Detection of anomalous consumers based on smart meter data

Joanna Kaleta[a], Konrad Wojdan[a], Konrad Świrski[a], Jan Dubiński[a]

[a] Institute of Heat Engineering, Warsaw University of Technology 21/25 Nowowiejska Street, 00-665 Warsaw, Poland

Abstract

The continuous smart grid development makes the advanced metering infrastructure an essential part of electricity management systems. Smart meters not only provide consumers with more economical and sustainable electricity consumption but also enable the energy supplier to identify suspicious behaviour or meter failure. In this work, a shape-based algorithm that indicates households with abnormal electricity consumption pattern within a given consumer group was proposed. The algorithm was developed under the assumption that the reason for unusual electricity consumption may not only be a meter failure or fraud, but also consumer's individual preferences and lifestyle. In the presented methodology, five unsupervised anomaly detection methods were used: K Nearest Neighbors, Local Outlier Factor, Principal Component Analysis, Isolation Forest and Histogram Based Outlier Score. Two time series similarity measures were applied: basic Euclidean distance and Dynamic Time Warping, which allows finding the best alignment between two time series. The algorithm's performance was tested with multiple parameter configurations on five different consumer groups. Additionally, an analysis of the individual types of anomalies and their detectability by the algorithm was performed.

1. Introduction

In recent years an increasing interest in renewable energy sources and prosumption can be observed among consumers [1], the concept of electromobility is gaining importance, and more and more electric devices are being connected to the grid. Aware consumers understand the concept of energy efficiency and want to have more control over their energy consumption structure and its costs. All these factors require energy suppliers to continuously invest in smart grids to ensure efficient and effective distribution of electric energy and consumer-supplier communication. Smart metering infrastructure is an integral part of a smart grid. It makes it easier for the consumer to manage his own consumption while the constant monitoring of the network ensures more stable supplies and more precise generation.

Additionally, monitoring the grid with smart meters can help detecting various energy losses caused by fraudsters manipulating the meter or stealing electricity using an illegal connection from the low voltage grid. Global monetary losses from nontechnical energy losses are estimated at $ 96 billion annually, according to the 2017 Northeast Group report [2]. Non-technical losses include electricity theft, fraud, billing errors, and other lost income. Low voltage grids are not fully protected against theft due to a large number of distributed endpoints available to the public. In such conditions, smart meters seem to be one of the most effective tools in fighting this problem. However, the observation of energy consumption patterns may also apply to ordinary consumers, who lead an unusual lifestyle. Accurate analysis and interpretation of abnormal consumption can bring such consumers real, tangible benefits. Firstly, the consumers may be offered a more favorable tariff from the supplier tailored to their activity. Secondly, detection of abnormal consumption can indicate periods of ineffective use of electricity.

The purpose of the study was to develop a methodology based on smart meter data, that allows indicating households with abnormal electricity consumption patterns within a certain group of consumers. It is worth emphasizing that we aim to detect consumers with repeated and/or long-lasting abnormal consumption, not a one-time anomaly in consumption history. Anomaly detection in the developed methodology is based on the shape of 24-hour energy consumption curves. This approach results from the assumption that households usually have similar periods of activity during the day and consumption significantly deviating from this shape may indicate smart meter manipulation or failure or atypical lifestyle. However, abnormal behaviours are problematic to detect because energy consumption of private consumers is highly variable (even within one household), and we do not have predefined patterns of

abnormal behaviour. Moreover, groups of electricity consumers may present different typical daily consumption patterns. For example, a party group of students could be compared with a calm group of seniors. For the first group high electricity consumption until long night hours is not surprising, while for the latter such cases are rather rare. This study was conducted under the assumption that grouping consumers with similar characteristics makes the existing abnormal consumers more visible to both the algorithm and the expert eye.

The main contributions of this study can be summarized as follows:

- We proposed an algorithm that detects abnormal electricity consumers based on historical readings from smart meters. Abnormal energy consumption can result from energy theft, meter failure or natural consumer activity deviating from the typical activity pattern in the analyzed group. Applicability and effectiveness of the developed algorithm in five different consumers groups are discussed.
- We compared the effectiveness of five unsupervised anomaly detection methods, upon which the algorithm is based: K Nearest Neighbours, Local Outlier Factor, Principal Component Analysis, Histogram Based Outlier Score and Isolation Forest are compared.
- Additionally, we compared two different similarity measures for consumption curves: Euclidean distance, which is a baseline and Dynamic Time Warping, which takes into account natural shifts in consumer activity hours during the day.

