Energy & Buildings 256 (2022) 111618

Contents lists available at ScienceDirect

Energy & Buildings

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

A novel methodology for identifying appliance usage patterns in buildings based on auto-correlation and probability distribution analysis

João Góis *, Lucas Pereira

ITI, LARSyS, Polo Científico e Tecnológico da Madeira, floor -2, 9020-105 Funchal, Portugal
Técnico Lisboa, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal

A R T I C L E I N F O

Article history:
Received 22 July 2021
Revised 19 October 2021
Accepted 25 October 2021
Available online 28 November 2021

Keywords:
Appliance consumption
Auto-correlation function
Probability distribution
User profiling

A B S T R A C T

In today's society, a current concern is to mitigate the risks of global climate change. Throughout the years there have been several initiatives to achieve more sustainable energy distribution in buildings. In this work, a new methodology is proposed for identifying appliance consumption patterns in buildings. It consists of, at first, conducting a seasonality analysis based on the Auto-Correlation Function for detecting the different appliance use patterns that arise in a given time window. Then, it is conducted a Probability Distribution Analysis based on the auto-correlation results and the calculation of an informative measure to select the prevailing consumption pattern. The methodology enables to distinguish between different use patterns for a given appliance for each building at specific time intervals, e.g., the seasons of the year. For the purpose of illustration, the methodology is applied to consumption data of four appliances selected from a domestic energy consumption dataset (REFIT) over one year. The results provide several insights on how a given appliance use evolves throughout the seasons for each household, and also highlighting use similarity for different appliances across the seasons. These results would be, otherwise, hidden away, and would require an individual analysis of consumption patterns of each appliance. Consequently, the methodology provides a consistent mechanism to identify different user profiles.

1. Introduction

A continuous reduction of greenhouse gas emissions is a growing need for achieving a more sustainable worldwide energy distribution. Throughout the years, there have been several initiatives to discuss the impact of improper energy usage. In [1], the authors emphasize that energy efficiency concerning the domestic sector should be improved to lower emissions and mitigate the risks of global climate change. To this end, a more detailed analysis of household electricity consumption should be employed. Such analysis should take into account a complex interaction of socio-economic, dwelling, and appliance-related factors, which ultimately leads to an increase in consumption volatility among users and households [1]. In particular, seasonal variation throughout the year appears to impact household consumption [2,3].

Smart-meters are thus a cornerstone for the realization of more sustainable electric power grids due to their ability to transfer consumption information to remote data processing systems, which enables the creation of novel smart-meter data-based services that go beyond the traditional bill at the end of the month [4–6]. Many countries have been investing in the installation of smart grids in distribution networks. In [7] it is indicated that electric companies in the US aimed at installing around 65 million smart meters by 2015 (covering more than 50% of households) and around 90 million by 2020. Also, European Union countries aimed to replace at least 80% of electricity meters with smart meters by 2020, representing close to 200 million smart meters.

Several research initiatives exist to explore the potential of smart-meter data for both the grid operators and the end-consumers, as highlighted by some literature reviews on the topic. For example, in [8], a review of smart-meter data analytics from the utility point of view is presented. The authors highlighted applications such as load forecasting, customer characterization, outage management, and demand response implementation. In contrast, in [6], the authors review the potential of smart-meter data from a consumer-centric perspective. Potential applications include user feedback, anomaly detection, and participation in demand-side flexibility programs.

Interestingly, even though seasonality is considered in many of the works, to the best of our knowledge, it is not possible to find

* Corresponding author at: ITI, LARSyS, Polo Científico e Tecnológico da Madeira, floor -2, 9020-105 Funchal, Portugal.
E-mail address: joao.gois@tecnico.ulisboa.pt (J. Góis).

https://doi.org/10.1016/j.enbuild.2021.111618
0378-7788/

any literature proposing data analytics methods to study seasonality in smart-meter data. Instead, the effects of seasonality are handled differently depending on the spatial dimension of the data. On the one hand, when it comes to data analytics on top of aggregated data, the classic approach is to identify and remove the seasonality component in a pre-processing stage and add it back on a post-processing stage [9]. On the other hand, when conducting data analytics on individual data, some researchers find it helpful to include seasonal information as input features [10–12], while others perform data analytics individually for each season [13].

1.1. Auto-correlation function for seasonality identification

The Auto-Correlation Function (ACF) [14] is a widely used technique to identify seasonality in time series data, with recent applications in a vast range of fields, such as energy consumption [15], financial markets [16], meteorology [17] and geo-spatial studies [18]. In broad terms, the ACF measures the correlation between the time series with itself at different time units (lags). For a time series y_t, it is defined as:

$$\rho_k = \mathrm{Cor}(y_t, y_{t-k}) = \frac{Cov(y_t, y_{t-1})}{\sqrt{Var(y_t)Var(y_{t-k})}}, \quad k = 1, 2, \ldots \tag{1}$$

where the k is the gap considered between observations. Thus, ρ_1 shows how the y values which are one lag apart (i.e., successive values) relate to each other, ρ_2 indicates how the y values, which are two lags apart are related to each other and so on. Spikes, i.e., values close to 1 or -1 indicate a high positive or negative auto-correlation, and values close to 0 indicate lack of auto-correlation.

For illustration purposes, Fig. 1 shows two weeks of consumption from a tumble dryer at one minute intervals, and the respective ACF. As it can be observed, consumption peaks occur around days 21 to 23 (June), 27 to 30 (June), and 4 to 6 (July). These dependencies are detected by the Auto-Correlation Plot (ACP), and are represented by the spikes at lags 0 to $20,000$ ($\approx 0 - 2$ days), $40,000$ to $70,000$ ($\approx 4 - 8$ days), and $120,000$ (≈ 14 days).

One limitation of the ACF, however, is that seasonality assessments suffer from the observer bias [19] since they rely on visual inspection of the generated ACP. Ultimately, different interpretations of the ACP would lead to the detection of inaccurate seasonal periods, mainly if the data granularity is sufficiently large.

1.2. Applications of the auto-correlation function to smart-meter data

The ACF has found its applications mostly in load forecasting, in some cases being the standard pre-processing technique to identify seasonality in the consumption data. In [21], autocorrelation analysis is employed in load forecasting, in a process that is also known as feature selection in the context of data analytics. To this end, the time-series features with larger auto-correlation values are selected for load forecasting, while the redundant or irrelevant features are removed. This method is considered, in a similar fashion, in [22,23,26].

In addition, in other cases autocorrelation analysis was incorporated into load forecasting models. In [20], the consideration of forecast error autocorrelation is suggested when reconciling forecasts in a temporal hierarchy, which improves forecast accuracy. In [24], the authors proposed a hybrid model that combines the ACF with Least Squares Support Vector Machine, whose experimental results improve forecast accuracy compared to benchmark approaches. In [25], the employed forecasting models for short-term prediction of electricity demand include an adjustment term for dealing with the first-order autocorrelation.

As for individual appliance studies in particular, applications of the ACF are very scarce, with less than a handful of works. In [27,28], the ACF is used for detecting periodicity and load patterns in Hilbert transforms of pulse waveforms from energy consumption data. The authors emphasized that statistical tools, particularly the ACF, provide valuable information to end-use disaggregation, such as level of demand and duty cycles of appliances for a given household and specific season.

In [29], the temporal structure of load demand time-series is analyzed via first-order auto-correlation to compare energy consumption between commercial and residential buildings. Among other results, the results highlight that commercial buildings have more evident seasonal patterns than residential ones due to time-scheduled equipment usage versus people's habits, which are likely to change among households.

In [13], a preliminary study was conducted to understand whether different ACF patterns affect the disaggregation performance of NILM algorithms. The ACF was used for detecting dishwasher consumption patterns across households from the REFIT dataset. More precisely, it was intended to assess to what extent similarity among dishwasher consumption patterns can affect the performance of Non-Intrusive Load Monitoring (NILM) algorithms.

1.3. Research contributions and paper organization

Noticeably, all the surveyed works relied on visual inspection of the ACP. In fact, to the best of our knowledge, there is no published literature aiming at developing alternatives to visual inspection. It seems that, although the characteristics of autocorrelation are widely acknowledged as part of time series modelling in a variety of research fields, the improvement of its inspection methods has often been ignored or neglected, which may be an explanation for the ACP still being mostly based on visual inspection. In fact, in [30] it is noted that autocorrelation is a well-known phenomenon that has been largely neglected in behavioral ecology, and if not accounted accordingly can lead to biased results. In other fields, such as data privacy, it is also noted the lack of research on autocorrelation analysis [31].

