Fault detection in district heating substations: a cluster-based and an instance-based approach

Jad Al Koussa [a,b], *Sara* Månsson[a,b,c,d],

[a] Unit Energy Technology, VITO, Boeretang 200, BE-2400 MOL, Belgium, jad.alkoussa@vito.be

[b] EnergyVille, Thor Park 8310, BE-3600, Genk

[c] Department of Energy Sciences, Faculty of Engineering, Lund University, P.O. Box 118, SE-221 00 Lund

[d] current affiliation: Utilifeed, Stora Badhusgatan 18-20 - 411 21 Gothenburg – Sweden, sara@utilifeed.com

Abstract. In district heating or collective heating substations, components can fail or can be inappropriately installed or configured (e.g. valves get broken, heat exchangers get fouled, controller parameters are inappropriately chosen, heat exchanger wrongly connected, internal heating system problems, etc.). The result of these faults is a reduced cooling of the supply water, and as such higher than necessary return temperatures to the grid and higher volume flows (to deliver the same needed power) occur, leading to higher OPEX for all stakeholders. In this work, two approaches for a fault detection routine for district heating substations are introduced, based solely on the energy meter data, with an application on a real-life district heating network in Sweden. The first approach is a cluster-based approach in which substations within the district heating are compared to each other using the overflow method and performance signatures to flag substations with sub-optimal performance compared to other substations in the network. The second method is an instance-based approach using a black-box model to predict the behaviour of the substation using an extended set of input variables and comparing the predictions to the measurements. The results from the two fault detection approaches show that both the cluster-based and the instance-based methods can detect deviating behaviours in DH customer installations.

Keywords. Fault detection and diagnosis, District heating, District heating substations
DOI: https://doi.org/10.34641/clima.2022.45

1. Introduction

District heating (DH) is a collective heating system in which heat is produced in one (or multiple) locations to meet the heat demand, usually of multiple customers at other locations. The heat is transferred using water circulating in a pressurized piping system. The heat demand consists of Space Heating (SH) demand as well as Domestic Hot Water (DHW) demand for sanitary purposes for residential and commercial buildings. A simplified illustration of a district heating system is shown in Fig. 1: the main components are the heat production where heat is produced, the heat distribution that transports the heat, the customer substation that transfers the heat from the heat distribution pipes to the internal heating system, and the internal heating system that consists of the SH and the DHW circuits.

Due to the inherent energy efficiency, heat source flexibility and the capability of utilizing residual heat, district heating has been identified by the European Union as an essential solution in future energy

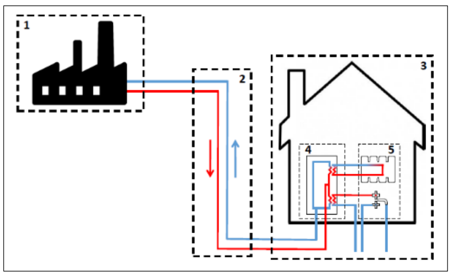

Fig. 1 - A simplified illustration of a district heating system, where 1 represents production, 2 represents distribution, 3 represents the customer installations, 4 represents the customer substation, and 5 represents the internal heating system.

systems to help decarbonizing Europe [1,2]. The district heating network can be supplied by a multitude of sources such as fossil fuel based boilers as well as more sustainable sources such as excess heat from industrial processes and Combined Heat and Power plants (CHPs) based on biomass or waste incineration, geothermal heat and solar thermal heat [1]. Moreover, large-scale heat pumps can be used as a heat source providing the possibility to increase the flexibility of the energy system by integrating the heating and electricity sectors [3].

In existing DH systems, multiple components can perform sub-optimally leading to inefficiency in the operation of the system due to an increase in the system's temperatures. Customers' installations (component 3 in Fig-1) is one of the major components that can perform sub-optimally, causing increased return temperatures in the DH system and consequently either the flow or the supply temperature need to be increased to provide the needed heat [4,5]. The customer's installation consists of: 1) the substation, which is a combination of heat exchangers, valves, sensors, actuators, control and a heat meter, as well as 2) the SH and DHW system that contains a number of valves, radiators and heat exchanger. The issues, or faults, in the customer installations may occur in a number of different components and include faults and problems such as fouling of heat exchangers, broken temperature sensors, control valves stuck in an open position, temperature sensors placed on the wrong pipe, DHW circulation connected before the pre-heater, high return temperatures from the customer's internal heating system, and high set point values in the customer system [7-9]. All of these faults may not be seen as an actual fault where something is broken, but they still lead to high return temperatures.

