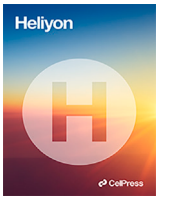
Heliyon 10 (2024) e30607

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Enhancing NILM classification via robust principal component analysis dimension reduction

Arbel Yaniv *, Yuval Beck

Department of Physical Electronics, School of Electrical Engineering, Tel Aviv University, Tel Aviv, Israel

A B S T R A C T

Non-intrusive load monitoring (NILM) techniques estimate the consumption of individual appliances in a household or facility, based on aggregated reading from a centralized meter. Usually, NILM techniques are shown to be improved when various power features and additional power quality parameters are included. However, adding power features leads to increased time complexity which is a disadvantage to real-time operation. Previous attempt to operate a principal component analysis (PCA) method to reduce the dimension of the problem managed to improve the run time but with considerably low accuracy. To this end, we utilize a robust PCA approach, to mitigate the influence of outliers in the data as a measure for improved performance. The proposed procedure achieves extraordinary results with accuracy over 96% for 600 hours long record of power quality measurements of the consumption of seven appliances from the standard AMPds dataset.

1. Introduction

Non-intrusive load monitoring (NILM) is a procedure to disaggregate the consumption of individual appliances by analyzing power parameters of the overall consumption in a facility from data measured at a centralized meter. The primary motivation behind NILM research is a more accurate forecasting and analysis of smart meter data, to provide actionable feedback for consumers who can, in turn, improve energy efficiency by up to 15% [1].

A recently explored approach for NILM introduces the application of computer vision techniques. This method includes transforming time series data of the power quality features, into a two-dimensional format. By doing so, the method leverages the robust capabilities of neural networks to reveal hidden patterns within the data [2]-[3]. Another artificial intelligence method being explored for NILM problems is the ensemble bagging tree architecture [4], where multiple decision trees are trained on diverse subsets of the training data, sampled with replacement. Their collective predictions are then aggregated to improve the model's overall accuracy and stability. Implementation of deep learning models for NILM problems, while powerful in disaggregating energy consumption, require significant computational resources for training. Moreover, considering the complexities and challenges of big-data management [5] is required. This includes data scarcity due to metering communication errors, absence of sufficiently labeled datasets necessary for the effective training of deep learning models, and the associated required high computing resources. Thus, adopting a comprehensive approach to guarantee scalability is essential. This includes everything from refining the algorithmic component to optimizing data management strategies, employing data splitting techniques for more manageable processing [6] and leveraging the computational power of graphical processing units (GPUs) [7].

This aspect poses challenges for real-time application, where rapid processing is essential for energy management [8]. The substantial computational demands during both the training phase and ongoing real-time analysis restrict the deployment of these

* Corresponding author.
E-mail address: arbelyaniv@mail.tau.ac.il (A. Yaniv).

https://doi.org/10.1016/j.heliyon.2024.e30607
Received 28 December 2023; Received in revised form 12 April 2024; Accepted 30 April 2024

models in large-scale use-cases. Furthermore, the energy required to run these complex models must not negate the energy savings they aim to achieve. Thus, the resources and energy associated with the deployment of deep learning models for NILM purposes requires further exploration to ensure their practicality and sustainability in real-time NILM applications.

Particularly, by integrating NILM-based algorithms directly into smart-meter equipment, the quality of the data as well as the computational complexity of the algorithm can become acute for several reasons:

1. Data quality: NILM algorithms rely on high-quality and accurate data to work effectively. If the smart meter's data is noisy or corrupted, the NILM algorithm's results may be unreliable.
2. Processing resources: Some NILM algorithms can be computationally intensive, requiring a significant amount of processing time and memory. This can be challenging for smart meters with limited resources.
3. Data storage: NILM algorithms often requires large amounts of data, which can quickly fill up the limited storage capacity of a smart meter.

Such integration of NILM technologies with real-time smart monitoring equipment could be used for several applications, such as intelligent home energy management systems [9] that provide real-time alerts and recommendations for energy savings actions, or ambient assisted living [10]-[11], to assist with the daily needs of the elderly population. Furthermore, real-time NILM technologies can enhance existing programs such as informed demand response enabling utilities to assess the capability for power reduction at any given time [12].

Utilizing a dimension reduction method as a preprocessing step can assist overcoming these limitations in three main aspects:

1. Reduce the complexity: Dimension reduction such as PCA can simplify the analysis and interpretation of NILM data, making it easier to identify patterns and trends in the data.
2. Accuracy: As dimension reduction methods are designed to find the components that capture the most variance in the data, they effectively compresses the information while retaining the most relevant features, enhancing the discriminative nature of the raw data.
3. Computation time: By reducing the dimension of the data, dimension reduction methods also help improve the performance of NILM techniques in terms of computational time, enabling NILM techniques to work more efficiently with a reduced set of features.

The importance of enhancing NILM algorithms efficiency cannot be overstated, especially when considering the vast computational resources necessary for deploying real-time, big-data reliant NILM applications at a large scale. Adopting dimension reduction as a preliminary processing step could be transformative, improving not only the accuracy, but also the computational efficiency. Integrating this data processing technique prior to the application of any NILM classification algorithm can significantly support and enhance the state-of-the-art energy disaggregation methods. This includes artificial-intelligence approaches, by serving as an essential and synergistic element of the overall analytical framework, driving the motivation behind this work.

In [13], a comparison of six dimensionality reduction techniques as preprocessing method for data of ten different domestic appliances is presented, demonstrating its contribution to NILM disaggregation results.

In [14], a dimensionality reduction scheme that relies on fuzzy-neighbors preserving analysis based QR-decomposition was developed to both reduce feature dimensionality, as well as to effectively decrease the intra-class distances and increase the extra-class distances of appliance features, thus providing better input data to the disaggregation classification process.

In [15], PCA dimensionality reduction was applied as an unsupervised NILM approach to identify power consumption patterns of home electrical appliances. In [16], PCA and k-means were used to detect the presence of appliance clusters, alongside a method to identify the appliances withing each cluster, followed by a minimum spanning tree as a dimension reduction for easier interpretation of the identified clusters. In [17], several pattern recognition algorithms for residential energy disaggregation were evaluated, including decision trees, support vector machine, optimum-path forest, multilayer perceptron, and k-nearest neighbors. The results show that decision trees, support vector machine and multilayer perceptron's confusion matrices were noticeably improved when using the PCA decomposition. In [18], PCA was applied on the public dataset AMPds and a private dataset. The original and modified dataset are then tested under two NILM methods of combinatorial optimization and modified cross-entropy [19] for the appliances disaggregation (classification) task. The results show that the run time is reduced and the accuracy is preserved compared to using the original measured features. However, the disaggregation quality is generally poor, with many appliances achieving F1-score (as detailed in section 6) of less than 50% (i.e. dishwasher- 0.5, heat pump- 0.2, kitchen fridge and oven- 0.3). This low accuracy is not sufficient for real-world applications. PCA has been widely used across various disciplines, prompting extensive research into its adaptation and modification to suit various objectives and data types. As the principal components are linear combinations of all original features, for high dimensional datasets, the interpretability of the PCs can be challenging. This invoked simplified PCs, which is a sparse variant of PCA. By imposing a constraint on the loadings [20], many of them are zeroed based on their value compared to a pre-defined threshold. In this variant, improved interpretability comes on the expense of the explained variance. Another variant of PCA is the symbolic data PCA, that was developed for dealing with grouped data measurements rather than an individual, both in space or time (interval-valued PCA) [21]-[22]. In the case where observations have a certain functional nature, functional PCA could be employed [23]. One approach could be to downsample the measurements, where the resulting components are sampled principal functions which can be smoothed to recover functional form. Another approach is to explicitly use the functional nature of the measurements, in which each row is referred to as a functional observation, that is a sample of the function at a specific time

