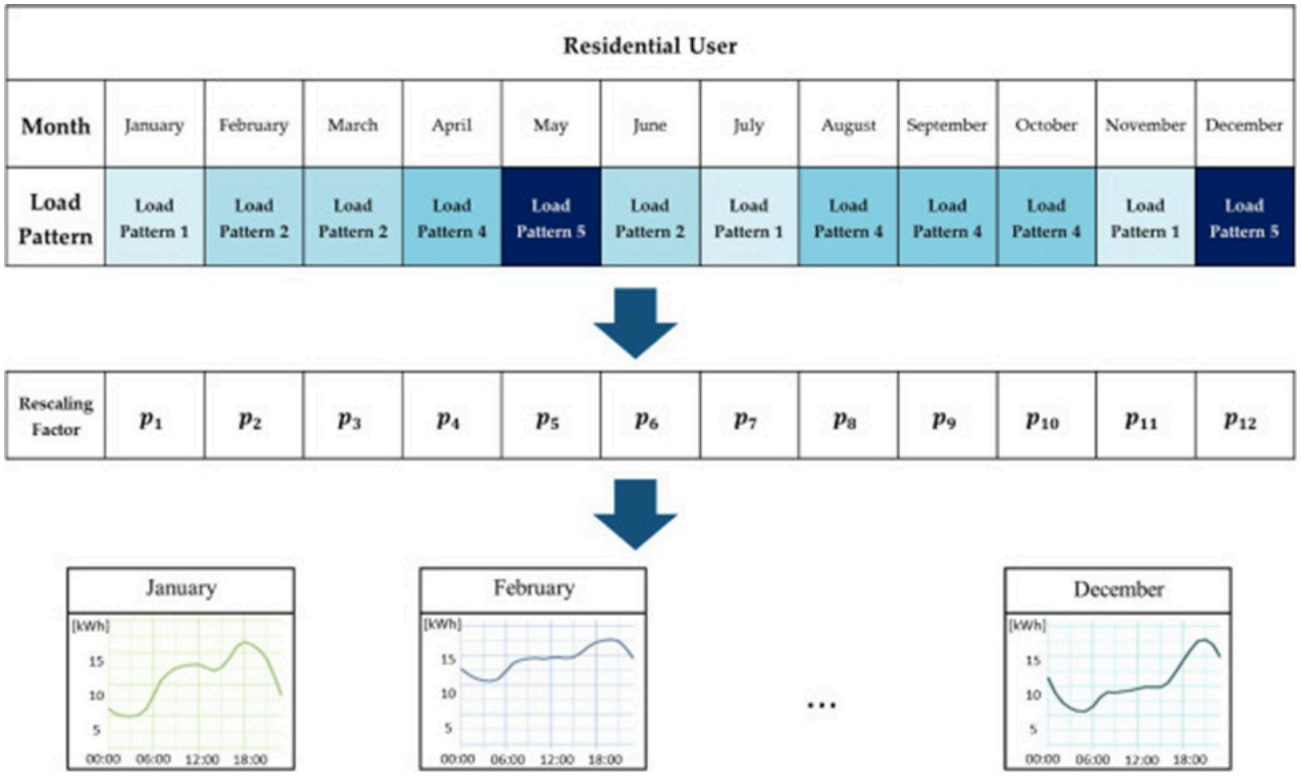


Figure 15. Execution diagram of the scaling coefficient procedure.

At this juncture, mean electrical load profiles are assembled for every user across each month. These profiles illustrate the typical behavioral patterns of the user throughout a given month. Consequently, the hourly electrical load profile covering the full year can be accurately mapped by assuming that the derived profile mirrors the daily pattern of every single day within that month.

Validation

The verification of the outcomes in this analysis was performed utilizing a collection of hourly and monthly recorded consumption statistics [32]. This involved reconstructing the load profiles via the outlined techniques and evaluating them against actual load curves. Köhler *et al*. [39] offered a comprehensive examination of the metrics routinely utilized to benchmark forecasted or artificial load profiles against real ones. Many of the similarity assessment metrics introduced in Köhler *et al*. [39] focus on the statistical traits of both actual and modeled profiles, including the minimum and maximum boundaries, median, standard deviation, and load duration curve error. Furthermore, complexity-driven metrics, like the peak count and fractal dimension, are also recommended.

