Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 149 (2018) 345–353

Energy Procedia

www.elsevier.com/locate/procedia

16th International Symposium on District Heating and Cooling, DHC2018,
9–12 September 2018, Hamburg, Germany

Ranking Abnormal Substations by Power Signature Dispersion

Ece Calikus, Sławomir Nowaczyk, Anita Sant'Anna, Stefan Byttner

Halmstad University, Depertmant of Intelligent Systems and Digital Design,
Kristian IV:s väg 3, Halmstad, 301 18 Halmstad

Abstract

The relation between heat demand and outdoor temperature (heat power signature) is a typical feature used to diagnose abnormal heat demand. Prior work is mainly based on setting thresholds, either statistically or manually, in order to identify outliers in the power signature. However, setting the correct threshold is a difficult task since heat demand is unique for each building. Too loose thresholds may allow outliers to go unspotted, while too tight thresholds can cause too many false alarms.

Moreover, just the number of outliers does not reflect the dispersion level in the power signature. However, high dispersion is often caused by fault or configuration problems and should be considered while modeling abnormal heat demand.

In this work, we present a novel method for ranking substations by measuring both dispersion and outliers in the power signature. We use robust regression to estimate a linear regression model. Observations that fall outside of the threshold in this model are considered outliers. Dispersion is measured using coefficient of determination R^2, which is a statistical measure of how close the data are to the fitted regression line.

Our method first produces two different lists by ranking substations using number of outliers and dispersion separately. Then, we merge the two lists into one using the Borda Count method. Substations appearing on the top of the list should indicate higher abnormality in heat demand compared to the ones on the bottom. We have applied our model on data from substations connected to two district heating networks in the south of Sweden. Three different approaches i.e. outlier-based, dispersion-based and aggregated methods are compared against the rankings based on return temperatures. The results show that our method significantly outperforms the state-of-the-art outlier-based method.

Selection and peer-review under responsibility of the scientific committee of the 16th International Symposium on District Heating and Cooling, DHC2018.

Keywords: Anomaly ranking, abnormal heat demand, power signature, district heating, fault detection

Selection and peer-review under responsibility of the scientific committee of the 16th International Symposium on District Heating and Cooling, DHC2018.
10.1016/j.egypro.2018.08.198

1. Introduction

Decreasing distribution temperatures is one of the most important steps to increase efficiency in district heating (DH) systems and plays vital role for the integration of 100% renewable energy supply [1]. Current distribution setups have high supply and return temperatures which lead to large heat losses in the network and inefficient use of heat sources.

One of the factors contributing to this situation is abnormal heat demand caused by faults in customer heating systems and substations [2 - 4]. In many cases, such problems do not directly cause noticeable decrease in customer comfort; however, they influence the performance of the network as a whole, leading to higher return temperatures and flow rates. Low temperatures in district heating can only be achieved if such abnormal demands are detected and eliminated.

Heat power signature models estimate the heat consumption of a building as a function of external climate data. They are typically presented as plots of total heat demand versus ambient temperature, showcasing the unique characteristics of each building (both physical and related to the behavior of the occupants). Many previous studies have been analyzing heat power signatures to diagnose abnormal heat demand.

Most methods are based on detecting outliers in the power signature by, either manually or statistically, setting a threshold on the power signature. However, setting a correct one is not always possible, since loose thresholds often allow outliers to go undetected, while too tight thresholds tend to cause too many false alarms [3, 5].

On the other hand, outliers are not the only symptom for abnormality in the power signature. High dispersion is also an indication of a problem such as faults or poor control [4], therefore must be taken into account. Existing methods that are based on counting outliers are not able to take dispersion into account. Combining both types of indicators requires a new approach.

In this work, we propose a novel method for ranking buildings by measuring both dispersion and outliers in their heat power signature and present large-scale analysis of district heating customers. Our method first produces two different lists by ranking substations using number of outliers and dispersion separately. Then, we merge the two lists into one using the Borda Count method.

Three different approaches, i.e., outlier-based, dispersion-based, and aggregated are evaluated against average and maximum return temperatures measured all the buildings in five different categories connected to two district heating networks in south of Sweden.

