Scientific Research Publishing

Open Journal of Energy Efficiency, 2021, 10, 37-49
https://www.scirp.org/journal/ojee
ISSN Online: 2169-2645
ISSN Print: 2169-2637

Comparison and Adaptation of Two Strategies for Anomaly Detection in Load Profiles Based on Methods from the Fields of Machine Learning and Statistics

Patrick Krawiec[1], Mark Junge[1], Jens Hesselbach[2]

[1]Limón GmbH, Kassel, Germany
[2]Department for Sustainable Products and Processes (Upp), University Kassel, Kassel, Germany
Email: krawiec@limon-gmbh.de

How to cite this paper: Krawiec, P., Junge, M. and Hesselbach, J. (2021) Comparison and Adaptation of Two Strategies for Anomaly Detection in Load Profiles Based on Methods from the Fields of Machine Learning and Statistics. *Open Journal of Energy Efficiency*, **10**, 37-49.
https://doi.org/10.4236/ojee.2021.102003

Received: January 29, 2021
Accepted: April 27, 2021
Published: April 30, 2021

Abstract

The Federal Office for Economic Affairs and Export Control (BAFA) of Germany promotes digital concepts for increasing energy efficiency as part of the "Pilotprogramm Einsparzähler". Within this program, Limón GmbH is developing software solutions in cooperation with the University of Kassel to identify efficiency potentials in load profiles by means of automated anomaly detection. Therefore, in this study two strategies for anomaly detection in load profiles are evaluated. To estimate the monthly load profile, strategy 1 uses the artificial neural network LSTM (Long Short-Term Memory), with a data period of one month (1 M) or three months (3 M), and strategy 2 uses the smoothing method PEWMA (Probalistic Exponential Weighted Moving Average). By comparing with original load profile data, residuals or summed residuals of the sequence lengths of two, four, six and eight hours are identified as an anomaly by exceeding a predefined threshold. The thresholds are defined by the Z-Score test, *i.e.*, residuals greater than 2, 2.5 or 3 standard deviations are considered anomalous. Furthermore, the ESD (Extreme Studentized Deviate) test is used to set thresholds by means of three significance level values of 0.05, 0.10 and 0.15, with a maximum of $k = 40$ iterations. Five load profiles are examined, which were obtained by the cluster method *k*-Means as a representative sample from all available data sets of the Limón GmbH. The evaluation shows that for strategy 1 a maximum F_1-value of 0.4 (1 M) and for all examined companies an average F_1-value of maximum 0.24 and standard deviation of 0.09 (1 M) could be achieved for the investigation on single residuals. In variant 3 M the highest F_1-value could be achieved with an average F_1-value of 0.21 and standard deviation of 0.06 (3 M) for summed

residuals of the partial sequence length of four hours. The PEWMA-based strategy 2 did not show a higher anomaly detection efficacy compared to strategy 1 in any of the investigated companies.

Keywords

Energy Efficiency, Anomaly Detection, Load Profiles, LSTM, PEWMA

1. Introduction

Establishing an environmentally compatible and sustainable energy supply is one of the central challenges of our time. For the implementation of the energy transition by 2030 [1] or 2050 [2], there are two main climate policy goals at the European level: increasing the share of renewable energy and increasing energy efficiency [3]. To realize these goals, comprehensive changes are needed in various sectors, such as energy supply, transport, and industry. Driven by political requirements and financial support programs, the efficient use of energy and resources is becoming more important for companies.

Limón GmbH is therefore developing algorithms for the automated analysis of energy data and load profile data in cooperation with the University of Kassel as part of the BAFA [4] savings meter funding program "Pilotprogramm Einsparzähler". An essential task is automated anomaly detection for the identification of saving potentials.

Anomaly detection has a wide range of applications, such as economics, network traffic, industrial process control, and statistics. Therefore, a number of books and survey papers exist that deal with anomaly detection [5] [6] [7] [8] [9].

Anomalies can be defined as follows: "Anomalies are patterns in data that do not conform to a well-defined notion of normal behavior" [5]. In the context of this work, normal behavior is modeled using methods from the fields of machine learning and statistics. The comparison of these methods for the estimation of time series and the resulting question of which method yields better forecasting accuracy has not yet been clarified and is the subject of current research [10]. The paper addresses this research question by comparing and adapting two anomaly detection strategies. The main difference between the two strategies is the way a typical monthly load profile is estimated and compared to the real load profile. Strategy 1 estimates the load profiles using the LSTM (Long Short-Term Memory) algorithm from the field of machine learning [11]. It has already been successfully applied in the estimation of time series [12] [13]. Strategy 2, on the other hand, uses the PEWMA (Probalistic Exponential Weighted Moving Average) estimation method from statistical time series analysis [14]. In comparison to the LSTM algorithm, it has much lower algorithm complexity as well as computational requirements. This method is evaluated based on these properties and promising results in the context of anomaly detection in time series [15]

[16]. The two methods will be applied to selected load profile data of companies from Germany provided by Limón GmbH. The work is further characterized by the fact that for the investigated companies only information from time stamps and the associated power values are known. In both strategies, the residuals obtained are examined for anomalies and compared to pre-defined reference anomalies. This makes it possible to quantify the anomaly detection efficacy.