For this research, the equivalence between the generated and actual load profiles was measured using the Normalized RMSE (NRMSE) and the Normalized MAE (NMAE). This was evaluated for both the full year and the specific timeframe featuring simultaneous aggregate consumption and aggregate power plant generation (NMAESE, NRMSESE), applying **Eqs. 12–15.**

$$NMAE = \frac{\sum_{i=1}^{n} \frac{|Q_i - \hat{Q}_i|}{n}}{\bar{Q}_i} \quad (12)$$

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

$$NMAE_{SE} = \frac{\sum_{i=1}^{N} \frac{\left|Q_i^* - \widehat{Q_i^*}\right|}{N}}{\bar{Q}_i^*} \quad (13)$$

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}(Q_i - \widehat{Q_i})^2}{n}}}{\bar{Q}_i} \quad (14)$$

$$NRMSE_{SE} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(Q_i^* - \widehat{Q_i^*})^2}{n}}}{\bar{Q}_i^*} \quad (15)$$

Where:

- Q_i indicates the actual aggregate consumption.
- $\widehat{Q_i}$ indicates the modeled aggregate consumption.
- Q_i^* represents the actual aggregate consumption during the simultaneous occurrence of consumption and energy generation.
- $\widehat{Q_i}^*$ represents the modeled aggregate consumption during the simultaneous occurrence of consumption and energy generation.
- n stands for the total count of observations.

Assessing the NMAE and NRMSE during periods of simultaneous aggregate consumption and aggregate power production helps gauge how well the proposed methodology reconstructs the aggregate electrical load during the critical hours needed for computing shared energy. This factor is intimately tied to the shared energy estimation, which serves as the core objective of this study. Additionally, the mean relative absolute error (MRAE) and relative absolute error (RAE) for shared energy are calculated monthly and yearly, respectively, via **Eqs. 16 and 17.**

$$MRAE = \frac{\sum_{m=1}^{M} \frac{\left|SE_m - \widehat{SE_m}\right|}{SE_m}}{M} \quad (16)$$

$$RAE = \frac{\left|SE_y - \widehat{SE_y}\right|}{SE_y} \quad (17)$$

Where:

· SE_m stands for the actual shared energy for the m-th month.

· $\widehat{SE_m}$ stands for the modeled shared energy for the m-th month.

· SE_y denotes the actual shared energy for the y-th year.

· $\widehat{SE_y}$ denotes the modeled shared energy for the y-th year.

· M indicates the overall count of months.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

The shared energy is assessed across hourly, monthly, and yearly intervals by utilizing **Eqs. 18–20.**

$$SE_h = min\left(E_{cons,h}; E_{prod,h}\right) \quad (18)$$

$$SE_m = \sum_{i=1}^{H} SE_i \quad (19)$$

$$SE_y = \sum_{i=1}^{M} SE_m \quad (20)$$

Where:

- $E_{cons,h}$ represents the aggregate usage for the h-th hour.
- $E_{prod,h}$ represents the power generation from the facilities linked to the aggregate usage for the h-th hour.
- SE_h indicates the shared energy for the h-th hour.
- SE_y indicates the shared energy for the y-th year.
- H is the total number of hours contained in a month.

Case Study Application

This section delineates the outcomes of implementing the recommended framework on the REC empirical study, which comprises 37 consumers selected at random from the dataset. To prevent an unrealistic evaluation of system performance, these consumers were drawn from a dataset segment that excluded the individuals utilized during the training phase of the classification step. The generation facility linked to the REC consists of a 70 kWp ground-mounted solar photovoltaic system situated on the outskirts of London. Its hourly generation potential is modeled using PVGIS [40], and the optimal tilt angle and orientation of the solar arrays are determined via the same software to maximize the system's energy output, set at 40 and −5°, respectively. Specifically, the orientation (or azimuth) represents the angle of the PV arrays relative to true South (where −90° represents East, 0° signifies South, and 90° denotes West). PVGIS is a simulation program that integrates solar radiation and ambient temperature records to precisely model the energy generation capability of photovoltaic installations.

Before assessing how effectively the recommended approach reconstructs the combined electrical demand profile, a qualitative assessment contrasting actual and modeled electricity patterns was conducted at the single-consumer scale for a selection of participants. Specifically, the modeled consumption curves for four distinct consumers are illustrated in **Figures. 16a–16d,** where they are contrasted against the reconstructed patterns for January, April, July, and October, respectively.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