Previous studies have shown that a large share of the substations in different DH systems are not performing well [6,7], indicating that substations have a large impact on the DH system's temperature and its efficiency. With the district heating utilites aiming at decreasing their system's temperatures to increase their efficiency and integrate low temperature renewable heat sources, and the emergence of the 4th generation district heating networks, it becomes increasingly important that faulty installations are detected and corrected. Traditionally, the poorly performing installations have been detected using manual analysis methods [10]. However, given the large number of substations that might be present in a DH network, the manual approach can be very time consuming and costly. For this reason, there is a need for more automated methods to help in the fault detection process.

In this paper, two methods to detect sub-optimally performing substations are presented: the first approach is a cluster-based approach in which substations are compared to each other based on their historical data and the second approach is an instance-based approach in which a substation is modelled using a black box and a deviation is detected by comparing the outcome of the model to the field measurements. In what follows, a review of exisiting methods for Fault Detection is presented followed by the explanation of the implemented methods and after that the results are presented. The paper finishes by a discussion on the implemented methods followed by concluding remarks and the prospect of future work.

2. Fault detection in district heating substations

Fault detection and diagnosis (FDD) is a collection of methods to monitor a system's behaviour, to determine if a fault is present in the system, and to determine the characteristics and root cause of the detected fault(s) [11]. FDD is applied in a variety of domains for fault detection and predictive maintenance.

The application of fault detection on DH customer installations is gaining momentum in recent years. The European Energy Efficiency Directive which became effective in 2012 states that all energy customers should be billed according to their actual energy consumption[12]. In order to do that, energy meters need to be placed at the primary side of the customer's district heating substation. With this data being available, the utilities have the chance to perform analysis for different purposes, such as FDD. Literature regarding this topic shows that different types of data analysis methods are applicable for the detection of faults in DH substations. The literature includes aproaches based on cluster-based fault detection techniques as well as instance-based techniques. Sandin et al in [13] provided an overview of different statistical and probabilistic methods for fault detection in substations, such as linear regression modeling, limit checking, correlation analysis and outlier detection, and provided examples of their application on hourly based data. Gadd and Werner in [5] used the overflow method, a common method that compares the actual flow over a period of time to an ideal flow derived from an ideal value of the primary temperature difference. The authors also use temperature difference signatures to detect faults. In [14], Calikus et al. use heat power signatures and their degree of abnormality to rank individual buildings. Farouq et al. present in [15] a reference-group based approach for detecting customer installations that display a deviating behaviour. The reference groups were based on a k-nearest neighbour approach, utilizing the return temperatures from the investigated installations to identify the reference groups for each installation in the data set. If an installation deviates significantly from its reference group, it is considered faulty.

Machine learning (ML) and instance-based methods have also been explored in the literature. In [16], Ingvarsson et al. explored the possibilities to use models for fault detection and to track slow drifts in the substations' performance. The results show that the best suited model is a SARIMAX (0, 1, 1)x(0, 1, 1)24, for any combination of variables. In the project

RELaTED [18], two tools have been developed for automatic fault detection in DH substations based on ML algorithms: DH doctor and DH Autotune. The first one exploits clustering in which anomalies can be detected by measuring the distance among the clusters and following the evolution of the centroids related to a particular variable over time. The second tool is based on hourly averaged readings and allows the prediction of the load. Alarms are activated if some KPIs exceed a threshold. The work done by Guelpa et al. in [17] focused on the detection of fouling in district heating substations. An automatic method using the most commonly collected metering variables, such as volumetric flow and temperatures in the substation primary and secondary circuits, has been developed.

The literature shows that there is a variety of successfully implemented methods for fault detection. However, many of the presented methods require an amount of manual handling and/or interpretation. Some of the methods also require a certain amount of understanding for more advanced computer science and data handling methods. In this study, the aim is to show two simple methods that help eliminating a lot of manual stages in the data analysis, which can be used by the DH industry without the need for prior advanced knowledge in computer science and data handling.