The rest of the paper is organized in the following order. In Section 2 related works are briefly discussed. In Section 3 used dataset is described and analyzed. The methodology is introduced in Section 4. In Section 5 results and findings are presented. Finally, Section 6 concludes the whole paper.

2. Related work

As our methodology is based on 24-hour consumption patterns, we first analyzed literature in which a similar approach to the problem was taken. Clustering of 24-hour consumption patterns is a good example. In [3] Lavin and Klabjan performed K-means clustering of 24-hour energy usage profiles of commercial and industrial customers to identify patterns and trends. Obtained profiles showed accurate grouping of similar customers, at the same time keeping the number of generated clusters at a reasonably low level, for example 12 clusters for 821 customers. Räsänen et al. [4] presented methodology capable of handling large amounts of electricity load data based on self-organizing maps (SOM) and clustering methods (K-means and hierarchical clustering). The presented approach resulted in better estimates of the customers' electricity use comparing to the load profiles predetermined by companies for different consumer categories. Teeraratkul et al. [5] used the K-means algorithm to group the 24-hour demand curves, however they employed the Dynamic Time Warping algorithm to calculate similarity between two curves. Using DTW as a dissimilarity measure for clustering resulted in a 50% reduction in the number of energy clusters, compared to traditional K-means.

The second aspect of our study is abnormal consumption detection. In [6] Hurst et al. compared two density-based clustering techniques: DBSCAN (Density-based spatial clustering of applications with noise) and OPTICS (Ordering Points To Identify the Clustering Structure) and one anomaly detection method LOF (Local Outlier Factor) on gas smart meter readings within groups of consumers with similar characteristics. Sial et al. [7] proposed four heuristics allowing for abnormal electricity consumption detection in a residential campus based on percentage change in consumption, distance from k nearest neighbour days, histogram buckets and principal component analysis. The proposed heuristics were evaluated on groups of smart meters having the same context. Although both works seem promising and the presented methodologies were well justified, the authors did not assess the total quality of obtained results, so it is impossible to tell how many of detected observations were anomalous or how many of all anomalous behaviours were detected.

3. Data

Data used in the study contains energy consumption readings for a sample of 5,567 London households that took part in the UK Power Networks led Low Carbon London project between November 2011 and February 2014. The dataset is freely available on the London Datastore website. The dataset contains unique household identifier, energy consumption in kWh (per half hour), date and time, CACI Acorn index [8], which accounts for consumer classification and tariff type (Standard or Dynamic Time of Use) for each household.

Due to the long duration of the project, we decided to define a representative period from which the readings for selected households were analyzed. Two months of 2013 - February and March - were

chosen for several reasons: 1) it is not a holiday period where longer trips and travels usually occur; 2) the period is long enough so that any trips should not be significant, while longer, suspicious inactivity will be visible; 3) the number of households is approximately constant over time; 4) the number of households in this period for the analyzed groups remains high - this is most important for less numerous groups.

Due to the multitude of consumer groups indicated by ACORN (17 types of private consumers), the next step was to limit them by selecting representatives. After analyzing the characteristics of individual groups, we decided to include the following groups in further analysis:

- Group A - Lavish Lifestyles - the wealthiest people in the UK, living in large, detached houses.
- Group D - City Sophisticates - young successful people (single or couples without children), who have their own apartments in most often large cities.
- Group J - Starting Out - young couples starting their careers, sometimes with small children. Most often they live in small, old flats below average market prices
- Group K - Students Life - students or recent graduates live in dormitories, shared flats or shared homes.
- Group P - Struggling Estates - low-income families (mainly large families or single parents) living in traditional urban settlements, mostly in small apartments or terraced houses.

To avoid the impact of energy prices on consumption pattern from the above-mentioned groups we selected only consumers with standard tariff. Example of typical electricity consumption in households that may be appropriate for each analyzed group is shown in Fig. 1.

The data used in this study was originally not labelled, so we followed an unsupervised anomaly detection approach. However, we manually verified all consumers from representative groups and then based on the expert knowledge we marked candidates with potentially suspicious/unexpected behaviours. Manually assigned labels should not serve as the ground truth but they helped assess the approximate effectiveness of the proposed algorithm for each group. As finding an atypical consumer is not only associated with fraud or malfunctions, we marked candidates such as 'night owls' people consuming electricity occasionally or in an irregular way. Using this approach, the rates of indicated anomalies are relatively high. The results after manual verification are presented in Table 1

Table 1: Ratio of manually selected anomalies

Group	Number of households	Number of anomalies	Ratio of anomalies
A	117	16	13.7%
D	213	21	9.6%
J	78	18	23.07%
K	139	16	11.5%
P	89	18	20.2%

Full list of anomalies with short explanation is to find in the Appendix [9]. There is a noticeable difference in the percentages of anomalies between groups. The most natural anomaly ratio occurs for groups D, K, A. High anomaly ratio in groups J and P can be explained with frequent occurrence of consumption pattern shown on Fig. 2, which we treated as anomalous. From an expert perspective this is rather unusual that a clear peak occurs every day at midnight, lasts 2-3 hours, and for the rest of the day consumption remains relatively low.