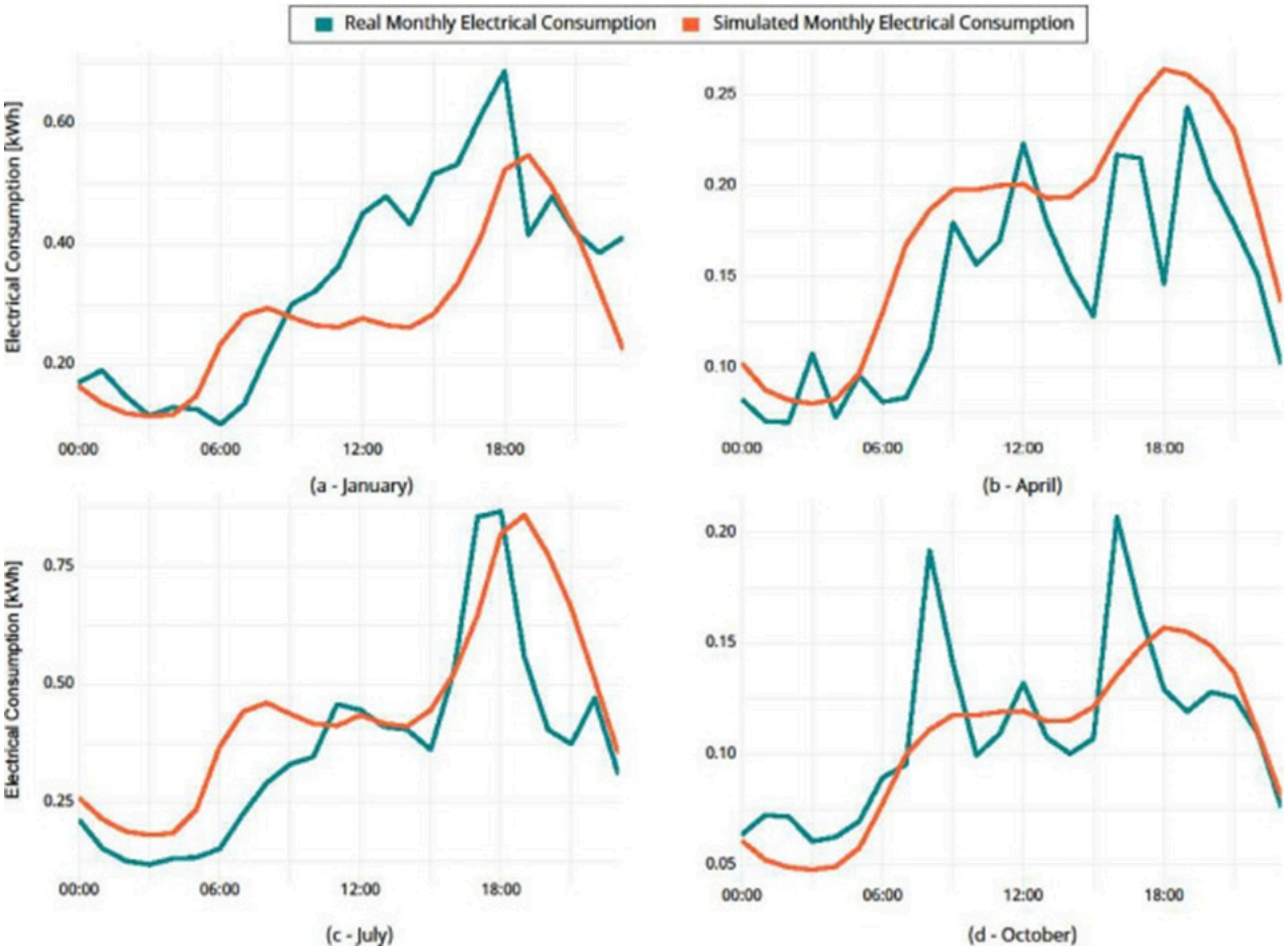


Figure 16. (a) Contrast between modeled and actual monthly mean power usage at the individual customer tier, for January, within a single REC participant; (b) contrast between modeled and actual monthly mean power usage at the individual customer tier, for April, within a single REC participant; (c) contrast between modeled and actual monthly mean power usage at the individual customer tier, for July, within a single REC participant; (d) contrast between modeled and actual monthly mean power usage at the individual customer tier, for October, within a single REC participant.

As illustrated, the actual and modeled profiles display prominent discrepancies during particular time blocks, primarily corresponding to demand peaks. For example, the peak loads in **Figures 16a and 16d** are notably under-predicted and temporally shifted within the simulated load curves. Conversely, certain consumption peaks in **Figures 16b and 16c** are over-predicted, or entirely missing from the simulated profiles. This occurs because the normalized centroids derived from the Load Pattern Recognition stage served as the baseline templates during the hourly profile reconstruction process to establish the actual electrical demand profile of a specific scale; consequently, they mirror the average electricity usage of a distinct statistical cohort. As a result, these modeled demand curves may reflect collective behavior rather than individual consumer patterns. Nonetheless, the primary objective of this methodology is the precise reconstruction of demand profiles at the collective scale. Thus, deploying normalized centroids as a baseline for the combined profile aligns logically with the goals of this study. The framework's capacity to model the aggregate demand curve—representing the combined total of all 37 consumers—is demonstrated in **Figures 17a–17d**. Here, the modeled demand trajectory is mapped against the actual trajectory for January, April, July, and December, revealing a strong alignment between the simulated collective profile and the real data.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