2. Theoretical Background

2.1. Point Anomalies & Partial Sequence Anomalies

Unusual data points or sequences in load profiles can be categorized according to different types. Point anomalies and partial sequence anomalies are relevant to this study, but additional types or further possibilities of systematizing anomalies exist [7] [9]. A point anomaly is a single time point that behaves unusually compared to other values in the time series. Point anomalies can be univariate or multivariate, depending on whether they affect one or more time-dependent variables. Partial sequence anomalies refer to the behavior of a sequence of consecutive time points that are considered unusual. Individual data points in this sequence do not necessarily represent a point anomaly. Anomaly subsequences can also occur in a time-dependent manner in one dimension (univariate subsequence) or in multiple dimensions (multivariate subsequence). Across the board, the classification of data as an anomaly is always context-dependent, *i.e.*, either the entire time series is seen as the context or the method only examines certain time windows, so that outliers are only valid locally or in close proximity [7]. However, anomalies can also occur in other contexts, such as when considering the temporal influences of seasons or weekday-weekend rhythms [5].

2.2. Residual-Based Anomaly Detection

In this work, a residual-based anomaly detection method is used for anomaly detection. In a first step, the normal behavior of a load profile is estimated using a mathematical model. In a second step, deviations (residuals) between estimated and observed values are formed and a decision is made whether an observation is anomalous by means of residual analysis.

2.2.1. Modeling Normal Behavior

Different regression models can be created from the known load profile data. Through these models, expected values can be formed and these represent a kind of normal behavior of a load profile, which is anomaly-free in the best case. In order to determine an expected value, a distinction is made between estimation and prediction models. The conceptual distinction between estimation and prediction is based on Blázquez-García *et al.* (2018) [7]. Estimation models use past data, the current time, and data temporally subsequent to the expected value to estimate an expected value. In contrast, prediction models use only past data to determine a temporally subsequent expected value.

The estimation model by LSTM algorithm uses two variants. The first variant uses a period of one month, abbreviated as 1 M in the following, to build the model. This corresponds to the period in which anomalies are to be identified. In the second variant, the model is created with data from three months (3 M), *i.e.*, two previous months and the month to be examined for anomalies are used. The additional analysis using a 3 M estimation model avoids the danger compared to the 1 M estimation model of estimating possible anomalies in the estimation model and thus hiding them.

The estimate based on the PEWMA method uses three smoothing parameters α_p = (0.3, 0.6 and 0.9) and are tested with regard to the possibility of successfully identifying anomalies. The larger α_p, the more the resulting curve is smoothed. Following Renshaw [17], the PEWMA procedure is used in this work with β = 0.5.

2.2.2. Residual Analysis

The estimated load profiles are compared with the observed load profiles of a month and the deviations are analyzed (residuals analysis). This analysis is carried out either on the level of individual residuals or based on the summation of successive residuals as partial sequence anomalies. The partial sequences of length two, four, six, eight hours are examined, each being sliding sequence windows. Residuals and residual sequences whose expression exceeded a defined threshold were scored as anomalies. Thresholds are systematically varied by using a generalized Extreme Studentized Deviate (ESD) [18] test and a Z-Score test procedure. In the ESD test, the three significance level values α = (0.05, 0.10, and 0.15) with a maximum of k = 40 iterations, are used as thresholds to identify anomalies. The Z-Score test uses z-standardized residuals. For identification as an anomaly, the 2, 2.5, and 3 multiples of the standard deviation of the data points are investigated as bounds. The identified anomalies can be evaluated using three metrics *Precision* (P), *Recall* (R) and F_1 measure [6].

The *Precision* is defined as the number of true positives (tp) divided by the sum of tp and the number of false positives (fp):

$$P = \frac{tp}{tp + fp} \tag{1}$$

The *Recall* R describes the ratio of tp to the sum of tp and false negatives (fn):

$$R = \frac{tp}{tp + fn} \tag{2}$$

The F_1 measure considers both P and R and forms the harmonic mean of both previous metrics:

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{3}$$

All metrics output numerical values between 0 and 1. In order to apply them, it is necessary to know the number of data points that have been classified as anomalous or normal by the anomaly detection method. Furthermore, the pop-

ulation of anomalies present in the data set must be known.