4. Methodology

4.1. Anomaly detection methods

In this section, all applied unsupervised anomaly detection methods are introduced. To implement all of them we used Python 3 library PyOD [10].

K Nearest Neighbours

K Nearest Neighbours (KNN) is a simple, global distance-based algorithm [11]. The anomaly coefficient for a given point is calculated as the average distance from the point to its k nearest neighbours. The higher the average distance, the greater the probability that the point is an anomaly.

Local Outlier Factor

The Local Outlier Factor (LOF) was proposed in [12]. This algorithm is based on the concept of local density defined by the nearest neighbours. By comparing the local density of a selected point with the local density of its neighbours, it is possible to identify regions of similar or lower density. Points with much lower density than their neighbours can be seen as outliers. This approach allows treating points as anomalies in one area of the dataset while the same points would not be outliers in another area. This is an advantage over the KNN algorithm.

Principal Component Analysis

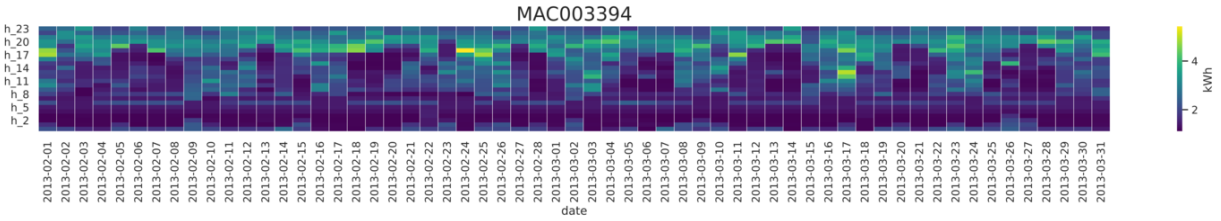

Figure 1: Typical consumption pattern on the example of household MAC003394

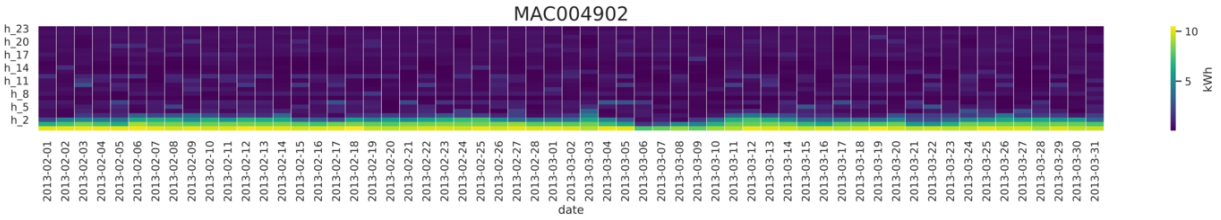

Figure 2: Popular anomalous consumption pattern in group J and P on the example of household MAC004902

The anomaly detection algorithm based on Primary Components Analysis (PCA) was proposed in [13] as an effective unsupervised method of anomaly detection. Principal components are constructed to maximize the explained variance - the first component explains the most variance, and each subsequent component - less and less. This can be thought of as if the first component is determined in the direction where the data is most "stretched". The described property of principal components can be successfully used in anomaly detection. Outliers can be located using point distances to the centroid of the dataset measured along each component. The greater the distance, the more likely the point is to be an anomaly. It is worth emphasizing that outliers should be most visible when the distance is measured along components explaining smaller variances.

Isolation Forest

The Isolation Forest (IF) algorithm was presentedby Liu et al. in [14]. Isolation Forest consists of N decision trees, each of which is constructed in the following way. First, the feature, according to which the points will be split, is randomly selected. Then the points are split into two sub-groups by a randomly selected split value between the maximum and minimum value of the selected feature. The procedure of splitting is continued recursively until the tree is fully constructed, i.e. the point is completely isolated from the other points. The main assumption of the algorithm is that anomalies should be easy to isolate because they require a lower number of splits than normal points that are closer to other points. For each data point, the normalized depth in the tree is computed. The anomaly coefficient for a point is defined as the mean of N depths of this point in the tree.