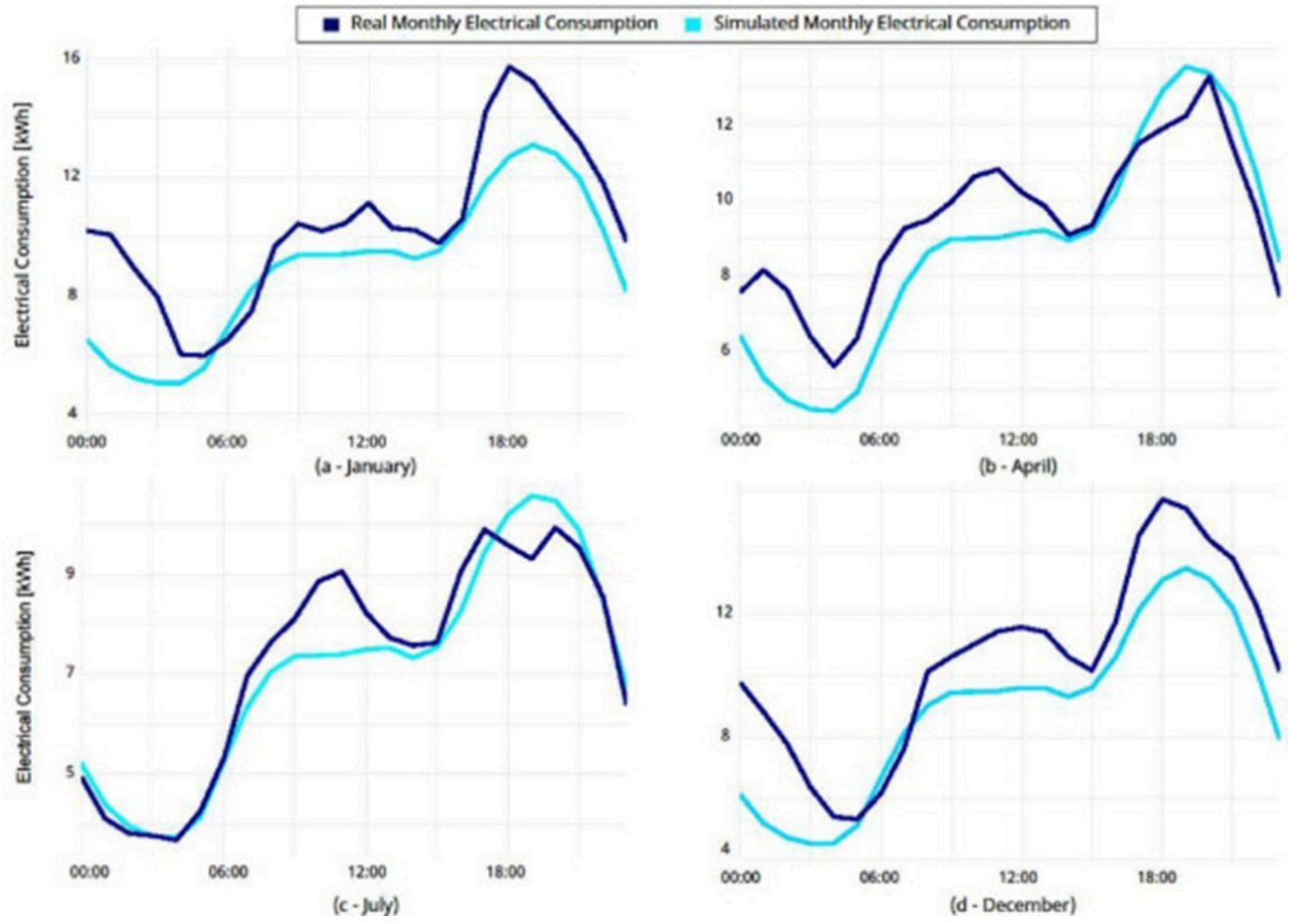


Figure 17. (a) Evaluation of modeled versus actual monthly mean power usage at the collective tier for January; (b) evaluation of modeled versus actual monthly mean power usage at the collective tier for April; (c) evaluation of modeled versus actual monthly mean power usage at the collective tier for July; (d) evaluation of modeled versus actual monthly mean power usage at the collective tier for December.

Utilizing the representative archetypes from the load pattern extraction process—which act as the cluster centroids established during that phase—yields consumption profiles that reflect group averages rather than individual consumer habits, as previously noted. The statistical outcomes of comparing the actual collective curve against the modeled version, using the parameters, are compiled in **Table 11.**

Table 11. Performance parameters assessed by contrasting the modeled demand curve against the actual curve.

Parameter	Value
NMAE	20.04%
$NMAE_{SE}$	18.34%
NRSME	26.17%
$NRSME_{SE}$	23.87%
MRAE	8.31%
RAE	0.12%

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

The analysis yields an NMAE of 20.04% and an NMAESE of 18.34%, alongside an NRMSE and NRSMESE of 26.17% and 23.87%, respectively. These metrics confirm that the proposed framework successfully enhances the reconstruction accuracy of the combined load profile, particularly during periods where REC electrical demand coincides with power plant generation. This performance is primarily driven by the scaling mechanism configuration, where the pivotal element driving the shift from standard archetypes to actual load curves is the monthly consumption within the “Red” time-of-use block, reflecting energy usage during peak midday hours.

The calculated error for shared energy registers at 8.31% on a monthly index (MRAE) and 0.12% on an annual timeline (RAE) demonstrates that the modeled curve provides highly accurate monthly estimates of shared energy. The monthly variance of the RAE, which informs the calculation of the MRAE, is presented in **Table 12.**

Table 12. Standardized profile metrics applied to determine scaling weights for monthly power usage.