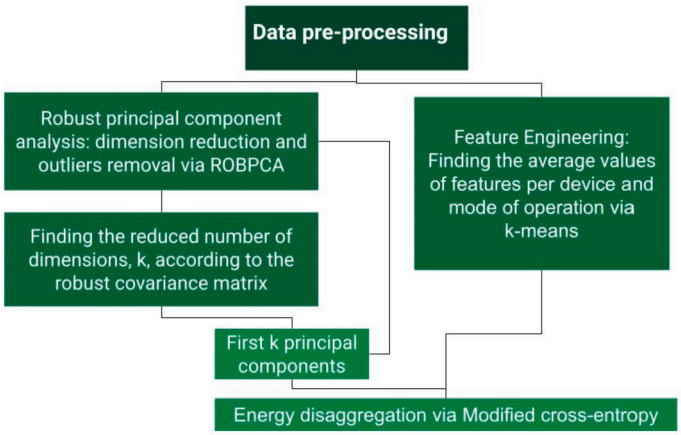


Fig. 1. Data pre-processing flow-chart for NILM.

(considering that each row represents a sample and each column a feature). Then, instead of a linear combination of the features with the loading coefficients to produce the transformed features, the new features are calculated as the functional inner product (integral) between the functional observation (original data) with loading functions. An approach to approximate the loading functions is to assume that the functional observations can be written as linear combinations of a set of basis functions that reflect the observations, such as Fourier series functions and spline functions. PCA's reliance on variance as a measure of significance means that it is inherently highly sensitive to the presence of outliers, which can significantly distort the resulting principal components. Consequently, developing robust versions of PCA that can withstand the effects of outliers and gross data errors has become one of the most active areas of research associated with PCA methodologies [24]-[25], and is applied also in this work.

2. Motivation and original contribution

Deploying NILM at a large-scale highlights the necessity for high computational efficiency, particularly when utility companies aim to roll out NILM services across a vast number of homes to achieve reliable consumption assessments. While dedicating a single server for NILM analysis for a single household does not pose a significant computational burden, scaling to utility-wide analysis presents a significant challenge, requiring substantial computational capabilities, including the adoption of parallel and multiprocessing techniques [26]. In deploying NILM software on such a large scale, it is vital to ensure that the computational costs associated with supporting such computational infrastructure, including energy consumption and associated carbon emissions, do not outweigh the energy savings achieved through NILM applications. Addressing the environmental impact of these computing systems is also of growing concern, emphasizing the need for sustainable computing solutions in large-scale NILM deployments [27]. This motivates this work, in which a robust PCA approach is proposed as a preprocessing step prior to the application of NILM disaggregation method. This preprocessing step reduces the dimension of the problem as well as mitigates the influence of outliers in the data, which could play an important role in big-data analysis for NILM datasets.

NILM accuracy and computation performance can be improved by the ability to attain measurements of numerous power features. This multidimensional feature scheme can add important information which is valuable for processing and disaggregating the devices. For improved computational efficiency, dimensionality reduction methods such as PCA could be utilized to reduce the dimension of the problem. Improvement in calculation time for multi-dimensional NILM algorithms was shown [18], but with no substantial improvement in the disaggregation results.

To this end, in this paper, the robust PCA method was applied on AMPds dataset as a preprocessing step, aiming to gain improved accuracy for the disaggregation task, based on MCE multi-label classification algorithm. The flowchart of the entire data pre-processing procedure is shown in Fig. 1. The process includes a data pre-processing step that is composed of two parts. The first part is finding the average values of the measured features per device and mode of operation, and is required for the disaggregation MCE algorithm as an input. The second part includes the dimension reduction and outlier removal by robust PCA as well as finding the reduced number of dimensions. The method was also applied on the iAWE dataset to further demonstrate the generalizability of the method. The results show that the proposed approach attains excellent accuracy of over 96.3% F1-score for the seven tested AMPDs appliances.

3. Robust PCA overview

PCA is a statistical method used to identify patterns in data, and represent the data in a lower dimensional space, while retaining as much of the original variability in the data as possible. This is achieved by transforming the original data set into a new coordinate system, where the new axes (known as "principal components") are orthogonal and ranked by the amount of variance they capture.

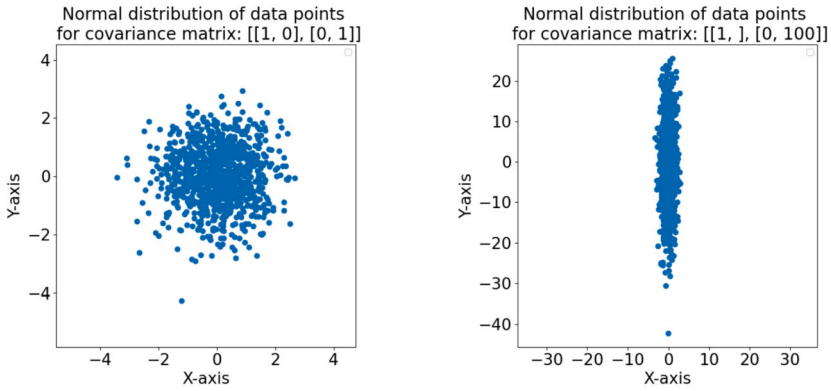


Fig. 2. A geometric interpretation of the covariance matrix.

Thus, the new data representation is a linear combination of the original variables. The first component corresponds to the direction of largest variance and each subsequent component is orthogonal to the previous and maximizes the variance.

However, this approach is sensitive to outliers, which may attract the first components, leading to unreliable data reduction.

The objective of robust PCA (RPCA) is to identify components that are not significantly influenced by outliers. However, since the term "outlier" lacks a precise mathematical definition, the RPCA problem is not well-defined. Nonetheless, various classical heuristic methods have been developed to tackle this problem, as highlighted in [28]. One common approach for handling outliers is to consider them as an additive sparse corruption [29]. This definition can be seen as valid since it acknowledges the sporadic and diverse nature of outliers, which can appear with varying magnitudes and frequencies. Corresponding to this definition, RPCA involves decomposing a given data matrix into a sparse (noise) and a low-rank component (signal of interest), where the column space of the low-rank matrix then gives the desired principal subspace (PCA solution). By doing so, it is possible to isolate the robust components from the potentially disruptive outliers. RPCA has been explored via the robust matrix completion [30]-[31] approach in which given some incomplete observations $\hat{X}$ of a corrupted and possibly noisy data matrix X, which is the superposition of a low-rank matrix L and a sparse matrix S, the task is to recover L and S. Aiming to find a low-rank approximation of the data matrix and the corresponding sparse noise matrix [32]-[33], could be formulated as the optimization problem in Eq. (1):

$$\min_{L,S} \ \mathrm{rank}(L) + \lambda \|S\|_{\ell} \qquad (1)$$

$$\text{s.t.} \quad X = L + S,$$

where $\mathrm{rank}(Z)$ denotes the function that returns the rank of the matrix Z, and $\|S\|_{\ell}$ denotes a regularization term like ℓ_0-norm for promoting the sparsity of E. The parameter λ is employed to balance the low-rank and sparse components in X. This problem is generally NP-hard as the rank function is discrete and nonconvex.