3. Description of the Load Profiles for Anomaly Detection

Limón GmbH has been able to evaluate data in the form of load profiles from several hundred companies over the last 10 years. This total material was grouped by using the *k*-Means clustering method to five prototypical load profiles. In order to predefine anomalies, the companies to be investigated were examined for anomalous profiles in expert interviews prior to the evaluation. The anomalies defined by the experts are used as references to determine the efficacy of the anomaly detection strategies.

3.1. Selected Load Profiles

Figure 1 represents three of the five load profiles used in the further work for the comparison and adaptation of the two strategies. In each case the total sum of predefined point and partial sequence anomalies is given.

The predefined anomalies of companies 1 to 4 can be described as temporally contextual anomalies, since the load profiles follow a temporal rhythm, *i.e.*, day-night and weekday-weekend dependencies exist, and unusual profiles exist in the context of the respective point in time. Company 5 (see Figure 3) is not subject to any of these dependencies. The anomalies predefined here can only be seen in the context of the overall progression, so unusual maximum or minimum power values are defined as reference anomalies.

3.2. Counting Method of the Anomalies

It is found that the majority of the companies have more predefined anomalous

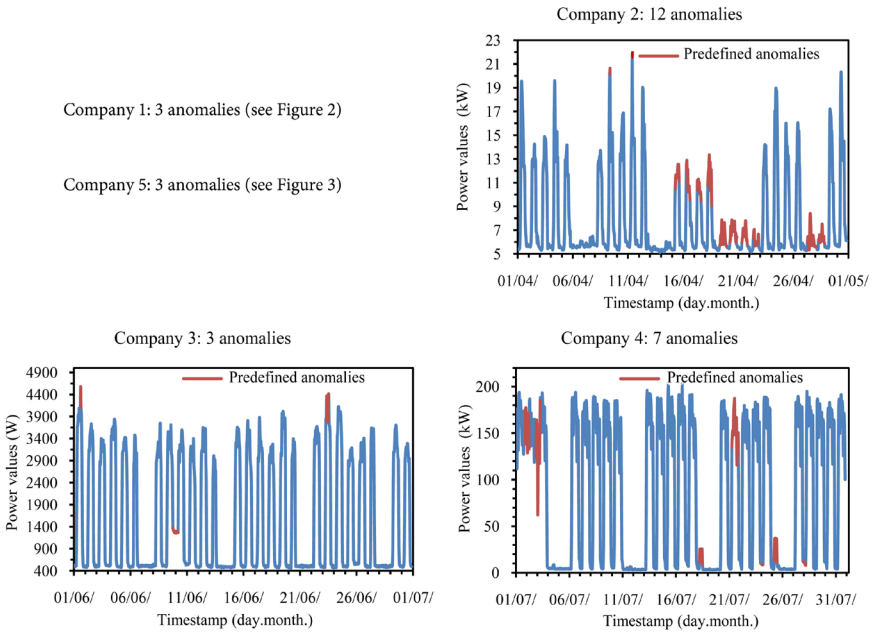

Figure 1. Display of the selected load profiles with predefined anomalies.

sequences than point anomalies. Point anomalies are identified by evaluating the residual values of individual time points. However, it is possible that single residual studies also identify anomalous subsequences. This is because the estimation in the range of anomalous sequences also differs from the given load profile data in part by only single high residual values. Therefore, the investigation of partial sequence anomalies using summed residual values can also identify single point anomalies. To address this difficulty, the following counting method for anomalies is established: 1) Anomalous sequence regions are counted as detected only once if identified more than once. 2) If point anomalies and anomalous sequences occur together in a load profile, then each anomalous sequence counts as one anomaly each.

4. Evaluation of the Two Strategy Approaches

Estimation using the two different strategies is followed by examination of the single and summed residuals to identify anomalies using threshold testing by Z-Score and ESD. In the following, two exemplary individual results and subsequent overarching results are discussed.

4.1. Exemplary Single Results

4.1.1. Single Residual Analysis-Strategy 1 (1 M)-Company 1

For the analysis of conspicuous single residuals for company 1, the LSTM model (1 M) is estimated and deviations between observed and estimated values are evaluated using the Z-Score and ESD test (see Figure 2). In general, many false positive assignments and a maximum of two out of three anomalies can be detected, *i.e.*, a maximum *Recall* of $R = 0.67$. The evaluation results not explicitly shown here reveal the reason for high number of *fp* results. At company 1, the estimation models 1 M & 3 M, are not able to model every peak in the load profile, so that high residual values also occur outside the predefined anomalous ranges. The analysis (see Table 1) with the ESD test at $\alpha = 0.05$ shows the highest *Precision* with $P = 0.17$, a *Recall* of $R = 0.33$, and the highest F_1 measure with