In the next two sections, the two approaches are presented.

3. Cluster-based approach

3.1 data used

The data used in this study was gathered from the business system of a DH utility in Sweden between April 2015 and March 2016, and it includes data from the 3 000 installations that had the largest energy consumption in the system. The data set contained typical energy meter data, such as the accumulated energy consumption, the accumulated volume passing through the installation, the primary return temperature, and the primary supply temperature for each of the 3 000 installations. Moreover, the installation ID, the postal code of the installation, and the outdoor temperatures were collected.

3.2 Method

The cluster-based method compares substations within a group or a cluster of substations to find faulty or poorly performing substations. A schematic of this method is shown in Figure 2.

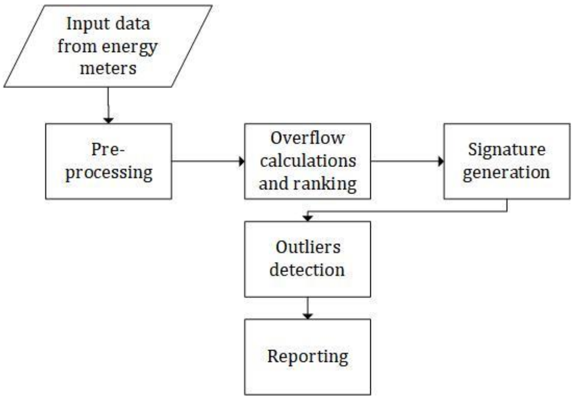

Fig. 2 – Schematic view of the cluster-based fault detection method.

First, a pre-processing and data handling is performed on the input data. During this step, the data is prepared and transformed into a suitable format for the data analysis step. Data processing may include several tasks, such as removal of duplicate values, feature extraction, and removal of outliers. In this first step, faults in sensors readings, communications and data logging can be detected. After that, data analysis is performed. Here, for each substation, an overflow value is calculated. The overflow represents the extra volume of water that passed on the primary side of the substation during a certain period compared to an ideal calculated volume. This ideal volume is calculated based on an ideal cooling value of the primary temperature. The overflow can be calculated using equation (1):

$$V_{overflow} = V_{actual} - V_{ideal} = V_{actual} - \frac{E_{actual}}{\rho . c_p . \Delta T_{ideal}} \text{ (m}^3\text{)} \quad (1)$$

Where

$V_{overflow}$	= overflow volume (m^3)
V_{actual}	= actual measured volume (m^3)
V_{ideal}	= ideal volume (m^3)
E_{actual}	= actual energy (J)
ρ	= fluid density (kg/m^3)
Cp	= specific heat capacity (J/kg.K)
ΔT_{ideal}	= ideal temperature difference between the primary supply and return (K)

The ΔT_{ideal} value varies depending on which DH system is being investigated. In this study, it was chosen to be 45 °C, in accordance with [3]. As equation (1) shows, the lower the primary ΔT, the higher the overflow. This is an indication that the customer's installation doesn't function optimally.

After the calculation of the overflow, a set of the substations with the lowest overflow values is picked and considered as the reference case: a representative set of the best performing substations. In this case, the top 25 % are considered for the reference case. Then, to determine from the remaining substations the poorly performing ones, three different criteria were used: the cooling performance, the return temperature level, and the energy consumed in the building. Two criteria were used to create two signatures: one cooling signature and one return temperature signature. The signatures consisted of a reference case and threshold values which were used for outlier detection. For each of the signatures, the reference case is determined by performing a piecewise linear regression on the reference substations chosen based on the overflow method. The mathematical relationship for a piecewise linear regression model with one breakpoint H is defined as follows [19]:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i - H)\, I(X_i, H) + \varepsilon_i\,, \qquad (2)$$
$$where\ I(X_i, H) \begin{cases} 1, & X_i > H, \\ 0, & X_i \leq H. \end{cases}$$

In the equation, β_n, n = 0, 1, 2 are the parameters of the regression model, Y_i is the dependent variable being modelled, X_i is the independent variable which is used to model the dependent variable, and ε_i is the model error. In the cooling signature, the cooling of the substation was modelled as a function of the outdoor temperature. The breakpoint for the piecewise linear regression was determined by visually inspecting the data set.

