Towards automated fault detection and diagnosis in district heating customers

generation and analysis of a labeled dataset with ground truth

Leiria, Daniel; Andersen, Kamilla Heimar; Melgaard, Simon Pommerencke; Johra, Hicham; Marszal-Pomianowska, Anna; Piscitelli, Marco; Capozzoli, Alfonso; Pomianowski, Michal Zbigniew

Published in:
Proceedings of Building Simulation 2023

DOI (link to publication from Publisher):
10.26868/25222708.2023.1576

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Towards automated fault detection and diagnosis in district heating customers: generation and analysis of a labeled dataset with ground truth

Daniel Leiria[*1], Kamilla Heimar Andersen[1], Simon Pommerencke Melgaard[1], Hicham Johra[1], Anna Marszal-Pomianowska[1], Marco Savino Piscitelli[2], Alfonso Capozzoli[2], and Michal Zbigniew Pomianowski[1]
[1] Aalborg University, Department of the Built Environment, Aalborg, Denmark
[2] Politecnico di Torino, Department of Energy, TEBE Research Group, BAEDA Lab, Turin, Italy

Abstract

This study aims to develop a framework for automated fault detection and diagnosis (AFDD) in district heating (DH) substations by comprehensively understanding typical faults. AFDD is presently dependent on manual detection and diagnosis, leading to limitations. To address this issue, the study utilized data from 158 fault reports and smart heat meter data from residential buildings in Denmark to investigate common faults and conduct a fault impact assessment. The study suggests additional indicators for use by DH utility companies to detect anomalies in the future. The findings indicate that greater attention to fault detection and diagnosis can decrease energy usage and return temperatures, demonstrating the significance of AFDD implementation.

Highlights

- The study analyzed 158 fault reports from DH systems, categorized according to the type of fault detected.
- A subset of 90 buildings with both faulty and non-faulty data points was used to benchmark different fault detection (FD) indicators.
- A novel FD indicator was proposed based on water volume and temperature difference in DH substations.
- DH system faults significantly impact energy use, and more efficient detection can reduce DH customers' energy usage on average by 14%.
- The proposed FD method shows promise in detecting anomalies for future use by DH utility companies and integration into an AFDD framework.

Introduction

District heating (DH) systems are becoming increasingly popular in many countries as an effective and centralized solution to provide heating to large agglomerates of buildings while reducing energy waste and integrating with other renewable energy sources. Nevertheless, like any complex system, DH networks are prone to faults that can lead to reduced energy efficiency and increased costs. By identifying and repairing these faults in DH systems, we can move towards more efficient and sustainable heating solutions, particularly in the context of 4th generation district heating (4GDH) systems prioritizing lower supply temperatures and higher energy efficiency (Li and Nord, 2018).

A successful transition towards 4GDH systems calls for the parallel transformation of the primary/supply and secondary/demand sides. Apart from the renovation of the building stock, the low-hanging fruits are the detection, diagnosis and fixing of the faults in the end-users' heating installations, which often cause high energy use and high return temperatures. According to Gadd and Werner (2015), from the analysis of a smart heat meters (SHM) dataset of 135 substations, 100 (74%) presented patterns that could indicate a faulty operation of the system. If this result is representative for the existing DH systems in operation today, it means that three-quarters of the heating distribution grid is underperforming. This presents a sizeable economic potential (approx. 0.05 to 0.5€/MWh·ºC) by fixing the existing faults in the DH end-user installations and consequently reducing the return temperature (Frederiksen and Werner, 2013; Gadd and Werner, 2014).

Background

There is significant research on fault detection and diagnosis (FDD) in DH systems, with studies focusing on developing automated algorithms to identify different symptoms and detect and diagnose faults in DH substations using SHM and other sensor data.

Gadd and Werner (2015) established a fault detection (FD) framework based on SHM and outdoor temperature measurements of a dataset of 135 Swedish substations. Månsson et al. (2019) presented a survey-based study to describe the current status of the Swedish DH utilities regarding FDD. The survey concluded that the common methods used by the utilities to evaluate the faulty customers are from within the Gadd and Werner (2015) framework. This FD framework finds customers with faulty substations through hourly SHM data when it is observed at least one of the three fault symptoms: a) *Unsuitable heat load pattern:* Any building with heating usage measurements different from what is expected from their occupancy profiles; b) *Low average annual temperature difference:* Any building with a yearly average ΔT significantly lower than 45ºC; c) *Poor substation control:* Any building with irregular energy oscillations and low correlation between heating usage and outdoor temperature.

