

Fault Detection and Condition Monitoring in District Heating Using Smart Meter Data

Felix Theusch¹, Patrick Klein¹, Ralph Bergmann^{1,2}, Wolfgang Wilke^{1,3}, Wolfgang Bock³ and Adrian Weber³

¹ *Business Information Systems II, University of Trier, 54296 Trier, Germany*
{theusch, klein, bergmann, wilke}@uni-trier.de

² *German Research Center for Artificial Intelligence (DFKI), Branch University of Trier, 54296 Trier, Germany*
ralph.bergmann@dfki.de

³ *msg systems AG, 50354 Hürth, Germany*
{wolfgang.wilke, wolfgang.bock, adrian.weber}@msg.group

ABSTRACT

Currently, large amounts of smart meter data are mainly used for billing purposes only, although they could also be valuable for decision support and business process optimization in customer service and maintenance. Therefore, this paper presents several relevant use cases for prognostics and health management based on a case study of a real meter data set of a medium-sized geothermal district heating network in southern Germany. First, we show the implementation of a machine learning algorithm for automatic fault detection based on cluster analysis and regression. Thereby, the district heating substation's control behaviour is learned and deviations due to a malfunction or failure can be detected before the customer notices them. In addition, we discuss the usefulness of two key performance indicators (return temperature and supply-return temperature difference) that can be computed relatively simple but resulting in very effective insights for condition-based maintenance and identifying substations with highly negative effects on the overall network. Our findings' correctness and usefulness were verified by the corresponding domain experts of the geothermal district heating company. Finally, we provide an outlook on smart meter data's role for the further development of intelligent district heating networks and the realization of highly complex approaches such as smart grids. To foster future research, we provide exemplary our RapidMiner processes.

1. INTRODUCTION

The anthropogenic climate change is the most pressing challenge of our time and is now showing itself in increasingly

Felix Theusch et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

dramatic effects e.g. rising seas, melting ice or extreme weather events. This development is caused by the steady increase of greenhouse gas emissions since the beginning of industrialisation. According to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), indirect CO₂ emissions for the generation of electricity and heat for building use accounts for 12% of global greenhouse gas emissions (Pachauri et al., 2014). Against this background, it is necessary to convert the heat supply to sustainable and resource-saving forms of energy.

The supply of heat to individual buildings by a central power plant is generally referred as district heating and contributes to the heat supply in almost all regions of the world. In many countries of Central, Northern and Eastern Europe, it is one of the most important types of heating, for example in Scandinavia, where its share is partly over 90% (Sayegh et al., 2017). In particular, by using non-fossil energy sources, for example geothermal energy (Rybach, 2003) or waste heat from industrial processes, district heating systems can make a significant contribution to reduce greenhouse gas emissions.

To further increase the acceptance and distribution of district heating networks (DHNs), they need to be operated efficiently and economically. Therefore (Lund et al., 2014) have defined the concept of the *4th Generation District Heating* (4GDH). 4GDH networks rely as much as possible on sustainable energy generation and highly efficient heat distribution. To achieve this, the main challenge is to reduce the network's temperature level drastically compared to today's systems. Especially, district heating substations (DHS), where heat is transferred from the DHN to the building's own heating system, can make a substantial contribution to this. However, faulty components can cause significant performance drops or the breakdown of DHS and must therefore be iden-

tified as soon as possible.

Since recently all DHSs will be equipped with smart meters that send data to the network operator for billing purposes, this paper examines whether and how these data can also be used for condition monitoring and network optimization. Our main contribution is the implementation of a fault detection approach that extends existing solutions with automated detection, as well as the application of two key performance indicators to determine inefficient and potentially faulty DHSs. Moreover, our findings are evaluated by means of case studies and the results are verified by several technical experts of the network operator. Furthermore, our implementation used for analysis is publicly available in the form of RapidMiner workflows at: <https://github.com/FeTheu/RM-fault-detection-in-dhs>.

The paper is structured as follows: we first describe the essential basics of district heating substations and smart meters in Sect. 2 and summarise selected related work. The focus of this paper in Sect. 3 is to demonstrate different data-driven use cases for fault detection and condition management using a real smart meter dataset of a medium-sized district heating network in Southern Germany.