Histogram Based Outlier Score

Histogram Based Outlier Score (HBOS) has been proposed in [15]. The algorithm assumes the independence of the features and calculates the degree of deviation of the values by building histograms. This is a fast algorithm especially useful for large data sets. A one-dimensional histogram using k intervals of equal width is constructed for each feature. The frequency of the values falling into each bin is reflected by its height. Then created histograms are normalized so that the maximum height of the bins is 1, which guarantees the same weight to be assigned to each variable. HBOS for the selected point is proportional to the sum of heights of the bins in which this point was found.

4.2. Distance Measures

We compared two similarity measures: Euclidean distance and Dynamic Time Warping.

Euclidean distance

The first metric that was used to measure the distance between two 24-hour electricity consumption curves was the Euclidean distance. We used it as a baseline - despite being very sensitive to distortions in the timeline, it is one of the most used metrics. It's advantages are ease of implementation and low computational complexity.

Dynamic Time Warping

Dynamic Time Warping algorithm (DTW) [16] finds the optimal alignment of two sequences by iterative warping of the time axis, which makes it insensitive to time shifts, scaling or different lengths of the compared series. The process of finding the optimal alignment can be represented in a 2D grid, where each grid cell contains the distance measured between the corresponding points of both sequences. The warping path minimizes the overall distance between the two curves. The procedure for calculating the total distance involves finding all possible routes in the grid and calculating the total distance for each of them. The warping paths are subjected to the following restrictions:

- Monotonicity condition: path must not reverse - both index values must remain the same or increase.
- Continuity condition: path moves step by step - both indexes can increase by a maximum of 1 for each step on the track.
- Boundary condition: the path starts in the lower left corner and ends in the upper right corner. Finding the optimal match can be time-consuming as it requires comparing each possible pair of elements from the first and second sequences.

Additionally, not limiting the maximum shift can lead to a bad alignment where a relatively small part of one time series is mapped to a large part of the other. To avoid this, global constraints narrowing the search area around the diagonal of the grid can be used [17]. Fig. 3 illustrates the comparison of Euclidean distance and DTW, warping path on the 2D grid and commonly used global Sakoe-Chiba constraint.

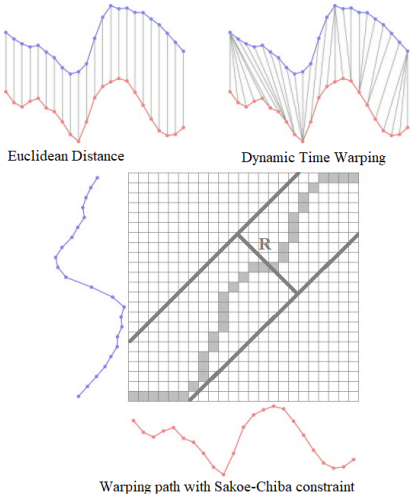

Figure 3: Dynamic Time Warping algorithm [17]

4.3. Global algorithm

Before applying the algorithm, we performed data preprocessing. First, we aggregated smart meter readings to 1-hour sum. Then for each household we scaled the readings for each day separately so that only shapes of daily consumption curves were taken into account, not the absolute values. Scaling was performed by dividing the consumption values by the peak value for each day. In the analyzed load curves no missing values were observed.

The algorithm is based on analyzing 24-hour consumption curves day by day. For each day electric loads from consumers from one group are collected and scored. Consumers, whose electric loads have the highest anomaly score, get their number of anomalous days increased by 1. Having all days analyzed, mean anomaly score for each household is calculated. If a given household has sufficiently high mean anomaly score, it becomes a candidate for later manual verification. The pseudocode is presented in Algorithm 1.

The main algorithm parameters are:

- anomaly detection method – one of the unsupervised methods mentioned in the abstract: KNN, LOF, PCA, IF, HBOS
- similarity measure - a choice between Euclidean and DTW metric, possible to use in algorithms based on point distances: KNN and LOF. For the DTW similarity measure, the global Sakoe-Chiba limitation was used and the window width was set to 2. We assume that among average, typical consumers, the shift of the peak activity should not be more than 2 hours.
- anomaly percentage r - a value describing what percentage of consumption curves in each day will qualify as anomalies.
- candidate selection threshold T – minimal mean anomaly score qualifying a household as anomalous. The threshold for indicating a candidate may range from 0 (the farm was not indicated on any day) to 1 (the household was indicated on every day of the representation period).