Based on those results, we conclude that outliers alone are not enough to identify abnormal heat demand in the buildings. The importance of also considering dispersion is clearly visible in analyzing high temperatures. The state-of-the-art outlier-based approach does not perform well alone for ranking abnormal buildings and it is significantly outperformed by dispersion-based and aggregated methods.

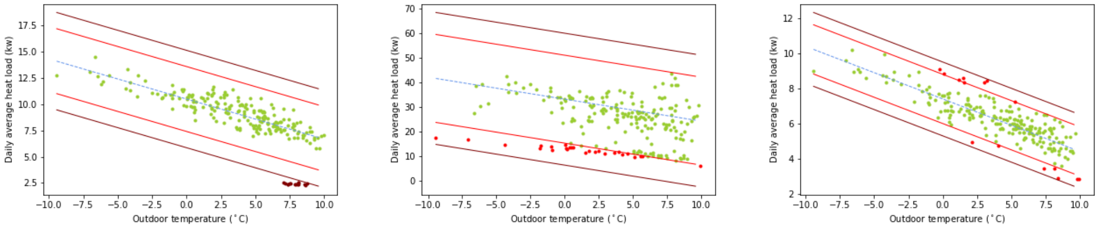
2. Related Work

The energy signature (ES) is a well-known method for the analysis of building energy consumption. It has been widely used for characterizing energy or heat demand behaviors of buildings in various studies.

[6] used ES methods for weather correction which aims to normalize energy consumption so that it becomes the representative of a building's expected long-term performance. In [7], a similar approach is applied to the entire DH network. Single heat power signature per year was plotted from heat load measurements of all the buildings connected to the network in order to compare different heat seasons.

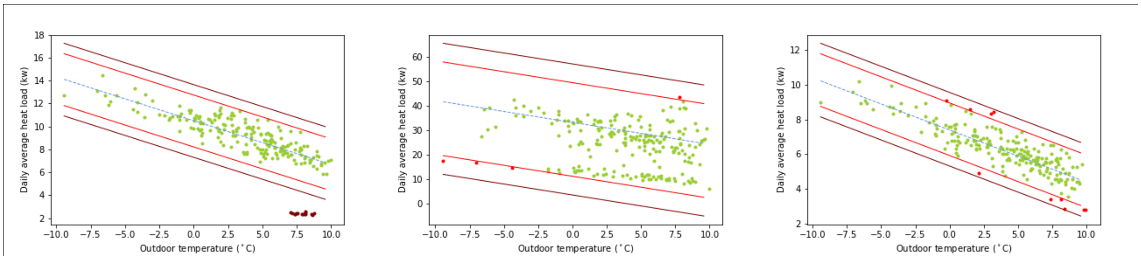
ES based methods have also been applied for the estimation of the amount of heat losses due to transmission and ventilation by quantifying the buildings' total heat loss coefficient [8 - 14]. In addition, they have been investigated to correctly estimate balance temperature in order to separate demand from space heating and domestic hot water [15].

ES methods provide useful information on buildings' energy performance in DH systems by analyzing correlation between the average heating power and outdoor temperature. Therefore, they have also been analyzed for the detection of anomalies or deviations in heat demand behaviors [4, 16]. Those approaches commonly implement outlier detection methods based on thresholding strategies to estimate errors in the energy signatures [3, 17-20].



(a) Abnormal Building: $T_a = 45$ºC, $T_m = 77$ºC (b) Abnormal Building: $T_a = 57$ ºC, $T_m = 72$ ºC (c) Normal Building: $T_a = 27$ ºC, $T_m = 30$ ºC

Figure 1: Thresholds based on standard deviation of the residuals.



(a) Abnormal Building: $T_a = 45$ºC, $T_m = 77$ºC (b) Abnormal Building: $T_a = 57$ ºC, $T_m = 72$ ºC (c) Normal Building: $T_a = 27$ ºC, $T_m = 30$ ºC

Figure 2: Thresholds based on median absolute deviation of the residuals.

However, defining the correct threshold is a difficult task. Identifying thresholds manually requires extensive human effort and domain knowledge. It is extremely time-consuming considering the number of buildings in a DH network. On the other hand, automatic determination of thresholds using statistical models is also very challenging since it requires finding an optimum strategy which will maximize the outlier detection performance while limiting false alarms. We demonstrate the difficulty of this task in Figure 1 and 2 by comparing two commonly used statistical thresholding strategies.