The present work intends to provide a new procedure that formalizes the ACF pattern analysis, particularly in the context of appliance consumption data. The new proposed methodology directly assesses, in any given case, whether the corresponding

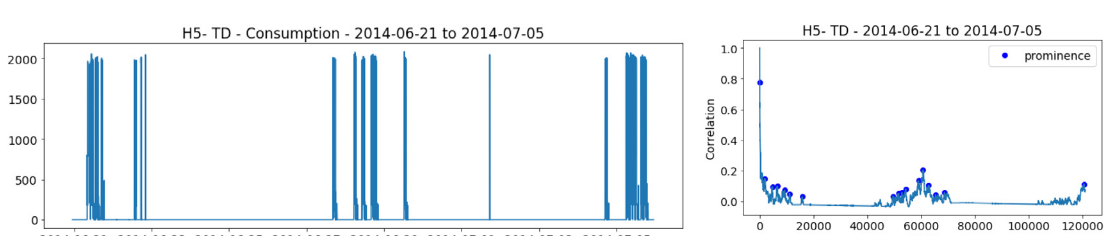


Fig. 1. Consumption data for the tumble dryer from household 5 for a time window ranging from 2014-06-21 to 2014-07-20 (left) and the corresponding ACF plots (right). Data granularity is set at 1 min. The ACF prominent peaks for $\tau = 0.05$ are represented by blue dots.

ACF shows a more frequent usage of an appliance in comparison to a moderated or sporadic use. To reduce the effects of the observer bias, the methodology does not rely on the visual inspection of the ACP. Instead, the seasonality is assessed by using an analytical procedure that includes a probability distribution analysis based on the extraction of consumption behaviors from the ACP and the calculation of an informative measure to identify the prevailing pattern. To illustrate the effectiveness of the methodology, a case study is presented where the methodology is applied to study individual appliance consumption on the REFIT [32] public dataset. The remainder of this paper is structured as follows. Section 2 provides a stepwise description of the proposed methodology. In Section 3 the methodology application is illustrated with a case study to a real-world energy consumption dataset, whose results and discussions are addressed in Section 4. Finally, the main conclusions, limitations, and an outlook on future work are presented in Section 5.

2. Methods

The proposed methodology is based on the stepwise procedure depicted in the diagram in Fig. 2. The first step consists of conducting a seasonality analysis of the time-series (for a given appliance being considered) at a sliding window level (Section 2.1). Then, the Energy Concentration Coefficient (C) is used for assessing whether, for a given sliding window, there is a consumption pattern that prevails among the other possible patterns (Section 2.2). Lastly, the analysis of the assigned patterns for each sliding window allows the identification of global usage profiles for the given appliance (Section 2.3).

2.1. Sliding window analysis of time-series

Sliding windows have been considered in a variety of areas with similar purposes. In financial time series analysis, it is commonly assumed that model parameters' constancy over time is not realistic due to the volatility of financial markets. To assess the stability of a model, the parameter estimates are computed over a sliding window of fixed size through the sample [33,34]. The sliding window approach was also incorporated in model procedures for predicting the relationship between renewable energy and industrial production [35], and economic growth [36]. In fields of research, such as NILM, sliding windows have been considered with the purpose of improving the performance of event detection, instead of considering the entire time horizon, e.g., [37,38]. Since ACF analysis roughly consists on the detection of spikes in theACP, the sliding window approach seems adequate in comparison to the whole time window analysis at once.

In previous works, the use of overlapping sliding windows was also suggested. In [39] overlapping sliding windows are used to improve the accuracy of human activity recognition systems. In [40] this strategy is used to improve anomaly detection performance in energy consumption datasets. Since overlap seems to improve model or algorithmic performance in comparison with non-overlap, an overlapping sliding window approach is useful to analyze auto-correlation in time windows that comprise a larger number of observations, such as seasonal and yearly data. Therefore, overlapping sliding windows are used in this work.

After dividing a time window into sliding windows, the following step is to analyze the corresponding ACFs. Notice that if the time series is not moderately large, the sliding windows will possibly contain a small number of observations (also depending on the chosen granularity). The most significant ACF peaks are obtained by exploring the concept of topographic prominence. This step is inspired by the real-world problem of mountain detection [41]. In simple terms, a prominent ACF peak occurs when the vertical distance between the peak and lowest contour line is greater than or equal to a given threshold, denoted by τ. In Fig. 3, the prominent ACF peaks for $\tau = 0.05$ are depicted for the ACF for the tumble dryer from household 5 for sliding window 1 during summer.

Based on the prominent ACF peaks for each sliding window, a difference set is obtained by calculating the difference between consecutive lags of prominent ACF peaks. More precisely, let each difference set be represented by a variable $X_i, i = 1, \ldots, n_s$ whose realizations are defined to be the respective differences, where n_s is the number of difference sets, also sliding windows, in which the larger interval was divided. For the sliding window in Fig. 3, the difference set, ω, is:

$$\omega = \{24, 1932, 4714, 6557, 9314, 11156, 15876,$$

$$49588, 51634, 52842, 54230, 58974, 60719, 62798, 65425, 68708, 120501\}$$

Sliding windows with a number of prominent peaks $\leqslant 1$ are discarded since at least 2 peaks are required for building up sets of differences.

2.2. Identification of relevant differences

In this step, the objective is to identify which differences are the most frequent on a difference set. For this purpose, it is proposed the calculation of the Probability Mass Function (PMF).

First, it is important to remark that for a particular sliding window, the respective difference set may contain differences that are very close in value, which in practice does not change usage interpretation. For such cases, it is adequate to aggregate the differences by probability, choosing for the representation of the aggregated differences, e.g., the lowest difference value. As such, an aggregation range, denoted herein by Δ, is defined, which should be selected according to the experiment requirements. To automatically detect if there is, at least, one significant difference on a difference set, it is proposed the comparison between the respective PMF (p_{Xi}) and the uniform PMF (p_{Ui}), in the same domain, through a quantitative measure. By considering PMFs, it is intended to quantify the frequency of different correlations (similarity behaviours) using probability and then compare with the uniform PMF in the same domain of similarity behaviours.

As an illustration, Fig. 4 depicts the PMF that is obtained for a sliding window (2014-06-21 to 2014-07-05) for the tumble dryer from household 5. The uniform PMF in the same domain is also computed. It can be seen that the most frequent difference value is 2079, with a probability of 0.375.

For comparison between p_{X_i} and p_{U_i}, the quantitative measure that was chosen is based on the Kullback–Leibler divergence (K-L divergence) [42], since K–L divergence accesses information loss from replacing the original distribution p for a sliding window with

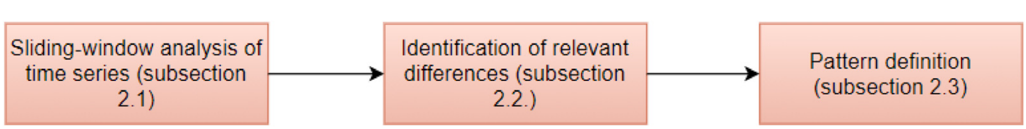


Fig. 2. Proposed methodology steps.