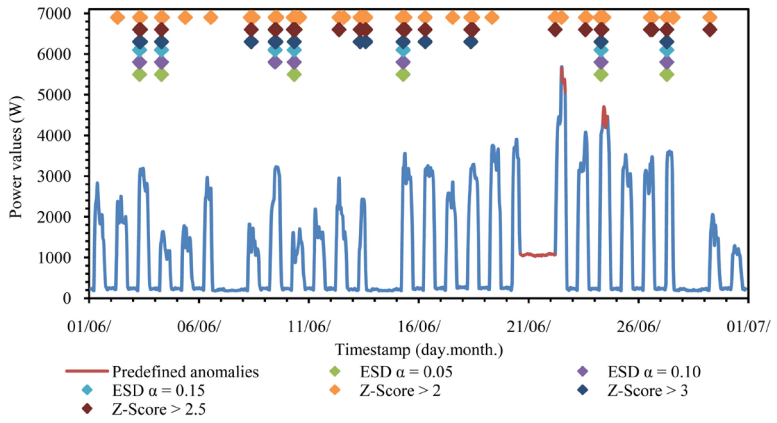

Figure 2. Plot of the identified anomalies of company 1 resulting from the three Z-Score values and the three significance level values of the ESD test ($k = 40$).

Table 1. Calculated *precisions*, *recalls*, and F_1 measures of company 1 (1 M) for three Z-Score thresholds and the three significance level values of the ESD test ($k = 40$).

Company 1 (1 M)	*Precision*	*Recall*	F_1 *measure*	*tp*	*fp*	*fn*
ESD $\alpha = 0.05$	0.17	0.33	0.22	1	5	2
ESD $\alpha = 0.10$	0.14	0.33	0.20	1	6	2
ESD $\alpha = 0.15$	0.14	0.33	0.20	1	6	2
Z-Score > 3	0.08	0.33	0.13	1	11	2
Z-Score > 2.5	0.10	0.67	0.17	2	19	1
Z-Score > 2	0.06	0.67	0.10	2	34	1

$F_1 = 0.22$. For less stringent significance levels ($\alpha = 0.10$ and 0.15), the values deteriorate. For the analysis by Z-Score thresholds, although two of the three reference anomalies can be identified as true positives under the two thresholds of Z = 2.0 and Z = 2.5, at the same time the number of false positive assignments increases considerably to 19 and 34, respectively.

4.1.2. Single Residual Analysis-Strategy 2-Company 5

The analysis of the data in Strategy 2 showed that the application of the PEWMA-based approach only seems to make more sense for company 5. The PEWMA method cannot utilize temporal contextual rhythms, such as weekday-weekend dependencies, and these are significantly responsible for the load profiles in companies 1 through 4.

For company 5, the highest anomaly detection quality results in the case of $\alpha_P = 0.9$, *i.e.*, with a very strongly smoothed curve. For two of the three reference anomalies with unusual maximum values, very large residuals arise. Their correct positive identification is thus facilitated (see **Figure 3**). **Table 2** provides a breakdown of the three metrics for Z-Score testing. The F_1 measures increase with increasing smoothing and tend to decrease with lower thresholds when the influence of the R values is considered. In general, the evaluation shows that more false positive assignments result with smaller threshold. The ESD was able to achieve a maximum *Recall* of $R = 0.33$, so an explicit presentation is omitted.

4.2. Strategy and Cross-Company Results

Residuals and residual sequences whose expression exceeds a defined threshold are evaluated as anomalies. The thresholds are systematically varied in an ESD test procedure or as a Z-Score test with three test thresholds each. An overarching evaluation of the threshold procedures allows the suitability for the residual-based anomaly detection approach to be assessed. **Figure 4** shows the average *Recall* of the Z-Score and ESD test per threshold of strategy 1 of variant 1 M and 3 M. The average *Recall* value of each ESD test is below the Z threshold of Z > 3. However, the significance levels chosen for the ESD test are set higher than average, since $\alpha = 0.5$ or smaller are usually used [18]. Preliminary analysis of the present data showed that no anomalies could be detected below an α-value