Those strategies are applied to the power signatures of three buildings estimated by linear regression. Two of the buildings are selected as abnormal examples whose return temperatures are high, and one is normal having low return temperature measurements. In Figure 1, thresholds are defined based on standard deviation (σ) of the residuals. The wider and the tighter thresholds respectively correspond to 3σ and 2σ above and below the regression line. The second strategy applies modified z-scores [21] which are computed using median absolute deviation of the residuals. The threshold is set so that modified z-scores do not exceed 3.5 for the wider case, and not exceed 2.5 for the tighter threshold.

The wider thresholds in both cases are able to identify outliers in the first abnormal buildings, but they fail detecting the second building showing high dispersion. Tighter thresholds are instead able to detect outliers in both the anomalous building; however, they also mistakenly mark some of the data in normal building as outliers, which leads to a high false alarm rate.

3. Method

In our approach, we rank buildings by measuring the “degree of abnormality” on heat power signatures with three different methods, i.e., outlier-based, dispersion-based and aggregated ranking. Power signatures of the buildings are estimated using robust regression in order to eliminate the influence of the outliers on model estimation.

3.1. Robust Regression

Ordinary Least Squares (OLS), a typical approach used to estimate energy signatures, is highly sensitive to outliers. Since parameter estimation is based on the minimization of squared error, even the presence of few outliers can have distorting influence and makes results of regression analysis including confidence intervals, prediction intervals, R2 values, t-statistics, p-values, etc. unreliable [21].

Robust regression methods try to overcome those issues by providing robust estimates when outliers are present in the data. In our work, we apply Random Sample Consensus (RANSAC) algorithm [22] to do robust estimation of model parameters of power signatures in the presence of outliers. RANSAC is an iterative approach which fits the regression line to subsets of data until the model with most inliers and the smallest residuals on the inliers is chosen. The process continues unless either user-defined fixed number of iterations or threshold for the minimum number of samples that would be accepted as inliers to generate a final model is reached.

RANSAC has been shown to be a very robust approach for parameter estimation, i.e., it can estimate the parameters with a high degree of accuracy even when a significant number of outliers are present in the data set [23]. However, there are several drawbacks that should be taken into account while applying this approach. For example, we do not have prior knowledge on the ratio of outliers in power signature for every building in our data set. Therefore, setting stopping criterion such as maximum number of iterations or inliers is not trivial. In our case, we set the ratio of outliers to 20% when fitting the regression line since having false positives does not wildly affect parameter estimation. We estimate the regression line and residuals with this method. However, we employed a different approach to compute the final number of outliers in order to avoid high number of false alarms produced by RANSAC and also computed R^2 measures removing those outliers.

In order to demonstrate the benefits of using robust models in this problem, we conducted an experiment comparing goodness-of-fit of each power signature estimated by traditional OLS and RANSAC method.

First, OLS and RANSAC methods are separately fitted to each power signature. Then, we measure R^2 scores of all the models estimated by OLS and RANSAC. In order to avoid influence of outliers, R^2 scores are computed on observations excluding outliers determined by both OLS and the RANSAC. We use a statistical threshold on residuals to detect outliers which is explained in section 3.2 explicitly.

According to R^2 results, RANSAC method has better goodness-of-fit score on 61% of the all power signatures. We also conduct Student's t-test [24] to conclude whether R^2 scores are significantly lower in the models estimated by OLS in comparison to RANSAC algorithm.