In this study, we focused on the influence of the four main parameters controlling the algorithm. Remaining parameters, corresponding to specific methods were set to values presented in Table 2 according to standards from literature.

Algorithm 1:
Pseudocode of global algorithm

Input : Smart meter readings for households from given group **(G)** in given measurement period **(P)**
Output: Label for each household in group **G**: typical consumer or anomalous consumer
Initilalization: Set candidate selection threshold *T*, anomalies percentage *r*, anomaly detection method and (if applicable for the chosen method) distance measure. Initialize number of anomalous days equal to zero for each household.
Procedure:
For each day in P:
1. Get 24 h electricity consumption curves from smart meter readings for the given day from all households
2. Compute anomaly scores for collected load curves using chosen anomaly detection method
3. Label *r* percent of curves with the highest anomaly scores as anomalous.
4. For those households, which consumption curves were labelled as anomalous, increase the number of its anomalous days by 1

End for
For each household in G:
1. Calculate mean anomaly score equal to number of its anomalous days divided by whole measurement period length
2. If mean anomaly score $>T$, mark the household as an anomalous consumer. Otherwise label the household as typical consumer.

End for
End Procedure

5. Results and findings

For all five described anomaly detection methods two values of percentage of anomalies in the dataset (0.1 and 0.25) and two values of candidate selection threshold (0.3 and 0.5) were tested. Additionally, for LOF and KNN two similarity measures (Euclidean and DTW) were applied, so we evaluated 28 configurations of the algorithm in total.

Having manually assigned labels we assessed Precision, Recall and F1-Score calculated in respect to the anomaly class for each configuration. Choosing the best configuration depends on whether the main goal is to detect as many anomalies as possible (then maximize Recall) or detect anomalies as precisely as possible (then maximize Precision). In this study, we decided to choose configurations with the highest F1-Score because we assumed that it indicates a configuration with sufficiently high both Precision and Recall. The best configurations are presented in Table 3, results for all configurations are to find in the Appendix [9].

Table 2: Values of anomaly detection methods parameters

Method	Parameter	Value
KNN	Number of neighbors	5
LOF	Number of neighbors	20
PCA	Considered components	all
IF	Number of trees	100
HBOS	Number of bins	$\sqrt{N}$, where N - number of samples

The results obtained for the best configurations are satisfactory – for each group we achieved Precision and Recall over 0.6. Despite the fact, that group J has the highest percentage of manually marked anomalous household, we obtained the highest F1-Score (0.83). The second best F1-Score was observed for group A (0.77), where all indicated households were real anomalies (Precision = 1.0).

5.1. Results summary

The results obtained for all configurations were very diverse, ranging from 0.176 to 1 for Precision, from 0.111 to 1 for Recall and from 0.182 to 0.813 for F1-Score for the anomaly class. It is impossible to choose a universal configuration that fits all consumer groups, so each case should be considered individually. The PCA and LOF with DTW were chosen several times as the best anomaly detection algorithms in terms of F1-Score, as well as in terms of Precision and Recall. The other algorithms never achieved the highest F1-Score. KNN and HBOS quite often scored the lowest on all three measures. Nevertheless, KNN was indicated several times as the best algorithm in terms of Precision or Recall, while HBOS only once - for group A in terms of Recall. IF never scored the lowest on all three measures, while several times scored the highest on Precision. It is worth mentioning that F1-Score is not always the best compromise between Precision and Recall - depending on the application higher Recall may be much more valuable, even at the cost of smaller Precision. In real conditions, excess candidates can be rejected manually at a low cost.

We also examined how applying DTW influenced the anomaly detection quality in comparison to Euclidean distance. To do so, we compared Precision, Recall and F1-Score for pairs of configu-

Table 3: Results for the best configurations

Consumer group	Method	Candidate selection threshold	Anomaly percentage	Precision for anomaly class	Recall for anomaly class	F1-Score for anomaly class
A	PCA	0.3	0.1	1.0	0.625	0.769
D	PCA	0.3	0.1	0.619	0.619	0.619
J	LOF DTW	0.5	0.25	0.929	0.722	0.813
K	LOF DTW	0.3	0.1	0.714	0.625	0.667
P	PCA	0.5	0.25	0.688	0.611	0.647

Table 4: Example pair of configurations to compare the impact of DTW on detection quality

Method	Candidate selection threshold	Anomaly percentage	Precision for anomaly class	Recall for anomaly class	F1-Score for anomaly class
LOF	0,5	0,1	0,75	0,188	0,3
LOF DTW	0,5	0,1	1	0,25	0,4