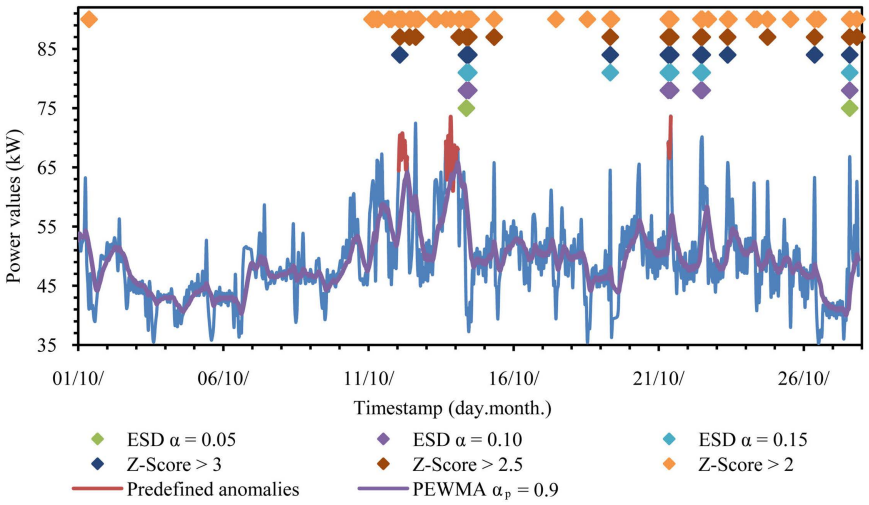

Figure 3. Plot of identified anomalies of company 5 for PEWMA $\alpha_P = 0.9$ resulting from Z-Score test and ESD test ($k = 40$).

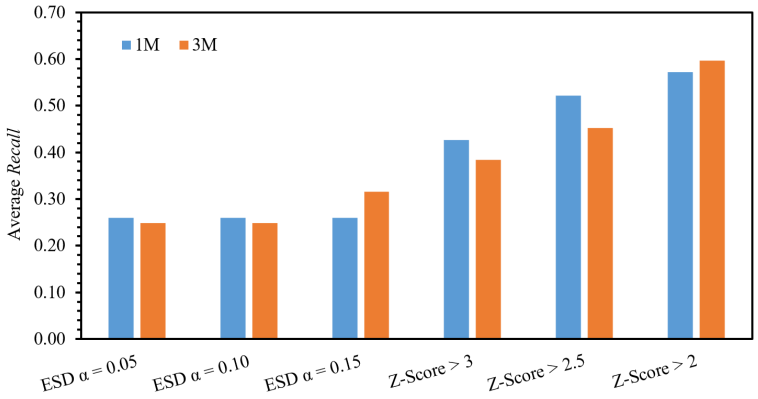

Figure 4. Plot of the average recall of companies 1 to 5 per threshold for the Z-Score and ESD test of strategy 1 (1 M & 3 M).

Table 2. Calculated *precisions*, *recalls*, and F_1 measures of company 5 for three Z-Score.

	$\alpha_P = 0.3$			$\alpha_P = 0.6$			$\alpha_P = 0.9$		
Company 5	P	R	F_1	P	R	F_1	P	R	F_1
Z-Score > 3	0.09	0.33	0.14	0.09	0.33	0.14	0.18	0.67	0.29
Z-Score > 2.5	0.05	0.33	0.09	0.11	0.67	0.19	0.12	0.67	0.20
Z-Score > 2	0.03	0.33	0.05	0.06	0.67	0.10	0.08	1.00	0.14

smaller than 0.05. Furthermore, the ESD test identifies anomalies depending on the amount of incoming data points. The high number of incoming data of $n = 720$, which corresponds to a period of about one month, seems to lead to the ESD test being too insensitive, since the λ_i value remains very stable with continuous iteration compared to R_i values of the test statistic [18]. Further, the evaluation showed that with increasing sequence length of the summed residuals as well as with increasing PEWMA alpha values, the frequency distribution of the

obtained residuals moves further and further away from a normal distribution. However, the normal distribution assumption is among the prerequisites of an ESD test. Accordingly, the Z-Score test, in contrast to the ESD test, can be considered more suitable to analyze the given data and is used for further discussion of the evaluation results.

By taking a cross-strategy view, the question of whether the efficacy of anomaly detection depends on the specific load profile of a company can be answered. Companies 1 to 4 have relatively similar load profile characteristics. Comprehensible temporal rhythms exist. Among other things, the experts identified deviations as reference anomalies that did not follow these rhythms. The LSTM network underlying strategy 1 can, in principle, recognize many of these input variables or temporal rhythms thanks to the information from the time stamp and take them into account in the regression model. Consequently, it could be assumed that the LSTM algorithm should be able to estimate the cyclic changes of the power values. In the evaluation, however, it is difficult to estimate individual load peaks in the maximum power range of the day or the slopes of suddenly rising or falling power values at the beginning or end of a core working time. In these areas, high residuals arise when comparing the expected values estimated by the LSTM network with the observed data. These are detected as anomalies in the residual analysis, but do not correspond to the reference anomalies resulting in low F_1 measures. However, the strategy based on the LSTM network can detect anomalies much more efficiently than strategy 2 for the first four companies. In contrast, the load profile of company 5 does not show temporal dependencies. Using both strategy 1 and strategy 2, similar anomaly detection qualities have resulted for this company. For the LSTM network, estimation based on the information from the timestamp is challenging, since these hardly influence the load profile. Other influencing factors not known for this work seem to play a role and cannot be used as input variables for the LSTM network. For the PEWMA based strategy 2 it can be observed that the efficacy of the anomaly detection is influenced by the context in which the anomalies are located. The higher the proportion of unusual power peaks or drops in the context of the overall course, the higher the anomaly detection efficacy.