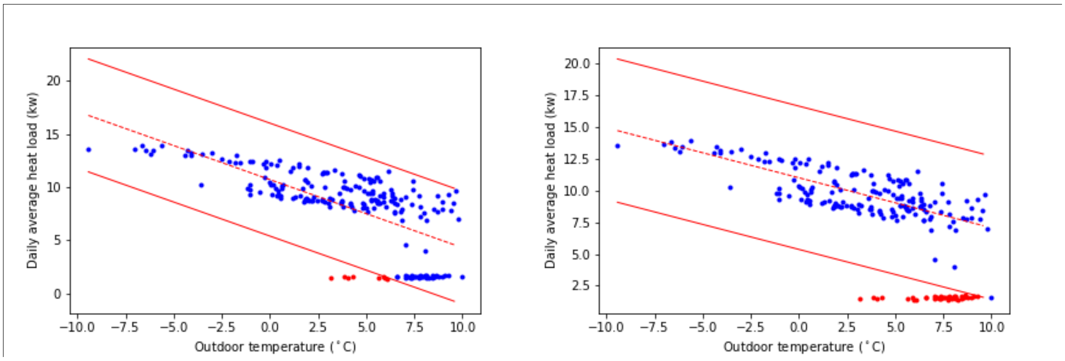
The null hypothesis is H_0: $\mu_1 - \mu_2 \geq 0$, alternative hypothesis is H_a: $\mu_1 - \mu_2 < 0$ and significance level is $\alpha = 0.05$. The t-statistics of single-tailed test is $T = -2.349835$ and $p = 0.00944$. The p-value (0.00944) is lower than significance level (0.05). Therefore, at 5% level of significance, the test provides sufficient evidence that the power signatures estimated by OLS have lower goodness-of-fit than the ones estimated by RANSAC algorithm.

Furthermore, we demonstrate that the estimation of power signatures between those two methods differs significantly for a large portion of substations. Figure 3 shows an example substation where the two methods result in significantly different models. Crucially, in our dataset, there are 250 more buildings that have higher difference between R^2 values from OLS and RANSAC than the example building represented in Figure 3. In other words, for almost 30% of the buildings, the importance of using robust regression is even more significant than for the example.

3.2. Outlier-based ranking

In the literature, a widely applied measure for determining the "degree of abnormality" in buildings is the "number of outliers" in the power signature.

"Number of outliers" is determined by setting a statistical threshold on the distribution of the residuals. Under the normality assumption, 95.45% and 99.73% of the values lie within two (2σ) and three (3σ) standard deviations of the mean, respectively. However, the presence of outliers and the effect of other factors on power signature lead to violation of the assumption of normally distributed residuals.



a) OLS (b) RANSAC

Figure 3: Difference between OLS and RANSAC in the presence of outliers

For non-normally distributed data, only 75% of the distribution's values are guaranteed to lie within (2σ) of the mean and 89% within (3σ), according to Chebyshev's inequality [25]. Considering that, we set the threshold to (3σ) around the mean of the residuals in order to ensure an upper-bound of approximately 11% on the false positive rate.

Given outdoor temperature x_i in power signature, let y_i be the actual heat load, $\hat{y}_i$ be the predicted heat load, $r_i = \hat{y}_i - y_i$ be the residual and μ and σ, be the mean and standard deviation of the distribution of the residuals $D(\mu, \sigma)$. Then, the outliers are determined as follows:

$$f(x_i, y_i) = \begin{cases} outlier, & if \ \ r_i - \mu \geq 3\sigma \\ inlier, & \quad otherwise \end{cases} \qquad r_i \sim D(\mu, \sigma) \tag{1}$$

Finally, all the buildings are sorted based on "the number of outliers" in decreasing order so that buildings that have "higher degree of abnormality" are placed higher in the list.

3.3. Dispersion ranking

In the second approach, the dispersion in the heat power signature is used as "degree of abnormality". In order to measure it, we use coefficient of determination (R^2). This is a statistical metric that evaluates the scatter of the data points around the fitted regression line. Since power signatures with lower R^2 values are more dispersed, we therefore define them as having higher "degree of abnormality".

As stated earlier, the outliers can also influence R^2, in particular, they can lead to low scores although the linearity of the model is satisfied. We reduce the effect of such misleading examples by removing the outliers detected with the thresholding strategy before calculating the R^2 scores. Then, the final ranking is produced by sorting all the buildings according to their R^2 values.

3.4. Aggregated approach

The Borda Count method [26] is a traditional voting method that was developed in the 18th-century and broadly applied as aggregation strategy to combine rankings produced by different algorithms.

Given a particular ranking as a sorted list of elements, the method works by assigning a score to each member of the list according to its relative position. Once the method is applied to different rankings, the final aggregated ranking is a sorted list based on the sum of scores of each element. This method can be seen as equivalent to combining ranks by their mean [27].