The *overflow* indicator, presented by Frederiksen and Werner (2013), is also used by few utilities for fault detection purposes. However, even though these methods have proven useful in detecting faulty systems based on SHM measurements, they lack the knowledge of what type of faults the different indicators detect.

https://doi.org/10.26868/25222708.2023.1576

Also, based on Swedish DH systems and taking into account the FD framework mentioned above, Månsson et al. (2018a) developed a statistical algorithm to detect faulty substations based on their patterns of energy, temperature difference between supply and return DH water (ΔT), and return temperature signatures (i.e., variables' variation due to outdoor temperature). A similar approach was proposed by Calikus et al. (2018), which analyzed the energy signature generated by a robust linear regression model of the substations to detect abnormal data points that are significantly deviated from the linear trend. Both methods relied on threshold settings and statistical methodologies to detect anomalous data points (outliers), which are most likely due to faulty heating system operations caused by technical or occupant behavior (Schaffer et al., 2022). Focusing on specific faults in components of DH substations, Guelpa and Verdi (2020), developed and tested a methodology based on measurements in the primary and secondary circuits to detect fouling in the heat exchanger in some buildings connected to the DH network of Turin, Italy.

In the machine learning (ML) discipline, Månsson et al. (2018b) proposed an ML algorithm using gradient boosting regressor to model the flowrate of a well-performing substation based on the external temperature and ΔT to compare it with the same dataset with faulty data points induced by the authors. Xue et al. (2017) proposed a data mining algorithm consisting of data cleaning, clustering, and analysis of two substations in Changchun, China, to extract more information from SHM data. On a larger scale, Calikus et al. (2019) advanced the field of FDD in the DH systems by applying unsupervised clustering methodologies in a 1,385 Swedish buildings dataset to detect buildings with abnormal heating profiles, and after they investigated the reasons behind the anomalous measurement profiles.

The described methods above are significant steps towards detecting and diagnosing the various faults in DH substations more efficiently than today's process. However, to a certain extent, all these methods lacked adequately labeled data regarding the occurring faults, thus hindering the further development and testing of these FDD methodologies. As Månsson et al. (2021) outlined, labeled data with ground truth is a historical register of specific faults that occurred in a particular substation. Furthermore, according to the authors, these ground truth datasets must be gathered on a large scale and be combined with SHM data to understand the difference between "faulty" and "well-performing" heating installation operations.

Contributions

Some DH utilities apply the indicators proposed by Gadd and Werner (2015) framework for automated FD while a human user still carries out the diagnosis process. This is due to, as explained above, the lack of adequately labeled data with ground truth or an infrastructure to record and gather this information. This study attempts to investigate this matter by:

- Coupling approximately 351 fault reports (labeled dataset with ground truth) issued by technicians with measurements from SHM data.
- Providing an overview of typical faults occurring in DH end-user heating installations.
- Assess the efficacy of the current indicators for FDD purposes and propose a new one to assess the DH customers.
- Perform a fault impact assessment of the different faults to investigate their consequences on the buildings' energy usage.

Methodology

Dataset description

The dataset used in this study combines SHM data from 351 residential buildings connected to the DH grid with faults assessment reports made by technicians when visiting the installations in Aalborg, Denmark. The SHM dataset consists of the energy use, water volume, and ΔT measurements with an hourly resolution for 2022 for the heating demand of space heating (SH) and domestic hot water (DHW). Each building has an individual meter ID associated with the customer, and its fault assessment report is generated after a visit by a utility company technician. These visits were triggered when the substation's SHM registers ΔT-values below 10ºC. The weather data is extracted from the Danish Meteorologic Institute (DMI) portal. The selected weather station is Tylstrup, the station nearest to Aalborg, available in the DMI database.

In this work, the data pre-processing consisted of analyzing each report individually to categorize the different faults. Some of the reports were found to be ambiguous in terms of describing the real cause of the fault and were therefore disregarded from this study. Later in the study, it is only assessed the installations where it is indicated by the technicians that the fault was fixed after the visit. Therefore, from the initial dataset of 351 buildings, only 90 were used for this work. The data treatment of SHM measurements was performed by following the framework described in Schaffer et al. (2022). Customer installations differ based on country and DH system. Some countries have direct connections without hydraulic separation in the space heating system, while others have a prevalent indirect connection where the SH and DHW systems are separated from the primary circuit by heat exchangers (Figure 1). In this study, the heating installation was divided into two sections, the primary circuit (substation) and the secondary circuit (systems inside the household).

https://doi.org/10.26868/25222708.2023.1576

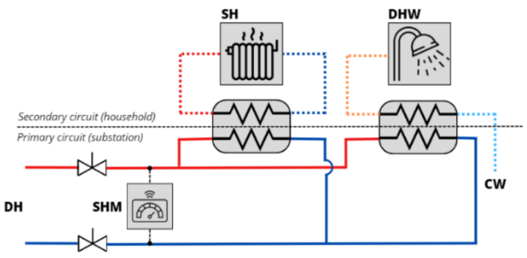

Figure 1: Illustration of a residential heating installation.

Definitions and FDD concepts

The immense body of literature on FDD employs several different naming conventions. In this study, the authors have adopted the ontology from Andersen et al. (2023) for definitions of a *symptom of a fault, fault, cause,* and *consequence* for the fault report analysis. Furthermore, labeled dataset with faulty ground truth is defined as annotations of each fault within the different defined fault categories presented in Table 1.