Furthermore, by comparing the cross-company results of strategy 1, it is possible to evaluate the effect of the length of the time period of the data used to estimate an LSTM model (1 M vs. 3 M) on the efficacy of anomaly detection. **Figure 5** shows the F_1 measures of the Z-Score tests of company 1 to 5 for the 1 M & 3 M variants. Additionally, the ratio of higher F_1 measures in the 1 M estimation compared to those from the 3 M estimation at the same threshold in each case is illustrated. In most cases, better F_1 measures result for the 3 M estimate.

A closer look at the data pool of the company from company 4 reveals that two of the previous months used for the 3 M estimate have very different load profiles. Consequently, the resulting estimate does not provide a good basis for identifying the reference anomalies. However, it also shows comparatively low F_1 values overall, since there is a high number of false positive assignments

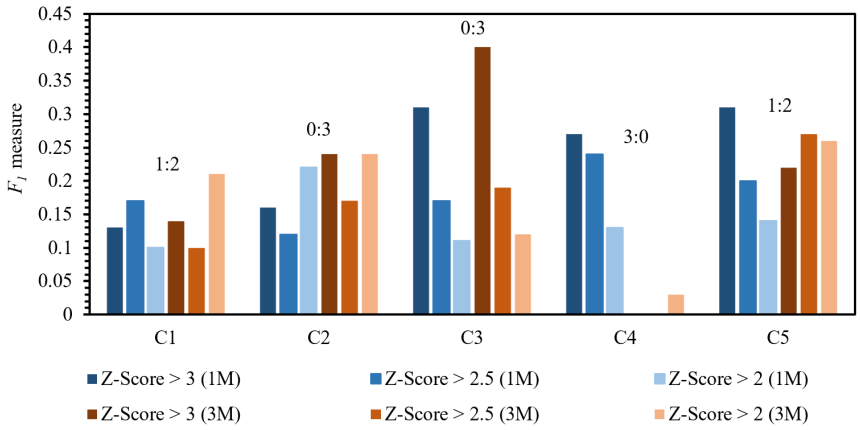

Figure 5. Plot of F_1 measures of Z-Score tests for company 1 to 5 in 1 M & 3 M variants and indication of the ratio of A:B with A: number of higher F_1 measures compared to 3 M and B: number of higher F_1 measures compared to 1 M.

depending on the threshold. Presumably, the use of significantly longer periods for training the LSTM network could serve to improve the anomaly detection efficacy.

By summarizing the results of single residual or residual sequence analysis and comparing them across strategies, the effects on the efficacy of anomaly detection can be evaluated. **Figure 6** presents the average F_1 measures (1 M: bluish shades, 3 M: reddish shades) and standard deviations for all companies for summed residuals (moving sequences) and single residuals in a comparative overview. For the two estimation variants 1 M and 3 M, different trends emerge. In variant 1 M, the average F_1 values of the individual residuals decrease as the threshold value becomes smaller. For summed residuals, there is also a slight tendency for the F_1 values to decrease with decreasing threshold values. Furthermore, under the 1 M estimation in all condition variations, higher F_1 values are found for single residuals than for summed residuals.

A more differentiated picture emerges for variant 3 M when the F_1 values are considered. The combination of threshold values of Z > 2.5 or Z > 3 and residual sequences of four and six summed hours (sum 4 & sum 6) shows the highest F_1 values in the overall comparison. Consequently, summation can lead to improved anomaly detection quality under certain conditions. In the analysis of the data, not explicitly presented here, using strategy 2, an F_1 measure of 0.40 was shown for company 5 with the residual summation of six hours as well as Z values greater than 3. Comparing these values with the strategy 1 analysis of the data of the company from company 5, identically high F_1 values are shown.

Considering all the data analyzed and using the Z-Score thresholds, it can be concluded that the use of strategy 1 based on the LSTM network is preferable to the use of strategy 2.