4. Results

4.1. Data

The dataset used in this study is comprised of smart meter readings from buildings connected to the district heating systems in Helsingborg and Ängelholm in the South-West of Sweden. The data set consists of hourly measurements of heat, flow, supply, and return temperatures on the primary side of the substations from the whole year 2016. In this study, we only use the heat measurements of the buildings in five customer categories: multi-dwelling buildings, industrial demands, health-care social services, public administration buildings, and commercial buildings. The number of buildings is approximately 1700.

Problems in smart meter devices can cause missing or erroneous readings in customer records. Therefore, we apply a data preprocessing step to deal with incorrect meter readings. Customers that have missing heat load or return temperature measurements for at least one consecutive day are excluded from the analysis. Shorter periods of missing values are filled using linear interpolation of surrounding values. Meter readings that have identical values for more than one day are also excluded. As a result, we include 896 buildings in this study.

For the buildings with good quality of data, heat power signatures are extracted based on the daily average heat load and average daily outdoor temperatures. We only consider days when the average outdoor temperature is below 10ºC. It has been shown that when the space heating is not the main source of the heat demand in a building, there is no strong correlation between outdoor temperature and the temperature difference between supply and return pipes. Instead of examining balance temperature for each signature, we simply set it to 10 ºC as stated in [3].

4.2. Evaluation

In this section, we conduct experiments to evaluate our novel method, which measures both dispersion and outliers by comparing with dispersion-based and outlier-based methods individually. Each method separately produces a ranking of the most anomalous buildings, and we evaluate those rankings using return temperature measurements of the buildings. For each building, we calculate average (denoted T_a) and maximum (denoted T_m) return temperatures measured on the same dates as the heat loads in the power signatures. Clearly, both are relevant from the optimization of DH networks perspective, but they capture different aspects. While T_a values indicate long-term return temperature behavior, the T_m captures the most extreme operation of a building.

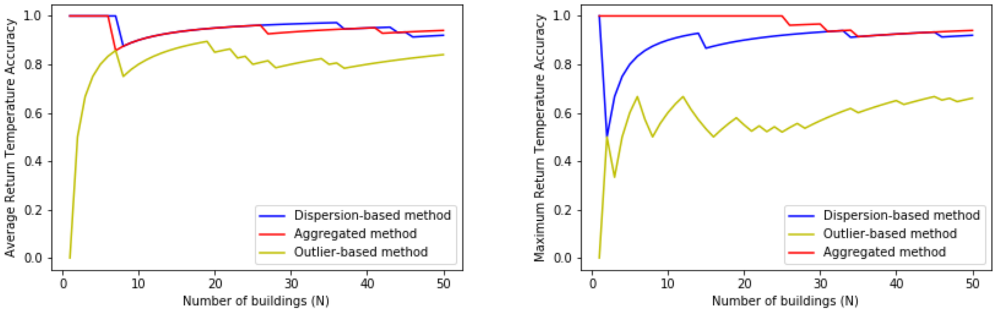
We present two different strategies to evaluate that buildings which have high rankings are actually problematic. The first strategy, "accuracy at the top" shows the ratio of abnormal buildings, compared to normal ones, near the top of the ranking. We compute "accuracy of top N buildings" as $(N_{abnormal}/(N_{normal} + N_{abnormal}))$ where N_{normal} and $N_{abnormal}$ are number of normal and abnormal buildings, respectively, among the top N ranked buildings. Buildings that have T_a higher than 35ºC or T_m higher than 45ºC are considered to be abnormal in this strategy.

The second strategy, "average temperature at the top", shows the average T_a and T_m values of top ranked buildings. We compute average T_a ($\hat{T}_a$) and average T_m ($\hat{T}_m$) of top N buildings as follow:

$$\hat{T}_a = \frac{\sum_{n=1}^{N} T_{a_n}}{N} \tag{2}$$

$$\hat{T}_m = \frac{\sum_{n=1}^{N} T_{m_n}}{N} \tag{3}$$

Figure 4 shows accuracies at top N buildings according their T_a and T_m values. Dispersion-based method and aggregated method show very similar performance in T_a accuracy (cf. Figure 4(a)). However, aggregated method converges at the level of 94% (red line), while dispersion-based method settles at 92% (blue line).