Furthermore, the framework proposed by Melgaard et al. (2022), originally developed for building systems, is applied in this study due to its overlap with the DH systems. The framework outlines three levels of FDD: fault detection, fault isolation and identification, and fault evaluation. Fault detection forms the fundamental basis of FDD, which involves determining the existence of a fault. Fault isolation and identification, in conjunction with fault detection, comprise the FDD process by pinpointing the fault location and identifying the cause of the fault. Finally, the fault evaluation stage assesses the impact of the fault, which can include estimating the excess energy or financial resources used due to the fault.

Applied FDD indicators

In this work, five different indicators are used to assess the installations' performance. Four of these are already in use by several DH companies, while the fifth indicator is a new one proposed. Moreover, to be able to compare the different buildings, the energy and volume measurements used in the indicators are divided by the building's heated floor area. Regarding the SHM data, the measurements have hourly resolution. However, the analysis was performed with a minimum resolution of one day due to measurement truncation problems raised when calculating the ΔT. Additionally, some of the indicators were assessed for the seasons of heating (from January to May and from September to December) and no-heating (from June to August). The indicators are the following:

Ind. 1: Overall substation's operation: Assessing the annual average measurements (energy, volume, and ΔT) to understand each customer's overall performance.

Ind. 2: Temperature difference intervals: Determining the number of days with different ΔT-values during the heating and no-heating seasons.

Ind. 3: Heating and outdoor temperature correlation (energy signature): Analysing the relationship between the daily heating demand and the outside temperature, as described in Gadd and Werner (2015).

Ind. 4: Overflow: Assessing the overconsumption (V_{over}) of the different buildings using equation 1 and considering the ΔT_{ideal} constant for all buildings equal to 45ºC (Gadd and Werner, 2014).

$$V_{over} = V_i - V_{ideal} = V_i - \frac{E_i}{\rho c_p \Delta T_{ideal}} \; [m^3/m^2] \qquad (1)$$

This work also proposes a novel indicator (equation 2) to identify faulty substations by calculating the daily ratio between the measured water volume per m² (V_i) and the measured temperature difference (ΔT_i) as a function of the daily average outdoor temperature (T_{out}).

Ind. 5: Volume-temperature ratio is the proposed indicator to assess the heating installation behavior, and it is calculated with equation 2:

$$V_{Temp} = f(T_{out}) = \frac{V_i}{\Delta T_i} \; [m^3/m^2\, ℃] \qquad (2)$$

According to the proposed indicator, a well-performing substation displays a ratio that changes linearly throughout the outdoor temperature conditions variation. Therefore, all data points that do not follow this linear profile are marked as faulty.

This research assesses how these indicators perform in analyzing the different symptoms measured by the SHM and attempts to correlate them with the faults identified during the visits to the installations. Furthermore, this study also endeavors to discuss the presented indicators' advantages and disadvantages and perform the fault impact assessment by estimating the energy savings obtained from the intervention made in the heating installations. The energy savings calculation is made through the difference in the sum of energy used before and after the visit divided by the building's heated area and the heating degree-days (in order to normalize it over the building size and outdoor temperature) and two case buildings are discussed using the indicator volume-temperature ratio.

Results and Discussion

Overview of the faults in DH substations

From a set of 351 fault assessment reports, only 158 reports of different residential buildings were used in the initial stage of the study. The visit to each installation was prompted because the SHM registered a ΔT below 10ºC, which according to the literature, is a symptom of an underperforming DH substation. Table 1 shows the types of faults, their frequency, and their identified causes. The categories are based on the ones proposed by Månsson et al. (2019).

Table 1: Types of faults and their description.

Fault categories	Nr. of faults	Cause of fault
Wrong general settings in the system	62 (39%)	Consists mainly of high settings in valves that control DHW heat exchanger/storage tank, SH system, etc.
SH system (Secondary side)	38 (24%)	Consisted mainly of defective components in SH

https://doi.org/10.26868/25222708.2023.1576

		systems inside the building (secondary side).
Control valves	31 (20%)	Consisted mainly of defective valves in the DHW heat exchanger/storage tank.
User behavior/practices	15 (10%)	Consisted mainly of sporadic usage of the SH systems by the occupants.
Controllers or sensors	7 (4%)	Consisted only of cases where the battery of the thermostat of SH systems in the building ran out.
DHW exchanger/tank	5 (3%)	Consisted only of defective DHW heat exchanger/storage tank.