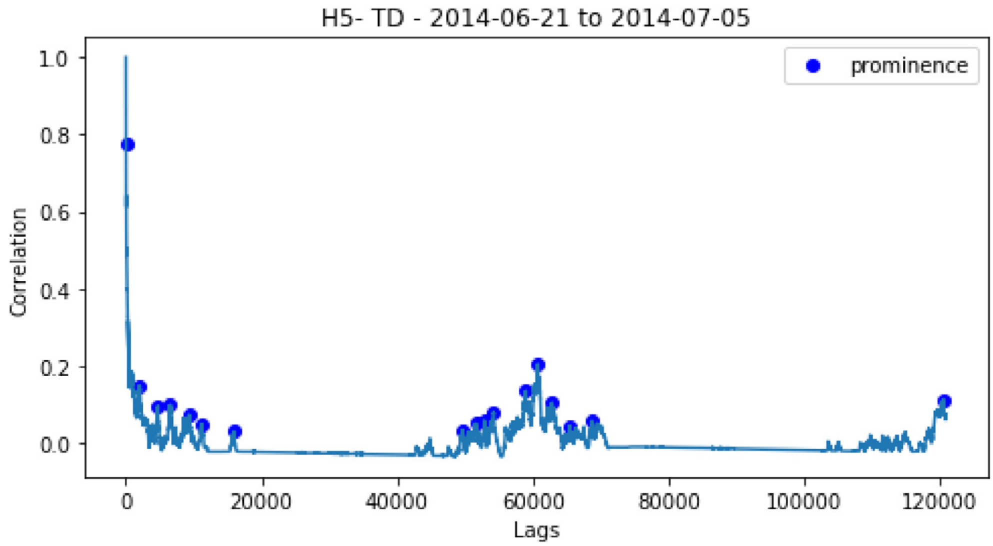
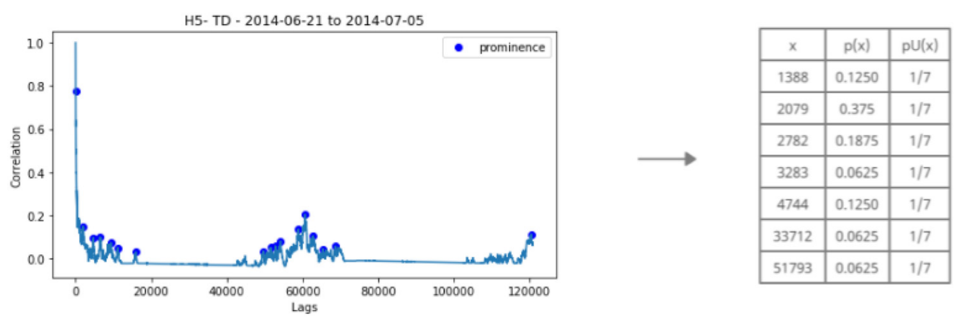


Fig. 3. The ACF for the tumble dryer from household 5 for a sliding window - ranging from 2014-06-21 to 2014-07-05. The ACF prominent peaks for $\tau = 0.05$ are represented by blue circles.



x	p(x)	pU(x)
1388	0.1250	1/7
2079	0.375	1/7
2782	0.1875	1/7
3283	0.0625	1/7
4744	0.1250	1/7
33712	0.0625	1/7
51793	0.0625	1/7

Fig. 4. The ACF for the tumble dryer from household 5 for a sliding window - ranging from 2014-06-21 to 2014-07-05. The ACF prominent peaks for $\tau = 0.05$ are represented by blue dots (left). The PMF is obtained along with the uniform PMF in the same domain (right).

an alternative distribution q. The KL-divergence between any well-defined PMFs p and q is given by Eq. 2:

$$D_{KL}(p||q) = \sum_{i=1}^{N} p(x_i) \log \frac{p(x_i)}{q(x_i)}, \quad i = 1, \dots, N \tag{2}$$

In this case, the alternative distribution is the uniform PMF (p_U) defined in the same domain. The greater the K-L divergence, the more non-uniform is p inside a sliding window since more extra information would be necessary if p_U were to be selected instead. In practice, it is this property that enables to assess the similarity or dissimilarity between p and p_U. More precisely, when there is similarity the extra information is small.

However, it is important to stress that the K-L divergence is not a distance measure since the symmetry and triangle inequality properties are not satisfied [42]. It is, instead, a measure of entropy increase when the true distribution is approximated by an alternative distribution [43] Since the K-L divergence does not possess a bounded scale, it becomes difficult to make meaningful comparisons. Therefore, the quantitative measure which is employed is a normalized version of the K-L divergence [44], known as the Entropy Concentration Coefficient (C), given by Eq. 3:

$$C\ (p,\ p_U) = \frac{D_{KL}\ (p||p_U)}{H\ (p) + D_{KL}\ (p||p_U)} \tag{3}$$

where $H(p) = -\sum_{i=1}^{N} p(x_i) \log p(x_i), i = 1, \dots, N$, is the entropy for PMF p. Note that the previous considerations on the K-L divergence also hold for C.

In this work, it is considered that the sliding windows in which $C \leqslant 0.1$ are not significant and therefore removed, since the proportion of loss is small. To illustrate the calculation of C in each sliding window, consider the PMFs p and p_U in Fig. 4. Then, both the entropy and KL-divergence are calculated considering the binary logarithm (base 2):

$$H(p) = 3/2 + \frac{3\log(8/3)}{8\log(2)} + \frac{3\log(16/3)}{16\log(2)} \text{bits} \tag{4}$$

$$D_{KL}(p||p_U) = -\frac{\log(8/7)}{4\log(2)} + \frac{3\log(21/16)}{16\log(2)} - \frac{3\log(16/7)}{16\log(2)} + \frac{3\log(21/8)}{8\log(2)} \text{bits} \tag{5}$$

By Eq. (3) we obtain that $C(p, p_U) \approx 0.11$, that is, on average it would require approximately 11% extra bits with respect to the original distribution.

As an illustration of the sliding window analysis, in Fig. 5 the most frequent difference, probability and C are obtained for each sliding window in the time window ranging from 2014-06-21 to 2014-09-21, concerning the tumble dryer consumption data from household 5. It can be seen that sliding windows 1 and 2 have similar most frequent differences, as well as for sliding windows 3 and 4.

2.3. Pattern definition

At this point, for a specified appliance data, a given time window was divided into sliding windows, and for each sliding window, it was obtained the most frequent difference (which can be denoted by D), the respective probability, and entropy concentration coefficient C.

Sliding window	Most frequent dif	Probability	C(pmf, uni)
1	2079	0.375	0.115374
2	2066	0.3	0.113033
3	2429	0.35	0.13234
4	2526	0.4	0.131254

Fig. 5. Sliding window analysis for the tumble dryer from household 5 in the time window ranging from 2014-06-21 to 2014-09-21. The respective most frequent differences, probabilities and C are obtained.

It is interesting to analyze how the D for each sliding window evolves throughout the seasons to identify possible different usage profiles. For instance, if for most sliding windows of a season the Ds are similar, denote it by a set D^* of Ds, it indicates that the differences in D^* prevail over the remaining Ds for a season. Thus, it suggests that there is a usage pattern that prevails in the season since the appliance is often used at multiples of lags in D^*.

According to the previous considerations, it is adequate to define a pattern criteria to distinguish between different types of usage for a given appliance. Let T denote the number of power samples in a day. Then, for k days there is a total of $T \times k$ power samples. For instance, for a given appliance the following frequency usage patterns can be defined.

- High frequency (H): most differences $d < T - \delta_1$.
- Medium frequency (M): most differences $T - \delta_1 < d < 3T - \delta_2$.
- Low frequency (L): most differences $d > 3T - \delta_2$.

where δ_1 and δ_2 are constants that smooth the boundaries between different patterns. It is adequate to smooth boundaries between H, M, and L since a given appliance is not necessarily activated at the same time when it is used. Although such fluctuations may occur, it does not change the usage frequency. The smoothers should be defined to accommodate those possible fluctuations. Notice that the choice of previous criteria is flexible. The researcher should choose the extremes that best fit the experiment. Also, it is adequate to compare usage frequencies with T since the time units in both cases are the same, which enables to define patterns, e.g., bi-daily, daily, or once every two days, with respect to T. From the previous criteria, other frequency patterns originate from the mixture of at least two of them, such as

- High-medium (H/M) frequency: a draw between H and M.
- Mediumlow (M/L) frequency: a draw between M and L.
- High-low (H/L) frequency: a draw between H and L.
- High–medium–low (H/M/L) frequency: a draw between H, M, and L.
- No pattern: zero frequency for both H, M, and L.

In the table presented in Fig. 5, since the granularity is set at 1 min, then $T = 24 \times 60 \times 1 = 1440$. It can be seen that the difference values are around 2000, thus for household 5 the predominant usage for the tumble dryer was mostly at multiples of 2000 and 2500 lags approximately. According to the previous criteria, medium frequency usage predominates. However, note that there are only significant ACF prominent peaks for 4 sliding windows since for the remaining no pattern was detected. Thus, the significant correlations are related to medium frequency, otherwise no pattern was detected.

3. Case study specification

In this section, it is described a case study to illustrate the applicability of the proposed methodology to a real-world dataset. The case study considers individual appliance data, and it is intended to show how the methodology can be used for analyzing a given appliance consumption over time across different households.

3.1. Dataset

The household appliance consumption data from the REFIT dataset [32] is used. REFIT is composed of electrical consumption measurements, both aggregate, and appliance levels from 20 households located in the UK. The data are originally timestamped and sampled at 8 s intervals. Note that the households are numbered from $1 - 13$ and $15 - 21$, which is the naming convention that we also use in this work.