The F_1 values obtained from the two strategies are rather low compared to examples from the literature. For example, Hundman *et al.* (2020) use an LSTM network to identify anomalies in multivariate time series of telemetry data from a spacecraft, where different types of thresholding procedures are tested [13]. The

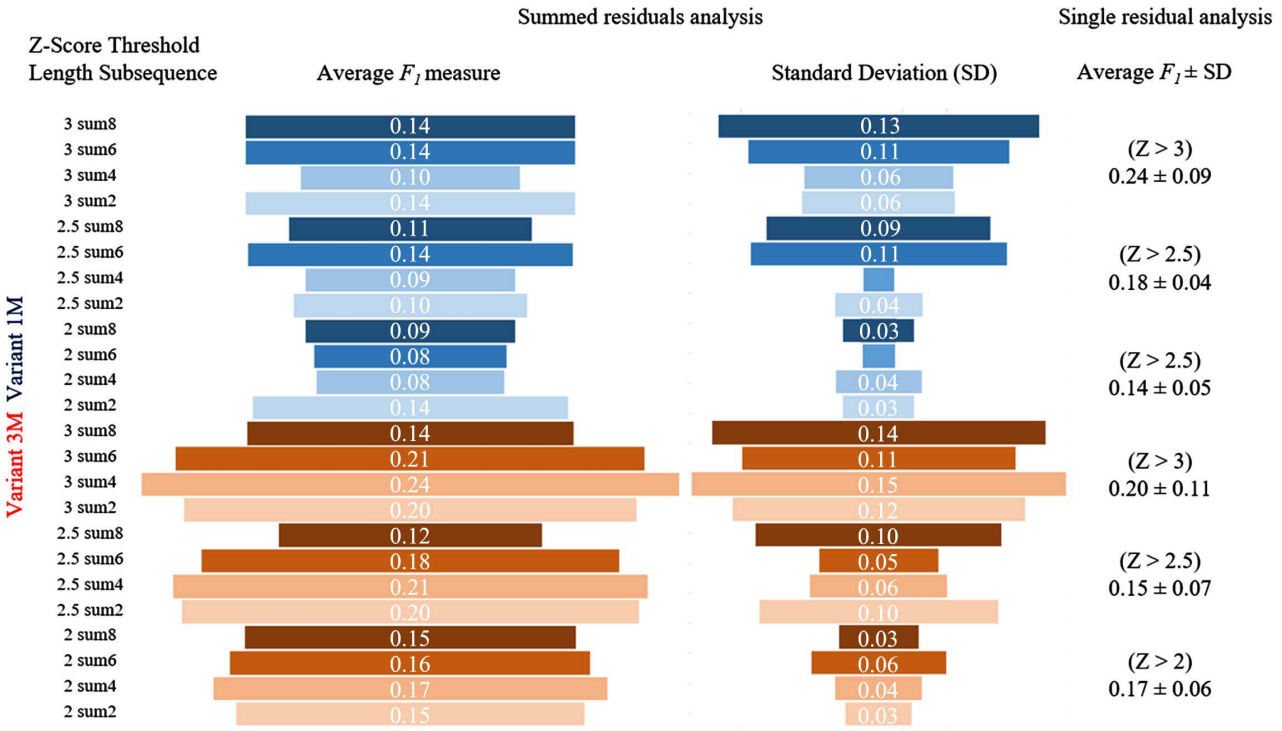

Figure 6. Plot of average F_1 measures and standard deviations (1 M: bluish shades, 3 M: reddish shades) for the companies 1 to 5 for summed subsequences and single residuals in comparison.

authors report *Precision* values of up to $P = 0.92$ in combination with a *Recall* of $R = 0.63$. Therefore, this publication suggests that other thresholding techniques should be tested for their usefulness, such as adaptive thresholding techniques. The publication leaves open which characteristics the curve progression of the described telemetry data has. Accordingly, the results cannot be compared with those of this work without exception. The PEWMA method has been successfully applied, for example, to streaming data from Twitter with resulting F_1 values of up to 0.80 [15] or to temperature time series with F_1 values of up to 0.84 [16]. However, the time series data analyzed in these cases have significantly different characteristics than the power time series analyzed here. Temperature time series, for example, are much lazier compared to power load profiles. Therefore, the elaborated strategy 2 could be applied in other areas to achieve satisfactory results.

5. Conclusions

This paper presents the comparison and adaption of two strategies for anomaly detection based on the LSTM network and the PEWMA method. The results show that anomaly can be partial to fully recognized by the chosen approaches previously defined, but with the maximum averaged F_1 value of 0.24. In many cases, the presence of a large number of false positive assignments leads to a decrease of F_1 values. In further efforts, approaches should be found to minimize the false positive results and thus improve the associated efficacy of the anomaly

detection.