Explicitly, the highest percentage of existing faults in the dataset are caused by high settings in the heating systems and defective components in the SH system indoors. This corroborates two important arguments in the field of FDD in the DH systems. Firstly, most of the existing faults on the end-user side can be easily solved by changing the systems' settings to better values representing a simple but rather impactful measure to increase the overall performance of the DH network (Gadd and Werner, 2015). Secondly, these values show that most faults occur on the secondary side of the heating circuit (i.e., inside the building instead of the substation). This implies that DH utilities need physical access to the customer's installation to fix most of its network faults. Utilities can accomplish this by establishing with the customers a service agreement on maintenance and repair of the heating installations, thus promoting good communication between the company and the customer and, in the long-term, maintaining a well-performing substation (Månsson et al., 2019).

After the intervention in each installation, some technicians were explicit in describing if the detected fault was solved. All the faults caused by occupants' abnormal use or high settings of the systems were fixed by the technicians, while the faults caused by a defective component that must be replaced were not fixed and must be addressed by a plumbing company hired by the customer. Figure 2, summarises the status of the heating installations after the visit, showing that 57% of the faults were solved, while only 27% could not be fixed by the technicians.

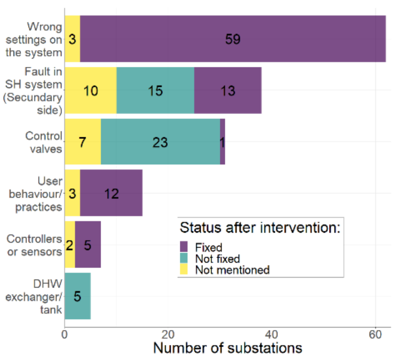

Figure 2: Type of each fault according to the installation's status after the intervention.

To summarize the described faults in the DH system at the end-user level, some labels are proposed in this study and applied further. Figure 3 shows these five proposed labels for each fault type.

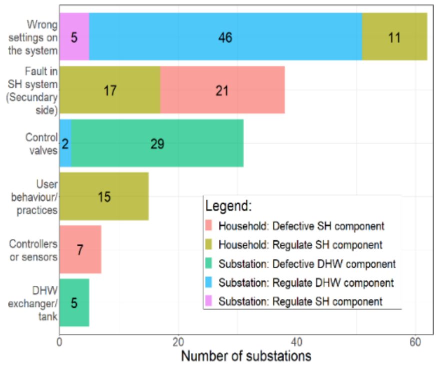

Figure 3: Type of each fault according to the proposed labels.

The labels are comprised of three parts. The first is whether the fault occurred in the central system (substation) or inside the building (household). The second part focuses on whether the fault refers to a broken component (defective) or the need to adjust the component's settings. The third part indicates which system the fault occurred in, either SH or DHW. This work also focuses in understanding which indicators are accurate in detecting and diagnosing the occurring faults in customer's level by assessing the system's symptoms through the SHM data. This task is performed by calculating and analyzing the indicators described in the *Methodology* section in the period prior and following the intervention.

FDD indicators performance assessment

Indicator 1: Overall substation's operation

The first tested indicator is the one used to understand the overall substation's operation. Figure 4, shows how the different fault labels are displaced according to the annual average volume and ΔT before and after the intervention. From these results, it is concluded that there is an overall increase of the ΔT and reduction of the energy and volume after the intervention – as expected.

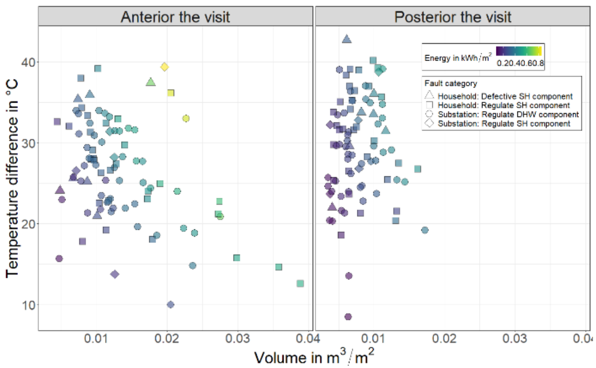

Figure 4: Representation of the annual average ΔT and water volume over energy usage and fault labels.

However, these results are as predicted when compared to the period before and after the intervention. There are a few remarks that must be highlighted regarding this indicator. As observed in the period before the intervention, a few buildings have good average measurements of energy, volume, and ΔT (right-top corner). Therefore, this indicator would not flag this group of buildings as faulty in the first instance. Another remark is the existence of buildings with low ΔT after the intervention that have low heating usage due to low water volume consumption (bottom-left corner). One of the reasons behind this symptom is the DHW settings that were readjusted after the visit to higher values by the occupants. Another drawback of this indicator is that it must have a large sample of measurements for the average values of the variables to be representative. Therefore, this indicator is relevant to detect faults for extreme cases (i.e., buildings with extremely low/high energy, volume, and ΔT-values) but might not be enough for detecting faults that occur sporadically during a few days of a year (as it happens in occupancy-based faults and high systems' settings).

Indicator 2: Temperature difference intervals

Focusing on the temperatures measured by the SHM, it is investigated the density distribution of the daily ΔT-values for the heating and no-heating seasons before and after the intervention in the heating installations by the graphs in Figure 5 and Figure 6.