Parameter	Month	Value
RAE_m	January	13.32 %
	February	12.83 %
	April	5.84 %
	May	1.05 %
	June	2.27 %
	July	5.62 %
	August	3.14 %
	September	8.48 %
	October	18.06 %
	November	11.38 %
	December	9.46 %

The data indicates that the framework is more adept at forecasting monthly shared energy during the summer season. In contrast, it shows elevated relative errors throughout the other periods of the year. This pattern stems largely from the consumer traits of the residential group under analysis. In this specific scenario, the electrical demand curves of the study participants exhibit a stronger reliance on energy usage during the “Red” time-of-use window throughout the summer. This correlation diminishes in other seasons, except in April and May, which post RAE metrics comparable to the summer months. This alignment is due to their proximity to summer, signaling a transition toward consumption habits tightly linked to the “Red” tariff period. Ultimately, the relative error for annual shared energy estimation is 0.12 %, dropping significantly below the MRAE. This discrepancy likely occurs because modeling errors and deviations between actual and simulated curves balance out over an extended twelve-month timeframe. Consequently, negative variances in certain months are countered by positive variances in others, culminating in a minimized cumulative error for the full year.

Results and Discussion

As highlighted in the introductory section of this study, most existing literature on demand profiling relies heavily on high-frequency temporal data, augmented by detailed parameters regarding household appliances, occupancy counts, and user behavioral traits. In contrast, this study introduces an approach that bypasses the need for such information-dense

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

datasets, which are frequently unavailable in practical scenarios. Furthermore, this investigation targets the integration of load profiling workflows within RECs to generate more precise projections of their core energy metric: shared energy. This focus introduces a novel perspective to the domain, contrasting with traditional profiling methods that center on high-resolution demand reconstruction for single users. Additionally, this framework can be deployed even within nations boasting advanced smart meter adoption rates [21] to bypass the scarcity of open-access data, thereby refining energy flow projections and strengthening REC financial viability assessments.

The resulting data indicates that the proposed approach successfully synthesizes a collective electrical load curve using only the monthly consumption figures of individual consumers. Broadly speaking, the model can undergo training with datasets sourced from varied geographical regions, allowing it to adaptively capture typical demand profiles and their interplay with energy usage across different time-of-use (ToU) price brackets. For this particular study, the LCL project dataset was chosen primarily due to its expansive volume of data [32]. Utilizing an alternative data source, such as one from another country, would expectedly alter the outcomes. This divergence would begin at the load pattern identification stage, where distinct electrical demand configurations would likely surface due to differing consumer routines, appliance ownership distributions, or structural building features.

Furthermore, the correlation between usage ratios across diverse ToU tiers and overall load configurations would likely vary, altering the results of the classification step. This shift could yield a different set of critical parameters for classification than those identified for this paper's specific case study. It is also critical to recognize that classifier accuracy depends heavily on the volume of typical curves to be categorized, their geometric profiles, and how they relate to consumption values across ToU tiers. The algorithm would face a substantial challenge in differentiating distinct load shapes if they map to nearly identical electricity usage figures. On the other hand, the classification sequence becomes straightforward when unique load shapes map to distinctly contrasting consumption volumes. In the former case, a different classification framework might be required instead of the one deployed here. In contrast, in the latter case, the current model would likely achieve superior classification accuracy due to the robust correlations emerging between the new load configurations and ToU structures.

The growth in the quantity of representative load curves might also happen due to the anticipated electrification of energy usage in the near future [41-44]. Truly, adopting electric heating systems and electric vehicles could substantially alter the power demand patterns of residential consumers. For example, integrating electric vehicles might create additional overnight power needs, thereby modifying the load profile's structure during those hours. However, how these novel technologies affect residential load patterns depends heavily on system configurations (e.g., whether battery storage is paired with photovoltaic arrays) and on how users interact with these systems based on daily routines, complicating forecasting efforts. Broadly speaking, it is reasonable to anticipate that the shift toward electrified consumption will expand the variety of standard load profiles identified during the pattern recognition stage. This would subsequently cause the classifier's performance to either improve or degrade, depending on the two previously discussed scenarios.

In practical applications where hourly or sub-hourly records are typically missing, a prevailing technique to calculate the shared energy produced by REC members involves applying standard power consumption models, frequently derived from country-wide statistical studies. These templates, reflecting the typical consumption patterns of a massive user base, often fail to accurately mirror the habits of consumers in specific, smaller regions. Since RECs represent localized, geographically concentrated groups of members, it is reasonable to anticipate that the energy demands of REC participants will normally diverge—sometimes drastically—from standard models. Therefore, relying on these generic templates, which usually offer just one or two profiles per consumer class (such as residential or industrial), can cause substantial inaccuracies when calculating shared energy. Conversely, the framework presented here employs typological curves (derived during the load pattern identification step) that reflect a highly specific, smaller group of consumers, namely the actual REC members. Thus, the typological curves pinpointed by this approach closely capture the habits of REC participants and offer a greater variety than the generic models derived from broad statistical reviews. Consequently, the introduced methodology improves the accuracy of shared energy projections by relying on a wider array of

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

representative load curves customized to the REC members, rather than using the generic templates standard across various nations.