In the optimization-based approach, the covariance matrix plays a critical role as it is used to estimate the low-rank and sparse components of the data. Examples of optimization-based approaches for RPCA include Linearized Alternating Direction Method (LADM) [34], Augmented Lagrange Multiplier (ALM) [35], and Minimum Covariance Determinant (MCD) [36] methods.

The determinant of a covariance matrix is related to the volume of the data structure's shape that defines the distribution of the data points. The larger the determinant, the more elongated the data structure is, which indicates that the data points are more spread out along certain directions, as illustrated in Fig. 2. In the figure, two normal distribution datasets of 1000 points are shown, with determinant of 10 in two different directions of x-axis and y-axis correspondingly. A subset of data points with a small determinant are likely to be a good representation of the main structure of the data, while ignoring outliers and other anomalies that may increase the determinant. By finding the subset with the smallest determinant, it can be ensured that the estimated covariance matrix is based on the most typical and informative data points.

Accordingly, the MCD method finds a subset of the data that has the smallest covariance determinant, subject to some constraints to ensure that the selected subset has a certain minimum size, and that it represents a statistically significant portion of the data. Once the subset has been identified, the covariance matrix is estimated based on this subset alone. By replacing the covariance matrix with the MCD estimator, which is robust to outliers, it is possible to obtain more accurate estimates of the low-rank and sparse components, as the MCD estimator is based on a subset of the data that is less influenced by outliers, and therefore produces a more reliable estimate of the underlying covariance structure of the data.

The MCD estimator enhances robustness and is practical for datasets with relatively low dimension compared to the number of observations, such as in this work. This is due to the fact that the MCD estimator identify a subset with the smallest determinant of the covariance matrix, and this computation will have a value greater than zero only if the number of features is smaller than the dimension of the subset (otherwise the covariance matrix of the subset is singular).

4. Modified cross-entropy overview

The NILM disaggregation method that is employed in this work is Modified Cross-Entropy (MCE) [19]. The NILM problem is modeled as a combinatorial optimization problem with the objective to decompose the aggregated consumption signal into its components $p_{i,j}$:

$$\hat{\mathbf{S}}[n] = \arg\min_{\hat{s_{i,j}}[n]} \left\| \mathbf{P}[n] - \sum_{i=1}^{m} \sum_{j=1}^{k} \hat{s_{i,j}}[n] p_{i,j} \right\| \tag{2}$$

where n is the number of time samples, m is the number of appliances, k is the number of modes per appliance (as appliances can have multiple states and not only on/off operation). $s_{i,j}$ and $p_{i,j}$ are the estimated state (operating/non-operating) and average power consumption (that is, the power consumption that corresponds to the operational mode if the appliance is working) of appliance i and mode j, correspondingly. $\mathbf{P}[n]$ is the measured power at time n, and $\hat{\mathbf{S}}[n] = [\hat{s}_1[n], \ldots, \hat{s}_m[n]]$ is a matrix of $k \cdot m$ where each column corresponds to an appliance and each row to a mode such that the elements are binary values representing if at time n, appliance i was operating at state j. Note that Eq. (2) is defined in relation to the active power but can be extended to encompass other electrical features that exhibit additive properties. The operation of the CE method [37] for combinatorial optimization involves an iterative procedure. In the first iteration, a random data of X samples (estimation of appliances mode of operation). Each sample is evaluated based on the objective (i.e. the sample performance). A certain group of samples with the nest performance are selected as the 'elite group'. For each such sample, statistics are collected (regarding the estimated values) and used to update the distribution from which the data will be generated in the next iteration. As in this process, the solution for each point in time is independent of the others, the solution does not incorporate the inherent time dependency of the NILM problem. Specifically, the search space for finding a solution is k^M (the number of modes time the number of appliances), which could be very large even for small-scale scenarios, out of which many options, even-though with good enough performance to be considered as a part of the elite group, are practically not feasible. Thus the modification penalizes the randomized solutions that are generated in the CE process based on the hamming distance [38] between the generated solution and the known previous states. This modification utilizes the nature of the problem to refine the search space to encompass reasonable options. For example, based on the examination of the REDD database, a valid number of changes between events is up to $\lfloor \log_2(M) \rfloor$. Moreover, the changes between modes for the same appliance are also taken into consideration and modeled in the penalty function for further refinement of the search process. The use of PCA, which is a linear dimension reduction technique, and MCE, which is additivity-based energy disaggregation method, preserve the underlying correlations of the actual consumption patterns. The consumption patterns as reflected by the appliance clusters reflect a distinct pattern of usage within households. These clusters are not abstract statistical groups but are directly tied to real-world appliances and their operational habits. For example, a cluster that encapsulates appliances with high energy demands. This methodological pathway from data pre-processing (PCA) to classification (MCE) retains a clear lineage back to the actual appliances, thus facilitating a direct interpretation of the energy usage patterns, preserving the underlying correlations of the actual consumption patterns.

5. Settings and test scenario

AMPds is an open data set proposed by Simon Fraser University [39], and is a popular domestic appliances' power measurements' dataset. The data is collected from a house in Canada over 1 year. AMPds contains low sample-rate recordings of 60 sec (sample per minute) for 21 sub-meters, and includes active, reactive and apparent powers, in addition to current and power factor signals. The simulations in this paper are based on 7 sub-meters that were monitored and labeled, six of them containing a single load, and one main smart-meter that measures all of the six loads together. The tested scenario is the same as in [18], which used MATLAB built-in PCA method and a classification method as detailed below for NILM disaggregation on AMPds dataset, as follows: The scenario contains 25 days (600 hours) of measurement at a sample rate of 1 minute. The scenario starts on April 05, 2012, 15:38:00. The MCE classification algorithm is used as the NILM disaggregation method and its implementation is identical to the one in [18], including the same configuration as well as its application on multiple features rather than classifying based on a single feature (active power) as was done in the original MCE version of the algorithm [19]. Figs. 3-4 include the first day (24 hours in minute resolution) and the fourth day of the dataset that is used in our simulations. Each sub-figure includes the current, active and reactive power graphs as a function of sample index (minutes) in yellow, red, green and blue accordingly. Note that there are two y-axis for clearer visualization of the features in the same sub-figure: the left y-axis is for the power values and the right y-axis is for the current values as measured in Amperes. As can be seen, each device works at different times. For example the oven was not operating on the first day while in the fourth day measurements of its operation were recorded.