3.1.1. Selected appliances

In this case study, four appliances are considered, namely, Washing Machine (WM), Dishwasher (DW), Kettle (KT), and Microwave (MW).

As indicated in [45], appliance consumption can be influenced by housing type, household characteristics, and appliance ownership. Furthermore, appliances can be categorized according to their pattern of use [46]. For instance, the WM, DW, and MW are standby appliances, which have 3 modes of operation: in use, on standby mode, or switched off. A KT or a Toaster (TT), on the other hand, are active appliances, which can be actively switched ON/OFF by the householders and do not have a standby mode. Cold appliances, such as the Fridge (FRI) and Freezer (FRZ), are in continuous use, with the power consumption cycling between zero and a set of power levels that is under thermostatic control. As such, cold appliances were not chosen for usage pattern analysis.

- WM: In [47] it is indicated that the intensity of WM usage increases when the number of householders increases. Since WM can largely contribute to significant water waste, efficient changes in usage can lead to notable water and electricity saving. In [48] it is suggested to run the WM on full loads. Some studies reported WM consumption patterns. In [49], it was noticed that users were prone to use WM only on a particular weekday. In [50] the WM consumption was not dependent on the season, the main variations depended on whether householders teleworked.

  In REFIT, the WM is available in 19 out of the 20 households. In particular, there are 2 WMs in household 4, in which the consumption data was jointly analyzed.
- DW: In [47] it is indicated that DW usage decreases as the number of family members increases. It is also verified that the presence of teenagers and increase in extra-salary positively affects the DW usage. In [51] similarly to the WM, it is stated that modern DWs should be run on full loads to reduce do electricity, water, and detergent consumption. Also, it is expected that conscious householders will use them more sparsely through time. As for the WM, the DW consumption is not dependent on seasonal variation but on whether householders teleworked or not [50].

  In REFIT, the DW is available in 15 out of the 20 households.

- KT: On a previous study using the REFIT dataset, [52] indicated the presence of well-defined patterns of use for weekdays during standard office hours. The pattern variation was mainly dependent on the type of occupancy and general daily schedule, and the vacation period. The KT usage appears to be regular at peak times (morning and evening around dinner) and sporadic otherwise. In [53] the KT is indicated as optionally used for preparing breakfast and relaxing/leisure activities. The KT is available in 14 out of the 20 REFIT households.
- MW: According to [47] the intensity of usage increases as the number of householders increases. It is also suggested that an increase in one of the household members income reduces usage, which derives from the family eating out more frequently. In [53] a study conducted in UK households indicates that the MW is used mostly for preparing breakfast and optionally for relaxing/leisure and preparing dinner. In REFIT, the MW is available in 16 out of the 20 households.

3.2. Time Window Splitting

To analyze the effects of seasonality in appliance consumption, the experiments are performed using a year-long period (from 2014-06-21 to 2015-06-20). The year-long data is further divided into meteorological seasons: 1) summer (from 2014-06-21 to 2014-09-21); 2) fall (2014-09-22 to 2014-12-20); 3) winter (2014-12-21 to 2015-03-19); and 4) spring (2015-03-20 to 2015-06-20).

The data for each season is further divided into 12 consecutive sliding windows, each of which consisting of a 2-week period, with a 1 week overlap. In fall and winter, the 12th sliding window contains less than 14 days due to calendar peculiarities. In the present experiment, the data granularity is set at 1 min, and invalid data entries (e.g., NaNs) were discarded.

3.3. Parameter selection

Throughout the description of the proposed methodology in Section 2, there is a set of parameters (ρ) that have to be defined prior to the experiment. The set can be denoted by:

$$\rho = \{\tau, \Delta, C_0, T, \delta_1, \delta_2\}$$

where τ is the threshold chosen for selecting the prominent peaks from the given appliance's ACF; Δ is the aggregation range in difference sets; C_0 is the threshold above which the energy concentration coefficients are considered to be significant and below which the sliding windows are removed; T is the number of appliance power samples per day; and δ_1, δ_2 are constants that are used for smoothing the boundaries between the different frequency patterns in the defined pattern criteria.

Notice that by setting a reasonably high τ, the number of selected ACF prominent peaks tends to decrease, thus the difference sets may be smaller or the frequency of difference values within the difference sets are reduced, and vice versa. By increasing Δ, the size of difference sets is reduced and since more difference values are aggregated, the probabilities assigned to difference values will change. The constants δ_1, δ_2 directly influence the number of sliding windows whose most frequent differences (or close differences) are assigned to a frequency pattern, and ultimately can change the chosen frequency pattern. For the cases in which many of the most frequent differences are close to the boundaries, the choice of the frequency pattern is very sensitive to δ_1, δ_2 or both.

For this case study in particular, the parameter values were selected for the purpose of illustration: $\tau = 0.05, \Delta = 400$ (i.e., activities separated by no more than approximately 6 h and 30 min are aggregated), $C_0 = 0.1, T = 1440, \delta_1 = 240$ and $\delta_2 = 820$. It is out of the scope of this paper to determine how to obtain the most adequate parameter values for the experiment, hence parameter selection is addressed in more detail in subSection 5.2. Then, the pattern identification criteria are set to:

- H: most differences are $d < 1200$.
- M: most differences are $1200 < d < 3500$.
- L: most differences are $d > 3500$.

The mixed frequency patterns that follow from the criteria obtained above are the same as represented in subSection 2.3.

3.4. Evaluation methodology

The results of this case study (presented in Section 4) are presented as follows. First, for a given appliance, the frequency patterns are obtained for each household throughout the seasons, which allows to check for seasonal variation among the households. Then, for a given season, the frequency patterns for each of the selected appliances are obtained for each household. This is also of interest since it is possible to compare the consumption of the different appliances in particular households.

The results are presented in the form of heatmaps and tables. The heatmaps provide a direct overview of the existing frequency patterns among households and across seasons. In this respect, heatmaps are more efficient than other plots, by enabling a visual understanding and interpretation. In the opposite direction, the tables are useful to inform on the totals of frequency patterns and thus ideal to make quantitative comparisons across seasons.

4. Results and discussion

In subSection 4.1, for the selected appliances, the usage patterns were compared through the seasons of the year. In subSection 4.2, the usage patterns of different appliances were compared for each season. In subSection 4.3, it was presented a use case of the proposed methodology on energy efficiency. The households that do not contain at least one of the appliances under study are colored with white on the heatmaps.

4.1. Individual appliance level

In the appliance level evaluation, for each appliance, the household consumption is assessed throughout the seasons. The results are summarized in Fig. 6 and Table 1. As it can be observed, the WM is an appliance that evidences seasonal variation since there is a large number of households with significant pattern change throughout the seasons (Fig. 6(a)). In fact, the M usage holds, in general, only for about 21% of the households. In general, the occurrence of different usage patterns is balanced across the seasons - H and M usage prevail, respectively, for about 42% (8/19) and 32% (6/19) of households.

From Table 1(a), in winter the M usage prevails over other usage patterns. In fact, there is an increasing tendency for M usage among the households as fall and winter arrive. With the beginning of spring, H and M usage patterns seem to balance out. Also, only for household 6, there is a number of seasons in which no significant pattern was detected according to the proposed methodology. Furthermore, according to Table 1(a), there is a total of 4 cases for which no pattern was assigned, which is not a large number in comparison with the total number of cases.

In contrast, for the DW, seasonal variation is less evident throughout the year. Still, some households exhibit pattern change, which is the case for households 1, 5, and 13.