Table 5: Improvement of anomaly detection quality after applying DTW

Group	Precision	Recall	F1-Score
A	7/8	6/8	6/8
D	6/8	6/8	6/8
J	7/8	6/8	6/8
K	5/8	3/8	4/8
P	2/8	2/8	3/8

Table 6: Improvement of Recall and F1-Score after increasing anomaly percentage

Group	Recall	F1-Score
A	14/14	8/14
D	14/14	4/14
J	14/14	9/14
K	14/14	8/14
P	14/14	7/14

Table 7: Improvement of Precision and F1 Score after increasing candidate selection threshold

Group	Precision	F1-Score
A	13/14	7/14
D	14/14	8/14
J	9/14	4/14
K	13/14	5/14
P	14/14	7/14

rations with different distance measures and other parameters unchanged. Example of such a pair is presented in Table 4.

In the example case, the values of all three measures increased after applying DTW. In total, we compared 8 pairs of configurations for each group and in Table 5 we presented for how many cases the improvement was observed. However, the application of DTW resulted in longer execution time and the improvement was not always significant.

Similarly, we investigated how increasing the anomaly percentage from 0.1 to 0.25 (leaving the other parameters unchanged) improves Recall and F1-Score. In total, we compared 14 pairs of configurations for each group and in Table 6 we presented for how many cases the improvement was observed. Although improvement in Recall was obtained in 100% of cases in each group, the Precision deteriorated so much, that as a result, the F1-Score improved only in some cases.

Finally, we examined how increasing the candidate indication threshold (leaving the other parameters unchanged) improved Precision and F1-Score. We compared 14 pairs of configurations which results are presented in Table 7.

5.2. Examples of anomalous households

In this section, we investigated three examples of anomalous households. It is easy to observe, how differently the households were evaluated depending on chosen configuration - mean anomaly scores for each configuration are presented in Table 8. The mean anomaly scores for all anomalous households are listed in the Appendix [9]. Then, for selected households we chose three examples of how different factors cause a significant increase in mean anomaly score:

- For household MAC003422 from group A (Fig. 4) using Dynamic Time Warping increases mean anomaly score from 0.43 to 0.73,
- For household MAC005340 from group P (Fig. 5) increasing anomalies percentage in the dataset increases the mean anomaly score from 0.07 to 0.97,
- For household MAC003025 from group J (Fig. 6) using HBOS instead of PCA increases mean anomaly score from 0.27 to 0.86

Table 8: Mean anomaly scores for example households

Configuration	MAC 003422	MAC 005340	MAC 003025
LOF 0.1	1.000	0.000	0.763
LOF DTW 0.1	1.000	0.000	0.780
LOF 0.25	1.000	0.707	0.847
LOF DTW 0.25	1.000	0.776	0.847
KNN 0.1	0.458	0.000	0.729
KNN DTW 0.1	0.729	0.000	0.763
KNN 0.25	0.661	0.000	0.763
KNN DTW 0.25	0.881	0.000	0.814
PCA 0.1	1.000	0.069	0.763
PCA 0.25	1.000	0.966	0.864
HBOS 0.1	0.881	0.000	0.169
HBOS 0.25	1.000	0.000	0.271
IF 0.1	1.000	0.000	0.763
IF 0.25	1.000	0.000	0.864

6. Conclusions

In this study, an unsupervised algorithm for detecting anomalous electricity consumers based on historical readings from smart meters was proposed. The algorithm was designed to detect consumers whose behaviour is atypical and stands out from a certain group of consumers for a long time. The proposed algorithm was tested for 28 configurations on 5 consumer groups. To measure the effectiveness of the proposed algorithm, abnormal households were annually marked. We proved that algorithm based on simple anomaly detection methods can successfully indicate atypical energy consumers. Therefore, our algorithm can be applied in real electricity management systems. Our study leads to an interesting observation, how differently the same energy consumption pattern can be scored depending on chosen anomaly detection method, similarity measure, other parameter values and also characteristics of the consumers group. For that reason, detecting anomalies in smart meter data is a non-trivial task and therefore any attempt to automate this process should be confronted with the expert knowledge.

References

[1] Institute for Renewable Energy EC BREC, Photovoltaic market in Poland 2021, 2021.

[2] Northeast Group, LLC, Electricity Theft and Non-Technical Losses: Global Markets, Solutions, and Vendors, 2017.

[3] A. Lavin, D. Klabjan, Clustering time-series energy data from smart meters, Energy Efficiency 8 (4) (2014) 681–689, doi:10.1007/s12053-014-9316-0.