5.1. Finding the reduced number of dimensions

For the baseline scenario, prior to operating the PCA method on the data, a training phase was conducted with an objective to calculate the principal components and choose the number k for the reduced dimensions. Utilizing robust PCA [40] MATLAB implementation from LIBRA toolbox [41], the robust covariance matrix based on the MCD method [42] is used for this purpose as follows. The robust covariance matrix, C, is computed according to Eq. (3):

$$C = P \cdot L \cdot P^T \tag{3}$$

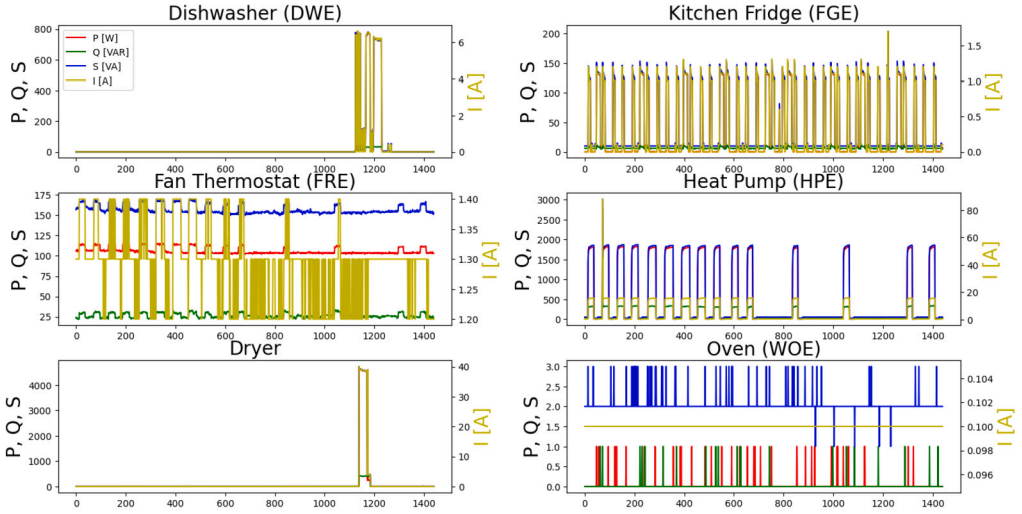


Fig. 3. AMPds sample data of the first day of the tested scenario.

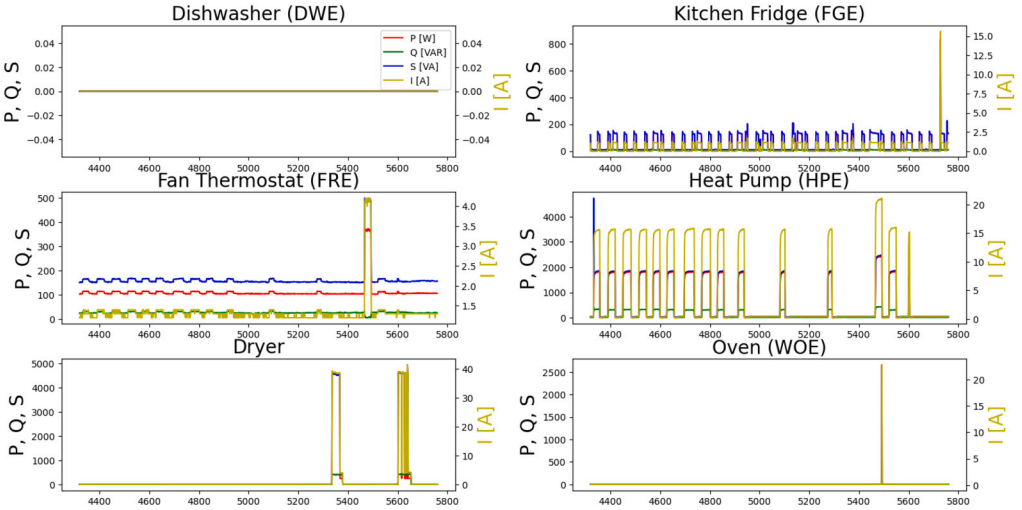


Fig. 4. AMPds sample data of the fourth day of the tested scenario.

Where P are the robust eigenvectors and L are the robust eigenvalues. As the objective of the PCA procedure applied on data with p features is to maximize the variance of the first k components and minimize the variance of the last $p-k$ components compared to all other orthogonal transformations, the number of principal components, k, is chosen based on the formula of the cumulative explained variance of the robust PCA procedure. k is chosen such that it captures a desired percentage of the total variance in the data, which is the cumulative explained variance. The chosen value is the smallest one for which the cumulative explained variance is greater than a threshold value t. This can be expressed mathematically as:

$$k = \min\{i : \frac{\sum_{j=1}^{i} \lambda_j}{\sum_{j=1}^{p} \lambda_j} \geq t\} \tag{4}$$

where i ranges from 1 to p, and λ_j is the eigenvalue of the j^{th} component.

A common threshold value for t is 0.95, which means that at least 95% of the total variance in the data is captured.

Examining the plot of the eigenvalues (Fig. 5), the selected number of principal components is one, based on Eq. (4). In this case, as for $i = 1$, the numerator consists of λ_1, while the denominator is the sum of the eigenvalues, essentially λ_1 plus a value of negligible significance. This yields a ratio near unity, specifically exceeding the 0.95 threshold.

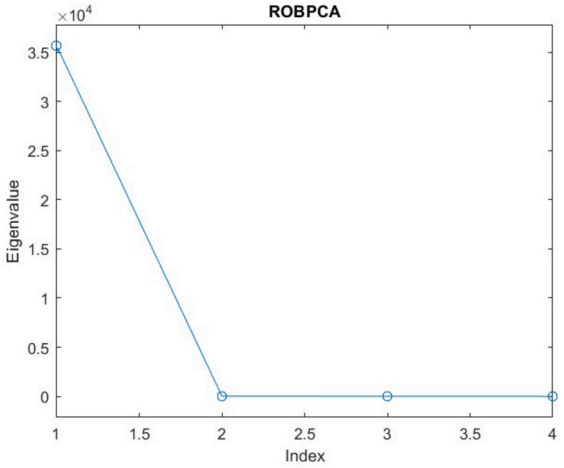


Fig. 5. Robust principal component eigenvalues vs. principal component number.

5.2. Finding the average values of features per device and mode of operation

Disaggregation by MCE algorithm is based on the knowledge of the average value of each feature per device, as the appliances combination search assumes that the examined electrical features are additive in the sense that the sum of consumption per device (measurements of electrical features by per-device meters) equals the total consumption of all working appliances (electrical features records by the main smart-meter measuring the total consumption of multiple appliances). This implies that any dimension reduction approach that is applied prior to the NILM disaggregation method which is based on this additivity assumption, such as MCE, must be linear. Then, the loadings coefficients (also known as weights) that represent the contribution of each original variable to the various principal components are found for the total appliances' consumption as measured by the main smart-meter, and is later applied to all per-device measurement values.

As some appliances can have multiple modes of operation, the algorithm requires information about the average value of the features for all possible modes of operation in the dataset. Finding this value was previously done manually [18], which is unpractical, unscalable and prone to errors. In order to find these values, automation of the procedure, in the form of k-means clustering method was used in this work, with 7 clusters related to on/off modes as well as additional modes (such as additional modes for the AMPDs recorded dishwasher of wash, rinse and dry). The number of clusters effects the results of the MCE classification algorithm, as elaborated later.

5.3. The dimension reduction robustness to outliers

The dimension reduction is done based on the MCD estimator (the loadings are computed as the eigenvectors of the MCD scatter matrix), which is robust to outliers in the data. The explanation of this outliers robustness is detailed next. Given n data points, the raw MCD estimator of those data points is based on the mean and covariance matrix based on the sample of size h, where $h < n$, which minimizes the determinant of the sample covariance matrix. The value h can be thought of as the minimum number of points that are not considered outliers. Based on this raw MCD estimator, $(\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0)$ it is possible to define the statistical distance, d, which is calculated for each point with index i, x_i, in the dataset, d_i, as shown in Eq. (5):

$$d_i = \sqrt{(\mathbf{x_i} - \hat{\boldsymbol{\mu}}_0)^t \hat{\boldsymbol{\Sigma}}_0^{-1} (\mathbf{x_i} - \hat{\boldsymbol{\mu}}_0)} \tag{5}$$

where $\hat{\boldsymbol{\mu}}_0$ is the mean of the h observations for which the determinant of the sample covariance matrix is as small as possible, and $\hat{\boldsymbol{\Sigma}}_0$ is the corresponding covariance matrix multiplied by a consistency factor. Now one can define the weighted MCD estimates (for better efficiency while retaining high robustness), $(\hat{\mu}_{MCD}, \hat{\Sigma}_{MCD})$:

$$\begin{aligned} \hat{\mu}_{MCD} &= \frac{\sum_{i=1}^{n} W(d_i^2) x_i}{\sum_{i=1}^{n} W(d_i^2)}, \\ \hat{\Sigma}_{MCD} &= c_1 \cdot \frac{1}{n} \sum_{i=1}^{n} W(d_i^2)(x_i - \hat{\mu}_{MCD})(x_i - \hat{\mu}_{MCD})^t \end{aligned} \tag{6}$$

where $W(d_i^2)$ equals 1 when the calculated distance is below the cutoff, $\sqrt{\chi^2_{\alpha,p}}$, and 0 otherwise. Note that the cutoff is the threshold value which is given based on the chi-squared distribution, the number of features, p, and the desired probability, α, for which the distance would be lower than the cutoff.