2. FOUNDATIONS AND RELATED WORK

2.1. District Heating Substation

Within the district heating substation (DHS), heat is transferred from the district heating network (DHN) to the heating systems on the building side. DHSs consist of a number of different components: The heart of any DHS is the *district heating controller*, which monitors and controls connected sensors and actuators, e.g. the *control valve* that regulates the flow and thus indirectly the heat consumption. The *heat exchanger* is used to transfer heat from the district heating pipe to the customer’s internal network. If the heating system consists of several heating circuits, a heat exchanger must be installed in each of them. To protect the heat exchangers and valves, *strainers* filter the water on both sides (Skagestad & Mildenstein, 2002). The *heat meter* records the consumed heat energy W_{th} for billing purposes with the customer as follows:

$$W_{th} = \rho \cdot V \cdot c \cdot \Delta T, \quad (1)$$

where V is the measured flow, $\Delta T = T_{ps} - T_{pr}$ is the temperature difference between supply T_{ps} and return flow T_{pr} , ρ is the density and c the heat capacity of the network’s liquid (typically water).

It is important to monitor the current condition of a DHS since malfunctioning components can lead to incorrect billing and waste of energy. (Sandin, Gustafsson, Delsing, & Eklund, 2012) identified valves, flow meters and temperature sensors as sources for potential failures. Moreover, faults are commonly caused by humans through incorrect installations, con-

figurations (of meters and control system) or during maintenance as well as intentionally by customers. Furthermore, degradation can lead to leakage in heat exchangers and pipes.

2.2. Smart Meter

Smart meters (SMs) record the consumption of electricity, gas, water or heat and are integrated through certified gateways into a network to share data in both directions. In addition to *customer billing*, smart meter data offers further areas of application such as *load analysis*, *load forecasting*, and (*peak*) *load management*. The ability to enable communication between energy suppliers and customers make SMs a central component in smart grids and enable novel innovative services such as variable tariffs or the optimisation of the entire DHN, for example by coordinating decentralised energy storage (Wang, Chen, Hong, & Kang, 2018). Moreover, with the amendment of the European Energy Directive in 2018, the European Union decided on a concrete timeline for the roll-out of the SM infrastructure in Europe. Since 2020, all newly installed district heating meters must be remotely readable and all existing meters must be retrofitted by 2027 (European Union, 2018).

The role of SMs is seen as a cornerstone of future *smart grids* and provide essential information about energy consumption for *smart home* applications (Jahn et al., 2010). Despite the different settings, the advantages and disadvantages of smart meters in the application fields of electricity, gas or heat are quite similar. Some of the key benefits of smart meters from the perspective of the customer and district heating provider are summarised in Table 1, based on (Sun et al., 2015) and (Zheng, Gao, & Lin, 2013).

Table 1. Benefits of smart meter

Stakeholder	Benefits
Customer	Higher transparency
	Variable tariffs
	Timely and more detailed consumption information
	Smart home enabler
	Lower energy costs
Network Operator	Reduced metering costs
	Improved load management and peak load reduction
	Automated billing processes
	Remote diagnosis
	New business models

2.3. Gathered Data

The DHN considered in this study was already equipped with modern SMs and a central meterdata management (MDM)

system during its development phase. All DHSs are equipped with a Modbus-capable controller that collects the data from the controller itself and from the heat meter. These data are retrieved at 3-minute intervals via the gateways.

The analysed data set is a subset from the MDM of a medium-sized southern German geothermal DHN with over 2000 connected office and household buildings. It includes data from 896 DHSs over a heating period (04.11.2019 to 31.03.2020). For each connected heating circuit (e.g. one circuit for heating and one for domestic hot water), different *actuator*, *parameter*, and *sensor* attributes are written in the MDM. Typical actuator values are the signal for the control valve position or the operating mode of the circulation pump (active/inactive), which are steered by the controller. Tab. 2 shows the sensor values used in the analyses during Sect. 3 and their notation.

Table 2. Selection of available sensor values

Feature	Notation	Unit
Power consumption	P	kW
Flow	V	m ³ /h
Primary supply temperature	T_{ps}	°C
Primary return temperature	T_{pr}	°C
Outdoor temperature	T_{out}	°C

Since the DHN's data acquisition was just launched, the recorded data still revealed some issues. The most common problem, which can be observed in different intensity in almost all DHSs, is an interrupted data transmission to the MDM. For example, in 137 out of 896 DHS, less than 90% of data is transmitted to the MDM via the gateway. This results in time series with missing and unequally sampled observations, which needs to be considered when choosing the algorithms for fault detection and is discussed in Sect. 3.1.