Once the piecewise linear regression for the reference case is determined, the deviating, or outlier, values can be identified. The outliers were identified using the mean and the standard deviations of the reference case values. Values located at a distance larger than 3 standard deviations from the mean are considered as outliers [13]. For the cooling, piecewise linear regression was used to model the mean of the reference case and so the thresholds were also linear. For the return temperature signature, the mean was modelled using a constant value, resulting in constant thresholds.

3.3 Results

Figures 3 and 4 show the cooling and return temperature signatures with one well performing and one poorly performing customer installation visually represented. In both Figures, the well performing installation is represented by the blue circles, while the poorly performing installation is represented by the red circles. As may be seen in the figures, the well performing installation has all its values located *inside* the thresholds, while the poorly performing installation has the main share of the values located *outside* of the thresholds.

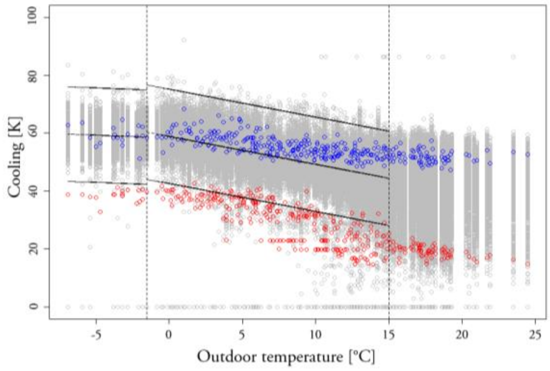

Fig. 3 – Cooling signature with one well performing (blue circles) and one poorly performing (red circles) installation represented.

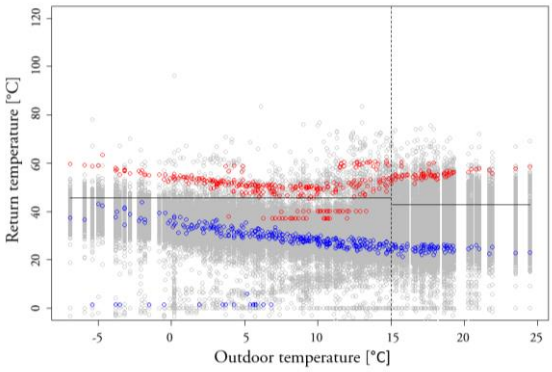

Fig. 4 – Return temperature signature with one well performing (blue circles) and one poorly performing (red circles) installation represented.

The developed fault detection approach identified 1 273 installations as being poorly performing, which corresponds to approximately 43 % of the investigated installations. Tab.1 presents the five installations that were identified as the substations that deviated the most from the expected behaviour.

Tab. 1 - The five substations with the highest overflows, and with outliers for both cooling and return temperature signatures.

Subs #	6	7	8	9	10
ID	X54	X93	X45	X41	X21
# of dT Outliers	277	277	169	201	277
# of Tr Outliers	366	366	316	363	366
$V_{overflow}$ (m^3)	124 581	102 885	64 824	57 121	55 927

4. Instance-based approach

4.1 data used

The dataset used in this study consisted of hourly values for one year (November-November) from the energy meter of one DH substation in a DH system in Sweden. The variables, or features, that were included in the dataset are: substation ID, average hourly outdoor temperature, hourly measurements of heat power consumption, mass flow, supply and return temperatures, as well as the average value of the outdoor temperature during the previous 24 hours.

4.2 Method

In this method, a well-performing substation is modelled and then its performance is predicted based on subsequent input data. Any faults are then detected as deviations from the normal behaviour in the model predictions. The installation was selected from a data set of approximately 1 000 customer installations. First the overflows (Section 3.2) for all installations were calculated and then 10 installations with the lowest values of overflow were visually inspected to identify the best performing one.