As noted earlier, numerous existing studies conduct load profiling utilizing datasets with fine-grained temporal data and supplementary details. Moreover, these approaches frequently concentrate on recreating the power demand of single consumers instead of a collective group. Within the examined literature, the research by Lazzeroni *et al*. [31] models the hourly consumption of single consumers using their monthly power usage statistics. Consequently, their study shares similarities with this paper but targets a distinct goal: recreating the power demand of an individual consumer utilizing inputs comparable to those in our study. Their findings at the single-consumer scale yielded an NMAE of roughly 26%. In contrast, our introduced framework, operating at the collective scale, attains a reduced NMAE of 20% when assessing annual profiles, and 18% during periods of overlap between collective power usage and generation from the energy facilities. This highlights the competitive load profiling capabilities of the framework at the aggregate tier.

Conclusion

In this study, we introduce a robust methodology to tackle the shortage of high-frequency energy usage records for REC participants during the “design” stage of an REC project. This is the period when simulations and feasibility studies are necessary to evaluate the project’s viability and establish the ideal REC configuration. Ultimately, we provide a framework that prospective aggregators can use to measure an REC’s energy dynamics during preliminary assessments precisely. We formulated a non-invasive technique to model the hourly power demand curves of household consumers, aiming to calculate the shared energy inside RECs. This strategy required thorough data preparation and the formulation of a load pattern identification procedure to pinpoint standard electricity demand curves. Subsequently, we built a classification mechanism capable of allocating load curve shapes by utilizing predictive features drawn from monthly utility invoices. We then modeled the hourly power demand curves for an REC case study through a data-backed rescaling mechanism, which relies on assessing the scaling factors linked to the period of highest energy consumption. Our investigation showed NMAE and NRMSE values of roughly 20% and 26%, respectively, when comparing the modeled collective load curve to the actual one.

Furthermore, we evaluated the NMAE and NRMSE during times of simultaneous collective consumption and power plant generation, achieving figures of approximately 18% and 23%. Additionally, we tested the method’s capacity to calculate shared energy inside an REC, yielding MRAE and RAE figures of 8.31% and 0.12%, respectively. These outcomes validate the effectiveness of our proposed approach, proving it can accurately project shared energy within RECs, even under the limited-data conditions typical of an REC project’s “design” phase.

Looking ahead, our prospective research initiatives will involve applying various datasets—both open-source and proprietary—to prove the framework’s versatility across diverse scenarios. We also plan to investigate different classification algorithms and scaling techniques, while analyzing shared energy estimation inaccuracies as they relate to the distinct features and scales of RECs. Moreover, adapting this proposed methodology to cover consumer categories outside of the residential domain offers exciting possibilities for future progress in the field of RECs.

Nomenclature

Q = Real Consumption (kWh)

$\hat{Q}$ = Simulated Consumption (kWh)

E = Energy (kWh)

T = Temporal extension (h)

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

q = Normalized consumption

TV = Threshold Value

C = Cluster

s = Scaling coefficient

w = Weight coefficient

NRMSE = Normalized Root Mean Square Error

NMAE = Normalized Mean Absolute Error

NRAE = Normalized Relative Absolute Error

Subscripts

h = Hour

m = Month

y = Year

t = Time of Use rate

u = User

p = Profile

max = Maximum

cons = Consumption

prod = Production

avg = Average

SE = Shared Energy

Acknowledgements

This publication is part of the project “Network 4 Energy Sustainable Transition – NEST”, Project code PE0000021, Concession Decree No. 1561 of 11.10.2022 adopted by Ministero dell’Università e della Ricerca (MUR), CUP E13C22001890001. Project funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.3 - Call for tender No. 341 of 15.03.2022 of MUR; funded by the European Union – NextGenerationEU.

This study was carried out within the Ministerial Decree no. 1062/2021 and received funding from the FSE REACT-EU - PON Ricerca e Innovazione 2014–2020. This manuscript reflects only the authors’ views and opinions; neither the European Union nor the European Commission can be considered responsible for them.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 30 Oct 2022 Revised: 05 Jan 2023 Accepted: 20 Jan 2023
Published online: 15 June 2024

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Dóci G, Vasileiadou E, Petersen AC. Exploring the transition potential of renewable energy communities. Futures. 2015;66:85-95. https://doi.org/10.1016/j.futures.2015.01.002.

2. Lowitzsch J, Hoicka CE, Van Tulder FJ. Renewable energy communities under the 2019 European clean energy package – governance model for the energy clusters of the future? Renew Sustain Energy Rev. 2020;122:109489. https://doi.org/10.1016/j.rser.2019.109489.