2.4. Related Work

Although the data of the MDM was designed especially for monitoring and billing energy consumption, related work e.g., (Seem, 2005), (Sandin, Gustafsson, & Delsing, 2013) have shown its suitability for fault detection and diagnosis. A proven approach to fault detection in heating, ventilation and air-conditioning (HVAC) is the detection of deviations from the regular intraday and intraweek load cycles. (Seem, 2005) and (Li, Bowers, & Schnier, 2009) therefore develop cluster analyses to classify the households with regard to their intraday and intraweek cycles. (Kiluk, 2012) further uses data on the size of the heated building area for a cluster analysis in combination with a piecewise regression. Through k-means clustering and logistic regression, (Gianniou, Liu, Heller, Nielsen, & Rode, 2018) segment meter data from single-family households into different consumption groups. (Xue

et al., 2017) combine clustering with association analysis and identify faulty DHSs and those with significant optimisation potential. Unlike the previous studies, (Månsson, Kallioniemi, Sernhed, & Thern, 2018) use the volume flow as the target variable in conjunction with a gradient boosting regressor. The analysis pipeline is additionally optimised by a tree-based pipeline optimisation tool (TPOT). This automatically generates different input constellations and evaluates which one provides the best results. To identify faulty DHSs, (Gadd & Werner, 2014) use the temperature difference between flow and return depending on the outdoor temperature. However, this approach can only be used at lower outdoor temperatures up to approx. 10 °C, as above that the domestic hot water heating becomes a dominant factor. In summary, the most approaches apply cluster analysis combined with regression analysis to learn the normal behaviour and use deviants to detect faults. However, to the best of the authors' knowledge, no regression analysis approach has yet been presented in previous work on how these deviants can be detected automatically.

3. CASE STUDIES ON FAULT DETECTION AND CONDITION MONITORING

The aim of our study is to demonstrate possible applications of smart meter data with regard to advanced maintenance in the field of district heating. The methods presented here can be implemented using only standard energy metering data and do not require any additional information. We first describe a machine learning (ML) approach to detect failures and malfunctions of the components of a DHS mentioned in subsection 2.1 and then discuss possibilities of condition monitoring.

3.1. Fault Detection

The heat demand of a DHS (measured via P) depends on several factors from which the outside temperature T_{out} as well as the individual user behaviour are the most dominant ones. Since user behaviour cannot be measured directly, the heat demand is subject to strong fluctuations in the course of the day and week. For this reason, the goal of the following ML approach is to learn a regular load profile that represents the normal behaviour and then to detect automatically deviations due to a malfunction or failure.

3.1.1. Approach

As shown in Fig. 1, the machine learning pipeline splits into four separate steps.

1) Preprocessing and transformation

The goal of preprocessing and transformation is to generate a suitable data set for the further analysis steps. Initially, an *outlier detection* using k -nearest neighbour (k -NN) classification is applied to the outdoor temperature T_{out} and con-

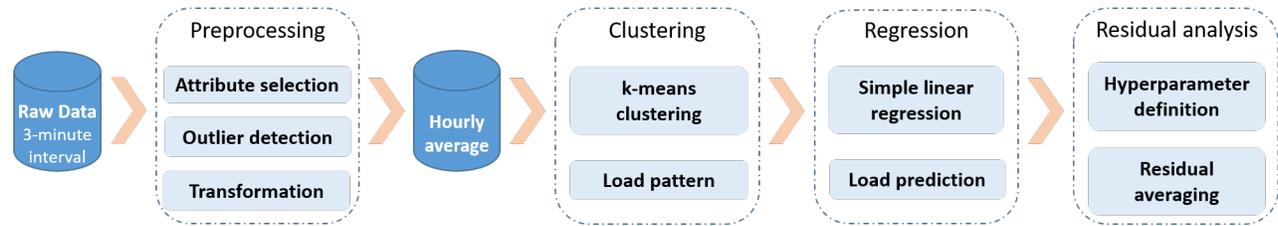


Figure 1. Illustration of the machine learning pipeline

sumed power P . Since both values should be in correlation, the outliers are eliminated on the basis of their Euclidean distance to the k -NN. In addition to outlier detection using k -NN, simple threshold values for temperature sensors are also used, for example by assuming that the supply temperature cannot rise above 110 °C and all values above this must be measurement errors. A replacement strategy for deleted observations is not necessary, because our approach does not require a time series sampled in equal intervals as input. Moreover, to reduce the computational effort and counteract fluctuations in the transmission rate between Gateway and MDM, we transform the sensor entries to *hourly averages*.