[4] T. Räsänen, D. Voukantsis, H. Niska, K. Karatzas, M. Kolehmainen, Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data, Applied Energy 87 (11) (2010) 3538–3545, doi:10.1016/j.apenergy.2010.05.015.

[5] T. Teeraratkul, D. O'Neill, S. Lall, Shape-Based Approach to Household Electric Load Curve Clustering and Prediction, IEEE Transactions on Smart Grid 9 (5) (2018) 5196–5206, doi:10.1109/tsg.2017.2683461.

[6] W. Hurst, C. A. C. Montañez, N. Shone, Time-Pattern Profiling from Smart Meter Data to Detect Outliers in Energy Consumption, IoT 1 (1) (2020) 92–108, doi:10.3390/iot1010006.

[7] A. Sial, A. Singh, A. Mahanti, Detecting anomalous energy consumption using contextual analysis of smart meter data, Wireless Networks 27 (6) (2019) 4275–4292, doi:10.1007/s11276-019-02074-8.

[8] Acorn - The smarter consumer classification | CACI, URL https://acorn.caci.co.uk/, accessed on Tue, September 14, 2021, 2013.

[9] Appendix on reproducibility, URL https://github.com/joaxkal/smart-meters-anomalous-consumers, 2021.

[10] Y. Zhao, Z. Nasrullah, Z. Li, PyOD: A Python Toolbox for Scalable Outlier Detection, Journal of Machine Learning Research 20 (96) (2019) 1–7, URL http://jmlr.org/papers/v20/19-011.html.

[11] P. Yang, B. Huang, KNN Based Outlier Detection Algorithm in Large Dataset, in: 2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing, IEEE, doi:10.1109/ettandgrs.2008.306, 2008.

[12] M. M. Breunig, H.-P. Kriegel, R. T. Ng, J. Sander, LOF: Identifying Density-Based Local Outliers, ACM SIGMOD Record 29 (2) (2000) 93–104, doi:10.1145/335191.335388.

[13] M.-L. Shyu, S.-C. Chen, K. Sarinnapakorn, L. Chang, A Novel Anomaly Detection Scheme Based on Principal Component Classifier, 2003.

[14] F. T. Liu, K. M. Ting, Z.-H. Zhou, Isolation-Based Anomaly Detection, ACM Transactions on Knowledge Discovery from Data 6 (1) (2012) 1–39, doi:10.1145/2133360.2133363.

[15] M. Goldstein, A. Dengel, Histogram-based Outlier Score (HBOS): A fast Unsupervised Anomaly Detection Algorithm, KI-2012: Poster and Demo Track .

[16] H. Sakoe, S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, IEEE Transactions on Acoustics Speech, and Signal Processing 26 (1) (1978) 43–49, doi:10.1109/tassp.1978.1163055.

[17] C. A. Ratanamahatana, E. Keogh, Making Time-series Classification More Accurate Using Learned Constraints, in: Proceedings of the 2004 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics, doi:10.1137/1.9781611972740.2, 2004.