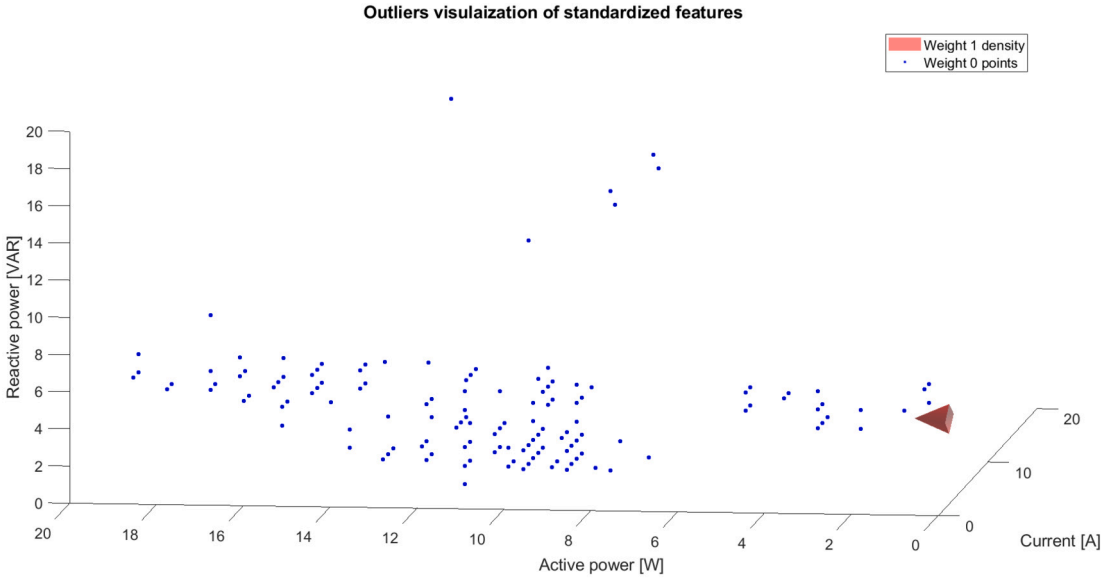


Fig. 6. AMPDs test scenario scaled standardized outliers visualization.

In Fig. 6, a visualization of the AMPds test scenario's observations is shown to illustrate the outlier detection via the robust MCD estimation. In the figure, each data point corresponds to a transformation of the three features. The transformation includes a third-order one-dimensional median filter to reduce noise. In addition, the data points in the visualization are standardized as part of the robust MCD estimation procedure according to Eq. (7):

$$\mathbf{X}' = \frac{\mathbf{X} - \mathbf{Median}}{\max(|\mathbf{X} - \mathbf{Median}|)} \tag{7}$$

where $\mathbf{X}'$ is the transformed dataset and **Median** is a vector with a median value per feature. Points that are assigned with weight 0, that are labeled as outliers and are not taken as part of the MCD estimates as shown in Eq. (6) are colored in blue.

The visualization incorporates an isosurface [43] for enhanced data representation as a significant number of the points are densely clustered in a small, high-density area. This method effectively highlights the region where points are not detected as outliers. Without the isosurface, the visualization would fail to clearly display this dense area, making it challenging to discern the non-outlier points in the broader spatial context that also includes outliers. Specifically, in the test scenario presented, 6,828 out of 36,001 points were identified as outliers. The isosurface visualization distinctly marks the area of non-outliers points as assigned with weight $W(d_i^2) = 1$ by the MCD procedure to effectively demonstrate those points compared to the scattered isolated outliers. In order to create this visualization, first the 3D points are organized into bins according to their spatial coordinates, with separate bins based on their associated weights (0 or 1). This process effectively quantifies the spatial distribution of points based on their weights. The isosurface function computes and renders a surface that encloses the volume where the density of such points exceeds a threshold that dictates the minimum density required for a region to be highlighted by the isosurface. In addition to the isosurface, locations of points with weight 0 (outliers) appear as a scatter.

6. Results

For evaluating the disaggregation results for robust PCA applied NILM dataset, the score metric used is F-measure, which is a common metric for binary classification algorithms:

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$

where β is a positive constant that controls the trade-off between precision and recall. Typically, $\beta = 1$ is used to calculate the F1-score, which is the harmonic mean of precision and recall.

Precision is important in NILM because it measures the accuracy of the disaggregation algorithm in correctly identifying the appliance that is responsible for a particular energy consumption pattern. On the other hand, recall is important because it measures the algorithm's ability to detect all instances of a particular appliance in the energy consumption data.

Thus, F-measure, balances the importance of precision and recall in evaluating the performance of a NILM algorithm. Precision and recall are defined as follows:

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

Table 1
Modified cross entropy F1-score (AMPds dataset).

Appliance/Feature	P,Q	PCA: PC_1, PC_2	RPCA: PC_1	RPCA: PC_1, PC_2	RPCA: PC_1 RDG
Basement (BME)	0.409	0.862	0.966	0.813	0.9213
Dishwasher (DWE)	0.517	0.5	0.963	0.948	0.9172
Dryer (CDE)	0.564	0.923	0.963	0.747	0.9215
Fan thermostat (FRE)	0.912	0.946	1	1	0.9883
Heat pump (HPE)	0.226	0.206	0.998	0.995	0.9708
Kitchen fridge (FGE)	0.742	0.313	0.966	0.954	0.8934
Oven (WOE)	0.390	0.317	0.963	0.946	0.9201
Average	0.617	0.654	0.975	0.922	0.93

Table 2
NILM computation time on the AMPds test-scenario.

Scenario	Avg. F-measure	computation time [sec]
P	0.617	6080.57
P, Q	0.654	6448.79
PC_1	0.975	5818.35

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

where true positives (TP) are the number of correct positive predictions, false positives (FP) are the number of incorrect positive predictions, and false negatives (FN) are the number of incorrect negative predictions.

The F1-score results obtained by applying the robust PCA on the AMPds dataset followed by NILM disaggregation via MCE algorithm are shown in Table 1. The dataset includes seven devices, including Type I devices: basement, fan thermostat and kitchen fridge, as well as Type II devices: dishwasher, dryer, heat pump and oven, as follows:

1. Type I: Appliances with only two operation states (ON/OFF). Type I loads consists of most household appliances, such as microwave, air conditioner, toaster.
2. Type II: Appliances with finite number of working states, also named as Finite State Machine (FSM). Type II loads has a repeatable switching pattern, such as coffee machine, washing machine, dishwasher and oven.