2) Cluster analysis

Cluster analysis is a common analysis approach for identifying time-dependent heat load patterns in HVAC systems. (Seem, 2005) and (Seem, 2007) as well as Li et al. (Li et al., 2009) apply it to daily averages to identify weekday cycles. Sandin et al. (Sandin et al., 2013) extend this idea to hourly recorded meter data, which is also the approach we have chosen.

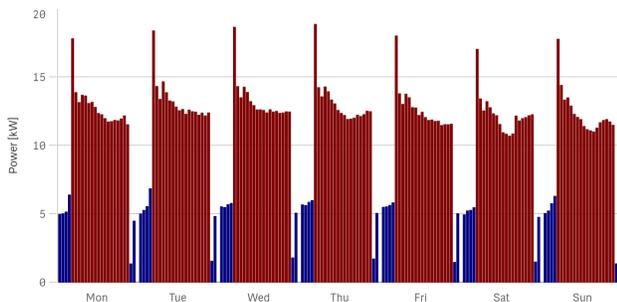


Figure 2. Heat load pattern with $k = 2$ for high (red) and low (blue) heat load

The aim of the cluster analysis is to identify an individual *heat load pattern* of a DHS on the basis of the historical records by assigning each time point (hourly) of a weekday to a cluster. For separation, we use *k-means clustering*, in which each observation of a DHS is divided into k clusters to obtain the most compact and separated clusters. This procedure is applied to P and the Euclidean Distance is used. The optimal number of clusters k may vary depending on the individual substation. The reference literature usually uses two (high and low load) or three (high, low and medium load) clusters

and defines k a priori. Similar to Xue et al. (2017) we do not strictly specify the number of clusters, but decide it based on the *Davies-Boulding Index* (Davies & Bouldin, 1979) individually for each DHS in order to obtain an optimal value for k .

In general, cluster analysis assumes that the behaviour of a DHS follows a regular pattern, which means that observations at a particular time but on a different week can be assigned to a cluster as clearly as possible. As a measure for the regularity of the heat load cycles, the *cluster consistency* is defined in the context of this work, which calculates how often a certain timestamp of the majority cluster was assigned compared to the total number of data points of this timestamp. For example, if 18 of 20 data points are assigned to cluster A (which therefore forms the majority cluster), the cluster consistency of this timestamp is 0.9. For instance, Fig. 2 shows hourly averaged heat load P for each day with $k = 2$.

3) Regression analysis

The cluster analysis groups different times of day based on their historical consumption. Using a *simple linear regression* for each of the previously defined cluster, the usually demanded thermal energy can now be predicted as a function of the outdoor temperature by calculating the linear dependence of the independent variable T_{out} on the dependent variable P . In this way, it is modelled that the heat demand increases with decreasing outdoor temperature because the indoor temperature is to be kept constant (usually 20-22 °C).

Compared to similar implementations in other DHNs, the simple linear regression already produces adequate results in our studies. This is because our data was recorded exclusively during the heating season. To model all year round dependency between outdoor temperature and heat load, a *piecewise linear regression* should be used to consider the dominating influence of tap water heating with increasing outdoor temperatures (Sandin et al., 2013).

4) Residual analysis

Using the regression analysis, a model that represents the DHS's regular behaviour was learned for each of the clusters. By applying a suitable residual analysis, deviations caused by a fault or failure must now be recognised as an alert. To the best of author's knowledge, previous work is limited to the manually visual detection of deviations and did not discuss

how to automatically analyse the residuals.