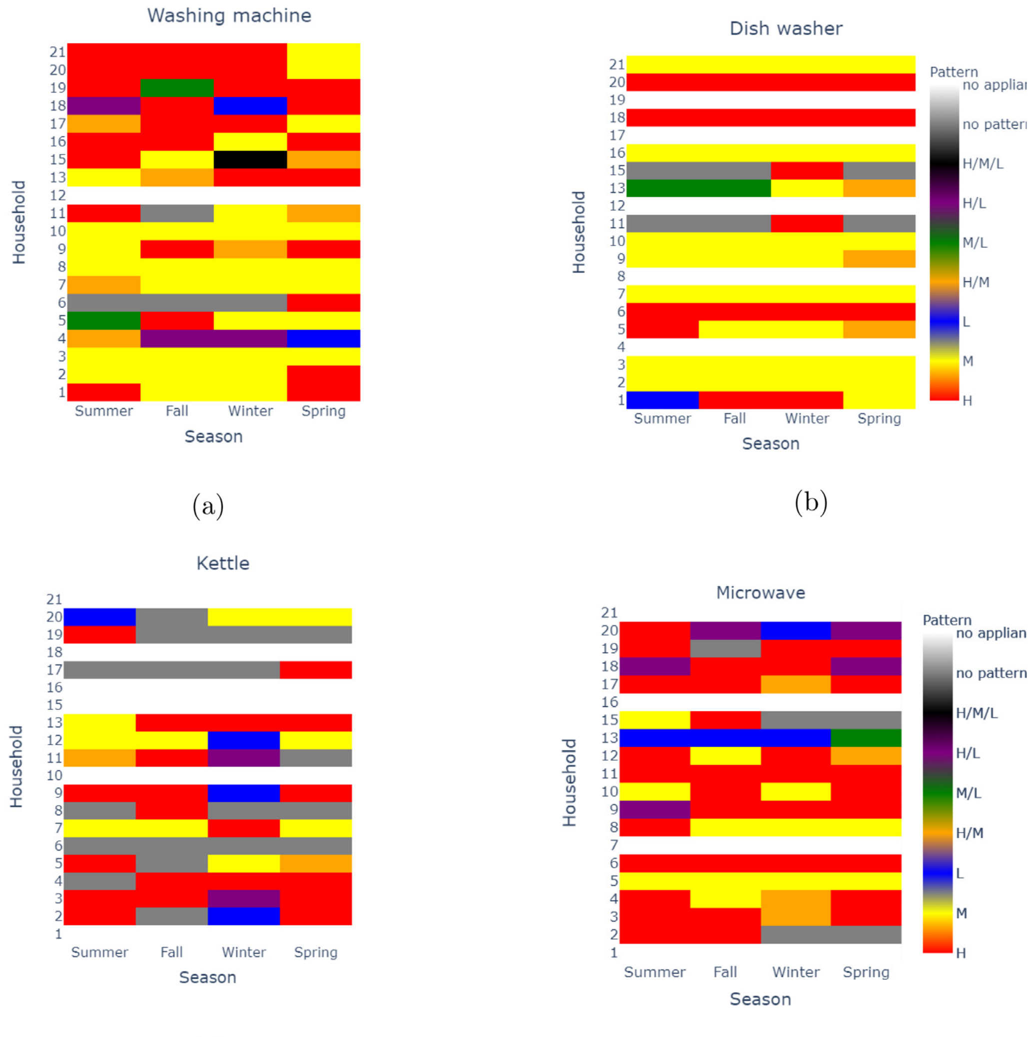


Fig. 6. Usage patterns detected by the developed procedure for the washing machine (a), dish washer (b), kettle (c) and microwave (d) throughout the seasons of the year.

Fig. 6(b) shows that the M usage prevails, in general, across the seasons for about 53% (8/15) of the households, while the H usage holds for all seasons for 20% of the households. From Table 1(b) it is clear that the M usage prevails over other usage patterns for all seasons. There are 2 households in which no pattern was detected for at least one season. Nevertheless, the number of no pattern cases is again not relevant in comparison with the total number of cases.

With respect to the KT, Fig. 6(c) indicates that the H usage prevails across seasons for about 50% (7/14) of households. From Table 1(c), by looking at each season separately, the H usage prevails over other usage patterns, except for winter. There is also relevant pattern change through seasons. Note that for about 64% of the households there is at least one season in which no pattern was detected. In this case, the number of no pattern cases is also relatively large in comparison with the total number of cases.

Concerning the MW, Fig. 6(d) indicates that there is relevant pattern change for about 37.5% of households. The H is the usage pattern that prevails across seasons – 62.5% (10/16) of the households. The M usage only prevails for 12.5% of households, while the L usage prevails only for 6.25% of households, which is substantially small in comparison to H usage. In fact, Table 1(d) shows that H usage prevails over other usage patterns throughout the seasons. Also, for about 18% of the households, there is at least one season in which no pattern was detected, but similarly to the DW and WM the total number of occurrences is not significant.

In general, it can be seen that the WM and the DW are not used with the same frequency as the KT and the MW. This was somewhat expected since the WM and DW, if not used efficiently, can lead to significant water waste and electricity consumption [48,51]. Nevertheless, it is possible that consumption patterns for WM and DW vary according to specific socio-economic factors that were mentioned in subSection 3.1.1. Hence, for a more in-depth analysis of the results, it would be important to access qualitative data regarding REFIT.

Although it was expected, in general, a larger frequency of the L usage pattern for the WM, the results may change when the parameters defined in subSection 3.3 are altered. There is a number of mixed pattern usages of the type M/L and H/L in which by altering δ_1 and δ_2, for instance, could increase the number of Ls. For the DW in particular, it was noticed a more consistent use, in general, throughout the year in comparison with the other studied appliances.

4.2. Season level

Similarly to the appliance level, the consumption patterns are analyzed at the season level. For each season, the appliance con-

Table 1
Total frequency of usage patterns through the seasons for the (a) washing machine, (b) dish washer, (c) kettle and (d) microwave. “Mixed” represents all mixed patterns possibilities. “No pattern” stands for the cases in which none of the defined frequency patterns were detected.

		(a) Washing machine			
Pattern	Su	F	W	Sp	Total
H	7	7	5	8	27
M	6	7	9	8	30
L	0	0	1	1	2
Mixed	5	3	3	2	13
No pattern	1	2	1	0	4
		(b) Dish washer			
H	4	4	6	3	17
M	7	8	9	7	31
L	1	0	0	0	1
Mixed	1	1	0	3	5
No pattern	2	2	0	2	6
		(c) Kettle			
H	5	6	3	6	20
M	3	2	2	3	10
L	1	0	3	0	4
Mixed	1	0	2	1	4
No pattern	4	6	4	4	18
		(d) Microwave			
H	10	9	6	8	33
M	3	3	3	2	11
L	1	1	2	0	4
Mixed	2	1	3	4	10
No pattern	0	1	2	2	5

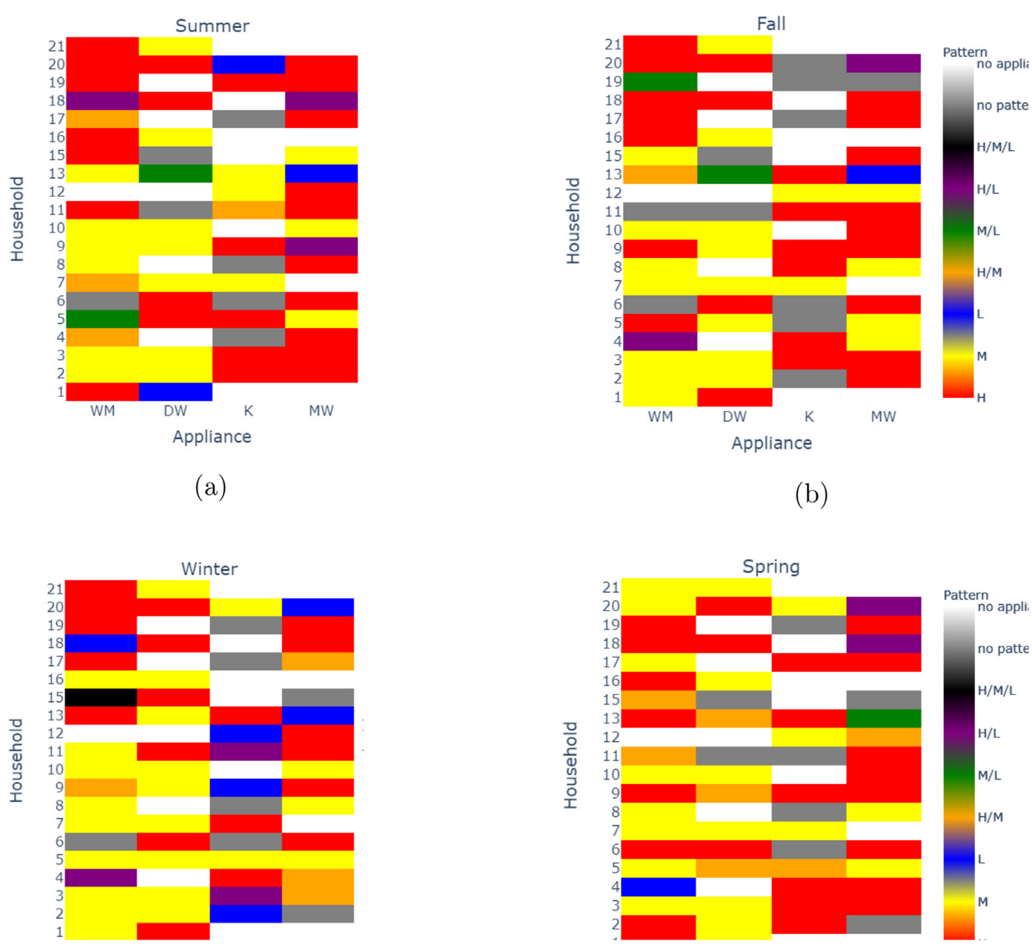


Fig. 7. Usage patterns detected by the developed procedure for the chosen appliances in summer (a), fall (b), winter (c) and spring (d).