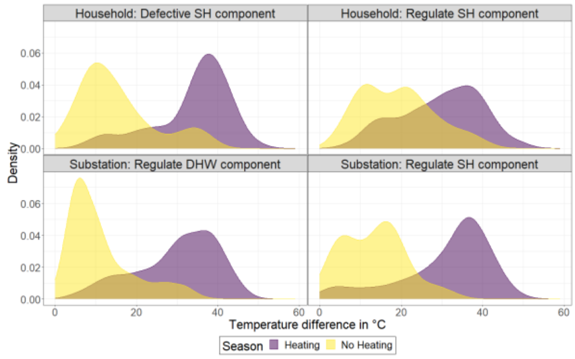

Figure 5: Density of ΔT during heating and no-heating seasons prior to the intervention.

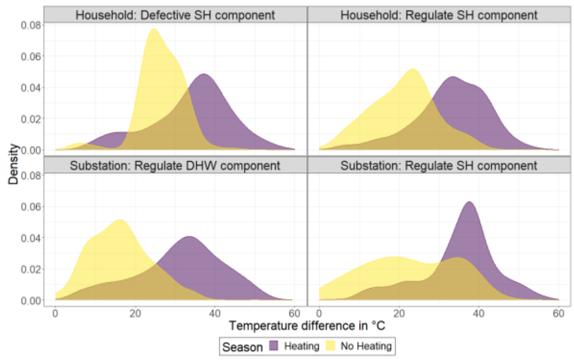

Figure 6: Density of ΔT during heating and no-heating seasons following the intervention.

According to the results, it is observed that all fault labels are characterised by lower daily ΔT-values for both seasons before the intervention. The main difference between both seasons is that during the heating season, the ΔT is higher than for warmer months. The main reason is that the SH and DHW systems are operating simultaneously during the heating season, while during the no-heating season, the heating demand is predominantly due to DHW usage. Therefore, it is also seen that the faults due to DHW decrease even more the ΔT measurements throughout both seasons, but mainly during the warmer months. It is also seen that after the visit, the overall ΔT increases for both seasons in all cases. However, following the intervention and the fault fixed, some buildings still have small ΔT, meaning that this indicator alone is not enough to assess the heating installation performance because it might highlight well-performing substations as faulty.

Indicator 3: Heating and outdoor temperature correlation

Gadd and Werner (2015) described that poor energy correlation with the outside temperature indicates poor substation control. According to Figure 7 and Figure 8, it is observed that after the visit, the correlation between the heating demand and the outdoor temperature increases, proving that this correlation is, in fact, related to the well or poor function of the substation.

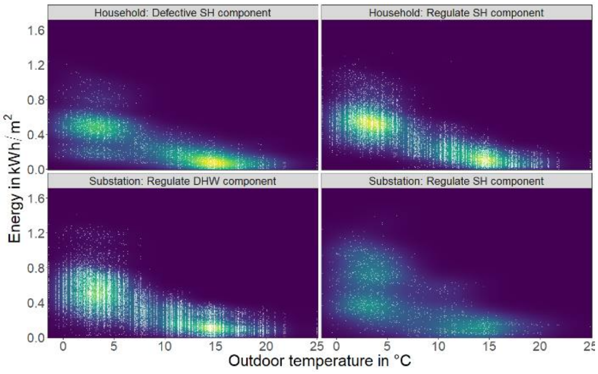

Figure 7: Energy signatures per fault category prior to the intervention.

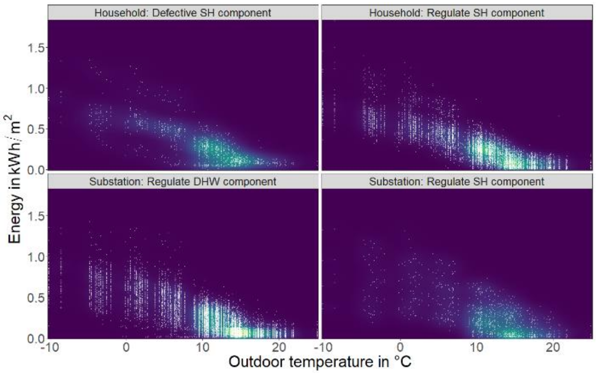

Figure 8: Energy signatures per fault category following the intervention.