The test scenario results of the original data features (P,Q) and with applying MATLAB built-in PCA are taken from the baseline work [18] on the same test scenario.

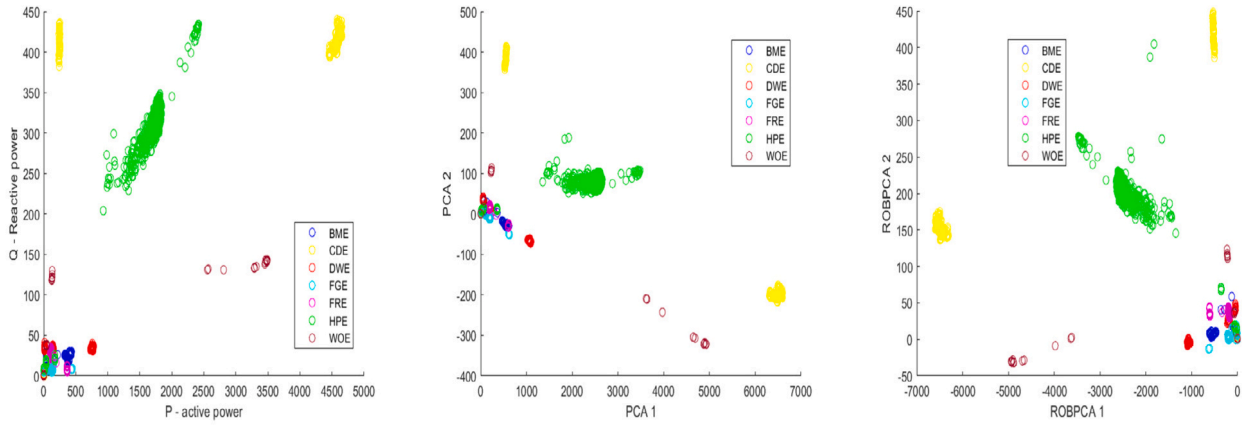
Regardless of using PCA or robust PCA, the results show improvement when principal components are used for classification compared to the original measured features. As presented in the baseline work, using two features achieved improved results both for the original data (P, Q better compared to P only), as well as for the principal components based on standard PCA. That is, the results based on the first two principal components, namely PC_1 and PC_2, are better compared to the results based on the first principal component, PC_1, only. However, for robust PCA, as demonstrated in section 5.1 and according to the cumulative explained variance, using a single robust feature PC_1 is better than using two robust components PC_1 and PC_2.

The proposed procedure of robust PCA with a single component yields improved results compared to all other options: the original measured features of the raw data (P,Q), regular PCA (PC_1 and/or PC_2) and two principal components of robust PCA.

For first principal robust PCA PC_1, the F1-score measure for all appliances is greater than 0.96, which is very accurate in comparison to other alternative approaches common in NILM disaggregation research. Using the original measured features or PCA yields minimum F1-score of 0.2 (for HPE), and most devices has average F1-score which is significantly smaller, by more than 30%.

The simulations performed in this work were carried out on an 11th Generation Intel(R) Core(TM) i7-1165G7 with 4 cores running at 2.80 GHz with 32.0 GB of RAM installed and using PyCharm. The computational time results are detailed in Table 2, containing the computation time for NILM disaggregation based on single or two parameters (P or P,Q) of raw data as well as the suggested best performing scenario of a single parameter PC_1 of robust PCA of the AMPds test-scenario. Each disaggregation was performed for the database of 36,000 samples (600 hours with 1 minute sample rate). The results show that the proposed approach has the best calculation time, with improvement of 9.7% compared to the results achieved with two raw parameters and 4.3% compared to the results achieved with a single raw parameter, along with a significant accuracy improvement.

The scatter plots of the (a) raw data of active and reactive power, (b) the first two principal components according to the proposed robust PCA procedure, and (c) the first two principal components according to the standard PCA procedure (same method used in [18]) are shown in Fig. 7. As can be seen, the level of separation between samples of different appliances is significantly higher for the proposed method, showing improved distinction between the clusters. Moreover, considering the scatter plot corresponding to



(a) P, Q of AMPDs samples (b) *PCA* first two PCs (c) *RobustPCA* first two PCs

Fig. 7. Left to right: P, Q of AMPDs samples, *PCA* first two PCs, and *RobustPCA* first two PCs.

Table 3
Modified cross entropy F1-score for two robust PCA features as a function of the number of clusters on the AMPds test scenario.

Appliance/Num. of clusters	5	6	7	8
Basement (BME)	0.915	0.964	0.813	0.968
Dishwasher (DWE)	0.312	0.953	0.948	0.962
Dryer (CDE)	0.915	0.954	0.747	0.962
Fan thermostat (FRE)	0.989	0.946	1	0.996
Heat pump (HPE)	0.2606	0.257	0.995	0.261
Kitchen fridge (FGE)	0.945	0.954	0.954	0.971
Oven (WOE)	0.915	0.961	0.946	0.96
Average	0.808	0.899	0.922	0.9

the proposed robust PCA method, it can be observed that the projections of the clusters on the first PC are more separated compare to the projections on the second PC, in alignment with the presented accuracy results.

7. Sensitivity analysis

In this section, sensitivity analysis was conducted considering various aspects of the model. This includes the selection of the number of clusters for both AMPds as well as an additional NILM dataset, providing a comprehensive assessment of the procedure under different average consumption inputs to the MCE algorithm for a variety of appliances. The analysis also includes robustness to noise examination via the incorporation of an extra noise model of the meter's measurements accuracy. Additionally, the method's scalability was validated through simulations that ran for longer periods, increasing the amount of considered operational states.

7.1. Number of clusters

Examining the simulation results for two components of robust PCA for various selections of the number of clusters as detailed in sub-section 5.2, it is possible to clearly observe the effect on the accuracy. This analysis was conducted on AMPds dataset for number of clusters between 5 to 8. As elaborated earlier, these clusters are used for the computation of the average values of features per device and mode of operation (inputs to the MCE disaggregation algorithm). The results are summarized in Table 3. As can be seen, the average consumption values as represented by 7 clusters per feature per device yields best results.

Additionally, the robust NILM procedure was applied on the iAWE dataset [44] to further demonstrate the generalizability of the method. This dataset includes measurements for 10 residential appliances in a home located in New Delhi India. The sampling rate is 1 Hz for the voltage, current, frequency and power features, and the recording period was 73 days in summer 2013. The collection of appliance level data was conducted via jPlugs, which collect data only when the appliance was ON. Additionally, the maximum amount of sensor data availability is for 23 days between 13-7-2013 and 4-8-2013. This period is a similar period to the examined test scenario simulated on the AMPds as described in the previous paragraph. Thus, a few data pre-processing were conducted before applying the simulations, as follows. The dataset segment from the start date of 13-7-2013 and 4-8-2013 was down-sampled to 1 minute resolution and the gaps of missing data was filled with zeros. The results for various selections of the number of clusters for

Table 4
Modified cross entropy F1-score for the first robust PCA feature as a function of the number of clusters on the iAWE test scenario.