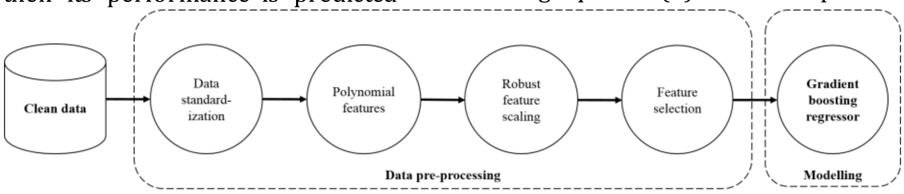

Fig. 5 – The TPOT pipeline used in the study

The fault detection method was developed in Python using the Tree-based Pipeline Optimization Tool (TPOT), an automated machine learning tool that creates combinations, or pipelines, of data transformations and machine learning models using genetic programming. This tool helps simplifying some of the requirements of ML methods regarding the data quality: the variables (or features) may have to be modified in some way, e.g., scaling or introduction of polynomial features, and a well-performing predictor must be chosen [20]. Given a set of data without missing or mislabelled values, the TPOT automatically optimizes feature selection, feature pre-processing, feature construction, model selection, and parameter optimization. The last step of the ML process, the model validation, is carried out by the user.

To choose the best TPOT pipeline to be used, 16 different training sets and 16 different test sets were used to produce 16 different pipelines. These pipelines were then introduced to the same training set, to evaluate which pipeline to further test and evaluate in the study. Training sets and tests set are subsets of the input data, with the ration training over test set equal to 80/20. To evaluate the performance of the TPOT pipeline, the coefficient of determination (R^2) and the Mean Absolute Error (MAE) were used. Equation (3) shows how R^2 is calculated. The closer R^2 is to one, the better is the performance of the model.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{3}$$

Where:

SS_{res} = the sum of squares of the residuals

SS_{tot} = the total sum of squares, which is proportional to the variance of the data

The mean absolute error, MAE, is the averaged sum of the absolute value of the residuals between actual values and predicted values, and can be calculated using equation (4). For a well performing model the MAE is expected to be as close to zero as possible.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |(actual\ value)_i - (predicted\ value)_i| \tag{4}$$

By investigating the values of MAE and R^2 the most accurate TPOT pipeline, shown in Figure 5, was chosen.

After choosing the TPOT pipeline, a choice of the parameter to be predicted using the substation model must be made. For that, different combinations of input/output variables were tested. The performance of the models was evaluated using the R2 and MAE values. The results are shown in Tab.2.

Tab. 2 - Parameter combinations that were tested with the model. Combination number 5, highlighted in grey, obtained the best R^2 and MAE value.

comb	Input	Output	R^2	MAE
1	T_{out}, T_r, T_s, t	ṁ	0.9703	0.1337
2	$T_{out,24}$, Tr, Ts, t	ṁ	0.9555	0.2027
3	T_{out}, T_s, t, ṁ	T_r	0.8839	1.1091
4	$T_{out,24}$, ṁ, T_s, t	T_r	0.8903	1.10348
5	$T_{out,24}$, T_s, t, T_{out}	ṁ	0.9740	0.1301
6	$T_{out,24}$, T_s, t, T_{out}	T_r	0.8841	1.0555

7	$T_{out,24}$, T_s, t, T_r, $\dot{m}$ T_{out}	0.9296	0.5857

To test the fault detection capability of the model, artificial faults were introduced to the data set. The model outputs for the faulty data sets were then compared to the outputs for the well-performing, original data set by calculating the hourly and accumulated residuals (rolling window of the previous 24 hours) between the model outputs and the actual measurements from the customer installation. By comparing these residuals to each other, it was possible to investigate whether the model performance changed when introduced to a data set containing faults.

Two faults were induced in the dataset. The first one represents a loss of communication between the energy meter and the database of the DH utility. To handle this type of fault, the utilities usually replace the missing values by a constant value. In this case, a value of 60 ˚C was chosen and inserted randomly in the data set for the supply temperature. Since this was an extremely low value compared to the original dataset, another fault was also induced where the original value of the supply temperature was decreased by 10 %. The second type of fault was induced as a gradual change in output over a time period, representing a drifting meter. This was done for the outdoor temperature sensor as well as the supply temperature sensor. These two faults are a common occurrence in DH metering data. In what follows only the results for a drifting meter will be presented.

4.3 Results

Figures 6 and 7 shows two different residual plots for two of the faults investigated in the study: the drifting outdoor temperature sensor (Figure 6) and the drifting supply temperature sensor (Figure 7). The residuals for the data sets containing a fault are represented by the red line in the figures. The blue line in both figures represents the residuals between the real and the predicted values for the data set that did not contain a fault.