The chosen technique of being able to identify anomalies using information from timestamps and associated performance values alone is probably one reason why the F_1 values are relatively low. The incorporation of additional factors affecting power values in the sense of a multivariate estimation model can possibly achieve higher anomaly detection efficacy. Furthermore, alternative methods for thresholding, such as adaptive methods, should be tested to determine their suitability. It must be considered that the anomalies identified by the expert interviews are not necessarily exhaustive. Thus, the false positive assignments may well have identified anomalies that were not noticed in the expert interviews. This will be considered further in future studies.

Overall, however, it can already be stated that the use of automated anomaly detection procedures can significantly reduce the search area for several hundred load profiles in a company and thus reduce the effort of manual analysis.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] European Commission (2014) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions a Policy Framework for Climate and Energy in the Period from 2020 to 2030. Document 52014DC0015. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52014DC0015&qid=1611915593867

[2] European Commission (2019) Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions the European Green Deal. Document 52014DC0015. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2019:640:FIN

[3] Krewitt, W., Nienhaus, K., Kleßmann, C., Capone, C., Stricker, E., Graus, W., Hoogwijk, M., Supersberger, N., Winterfeld, U. and Samadi, S. (2009) Role and Potential of Renewable Energy and Energy Efficiency for Global Energy Supply. https://www.umweltbundesamt.de/publikationen/role-potential-of-renewable-energy-energy

[4] Federal Office of Economics and Export Control (2021) Federal Funding for Pilot Program on Energy-Saving Meters. https://www.bafa.de/DE/Energie/Energieeffizienz/Einsparzaehler/einsparzaehler_node.html

[5] Chandola, V., Banerjee, A. and Kumar, V. (2009) Anomaly Detection: A Survey. *ACM Computing Surveys*, **41**, 1-58. https://doi.org/10.1145/1541880.1541882

[6] Aggarwal, C.C. (2017) Outlier Analysis. 5th Edition, Springer International Publishing, Cham.

[7] Blázquez-García, A., Conde, A., Mori, U. and Lozano J.A. (2020) A Review on outlier/Anomaly Detection in Time Series Data. *arXiv preprint*, arXiv: 2002.04236v1, 1-32. https://arxiv.org/abs/2002.04236

[8] Gupta, M., Gao, J., Aggarwal, C.C. and Han, J. (2014) Outlier Detection for Temporal Data: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, **26**, 2250-2267. https://doi.org/10.1109/TKDE.2013.184

[9] Wang, X., Lin, J., Patel, N. and Braun, M. (2018) Exact Variable-Length Anomaly Detection Algorithm for Univariate and Multivariate Time Series. *Data Mining and Knowledge Discovery*, **32**, 1806-1844. https://doi.org/10.1007/s10618-018-0569-7

[10] Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2018) Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward. *PLoS ONE*, **13**, e0194889. https://doi.org/10.1371/journal.pone.0194889

[11] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural computation*, **9**, 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[12] Siami-Namini, S., Tavakoli, N. and Namin, A.S. (2018) A Comparison of ARIMA and LSTM in Forecasting Time Series. 2018 17*th IEEE International Conference on Machine Learning and Applications*, Orlando, 17-20 December 2018, 1394-1401. https://doi.org/10.1109/ICMLA.2018.00227

[13] Hundman, K., Constantinou, V., Laporte, C., Colwell, I. and Soderstrom, T. (2018) Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. *Proceedings of the* 24*th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, 19-23 August 2018, 387-395. https://doi.org/10.1145/3219819.3219845

[14] Carter, K.M. and Streilein, W.W. (2012) Probabilistic Reasoning for Streaming Anomaly Detection. 2012 *IEEE Statistical Signal Processing Workshop* (*SSP*), Ann Arbor, 5-8 August 2012, 377-380. https://doi.org/10.1109/SSP.2012.6319708

[15] Patel, K., Hoeber, O. and Hamilton, H.J. (2015) Real-Time Sentiment-Based Anomaly Detection in Twitter Data Streams. In: Barbosa, D. and Milios E., Eds., *Advances in Artificial Intelligence*, Springer, Cham, 196-203. https://doi.org/10.1007/978-3-319-18356-5_17

[16] Novacic, J. (2019) Implementation of Anomaly Detection on a Time-Series Temperature Data Set. Thesis, Malmö Universitet, Malmö.

[17] Renshaw, J. (2016) Anomaly Detection Using AWS IoT and AWS Lambda. https://aws.amazon.com/de/blogs/iot/anomaly-detection-using-aws-iot-and-aws-lambda

[18] Rosner, B. (1983) Percentage Points for a Generalized ESD Many-Outlier Procedure. *Technometrics*, **25**, 165-172. https://doi.org/10.1080/00401706.1983.10487848