3. Hanke F, Guyet R, Feenstra M. Do renewable energy communities deliver energy justice? Exploring insights from 71 European cases. Energy Res Soc Sci. 2021;80:102244. https://doi.org/10.1016/j.erss.2021.102244.

4. Minuto FD, Lanzini A, Giannuzzo L, Borchiellini R. Digital platforms for renewable energy communities projects: an overview. Int J Sustain Dev Plann. 2022;17(7):2007-13. https://doi.org/10.18280/ijsdp.17070.

5. Heldeweg MA, Saintier S. Renewable energy communities as 'socio-legal institutions': a normative frame for energy decentralization? Renew Sustain Energy Rev. 2020;119:109518. https://doi.org/10.1016/j.rser.2019.109518.

6. Caramizaru A, Uihlein A. Energy communities: an overview of energy and social innovation. Luxembourg: Publications Office of the European Union; 2020. Report No.: 30083. https://doi.org/10.2760/180576.

7. Gjorgievski V, Cundeva S, Georghiou GE. Social arrangements, technical designs and impacts of energy communities: a review. Renew Energy. 2021;169:1138-56. https://doi.org/10.1016/j.renene.2021.01.078.

8. Hoicka CE, Lowitzsch J, Brisbois MC, Kumar A, Camargo LEA. Implementing a just renewable energy transition: policy advice for transposing the new European rules for renewable energy communities. Energy Policy. 2021;156:112435. https://doi.org/10.1016/j.enpol.2021.112435.

9. Bashi MH, De Tommasi L, Cam AL, Relaño LS, Lyons P, Mundó J, et al. A review and mapping exercise of energy community regulatory challenges in European member states based on a survey of collective energy actors. Renew Sustain Energy Rev. 2023;172:113055. https://doi.org/10.1016/j.rser.2022.113055.

10. Volpato G, Carraro G, Cont M, Danieli P, Rech S, Lazzaretto A. General guidelines for the optimal economic aggregation of prosumers in energy communities. Energy. 2022;258:124800. https://doi.org/10.1016/j.energy.2022.124800.

11. Weckesser T, Dominković DF, Blomgren EMV, Schledorn A, Madsen H. Renewable energy communities: optimal sizing and distribution grid impact of photo-voltaics and battery storage. Appl Energy. 2021;301:117408. https://doi.org/10.1016/j.apenergy.2021.117408.

12. Bartolini A, Carducci F, Muñoz C, Kyrki V. Energy storage and multi energy systems in local energy communities with high renewable energy penetration. Renew Energy. 2020;159:595-609. https://doi.org/10.1016/j.renene.2020.05.131.

13. Bianchi FR, Bosio B, Conte F, Massucco S, Mosaico G, Natrella G, et al. Modelling and optimal management of renewable energy communities using reversible solid oxide cells. Appl Energy. 2023;334:120657. https://doi.org/10.1016/j.apenergy.2023.120657.

14. Fioriti D, Frangioni A, Poli D. Optimal sizing of energy communities with fair revenue sharing and exit clauses: value, role and business model of aggregators and users. Appl Energy. 2021;299:117328. https://doi.org/10.1016/j.apenergy.2021.117328.

15. Minuto F, Lanzini A. Energy-sharing mechanisms for energy community members under different asset ownership schemes and user demand profiles. Renew Sustain Energy Rev. 2022;168:112859. https://doi.org/10.1016/j.rser.2022.112859.

16. Garavaso P, Bignucolo F, Vivian J, Alessio G, De Carli M. Optimal planning and operation of a residential energy community under shared electricity incentives. Energies. 2021;14(8):2045. https://doi.org/10.3390/en14082045.

17. Mignoni N, Scarabaggio P, Carli R, Dotoli M. Control frameworks for transactive energy storage services in energy communities. Control Eng Pract. 2023;130:105364. https://doi.org/10.1016/j.conengprac.2022.105364.

18. Conte F, D'Antoni F, Natrella G, Merone M. A new hybrid AI optimal management method for renewable energy communities. Energy AI. 2022;10:100197. https://doi.org/10.1016/j.egyai.2022.100197.

19. Zhou Y. Advances of machine learning in multi-energy district communities – mechanisms, applications and perspectives. Energy AI. 2022;10:100187. https://doi.org/10.1016/j.egyai.2022.100187.

20. Liu Z, Ma J, Xing C, Liu J, He Y, Zhou Y, et al. Artificial intelligence powered large-scale renewable integrations in multi-energy systems for carbon neutrality transition: challenges and future perspectives. Energy AI. 2022;10:100195.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

https://doi.org/10.1016/j.egyai.2022.100195.

21. European Commission. Benchmarking smart metering deployment in the EU-27 with a focus on electricity. Available at: https://ses.jrc.ec.europa.eu/publications-list/benchmarking-smart-metering-deployment-eu-27-focus-electricity. Accessed October 23, 2023.