For a better overview of the correlation increase after the visit, the coefficient of determination (R^2) obtained from the energy signature fitting was determined for each building before and after intervention for the heating and no-heating season (Figure 9).

https://doi.org/10.26868/25222708.2023.1576

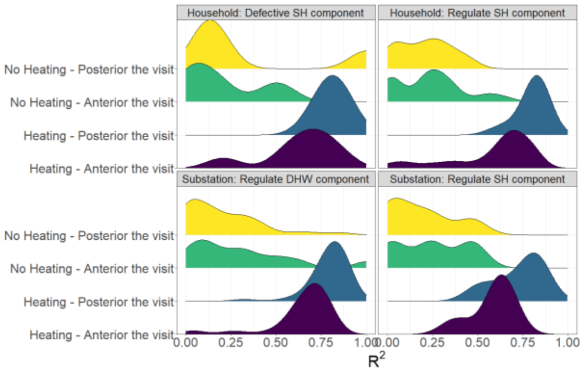

Figure 9: The R^2 of the energy signatures of each building per fault category for the different seasons prior to and following the intervention.

In Figure 9, one can see an increase of R^2 after the intervention for the heating season, while in the no-heating season, the R^2 tends to decrease toward zero. It is also observed that the correlation is not only dependent on the substation level but on the components in the building (e.g., underfloor heating control) because in all fault labels, a significant change in the R^2 is observed in both seasons after the visit. Therefore, it can be concluded that a well-performing system must have R^2 closer to one during the colder months while an R^2 closer to zero when only the DHW is mainly operating in the no-heating season. If this is not observed, it shows that the weather conditions are not the main driver of the substation's performance, but also a faulty component or occupants controlling the system poorly.

Even though this indicator is relevant for detecting faulty heating installations, it also requires a large sample of data points to describe this correlation accurately. Therefore, this indicator can only be used for long-time periods of collected data but not for short-term measurements. Another essential factor to consider for this indicator is that it can cause false alarms while attempting to detect faults due to the correlation being affected by two systems (SH and DHW) operating simultaneously. This is seen in the no-heating season, where few buildings have an $R^2 >$ 0.5 after intervention due to the SH still operating (e.g., bathrooms with underfloor heating or water-heated towel dryers operating during summer).

Indicator 4: Overflow

The overflow indicator can be calculated as the difference between the measured volume by the SHM and the ideal volume when considered an ideal ΔT (usually 45 ºC). From all indicators above, the overflow does not need large data samples to be calculated, making it a great performance value for short-term measurements. As one can see in Figure 10, the overflow of four different buildings is calculated over time. From the results, it is clear that the overflow is higher for faults involving DHW systems when this system is mainly operating (non-heating season). While it is observed when the fault is due to a defective component, the overflow has a rapid large spike over a short period of time.

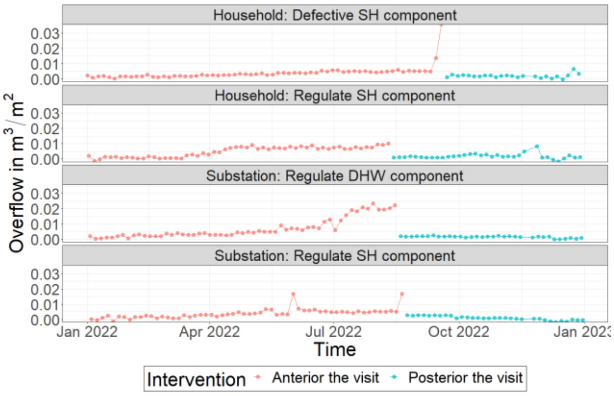

Figure 10: Overflow over time of four building cases per fault category prior to and after the intervention.

As expected, it is observed that there is a significant reduction of the overflow when a fault is fixed. This reduction occurs because the water volume measured in the building decreases while the ΔT increases. As seen, the overflow indicator is effective for DH customers' analysis for large and small sample measurements. Additionally, it can evaluate the installations and rank them based on their overflow (Månsson et al., 2019). However, this indicator has two drawbacks when applying it. Firstly, it is a value highly dependent on the ΔT variation without considering possible changes in the volume that are not dependent on ΔT. This is due to equation 1, where the ideal volume is calculated only considering an ideal ΔT. The second drawback is the predefinition of an ideal ΔT as a constant value. This preselection may hinder the comparison between buildings where their ideal ΔT might be different due to their location in the network, and for the same reason, buildings with lower ideal ΔT may be accounted as faulty when in reality, their substations are performing well.

Indicator 5: Volume-temperature ratio

Because of the necessity of comparing several buildings in the network and knowing that each building might have different ideal standards of volume and ΔT, this study proposes a new indicator to be used by the DH companies when assessing their customers. The indicator is called *volume-temperature ratio* and is based on the fact that there is a direct proportion between the volume usage and the ΔT, which is linear throughout the outdoor temperature variation when the substation is well-performing. This relation can be observed for four different building cases in Figure 11.

https://doi.org/10.26868/25222708.2023.1576

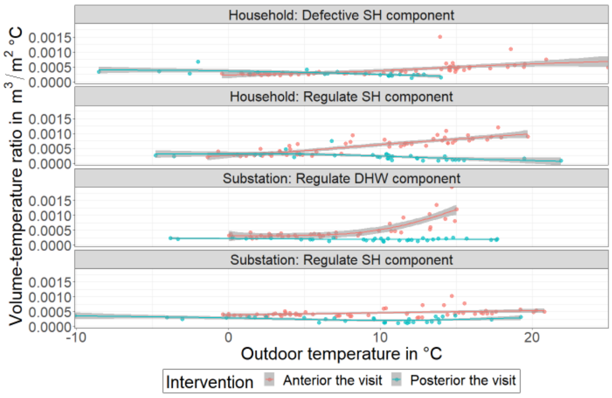

Figure 11: Volume-temperature ratio over outdoor temperature of four building cases per fault category prior to and after the intervention.