Figure 6.a) shows the residuals between the real values and the values that were predicted by the model. Figure 6.b) shows the cumulative sum over a 24 h interval of the same residual values. As may be seen in Figure 6.a), the model prediction changes when this fault is present in the data set, but the deviation is not clearly distinguishable. The cumulative sum of residuals in Figure 6.b) displays a clear deviation approximately one month after the fault was induced in the data, indicating that this method may be well suited for automated detection of a drifting outdoor temperature meter.

Figure 7.a) shows the residuals between the real values and the values that were predicted by the model. Figure 7.b) shows the cumulative sum over a 24 h interval of the same residual values. As may be seen in Figure 7.a), the model prediction does not change significantly when this type of fault is present in the data set. This may also be seen in Figure 7.b).

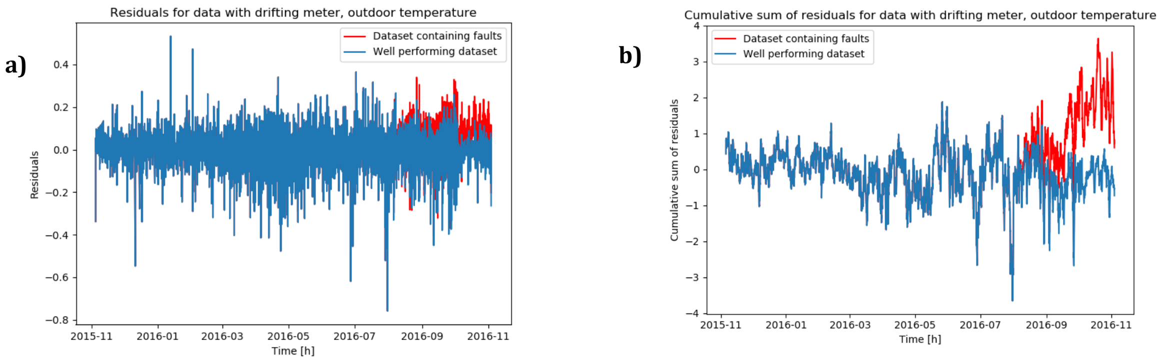

Fig. 6 – Analysis of the residuals between the real and the predicted values for the data set without a fault (blue line) and the data set with a drifting outdoor temperature meter (red line). **a)** Residuals between real and predicted values as a function of time. **b)** Cumulative sum of residuals as a funcion of time.

4. Discussions and conclusions

The results from the two fault detection approaches show that both can detect deviating behaviours in DH customer installations. The cluster-based method

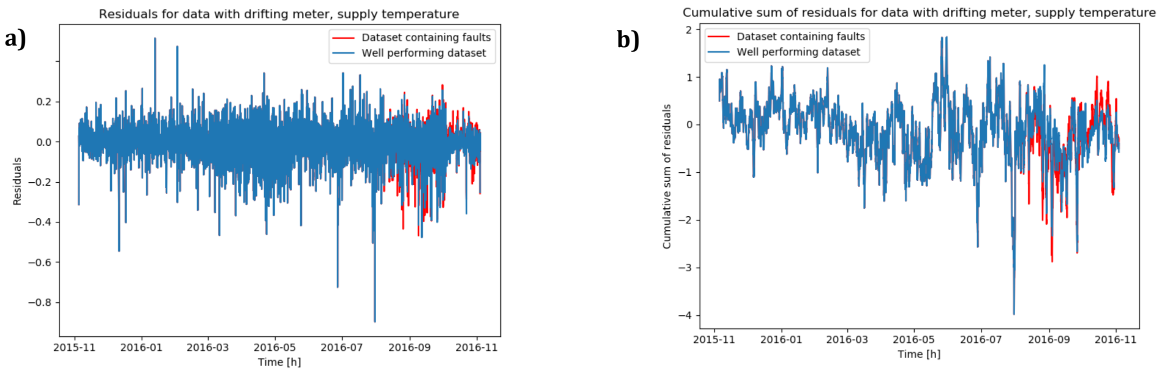

Fig. 7 – Analysis of the residuals between the real and the predicted values for the data set without a fault (blue line) and the data set with a drifting supply temperature meter (red line). **a)** Residuals between real and predicted values as a function of time. **b)** Cumulative sum of residuals as a function of time.

rapidly detects several installations that have a high overflow. Since a high overflow is an indication that the installation is not working as it should, this further implies that the cluster-based approach detects poorly performing installations.