(a) T_a accuracies among top ranked buildings (b) T_m accuracies among top ranked buildings

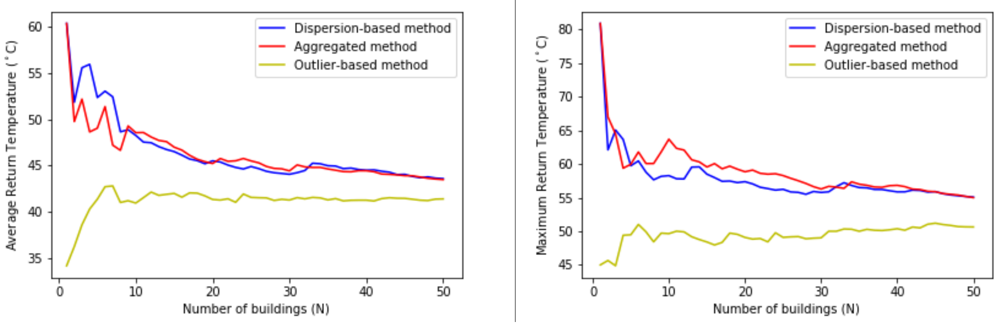
Figure 4: Accuracies at the top

There are minor differences along the way, but they are not significant. On the other hand, the outlier-based method achieves significantly lower accuracy of 83% (yellow line), and actually fails to detect the most severely abnormal building.

In terms of accuracy based on T_m, the aggregated approach significantly outperforms both dispersion-based and outlier-based methods. It perfectly identifies (achieving 100% accuracy) the top 27 abnormal buildings, and flattens at 94% similar (cf. red line in Figure 4b). Dispersion-based approach is hindered by several false positives near the top of the ranking, but its accuracy increases with the number of customers and reaches to 92% (blue line). The outlier-based approach, again, shows by far the worst performance, with final accuracy of only 66% (yellow line).

In Figure 5, the results of average temperatures at the top are shown. Dispersion-based (blue line) and aggregated (yellow line) methods start and get flattened at the same temperature in both cases (cf. Figure 5(a) and Figure 5(b)). Dispersion-based method shows slightly higher $\hat{T}_a$ until top 10 buildings (cf. blue line in Figure 5(a)), while aggregated method is almost constantly better at $\hat{T}_m$ until convergence (cf. red line in Figure 5(a)). Buildings that got higher rankings by the outlier-based approach show significantly lower return temperatures (cf. yellow line in Figure 5(a) and Figure 5(b)).

We present two different results, since they capture different types of abnormality, both of which can be important. It is crucial to notice that our proposed method outperforms state-of-the-art in either case. In particular, the buildings that are experiencing long-term problems are likely to be characterized by large T_a, while abrupt failures will cause unusually high T_m; however, those latter ones may not affect T_a significantly if they are quickly fixed. An anomaly detection method needs to detect both kinds of problems -- and as shown in Figures 4 and 5, the proposed method exactly provides that.



(a) Average T_a of top ranked buildings (b) Average T_m of the top ranked building

Figure 5: Average temperature at the top.

5. Conclusion

In this work, we postulate that ranking abnormal buildings should be done based on power signatures estimated using robust regression and needs to include measuring both dispersion and outliers. We propose a novel method for doing that, based on combining building rankings to measure their "degree of abnormality".

We've conducted experiments on large-scale data from substations connected to two district heating networks in south of Sweden. We have compared our method against the state-of-the-art outlier-based approach. Return temperatures of the buildings have been taken as reference for two different types of evaluation which we referred as "accuracy at the top" and "average temperature at the top". The first one has shown the ratio of abnormal buildings ranked on the top, while the second one has shown average return temperatures of top abnormal buildings.

The results demonstrate that dispersion-based and aggregated methods significantly outperform the state-of-the-art approach.