Figure 11 shows that for all the assessed buildings, the ratio of volume and ΔT should be constant for well-performing substations, regardless of the system. Besides, it is also noted that there are no significant differences between heating and no heating seasons. To better overview the values calculated from this ratio, Figure 12 shows the distribution of the volume-temperature ratio for each fault label before and after the visit. Compared to the overflow, the volume-temperature ratio attempts to solve the drawbacks of the former indicator while maintaining its benefits. Because there is no requirement to establish an ideal volume or ΔT, the ratio can be applied to compare all the buildings regardless of their location in the grid. It also, as the overflow, can be used to detect faulty singular data points without needing a large data sample.

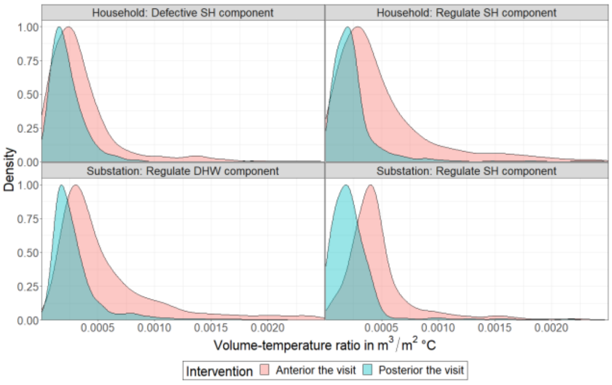

Figure 12: Density of the volume-temperature ratio for 158 buildings dataset prior to and following the intervention.

Energy savings due to FDD intervention

As previously shown above, many are the types of faults that can occur in the heating installations in a household connected to the district heating leading consequently to significant energy bills. From the initial dataset of 351 buildings, in 90 of them, it is mentioned in their reports that the fault was fixed by the technician or by the customer himself. Of these 90 buildings, 59 of them have positive energy saved after the intervention (refer to Figure 13), with an average energy saving of 14%, while 31 buildings do not report energy savings (negative values)

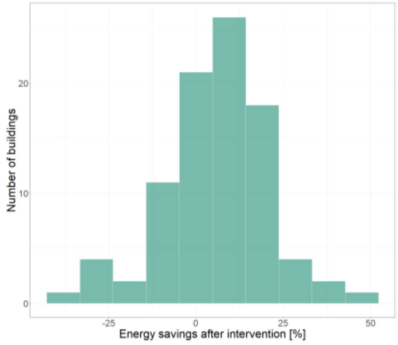

Figure 13: Energy saving after the intervention.

The main reason behind these negative values is due to some of the technical interventions were not enough to improve the system's performance, e.g., occupants continuing using the system suboptimal after the intervention, or operation changes.

It is selected two buildings where one had a large saving post-intervention (+25%), and the other displayed negative energy savings after the technician's visit (-30%). Both cases are analyzed with the indicator volume-temperature ratio, as one can see in Figure 14 and Figure 15.

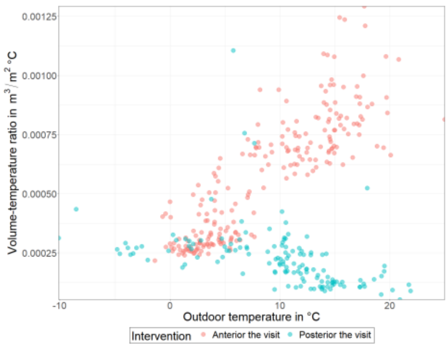

Figure 14: Volume-temperature ratio of a building with positive energy savings.

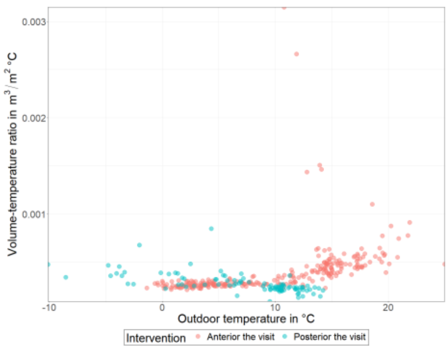

Figure 15: Volume-temperature ratio of a building with negative energy savings.