For the instance-based approach, the model behaviour changes when a fault is present in the data, as may be seen in Figures 6 and 7. This indicates that the method can detect the deviations induced in the data set. However, when analysing the results, it can be concluded that different faults have different impacts on the model performance. This is especially clear for the two faults that are related to drifting meters. Although the fault is the same (a gradual increase of the meter readings over time), a drifting outdoor temperature sensor in the meter readings has a much larger impact on the model performance than when a drifting supply temperature sensor is present. This may be seen in Figures 6 and 7. In Figure 6.b), the drifting outdoor temperature sensor causes the cumulative sum of residuals for the data set known to contain a fault (red line) to deviate from the expected behaviour (blue line) soon after the fault has been induced in the data set. This shows that the model predictions have changed, thus indicating that a fault is present in the data set. When investigating the cumulative sum of residuals for the predictions of the drifting supply temperature sensor (Figure7.b)), it can be noticed that the deviation from the expected behaviour is not as large as for the data set in Figure 6. This indicates that the model used in the instance-based fault detection approach may not be able to detect *all* faults that are present in DH customer installations. However, this also indicates that the method may be suitable for fault diagnosis as it seems to be capable of distinguishing between different faults.

The two methods are based on two different fault detection approaches. The cluster-based approach is based on a reference case of several installations, to which the behaviour of each individual installations is compared. This means that the method can be generalized between different installations and different DH systems rather easily. The change that must be made when investigating a new DH system is to change the reference case that the fault detection signatures are based on. However, the generality of the method may be a problem since the current method does not differentiate between different types of customer installations which may have a wide variety of different behaviours in terms of heat use, e.g., if comparing a school building to a single-family household. One way to solve this problem is to introduce another, initial step in the fault detection method, right after the data validation step. In this step, the buildings would be divided into smaller groups or clusters, representing different types of buildings (in terms of heat use). The fault detection method would then be applied to each of these groups. Another improvement on the method would be to introduce extra Key Performance Indicators (KPIs) in addition to the absolute overflow calculated. For example, using a relative or an energy weighted overflow can enhance the accuracy of the method.

The instance-based approach uses the model of *one* installation. This means that one model for each installation in the DH system should be developed to obtain a well-performing fault detection method. However, this may be a very time-consuming task that would have to be repeated each time a new customer installation is installed in the DH system. One way to reduce the number of models needed could be to implement the same type of grouping of installation types as suggested for the cluster-based fault detection approach. The behaviour of each group would then be modelled in the way suggested in this paper. This approach would probably reduce the accuracy of the fault detection model, since the model of a *group* of installations will be less accurate for *individual* installations.

A natural step in the process of developing a fault detection method is to validate the results. However, obtaining such data sets from DH industry has been a challenge. Most DH utilities do not log when a fault has occurred in their systems, only when they have corrected a fault. Neither do they connect the records of the corrected faults to deviations in customer data. This may be referred to as the lack of labelled data. Labelled data may in this case be described as data where a specific fault is known to have occurred at a specific time, in a specific customer installation. The lack of labelled data has made it challenging to validate the developed fault detection methods properly. The lack of labelled data also introduced a problem that may be present for both the cluster-based and the instance-based approach: how to know that the data from the installations represent a well performing installation that can be either modelled or included in the reference case? Since no labelled data is available, it may be that some installations that contain a fault today are included in the reference case for the cluster-based approach. It is also possible that some installations seem to be well performing since no larger deviations are present in the data set. However, they may in fact contain a fault that has been present during a long time, which makes it look like their "faulty" behaviour is their normal behaviour. Using the instance-based approach on such data sets would generate an incorrect model of the expected behaviour of the installation.