Appliance/Num. of clusters	2	3	7	9
Air-conditioner	0.565	0.612	0.597	0.598
washing-machine	0.863	0.892	0.975	0.703
Laptop	0.602	0.599	0.596	0.597
Iron	0.637	0.652	0.975	0.7
Kitchen outlets	0.863	0.892	0.6	0.605
Television	0.8646	0.893	0.975	0.975
Water filter	0.6	0.602	0.603	0.603
Average	0.71	0.73	0.76	0.683

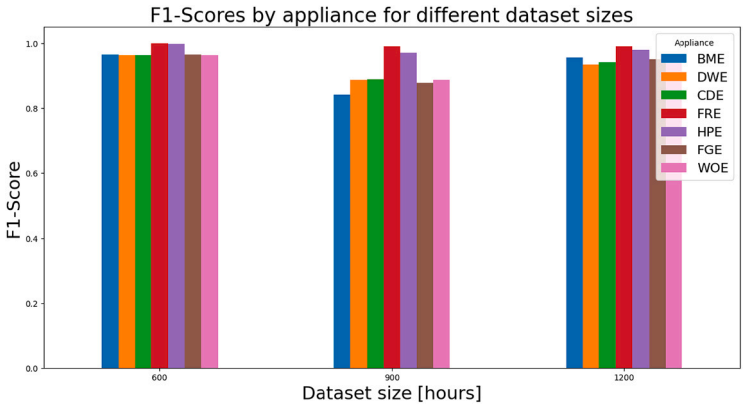


Fig. 8. AMPDs scenario size vs. f1-score.

the MCE accuracy based on the first principal component are summarized in Table 4. As can be seen, the best results are attained for the same selection of number of clusters as for the AMPds dataset.

7.2. Robustness to noise

AMPds electricity meters are all DENT Class 0.5 with high accuracy with typical full scale error lower than 0.5-1%, following ANSI C12.20 and IEC 62053 accuracy standards. Full scale accuracy refers to the meter's precision relative to its full-scale value, and is defined as a percentage of the total meter's scale capacity. For example, for a voltmeter with a full-scale value of 200 volts and a full-scale accuracy of $\pm 0.5\%$, the true measurement can deviate by as much as ± 1 volts across the entire scale, regardless of the actual reading. Yet, the error is still highly correlated with the reading [45]. Thus, a straight-forward model for sensitivity analysis would be to contaminate the signals with a Gaussian white noise, with variance proportional to the measured signals. This kind of measurement error (which is proportional to the actual reading) is referred to as RDG accuracy, for which the error is a percentage of the actual reading value. For example, for a voltmeter with an RDG accuracy of $\pm 0.5\%$, and for a measurement of 100 volts, the true value can be anywhere between 99.5 volts and 100.5 volts. The results for the best use-case (first robust principal component based MCE) contaminated signal with 1% RDG error are shown in Table 1. As can be seen, the accuracy is slightly lower, but still very high, with an average accuracy reduction of 4% for test case of both full-scale as well as synthetic RDG measurement error.

7.3. Scalability

The f1-score accuracy of the robust NILM procedure as simulated on various durations of recordings of the AMPds dataset is shown in Fig. 8, up to simulation of 1-minute resolution appliances classification for 1,200 hours long recording, corresponding to energy disaggregation for 72,000 samples, demonstrating the approach's scalability, with persistent accuracy results.

8. Conclusions and future work

The results demonstrate the dominant benefit of using a robust PCA method for NILM purposes with an optimization-based disaggregation method of MCE. The results are indubitably very accurate compared to other NILM approaches including other optimization methods, pattern recognition approaches as well as supervised and unsupervised learning [8]. Performing disaggregation with the

MCE algorithm on the original measured features compared to robust PCA features shows a significant difference in accuracy, making the MCE applicable for real-world purposes, otherwise practically unusable due to insufficient accuracy performance.

Robust PCA, being a linear dimensionality reduction method, does not account the non-linear relationships between electrical features. Therefore, future work could involve comparing NILM classification results from robust PCA pre-processing with those attained using non-linear dimension reduction algorithms such as vector diffusion maps [46]. It should be noted that for a fair comparison of robust PCA to any nonlinear dimension reduction pre-processing procedure, the appliance classification algorithm can not be the MCE used here. Rather, it will require other disaggregation approaches such as machine-learning ones, as they focus on uncovering hidden patterns within the data rather than employing a direct decomposition. In contrast, white-box models, which aim to decompose the observed power measurement into potential combinations of appliance power signals, hinge on the additivity assumption. This assumption does not hold up under non-linear transformations. Such a comparison to the disaggregation results with a non-linear dimension reduction pre-processing step could be beneficial for several factors, including the potential for improved accuracy and interpretability as nonlinear methods can capture more complex relationships between different features of the data. However, these possible improvements may result in an increase in computational complexity, which is an important factor for NILM applications as discussed earlier. Thus, a comprehensive comparison of the various measures could be valuable.

As the robust PCA approach identifies outliers, it can be also used for outlier-detection based analysis as a complimentary process to the NILM technology, such as for demand-side management companies to help monitor consumer appliances and identify and alert regarding any unusual or anomalous behavior. This may be also useful for further research in the domain of NILM outlier cause analysis, and practical for hardware and manufacturing improvements as well as predictive maintenance applications of the monitored devices.

This work is the first attempt to perform robust dimension reduction pre-processing techniques to enhance NILM accuracy and computational efficiency. It serves as an important milestone for the deployment of real-time large-scale NILM applications, as a suggested preliminary building block of any NILM solution. The results highlight the viability of the approach to serve as a complementary step in any other state-of-the-art approaches.

CRediT authorship contribution statement

Arbel Yaniv: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Yuval Beck:** Writing – review & editing, Validation, Supervision, Conceptualization.

Declaration of competing interest

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

Data availability

The datasets utilized in this study are publicly available.

Acknowledgements

This research was supported by the Israel Ministry of Infrastructure, Energy and Water Resources.

References

[1] B. Neenan, J. Robinson, R. Boisvert, Residential electricity use feedback: a research synthesis and economic framework, Elect. Power Res. Inst. 3 (2009) 123–129.
[2] Y. Himeur, A. Alsalemi, F. Bensaali, A. Amira, Appliance identification using a histogram post-processing of 2d local binary patterns for smart grid applications, in: 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021, pp. 5744–5751.
[3] Y. Himeur, A. Alsalemi, F. Bensaali, A. Amira, Smart non-intrusive appliance identification using a novel local power histogramming descriptor with an improved k-nearest neighbors classifier, Sustain. Cities Soc. 67 (2021) 102764.
[4] Y. Himeur, A. Alsalemi, F. Bensaali, A. Amira, Robust event-based non-intrusive appliance recognition using multi-scale wavelet packet tree and ensemble bagging tree, Appl. Energy 267 (2020) 114877.
[5] Y. Himeur, M. Elnour, F. Fadli, N. Meskin, I. Petri, Y. Rezgui, F. Bensaali, A. Amira, Next-generation energy systems for sustainable smart cities: roles of transfer learning, Sustain. Cities Soc. (2022) 104059.
[6] G. Bathla, H. Aggarwal, R. Rani, Scalable recommendation using large scale graph partitioning with pregel and giraph, Int. J. Cogn. Inf. Nat. Intell. 14 (4) (2020) 42–61.
[7] H. Li, K. Li, J. An, W. Zheng, K. Li, An efficient manifold regularized sparse non-negative matrix factorization model for large-scale recommender systems on gpus, Inf. Sci. 496 (2019) 464–484.
[8] G.-F. Angelis, C. Timplalexis, S. Krinidis, D. Ioannidis, D. Tzovaras, Nilm applications: literature review of learning approaches, recent developments and challenges, Energy Build. (2022) 111951.
[9] R. Ramadan, Q. Huang, O. Bamisile, A.S. Zalhaf, Intelligent home energy management using Internet of things platform based on nilm technique, Sustain. Energy Grids Netw. 31 (2022) 100785.
[10] H. Bousbiat, C. Klemenjak, G. Leitner, W. Elmenreich, Augmenting an assisted living lab with non-intrusive load monitoring, in: 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), IEEE, 2020, pp. 1–5.
[11] A. Ruano, A. Hernandez, J. Ureña, M. Ruano, J. Garcia, Nilm techniques for intelligent home energy management and ambient assisted living: a review, Energies 12 (11) (2019) 2203.