The first case pertains to a building where an incorrect setting was detected in the SH system, while the second case involved a defective radiator thermostat valve. These instances highlight the significance of identifying and diagnosing faults to ensure the efficient and safe operation of buildings and their systems. It was observed that after the fault being repaired, the data points conform to a linear trend, as expected for this indicator. Notably, the significant difference in energy savings between the two

https://doi.org/10.26868/25222708.2023.1576

cases is primarily attributable to the first case having a significantly greater number of data points that deviated from the trend compared to the second. This indicates that certain faults in DH customer installations may persist over extended periods while others may be momentary. Thus, the fault impact assessment per building must be performed on a data point basis rather than by time periods (e.g., pre- and post-intervention).

Conclusion

This study has demonstrated the importance of fault detection in DH systems and its potential impact on the buildings' energy usage. Through the analysis of 158 fault reports, it was concluded that the faults in the highest amount concerned high settings on the system by poor regulation or occupants' practices and defective SH system components indoors. This study also benchmarked different FD indicators using a subset of 90 buildings with faulty and non-faulty data points. A novel FD indicator based on water volume and temperature difference in DH substations was proposed, showing promise in detecting anomalies for future use by DH utility companies and integration into an automated fault detection and diagnosis (AFDD) framework. The findings suggest that DH system faults significantly impact energy use, and more efficient FD has the potential to reduce energy usage by customers by an average of 14%.

Suggestions for further work

There are numerous potential directions for further research in FDD in DH systems based on the findings of this study. One possible avenue is to explore the use of ML supervised algorithms to train classification models to analyze the relation of the volume-temperature ratio with its magnitude and occurrence period to diagnose different types of faults at the DH customer's level. Furthermore, researchers could analyze these indicators (e.g., energy signature, volume-temperature ratio, etc.) in the different existing heating system solutions implemented in buildings worldwide (e.g., direct/indirect connection, storage/instantaneous system, underfloor/radiator SH systems, etc.), and further on explore the potential for integrating FDD with other smart technologies in DH systems. However, to continue with this work, the DH companies must make a more extensive effort to collect good quality datasets with ground truth to progress in developing automated and implementable FDD methodologies.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 893945 (E-DYCE). This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 958345 (PRELUDE). The authors would like to acknowledge Aalborg Forsyning for their support in data collection.

References

Andersen, K. H., Melgaard S. P., and Leiria D. (2023). *Summary of Existing FDD Frameworks for Building Systems*. DCE Technical Reports . Department of the Built Environment, Aalborg University (DK).

Calikus, E., S. Nowaczyk, A. Sant'Anna, and S. Byttner (2018). Ranking Abnormal Substations by Power Signature Dispersion. *Energy Procedia 149*.

Calikus, E., S. Nowaczyk, A. Sant'Anna, H. Gadd, and S. Werner (2019). A data-driven approach for discovering heat load patterns in district heating. *Applied Energy 252*.

Frederiksen S., Werner S. (2013). *District Heating and Cooling*. Studentlitteratur AB. Lund (SE).

Gadd, H., and S. Werner (2014). Achieving low return temperatures from district heating substations. *Applied Energy 136*.

Gadd, H., and S. Werner (2015). Fault detection in district heating substations. *Applied Energy 157*.

Guelpa, E., and V. Verda (2020). Automatic fouling detection in district heating substations: Methodology and tests. *Applied Energy 258.*

Li, H., and N. Nord (2018). Transition to the 4th generation district heating - possibilities, bottlenecks, and challenges. *Energy Procedia 149*.

Melgaard, S. P., K. H. Andersen, A. Marszal-Pomianowska, R. L. Jensen, and P. K. Heiselberg (2022). Fault Detection and Diagnosis Encyclopedia for Building Systems: A Systematic Review. *Energies 15*(12).

Månsson, S., K. Davidsson, P. Lauenburg, and M. Thern (2018a). Automated Statistical Methods for Fault Detection in District Heating Customer Installations. *Energies 12*(1).

Månsson, S., P. J. Kallioniemi, K. Sernhed, and M. Thern (2018b). A machine learning approach to fault detection in district heating substations. *Energy Procedia 149*.

Månsson, S., P. J. Kallioniemi, M. Thern, T. V. Oevelen, and K. Sernhed (2019). Faults in district heating customer installations and ways to approach them: Experiences from Swedish utilities. *Energy 180*.

Månsson, S., I. L. Benzi, M. Thern, R. Salenbien, K. Sernhed, P. J. Kallioniemi (2021). A taxonomy for labeling deviations in district heating customer data. *Smart Energy 2*.

Schaffer, M., T. Tvedebrink, and A. Marszal-Pomianowska (2022). Three years of hourly data from 3021 smart heat meters installed in Danish residential buildings. *Scientific Data 9*(420).

Xue, P., Z. Zhou, X. Fang, X. Chen, L. Liu, Y. Liu, and J. Liu (2017). Fault detection and operation optimization in district heating substations based on data mining techniques. *Applied Energy 20.*

https://doi.org/10.26868/25222708.2023.1576

https://doi.org/10.26868/25222708.2023.1576