Table 2
Pattern similarity detected across the seasons for (a) washing machine and dish washer, (b) kettle and microwave and (c) washing machine and microwave.

(a) Washing machine – Dish washer			
Summer	Fall	Winter	Spring
53%	40%	53%	60%
(b) Kettle – Microwave			
38%	31%	23%	46%
(c) Washing machine – Microwave			
53%	33%	47%	47%

sumption is assessed, which may provide insights on how the consumption differs or not among different appliances. The results are summarized in Fig. 7, and Table 2.

By analyzing Fig. 7 it was noticed some important relationships among the appliances throughout the seasons. The WM and the DW are similarly used for a minimum of approximately 53% of households in fall and a maximum of 60% of households in spring (Table 2(a)). For the KT and MW, there is a similar usage for a minimum of approximately 23% of households in winter and a maximum of 46% of households in spring (Table 2(b)). Also, the WM and microwave exhibit similar usage for a minimum of approximately 33% of households in fall and a maximum of 53% of households in summer (Table 2(c)).

Like with the case presented in Section 4.1, it was expected that the WM and DW exhibited, in general, similarity in terms of usage. The same applies to the KT and MW. For the WM and MW it was noticed significant percentages in terms of similarity, which at first was not expected due to the need to fit WM usage more efficiently in housework routine than that of the MW.

4.3. Use case: energy efficiency assessments

Herein is illustrated a simple use case of the proposed methodology regarding the energy efficiency of individual appliances. In this case, the chosen appliance was the WM. It was intended to assess the efficiency of WMs across households considering the experiment duration in the previous case study, i.e., throughout the four seasons.

To measure the efficiency of appliance usage, it was defined that the pattern results from the case study would be considered. In particular, for the WMs, the corresponding results are shown in Fig. 6(a). For each household's WM a global usage pattern was assigned in order to cover the four seasons, which would represent the pattern that prevailed for two or more seasons (in the former, provided the remaining seasons were assigned different usage patterns). In the case of ties, such households were classified as undefined. In Table 3, it is shown how households were grouped in terms of the definition of WM global patterns from Fig. 6(a). Since global usage patterns H or undefined indicate a high frequency of WM usage or simply an irregular usage frequency throughout the experiment, it was considered that WM usage in these circumstances was less efficient, especially if the number of occupants is small. In contrast, since global patterns M or L indicate a moderate or sporadic usage, such WM usage was considered efficient.

Table 3
Households aggregated by global pattern from pattern in Fig. 6(a).

Pattern	Households
High	9, 13, 16, 17, 18, 19, 20, 21
Medium	2, 3, 5, 7, 8, 10
High/Low	4
Undefined	1, 6, 11, 15

Table 4
Definition of categories for REFIT dataset features (a) Occupancy and (b) Housesold size.

(a)	
Occupancy	Households
1 to 2	1, 3, 4, 6, 8, 9, 11, 15, 18, 20
3 to 4	2, 5, 7, 10, 13, 17, 19, 21
More than 4	16
(b)	
Household size	Households
2 to 3	2, 3, 7, 8, 9, 10, 11, 15, 17, 18 19, 20, 21
4 or more	1, 4, 5, 6, 13, 16

The next step was to analyze the available household features from REFIT and relate them to the global patterns in Table 3. This enabled to check whether there was any tendency between the global pattern results and household features. Among the available features, it was considered the number of occupants (occupancy) and the number of household beds (household size) since these features provided a more representative input on energy efficiency.

Overall, occupancy takes values between 1 and 6, and household size between 2 and 5. Since some feature values are less frequent than others, occupancy and household size were divided into categories, as shown in Table 4. Notice that for a broader study it would be recommended that more features regarding the dataset were available.

Then, global patterns were grouped by feature categories (Figs. 8 and 9), which enabled to depict the most relevant relationships for assessing the efficiency of WM usage. In terms of occupancy, Fig. 8 indicates that households 2, 5, 7, and 10 have 3 to 4 occupants and M pattern for the WM, as opposed to households 3 and 8, whose number of occupants lies in 1 to 2. Also, for households 13, 17, 19, and 21 the pattern is H, and the occupancy lies in 3 to 4, in contrast with households 9, 18, and 20 with occupancy in 1 to 2. Notice that all households with an undefined pattern have 1 to 2 occupants. Thus, it seems that there is a more volatile consumption for households with 1 to 2 occupants, with more incidence of H and undefined patterns, i.e., undefined (40% of households), H (30%), M (20%), and H/L (10%). Hence, for the chosen parameters, households with 3 to 4 occupants seem more energy-efficient since there is 50% of households for both patterns H and M, which indicates a more moderate usage. Since there is only one household with more than 4 occupants, this category was not studied in detail.

For household size, Fig. 9 indicates that households 2, 3, 7, 8 and 10 have 2 to 3 beds and M pattern, which is almost all households with M pattern, except for household 5. Also, it can be noticed that most households with the H pattern, i.e., households 9, 17, 18, 19, 20 and 21, have 2 to 3 beds, except for households 13 and 16 which have 4 or more beds. Hence, it seems that for households with 2 to 3 beds there is more energy efficiency, with a larger incidence of the M pattern, i.e., H (about 46% of households), M (about 38%), and undefined (about 15%), than households with 4 or more beds size, i.e., H (about 33%), M (about 17%), and undefined (about 33%).

5. Conclusion

In this work, a formal procedure based on the ACF was proposed to complement the analysis of household consumption patterns

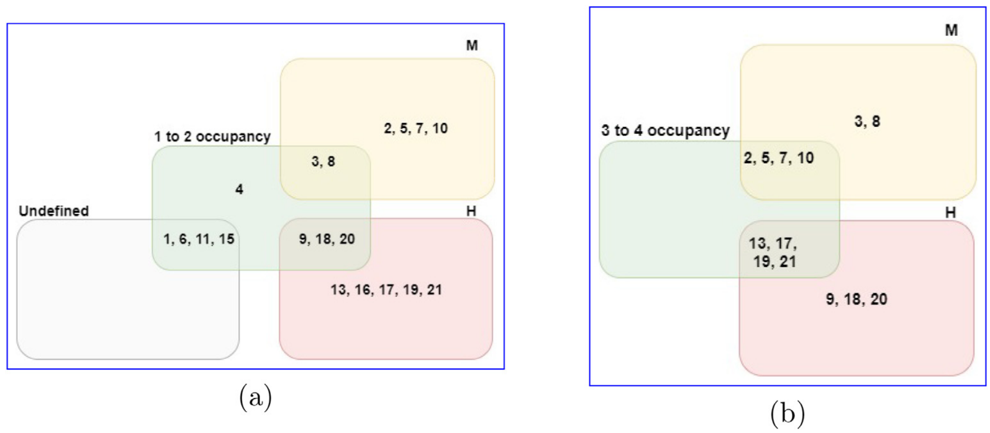


Fig. 8. Identification of households with a given global pattern result for the WM which have occupancy corresponding to (a) 1 to 2 occupants, and (b) 3 to 4 occupants.