[12] S. Barker, S. Kalra, D. Irwin, P. Shenoy, Nilm redux: the case for emphasizing applications over accuracy, in: NILM-2014 Workshop, 2014.
[13] V. Isanbaev, R. Baños, F.M. Arrabal-Campos, C. Gil, F.G. Montoya, A. Alcayde, A comparative study on pretreatment methods and dimensionality reduction techniques for energy data disaggregation in home appliances, Adv. Eng. Inform. 54 (2022) 101805.
[14] Y. Himeur, A. Alsalemi, F. Bensaali, A. Amira, Effective non-intrusive load monitoring of buildings based on a novel multi-descriptor fusion with dimensionality reduction, Appl. Energy 279 (2020) 115872.
[15] A. Moradzadeh, O. Sadeghian, K. Pourhossein, B. Mohammadi-Ivatloo, A. Anvari-Moghaddam, Improving residential load disaggregation for sustainable development of energy via principal component analysis, Sustainability 12 (8) (2020) 3158.
[16] F.S. Villar, P.H.J. Nardelli, A. Narayanan, R.C. Moioli, H. Azzini, L.C.P. da Silva, Noninvasive detection of appliance utilization patterns in residential electricity demand, Energies 14 (6) (2021) 1563.
[17] W.A. Souza, F.P. Marafão, E.V. Liberado, M.G. Simões, L.C. Da Silva, A nilm dataset for cognitive meters based on conservative power theory and pattern recognition techniques, J. Control Autom. Electr. Syst. 29 (2018) 742–755.
[18] R. Machlev, D. Tolkachov, Y. Levron, Y. Beck, Dimension reduction for nilm classification based on principle component analysis, Electr. Power Syst. Res. 187 (2020) 106459.
[19] R. Machlev, Y. Levron, Y. Beck, Modified cross-entropy method for classification of events in nilm systems, IEEE Trans. Smart Grid 10 (5) (2018) 4962–4973.
[20] H. Zou, L. Xue, A selective overview of sparse principal component analysis, Proc. IEEE 106 (8) (2018) 1311–1320.
[21] L. Rouani, M.F. Harkat, A. Kouadri, S. Mekhilef, Shading fault detection in a grid-connected pv system using vertices principal component analysis, Renew. Energy 164 (2021) 1527–1539.
[22] E. Diday, Principal component analysis for categorical histogram data: some open directions of research, in: Classification and Multivariate Analysis for Complex Data Structures, Springer, 2011, pp. 3–15.
[23] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philos. Trans. R. Soc. A, Math. Phys. Eng. Sci. 374 (2065) (2016) 20150202.
[24] C. Peng, Y. Chen, Z. Kang, C. Chen, Q. Cheng, Robust principal component analysis: a factorization-based approach with linear complexity, Inf. Sci. 513 (2020) 581–599.
[25] J.-H. Yang, X.-L. Zhao, T.-Y. Ji, T.-H. Ma, T.-Z. Huang, Low-rank tensor train for tensor robust principal component analysis, Appl. Math. Comput. 367 (2020) 124783.
[26] Y.-C. Hu, Y.-H. Lin, C.-H. Lin, Artificial intelligence, accelerated in parallel computing and applied to nonintrusive appliance load monitoring for residential demand-side management in a smart grid: a comparative study, Appl. Sci. 10 (22) (2020) 8114.
[27] M.W. Asres, L. Ardito, E. Patti, Computational cost analysis and data-driven predictive modeling of cloud-based online-nilm algorithm, IEEE Trans. Cloud Comput. 10 (4) (2021) 2409–2423.
[28] F. De La Torre, M.J. Black, A framework for robust subspace learning, Int. J. Comput. Vis. 54 (2003) 117–142.
[29] J. Wright, Y. Ma, Dense error correction via l-minimization, Tech. Rep., Tech. Rep. UILU-ENG-08-2210, DC 237 http://perception.csl.uiuc.edu, 2008.
[30] C. Hage, M. Kleinsteuber, Robust pca and subspace tracking from incomplete observations using ℓ 0-surrogates, Comput. Stat. 29 (3–4) (2014) 467–487.
[31] N. Vaswani, P. Narayanamurthy, Static and dynamic robust pca and matrix completion: a review, Proc. IEEE 106 (8) (2018) 1359–1379.
[32] E.J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, J. ACM 58 (3) (2011) 1–37.
[33] T. Zhang, Y. Yang, Robust pca by manifold optimization, J. Mach. Learn. Res. 19 (1) (2018) 3101–3139.
[34] C. Guyon, T. Bouwmans, E.-H. Zahzah, Foreground detection by robust pca solved via a linearized alternating direction method, in: Image Analysis and Recognition: 9th International Conference, ICIAR 2012, Proceedings, Part I 9, Aveiro, Portugal, June 25–27, 2012, Springer, 2012, pp. 115–122.
[35] Z. Lin, M. Chen, Y. Ma, The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices, arXiv preprint arXiv:1009.5055, 2010.
[36] B.B. Alkan, Robust principal component analysis based on modified minimum covariance determinant in the presence of outliers, Alphanumeric J. 4 (2) (2016) 85–94.
[37] S. Mannor, D. Peleg, R. Rubinstein, The cross entropy method for classification, in: Proceedings of the 22nd International Conference on Machine Learning, 2005, pp. 561–568.
[38] K. Labib, P. Uznanski, D. Wolleb-Graf, Hamming Distance Completeness, 30th Annual Symposium on Combinatorial Pattern Matching (CPM 2019), vol. 128, Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2019, p. 14.
[39] S. Makonin, F. Popowich, L. Bartram, B. Gill, I.V. Bajić, Ampds: a public dataset for load disaggregation and eco-feedback research, in: 2013 IEEE Electrical Power & Energy Conference, IEEE, 2013, pp. 1–6.
[40] M. Hubert, P.J. Rousseeuw, K. Vanden Branden, Robpca: a new approach to robust principal component analysis, Technometrics 47 (1) (2005) 64–79.
[41] S. Verboven, M. Hubert, Matlab library libra, Wiley Interdiscip. Rev.: Comput. Stat. 2 (4) (2010) 509–515.
[42] M. Hubert, M. Debruyne, P.J. Rousseeuw, Minimum covariance determinant and extensions, Wiley Interdiscip. Rev.: Comput. Stat. 10 (3) (2018) e1421.
[43] J. López, A. Esteban, J. Hernández, P. Gómez, R. Zamora, C. Zanzi, F. Faura, A new isosurface extraction method on arbitrary grids, J. Comput. Phys. 444 (2021) 110579.
[44] N. Batra, M. Gulati, A. Singh, M.B. Srivastava, It's different: insights into home energy consumption in India, in: Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, 2013, pp. 1–8.
[45] S. Makonin, B. Ellert, I.V. Bajić, F. Popowich, Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014, Sci. Data 3 (1) (2016) 1–12.
[46] A. Singer, H.-T. Wu, Vector diffusion maps and the connection Laplacian, Commun. Pure Appl. Math. 65 (8) (2012) 1067–1144.