A DHS is identified as potentially faulty if a relatively large number of consecutive observed values have relatively high residuals measured by the error term ε . The error term's behaviour is heteroskedastic, which means its variance decreases with increasing T_{out} , i.e. with higher values of the abscissa. For example, an error term of 2 kW is more significant at an outdoor temperature of 15 °C than at 5 °C, since less energy is consumed at higher outdoor temperatures and thus the relative deviation is higher. Therefore, the error term ε is set in relation to the expected value E and calculate the deviation measure D .

$$D = \frac{\varepsilon}{E} \quad (2)$$

A value of $D = 0.3$ therefore means that the actual heat demand deviates by 30% from the regression forecast. For an automatic processing, let us assume that an algorithm analyses the successively arriving values for D . To increase the robustness of the deviation measure against single, short-term fluctuations of the sensor values, the deviation measurements of the previous five hours are arithmetically averaged to \bar{D} . If \bar{D} is outside the threshold values S_1 (upper threshold) or S_2 (lower threshold) for a longer period of time A_f , the system's condition is interpreted as faulty. Different strategies can be used for the choice of the threshold values S_1 and S_2 :

1. The upper and lower thresholds are defined a priori for all substations as a fixed value, e.g. $S_1 = S_2 = 0.5$.
2. Individual threshold values are learned for each DHS:
 - (a) Uniform thresholds for all timestamps, or
 - (b) individual thresholds for each timestamp.

The learning of DHS-specific threshold values (2a) has the advantage that the accuracy of predictions can be significantly improved on the basis of historical data through learning effects. If a DHS behaves relatively unicyclically, larger fluctuations in heat demand usually occur and the threshold values should be chosen more broadly. When learning individual thresholds at the time stamp level (2b), it can also be considered that certain times of the day are also more irregular in normal behaviour (e.g. when switching between day and night mode).

Fig. 3 outlines the scenario proposed here. In the regression analysis on the left, the last five data points deviate strongly from the regression line (red data points). In the residuals plot (right), the averaged deviation measure is now entered in the chronological order and exceeds the previously described threshold value S_1 from entry value 9 onwards. In this case, averaging delays the alarm slightly, but our experiments show that attenuating random peaks makes the result of the ML model much more robust.

Finally, we define the threshold parameter A_f for generat-

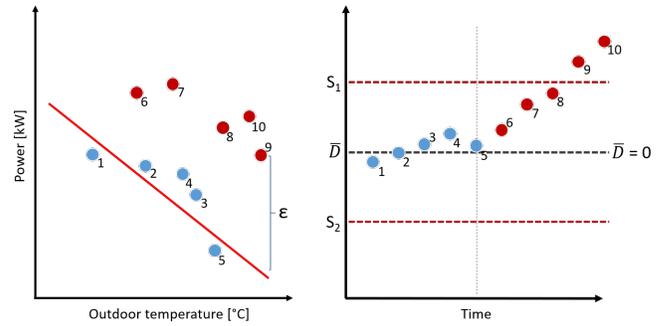


Figure 3. Regression and residual plot over time

ing an alert if the number of successive \bar{D} values exceed this threshold. Typically, A_f is DHS specific and for defining it, for instance, the maximum value from a validation data set can be used. If no \bar{D} are found on the validation data set, a value can be defined based on expert knowledge or DHS with similar characteristics (load behaviour, DHS type, etc.) can be used.

3.1.2. Results

In this section, our proposed approach is demonstrated on two DHSs (A and B) for which a failure was encountered and a maintenance action was necessary. For demonstration purposes, two substations with an opposite pattern of normal behaviour are selected. Fig. 4 shows the usual plant behaviour on a randomly selected day (11th February 2020).