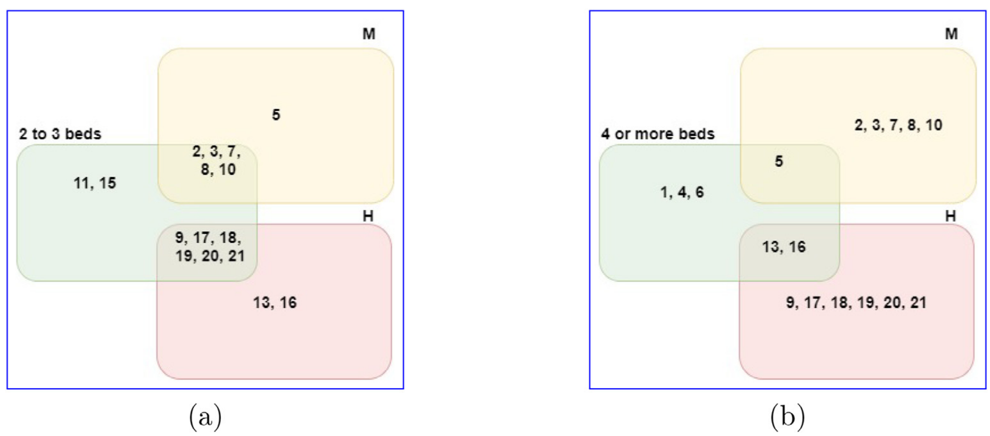


Fig. 9. Identification of households with a given global pattern result for the WM which have size corresponding to (a) 2 to 3 beds, and (b) 4 or more beds.

from energy consumption datasets. The procedure enables a straightforward and rigorous distinction between different usage profiles across the seasons of the year – seasonal variation. It also enables to easily check for similarity of usage between different household appliances. The procedure was applied to the REFIT dataset for illustration purposes, in particular, the washing machine, dishwasher, kettle, and microwave consumption data were considered.

5.1. Research implications and potential applications

The methodology presented in this work improves the existing state-of-the-art on consumption pattern analysis. The current research on the topic would benefit from an easier and more rigorous detection of consumption patterns, that ultimately leads to the definition of user profiles, and a better energy distribution among buildings.

In addition, the results obtained in the form of heatmaps and tables are useful for assessing the similarity of usage between appliances, which varies across households and seasons of the year, and also for an appliance in particular. In particular, the methodology would improve research on the creation of sustainable energy communities by enabling easier identification of buildings in which new strategies and practices should be implemented at the housework level to reach sustainable consumption. For instance, in the case of washing machines and dishwashers, developing better practices would ensure a reduction in electricity and water waste. Thus, a potential application would be the automatic assessment of efficient and inefficient appliance usage, as illustrated in the use case in subSection 4.3 for the washing machine.

This methodology can also be of interest for studies concerning dataset comparability, e.g. [54–56], in which datasets with varying complexities and different characteristics affect disaggregation performance. In the context of load disaggregation, it would enable checking how consumption patterns impact the disaggregation performance of appliances from the aggregate time series data.

5.2. Limitations

Even though this work proposes a flexible methodology to the type of experiment and researcher choices, particularly in terms of parameter setting, it comes with its caveats.

More precisely, there can be high volatility in the results when the pattern criteria are altered. For instance, if the most frequent differences are close to the boundaries between frequency patterns, any changes concerning the parameters δ_1 and δ_2 (remaining parameters fixed) can greatly impact the frequency pattern that results for a given case. There is a similar effect for the parameters τ and C_0, which affect, respectively, the number of selected ACF prominent peaks for a sliding window and the number of sliding windows that are significant for choosing the frequency pattern. Furthermore, if the number of power samples per day, T, is large enough, not only there can be changes in terms of the interpretation of the ACF prominent peaks, but also the computation time for calculating the ACF increases.

Therefore, before running the experiment, it is recommended to carry out a sensitivity analysis for the parameters. Depending on the research topic, a sensitivity analysis may be conducted differently, but, in general, it consists of studying how the output of a model varies with uncertainty in the model input, check, e.g., [57]. This enables the researcher to analyze how the pattern identification is influenced by different choices of parameter values and ultimately choose the more adequate values for the dataset being considered. In practice, the selection of parameter values is optimized for that case. It is also clear that the assessment of energy efficiency as depicted in subSection 4.3 is limited by the choice of the methodology parameters, for which a small variation may lead to a different conclusion in terms of an appliance efficiency. A more detailed analysis of energy efficiency would also benefit from access to more information on household characteristics.

Note that, by changing the dataset or the experiment conditions, the choice of parameters may change. Naturally, sensitivity analysis results may depend on the number of households in the dataset, which are expected to be more accurate for larger datasets. In this work, the methodology was illustrated considering a dataset with 20 independent households. It would be interesting to consider datasets from different countries or regions. If it was possible, the experiment could be conducted for a longer period, e.g., one more year of data. Doing this would enable checking how the pattern results change across different years, enabling not only the analysis of seasonal variation but also annual variation.

5.3. Future work directions

A future work direction is to demonstrate the applicability of the proposed methodology to the comparability of datasets. For such accomplishment, the next step would be to show how to conduct a sensitivity analysis to the presented case study, which would be a landmark to new studies considering the proposed methods. From that point onward, it is adequate to assess the relationship between the appliance consumption analysis and the respective disaggregation performance from the aggregate data.

Additionally, future work should seek to benchmark the proposed method with alternative approaches. For example, time–frequency analysis methods such as the short-term Fourier transform and wavelets can be used to characterize the appliance consumption by exploring the time and frequency domains in simultaneous [58].

Lastly, the present methodology could be applied to other domains beyond electricity consumption. For instance, it could be extrapolated to water consumption studies, as it would be enough that the dataset kept a time series structure and instant water flow records. This could be an important tool for improving water monitoring and management in buildings, as pointed out in [59]. Interestingly, it would also enable a joint analysis of electricity and water consumption in buildings [60,61], in particular for appliances such as the washing machine and the dishwasher.

Declaration of Competing Interest

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

Acknowledgements

This work received funding from the Portuguese Foundation for Science and Technology (FCT) under project UIDB/50009/2020, and grants DFA/BD/8075/2020 (J.G.) e CEECIND/01179/2017 (L.P.).