Therefore, an important initial step to further improve the fault detection methods developed would be to identify several installations where specific faults are known to have occurred, at a specific time. It is also important to identify several installations that are known to *not* have had any faults during the same time. This would provide the possibility to validate the methods, and to make sure that they detect the installations that *actually* contained a fault during the investigated period of time. To also overcome the lack of labelled data, both methods can be used simultaneously by the district heating utility. Starting from historical data, the cluster-based approach can be used to gradually detect malfunctioning substations. It can be re-evaluated on a periodic basis, for instance monthly. The instance-based approach can be used in parallel. Given that the instance-based method uses historical data, it would not be known from the start if the behaviour modelled is for the well-behaving or faulty substation. For that, the cluster-based method can help. As a faulty substation is corrected, the model is recalibrated, and any further deviations can be an indication of a fault in the substation. Another point of improvement is how the deviating behaviours are detected for the instance-based approach. Currently, this is done by visual inspection of the residuals. To be able to use the instance-based approach on a larger scale, the identification of deviating residuals must be automated. This could be done by introducing, e.g., threshold values that the deviations may not exceed (in similarity to the threshold values used in the cluster-based approach). Overall, the two approaches show great promise for fault detection. The next steps include testing the methods on labelled data sets, and to develop a solution for fully automating the fault detection methods.

5. Acknowledgement

This work is part of the TEMPO project which is funded by the European Union's Horizon 2020 Programme under Grant Agreement no. 768936.

6. Data access statement

The datasets analysed during the current study are not publicly available because of privacy requirements.

7. References

[1] Lund H. et al. "4th Generation District Heating (4GDH) - Integrating smart thermal grids into future sustainable energy systems". Energy.2014;68:1–11.

[2] Kavvadias K., Jiménez-Navarro J., Thomassen G. Decarbonising the EU heating sector -Integration of the power and heating sector. EUR 29772 EN. Luxembourg: European Union, 2019.

[3] Lund H. et al. "Smart energy and smart energy systems". Energy.2017;137:556–565.

[4] Sandin F., Gustafsson J., Delsing J. Fault Detection with Hourly District Energy Data; Research Report 2013;27.

[5] Gadd H., Werner S. Achieving Low Return Temperature from District Heating Substations. Appl. Energy. 2014;136:59–67.

[6] Gadd H., Werner S. Fault Detection in District Heating Substations. Appl. Energy. 2015;157:51–59.

[7] Werner S. Feltyper i fjärrvärmecentraler strukturerade (Types of Faults in District Heating Substations Structured); Halmstad Högskola: Halmstad.2013. (In Swedish)

[8] Yliniemi K. Fault Detection in District Heating Substations. Licentiate Thesis, Luleå University of Technology, Luleå, Sweden.2005.

[9] Alsbjer M., Wahlgren P. Fjärrvärmecentraler 10 år—Håller de måttet? (District Heating Substations 10 Years—Do They Measure up?); Research Report 2011;9. (In Swedish)

[10] Frederiksen, S.,Werner S. District Heating and Cooling; Studentlitteratur: Lund, Sweden, 2014.

[11] Park YJ., Shu-Kai F., Chia-yu H. "A Review on Fault Detection and Process Diagnostics in Industrial Processes". Processes.2020.

[12] The European Commission. Annex to the Commission recommendation on the implementation of the new metering and billing provisions of the Energy Efficiency Directive 2012/27/EU no C(2019) 6631 final. 2019.

[13] Sandin F., Gustafsson J., Delsing J. Fault detection with hourly district data. 2013.

[14] Calikus E et al. "Ranking Abnormal Substations by Power Signature Dispersion". Energy Procedia.2018; 149: 345–353.

[15] Farouq S. et al. "Large-scale monitoring of operationally diverse district heating substations: A reference-group based approach". Engineering Applications of Artificial Intelligence 2020.

[16] Ingvarsson S. Statistical Modelling of Individual Substations in a District Heating System. 2018.

[17] Guelpa E., Verda V. Automatic fouling detection in district heating substations: Methodology and tests. Applied Energy, 2020;258.

[18] Garay R., Fester J., Mugaguren, L., Martinez R. RELaTED D2.4 – Energy flexibility and DH Control. 2019.

[19] Cogger K.O. Piecewise linear modeling: Theory, guidelines, and applications. In Management Science & Technology Symposium. 2006.

[20] Olson R., Urbanowicz R., Andrews P., Lavender N., Kidd L., Moore J. "Automating Biomedical Data Science Through Tree-Based Pipeline Optimization," European conference on the Applications of evolutionary computation. 2016.