References

[1] H. Lund, S. Werner, R. Wiltshire, S. Svendsen, J. E. Thorsen, F. Hvelplund, B. V. Mathiesen, 4th generation district heating (4gdh): Integrating smart thermal grids into future sustainable energy systems, *Energy* 68 (2014) 1–11.
[2] H. Gadd, S. Werner, Heat load patterns in district heating substations, *Applied energy* 108 (2013) 176–183.
[3] H. Gadd, S. Werner, Achieving low return temperatures from district heating substations, *Applied energy* 136 (2014) 59–67.
[4] H. Gadd, S. Werner, Fault detection in district heating substations, *Applied Energy* 117 (2015) 51–59.
[5] F. Sandin, J. Gustafsson, J. Delsing, R. Eklund, Basic methods for automated fault detection and energy data validation in existing district heating systems, In *International Symposium on District Heating and Cooling:* 03/09/2012-04/09/2012, District Energy Development Center, 2012.
[6] L. Lundström, Adaptive weather correction of energy consumption data, *Energy Procedia* 105 (2017) 3397–3402.
[7] M. Noussan, M. Jarre, A. Poggio, Real operation data analysis on district heating load patterns, *Energy* 129 (2017) 70–78.
[8] L. Belussi, L. Danza, I. Meroni, F. Salamone, Energy performance assessment with empirical methods: Application of energy signature, *OptoElectronics Review* 23 (1) (2015) 85–89.
[9] I. Allard, T. Olofsson, G. Nair, Energy performance indicators in the swedish building procurement process, *Sustainability* 9 (10) (2017) 1877. 14
[10] G. Nordström, H. Johnsson, S. Lidelöw, Using the energy signature method to estimate the effective u-value of buildings, in: Sustainability in Energy and Buildings, Springer, 2013, pp. 35–44.
[11] S. Danov, J. Carbonell, J. Cipriano, J. Mart´ı-Herrero, Approaches to evaluate building energy performance from daily consumption data considering dynamic and solar gain effects, *Energy and Buildings* 57 (2013) 110–118.
[12] C. Ghiaus, Experimental estimation of building energy performance by robust regression, *Energy and buildings*, 38 (6) (2006) 582–587.
[13] J.-U. Sj¨ogren, S. Andersson, T. Olofsson, Sensitivity of the total heat loss coefficient determined by the energy signature approach to different time periods and gained energy, *Energy and Buildings* 41 (7) (2009) 801–808.
[14] J.-U. Sj¨ogren, S. Andersson, T. Olofsson, An approach to evaluate the energy performance of buildings based on incomplete monthly data, Energy and Buildings 39 (8) (2007) 945–953.
[15] Averfalk H, Werner S. Novel low temperature heat distribution technology. *Energy*. 2018 Jan 2.
[16] L. Pistore, G. Pernigotto, F. Cappelletti, P. Romagnoni, A. Gasparella, From energy signature to cluster analysis: an integrated approach.
[17] L. Farinaccio, R. Zmeureanu, Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major enduses*, Energy and Buildings* 30 (3) (1999) 245–259.
[18] L. Danza, L. Belussi, I. Meroni, M. Mililli, F. Salamone, Hourly calculation method of air source heat pump behavior, *Buildings* 6 (2) (2016) 16.
[19] L. Belussi, L. Danza, Method for the prediction of malfunctions of buildings through real energy consumption analysis: Holistic and multidisciplinary approach of energy signature, *Energy and* Buildings 55 (2012) 715–720.
[20] Acquaviva A, Apiletti D, Attanasio A, Baralis E, Bottaccioli L, Castagnetti FB, Cerquitelli T, Chiusano S, Macii E, Martellacci D, Patti E. Energy signature analysis: Knowledge at your fingertips. InBig Data (BigData Congress), 2015 IEEE International Congress on 2015 Jun 27 (pp. 543-550). IEEE.
[21] R. Maronna, R. D. Martin, V. Yohai, Robust statistics, Vol. 1, John Wiley & Sons, Chichester. ISBN, 2006.
[22] Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. InReadings in computer vision 1987 (pp. 726-740). [23] D. R. Kisku, P. Gupta, J. K. Sing, Advances in biometrics for secure human authentication and recognition, CRC Press, 2013.
[24] G. W. Snedecor, Statistical Methods: NY George W. Snedecor and William G. Cochran, Iowa State University Press, 1967.
[25] G. Grimmett, D. Stirzaker, Probability and random processes, 2001.
[26] J. C. de Borda, M´emoire sur les ´elections au scrutin.
[27] P. A. Jaskowiak, D. Moulavi, A. C. Furtado, R. J. Campello, A. Zimek, J. Sander, On strategies for building effective ensembles of relative clustering validity criteria, Knowledge and Information Systems 47 (2) (2016) 329–354