References

[1] R.V. Jones, A. Fuertes, K.J. Lomas, The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings, Renew. Sustain. Energy Rev. 43 (2015) 901–917.
[2] A. Fazeli, M. Gillott, Analysing the effects of seasonal variation on occupancy in an electricity demand model, Int. J. Low-Carbon Technol. 8 (2013) 282–288.
[3] M. Li, D. Allinson, M. He, Seasonal variation in household electricity demand: a comparison of monitored and synthetic daily load profiles, Energy Build. 179 (2018) 292–300.
[4] B. Neenan, R.C. Hemphill, Societal benefits of smart metering investments, Electricity J. 21 (2008) 32–45.
[5] J. Batalla-Bejerano, E. Trujillo-Baute, M. Villa-Arrieta, Smart meters and consumer behaviour: Insights from the empirical literature, Energy Policy 144 (2020) 111610.
[6] B. Völker, A. Reinhardt, A. Faustine, L. Pereira, Watt's up at Home? Smart meter data analytics from a consumer-centric perspective, Energies 14 (2021) 719.
[7] S. Chakraborty, S. Das, T. Sidhu, A. Siva, Smart meters for enhancing protection and monitoring functions in emerging distribution systems, Int. J. Electr. Power Energy Syst. 127 (2021) 106626.
[8] Y. Wang, Q. Chen, T. Hong, C. Kang, Review of smart meter data analytics: applications, methodologies, and challenges, IEEE Trans. Smart Grid 10 (2019) 3125–3148.
[9] B.A. Høverstad, A. Tidemann, H. Langseth, P. Öztürk, Short-term load forecasting with seasonal decomposition using evolution for parameter tuning, IEEE Trans. Smart Grid 6 (2015) 1904–1913.
[10] C. Dinesh, S. Makonin, I.V. Bajić, Residential power forecasting based on affinity aggregation spectral clustering, IEEE Access 8 (2020) 99431–99444.
[11] N. Batra, A. Singh, K. Whitehouse, Neighbourhood nilm: a big-data approach to household energy disaggregation, arXiv preprint arXiv:1511.02900 (2015)..
[12] M. Kaselimi, N. Doulamis, A. Voulodimos, E. Protopapadakis, A. Doulamis, Context aware energy disaggregation using adaptive bidirectional lstm models, IEEE Trans. Smart Grid 11 (2020) 3054–3067.
[13] J. Gois, C. Klemenjak, L. Pereira, On the relationship between seasons of the year and disaggregation performance, in: Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring, NILM'20, Association for Computing Machinery, New York, NY, USA, 2020, pp. 70–74..
[14] G. Box, G. Jenkins, G. Reinsel, Time Series Analysis: Forecasting & Control, Prentice Hall, third ed., 1994.
[15] A. Tureczek, P.S. Nielsen, H. Madsen, Electricity consumption clustering using smart meter data, Energies 11 (2018) 859.
[16] I. Martin, On the autocorrelation of the stock market, J. Finan. Econometr. 19 (2021) 39–52.
[17] N. Khanmohammadi, H. Rezaie, J. Behmanesh, The effect of autocorrelation on the meteorological parameters trend, Meteorol. Atmos. Phys. (2020) 1–13.
[18] D. Balaji, V. Saravanabavan, Geo spatial variation of dengue risk zone in madurai city using autocorrelation techniques, Geojournal (2020) 1–21.
[19] E.C. Poulton, Observer bias, Appl. Ergon. 6 (1975) 3–8.
[20] P. Nystrup, E. Lindström, P. Pinson, H. Madsen, Temporal hierarchies with autocorrelation for load forecasting, Eur. J. Oper. Res. 280 (2020) 876–888.
[21] R. Sood, I. Koprinska, V.G. Agelidis, Electricity load forecasting based on autocorrelation analysis, in: The 2010 International Joint Conference on Neural Networks (IJCNN), IEEE, pp. 1–8..
[22] I. Koprinska, M. Rana, V.G. Agelidis, Yearly and seasonal models for electricity load forecasting, in: The 2011 International Joint Conference on Neural Networks, IEEE, pp. 1474–1481..
[23] I. Koprinska, M. Rana, V.G. Agelidis, Correlation and instance based feature selection for electricity load forecasting, Knowl.-Based Syst. 82 (2015) 29–40.
[24] A. Yang, W. Li, X. Yang, Short-term electricity load forecasting based on feature selection and least squares support vector machines, Knowl.-Based Syst. 163 (2019) 159–173.
[25] J.W. Taylor, An evaluation of methods for very short-term load forecasting using minute-by-minute british data, Int. J. Forecast. 24 (2008) 645–658.
[26] I. Koprinska, M. Rana, V.G. Agelidis, Electricity load forecasting: a weekday-based approach, in: International Conference on Artificial Neural Networks, Springer, pp. 33–41..
[27] A. Gabaldón, M. Ortiz-García, R. Molina, S. Valero-Verdú, Disaggregation of the electric loads of small customers through the application of the hilbert transform, Energy Eff. 7 (2014) 711–728.
[28] A. Gabaldon, R. Molina, A. Marin-Parra, S. Valero-Verdu, C. Álvarez, Residential end-uses disaggregation and demand response evaluation using integral transforms, J. Mod. Power Syst. Clean Energy 5 (2017) 91–104.
[29] S. Henriet, U. Simsekli, G. Richard, B. Fuentes, Energy disaggregation for commercial buildings: a statistical analysis, in: International Workshop on Non-Intrusive Load Monitoring (NILM2018), Austin, TX, USA, pp. 1–4..
[30] D.J. Mitchell, A.M. Dujon, C. Beckmann, P.A. Biro, Temporal autocorrelation: a neglected factor in the study of behavioral repeatability and plasticity, Behav. Ecol. 31 (2020) 222–231.
[31] J. Zhao, S. Liu, X. Xiong, Z. Cai, Differentially private autocorrelation time-series data publishing based on sliding window, Secur. Commun. Networks 2021 (2021).
[32] D. Murray, L. Stankovic, V. Stankovic, An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study, Sci. Data 4 (2017) 160122.

[33] E. Zivot, J. Wang, Modeling Financial Time Series with S-Plus, vol. 191, Springer Science & Business Media, 2007..
[34] L. Hällman, The rolling window method: precisions of financial forecasting, 2017.
[35] T.H. Van Hoang, S.J.H. Shahzad, R.L. Czudaj, Renewable energy consumption and industrial production: a disaggregated time-frequency analysis for the us, Energy Econ. 85 (2020) 104433.
[36] M. Balcilar, Z.A. Ozdemir, Y. Arslanturk, Economic growth and energy consumption causal nexus viewed through a bootstrap rolling window, Energy Econ. 32 (2010) 1398–1410.
[37] K. Basu, V. Debusschere, S. Bacha, A. Hably, D. Van Delft, G.J. Dirven, A generic data driven approach for low sampling load disaggregation, Sustain. Energy, Grids Networks 9 (2017) 118–127.
[38] L. Pereira, Developing and evaluating a probabilistic event detector for non-intrusive load monitoring, in: Proceedings of the Fifth IFIP Conference on Sustainable Internet and ICT for Sustainability, IEEE/ IFIP, Funchal, Portugal, 2017, pp. 1–10..
[39] A. Dehghani, O. Sarbishei, T. Glatard, E. Shihab, A quantitative comparison of overlapping and non-overlapping sliding windows for human activity recognition using inertial sensors, Sensors 19 (2019) 5026.
[40] D.B. Araya, K. Grolinger, H.F. ElYamany, M.A. Capretz, G. Bitsuamlak, An ensemble learning framework for anomaly detection in building energy consumption, Energy Build. 144 (2017) 191–206.
[41] A. Kirmse, J. de Ferranti, Calculating the prominence and isolation of every mountain in the world, Prog. Phys. Geogr. 41 (2017) 788–802.
[42] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1951) 79–86.
[43] T.M. Cover, J.A. Thomas, Elements of Information Theory, second ed., 2006..
[44] C. Bandt, Entropy ratio and entropy concentration coefficient, with application to the covid-19 pandemic, Entropy 22 (2020) 1315.
[45] G.K. Tso, K.K. Yau, A study of domestic energy usage patterns in hong kong, Energy 28 (2003) 1671–1682.
[46] S. Firth, K. Lomas, A. Wright, R. Wall, Identifying trends in the use of domestic appliances from household electricity consumption measurements, Energy Build. 40 (2008) 926–936.
[47] S. Matsumoto, How do household characteristics affect appliance usage? application of conditional demand analysis to japanese household data, Energy Policy 94 (2016) 214–223.
[48] A. Ponniran, E. Sulaiman, S.A. Jumaat, M. Ishak, M.A. Chulan, S. Saiman, A study on electric energy usage at the residential area, in: Engineering Conference, p. 117..
[49] I. Laicane, D. Blumberga, A. Blumberga, M. Rosa, Evaluation of household electricity savings. Analysis of household electricity demand profile and user activities, Energy Procedia 72 (2015) 285–292.
[50] K. Cetin, P. Tabares-Velasco, A. Novoselac, Appliance daily energy use in new residential buildings: use profiles and variation in time-of-use, Energy Build. 84 (2014) 716–726.
[51] R. Stamminger, R. Badura, G. Broil, S. Dörr, A. Elschenbroich, A european comparison of washing dishes by hand, in: University of Bonn, Proceedings of the International Conference on Energy Efficiency in Domestic Appliances and Lighting (EEDAL), Turin, Italy..
[52] D. Murray, J. Liao, L. Stankovic, V. Stankovic, Understanding usage patterns of electric kettle and energy saving potential, Appl. Energy 171 (2016) 231–242.
[53] J. Liao, L. Stankovic, V. Stankovic, Detecting household activity patterns from smart meter data, in: 2014 International Conference on Intelligent Environments, IEEE, pp. 71–78..
[54] D. Egarter, M. Pöchacker, W. Elmenreich, Complexity of power draws for load disaggregation, arXiv preprint arXiv:1501.02954 (2015)..
[55] C. Klemenjak, S. Makonin, W. Elmenreich, Towards comparability in non-intrusive load monitoring: on data and performance evaluation, in: 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), IEEE, pp. 1–5..
[56] C. Nalmpantis, D. Vrakas, Machine learning approaches for non-intrusive load monitoring: from qualitative to quantitative comparison, Artif. Intell. Rev. 52 (2019) 217–243.
[57] A. Saltelli, Sensitivity analysis for importance assessment, Risk Anal. 22 (2002) 579–590.
[58] J.W. Pitton, The statistics of time–frequency analysis, J. Franklin Inst. 337 (2000) 379–388.
[59] R.E. Amaral, J. Brito, M. Buckman, E. Drake, E. Ilatova, P. Rice, C. Sabbagh, S. Voronkin, Y.S. Abraham, Waste management and operational energy for sustainable buildings: a review, Sustainability 12 (2020) 5337.
[60] A. Movahedi, S. Derrible, Interrelated patterns of electricity, gas, and water consumption in large-scale buildings (2020)..
[61] C.-L. Cheng, Study of the inter-relationship between water use and energy conservation for a building, Energy Build. 34 (2002) 261–266.