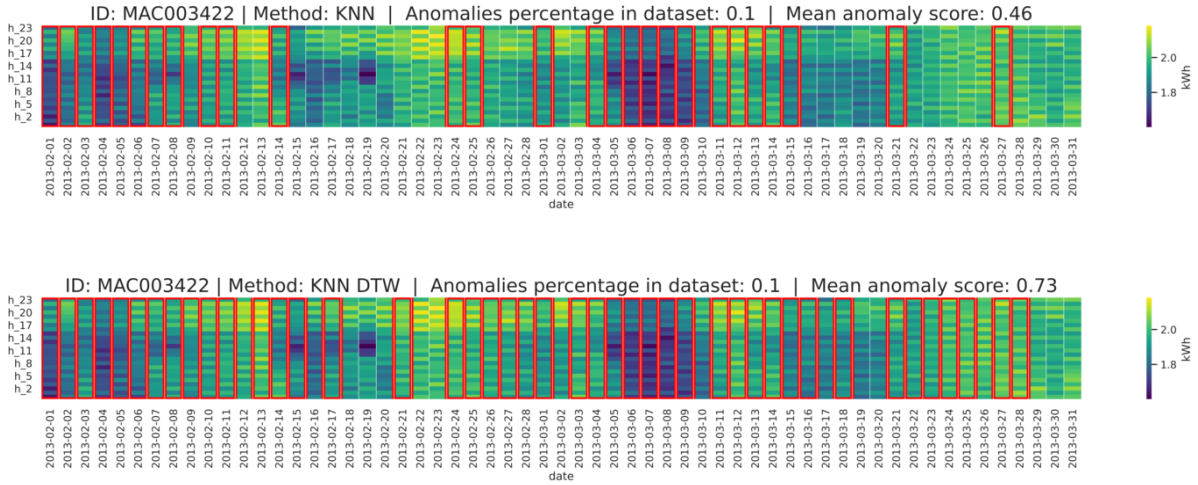

Figure 4: Mean anomaly scores 0.46 and 0.73 for MAC003422

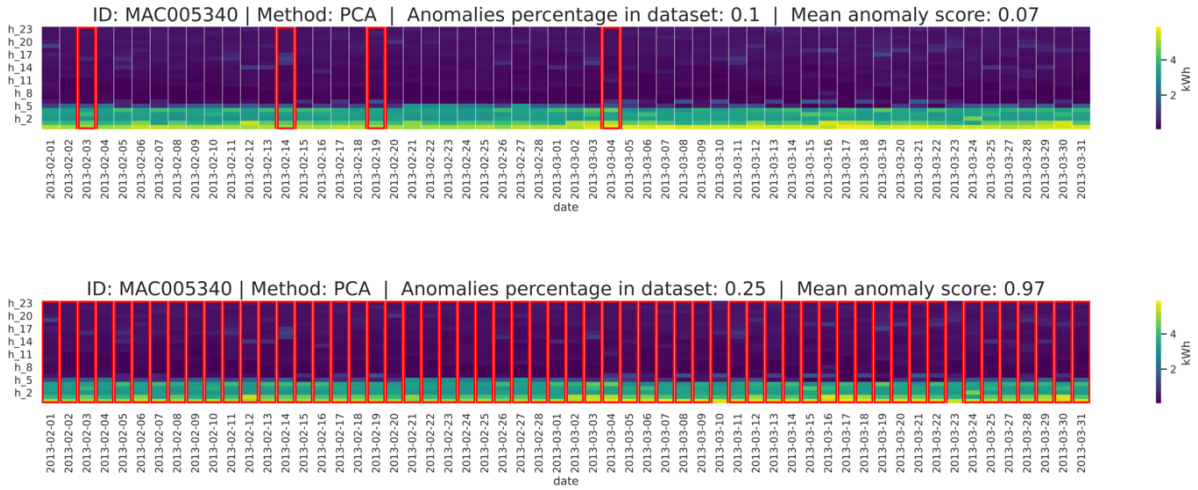

Figure 5: Mean anomaly scores 0.07 and 0.97 for MAC005340

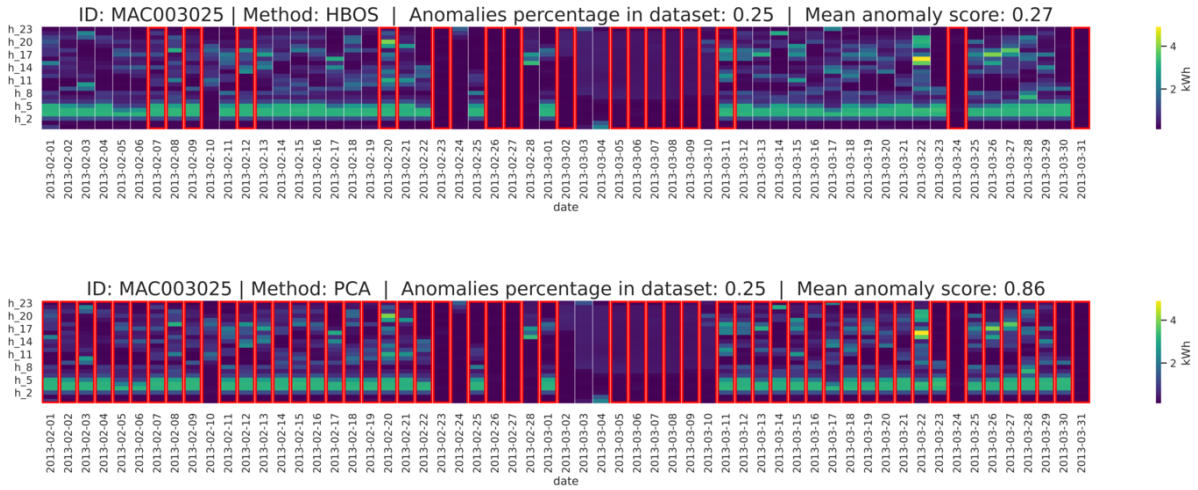

Figure 6: Mean anomaly scores 0.27 and 0.86 for MAC003025