22. Agency for the Cooperation of Energy Regulators (ACER). CEER market monitoring report (MMR). Available at: https://www.acer.europa.eu/electricity/market-monitoring-report. Accessed October 23, 2023.

23. 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. https://doi.org/10.1016/j.apenergy.2018.12.025.

24. Pérez-Chacón R, Luna-Romera JM, Troncoso A, Martínez-Álvarez F, Riquelme JC. Big data analytics for discovering electricity consumption patterns in smart cities. Energies. 2018;11(3):683. https://doi.org/10.3390/en11030683.

25. Wang QC, Ding Y, Kong X, Tian Z, Xu L, He Q. Load pattern recognition based optimization method for energy flexibility in office buildings. Energy. 2022;254:124475. https://doi.org/10.1016/j.energy.2022.124475.

26. Fang M, Xiang Y, Xu BH, Wang T, Pan L, Liu J, et al. Data-driven load pattern identification based on R-vine copula and random forest method. IEEE Trans Ind Appl. 2022;58(6):7919-29. https://doi.org/10.1109/tia.2022.3200920.

27. Pérez-Ortiz M, Jiménez-Fernández S, Gutiérrez PA, Alexandre E, Hervás-Martínez C, Salcedo-Sanz S. A review of classification problems and algorithms in renewable energy applications. Energies. 2016;9(8):607. https://doi.org/10.3390/en9080607.

28. Neelamegam S, Ramaraj E. Classification algorithm in data mining: an overview. Int J P2P Netw Trends Technol. 2013;4(8):369-74.

29. Alrawi O, Bayram IS, Al-Ghamdi SG, Koç M. High-resolution household load profiling and evaluation of rooftop PV systems in selected houses in Qatar. Energies. 2019;12(20):3876. https://doi.org/10.3390/en12203876.

30. 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. https://doi.org/10.1016/j.apenergy.2019.113727.

31. Lazzeroni P, Lorenti G, Repetto M. A data-driven approach to predict hourly load profiles from time-of-use electricity bills. IEEE Access. 2023;11:60501-15. https://doi.org/10.1109/ACCESS.2023.3286020.

32. Schofield JT, Carmichael R, Tindemans SH, Bilton M, Woolf M, Strbac G. Low carbon London project: data from the dynamic time-of-use electricity pricing trial, 2013-2016. [data collection]. UK Data Service. https://doi.org/10.5255/ukda-sn-7857-2.

33. Dobre C, Xhafa F. Pervasive computing: next generation platforms for intelligent data collection. Cambridge, MA: Morgan Kaufmann; 2016.

34. Walfish S. A review of statistical outlier methods. Pharm Technol. 2006;30(11):82.

35. UK Government. National statistics – quarterly energy prices: December 2014. Available at: https://webarchive.nationalarchives.gov.uk/ukgwa/20141208232721/http://www.decc.gov.uk/en/content/cms/statistics/energy_pri

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press

ces/quarterly/quarterly.aspx. Accessed October 23, 2023.

36. Davies DB, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell. 1979;1(2):224-7. https://doi.org/10.1109/tpami.1979.4766909.

37. Hartigan JA, Wong MA. Algorithm AS 136: a K-means clustering algorithm. Appl Stat. 1979;28(1):100-8. https://doi.org/10.2307/2346830.

38. Lloyd S. Least squares quantization in PCM. IEEE Trans Inf Theory. 1982;28(2):129-37. https://doi.org/10.1109/tit.1982.1056489.

39. Köhler S, Rongstock R, Hein M, Eicker U. Similarity measures and comparison methods for residential electricity load profiles. Energy Build. 2022;271:112327. https://doi.org/10.1016/j.enbuild.2022.112327.

40. European Commission, Joint Research Centre (JRC). PVGIS user's manual. Available at: https://ec.europa.eu/jrc/en/PVGIS/docs/usermanual. Accessed April 3, 2020.

41. Grottera C, Barbier C, Sanches-Pereira A, Abreu MW, Uchôa C, Tudeschini LG, et al. Linking electricity consumption of home appliances and standard of living: a comparison between Brazilian and French households. Renew Sustain Energy Rev. 2018;94:877-88. https://doi.org/10.1016/j.rser.2018.06.063.

42. Fischer D, Härtl A, Wille-Haussmann B. Model for electric load profiles with high time resolution for German households. Energy Build. 2015;92:170-9. https://doi.org/10.1016/j.enbuild.2015.01.058.

43. Besagni G, Premoli Vilà L, Borgarello M. Italian household load profiles: a monitoring campaign. Buildings. 2020;10(12):217. https://doi.org/10.3390/buildings10120217.

44. Zapata Castillo V, De Boer H, Maícas Muñoz R, Gernaat DEHJ, Benders R, van Vuuren D. Future global electricity demand load curves. Energy. 2022;258:124741. https://doi.org/10.1016/j.energy.2022.124741.

© 2024 The Author(s). | Published by Institute of Smart Energy Systems Press