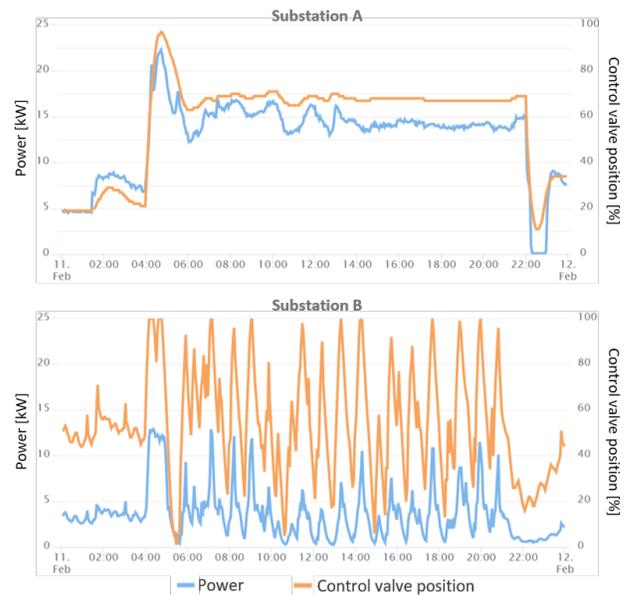


Figure 4. Comparison of the two examples on 11 February

DHS A: circulator pump breakdown

On this substation, the circulation pump suddenly fails on 21.03.2020, which was noticed by the customer the next day when the room temperature cooled down and was remedied by replacing the pump. This system does not heat domestic hot water. As can be seen in the upper time series in Fig. 4, the substation behaves very steadily, with the exception of the changes between day and night mode (between 4 and 5 o'clock). This means that only a few consecutive observation points with high error terms should be significant to detect a potential failure, which can also be seen in the high *cluster consistency* of 0.94. Fig. 5 shows the regression analysis for both clusters. From the pump defect onwards, the measured sensor values are far away from the predicted value, as can be seen from the red data points at the bottom of the diagram. Also, for the "high load" cluster (red triangle) as well as for the "low load" cluster (blue), individual data points with larger deviations from the regression line can be recognised. These can be explained by the changes between day and night mode, where small deviations occur for a short time.

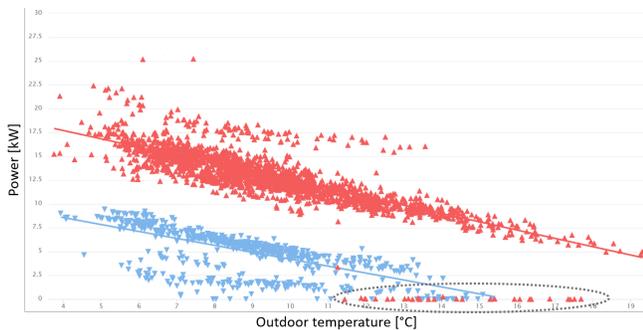


Figure 5. Regression analysis for pump failure of DHS A

Due to the averaging of the deviation measure \bar{D} (cf. Sect. 3.1.1), these short-term deviations do not lead to any false alarm, as can be seen in the upper time series in Fig. 6. In the entire time series, the algorithm only detects a deviation in the period in which the circulator pump breaks down. The algorithm detects the failure already three hours after the pump has broken down. This means that this fault can be detected much earlier than it is noticed by the customer.

DHS B: leaking control valve

Substation B comprises a heating circuit for building heating and domestic hot water heating including a buffer tank. On 16.01.2020, the customer reports an insufficient temperature of the tap water and the radiators. A close look at the sensor data reveals that the leakage already exists from 10.01.2020 and thus a few days before the customer noticed it. During the inspection by the service technician, a leaking screw connection on the control valve was identified and repaired on the same day by replacing the seal. Based on the Davies-Bouldin Index, the heat load pattern of this DHS is divided

into three clusters (high, medium, low heat demand). In contrast to A, the control signal of the volume flow control fluctuates strongly between the extreme states 100% (maximum flow) and 0% (no flow), which is expressed in a poor *cluster consistency* of 0.57 (vs. 0.94 for DHS A). This is demonstrated in the lower time series in Fig. 4, which shows the performance of a randomly selected day. The reason for the unsteady behaviour is that the buffer tank is heated with an additional heat source, which massively affects the control loop of the DHS controller.

The time series at the bottom in Fig. 6 shows the sensor value P and the intervals identified as potentially faulty by the algorithm. It can be seen that many false alarms occur, e.g. when a certain level of P is predicted according to the learned model, but the DHS takes much less energy from the district heating network due to the additional heat source. For clarification, the length of the longest intervals leading to a false alarm is given (square brackets; in hours). To avoid false alarms caused by unstable system behaviour, the threshold value A_f can be defined as in Sect. 3.1.1 before a signal is considered as a fault. Even if the DHS failure can be noticed before the customer due to the long failure interval (130 hours), statistical methods only work to a limited extent for DHS with such unsteady behaviour.

3.1.3. Discussion

Our approach combines various methods from related literature and extends them at different points, for example, the automatic analysis of the regression result by using residuals analysis. Compared to deep learning methods, which often lack explainability, the method proposed here makes it easier to understand the results of the algorithm. For example, in the case of a failure notification, the expert can see from the regression plot why the algorithm is reporting an alarm. In addition to power P , the designed ML pipeline can also be applied to all other sensors whose values correlate with the outdoor temperature T_{out} . For example, if the heating system consists of several heating circuits, the power consumption can be analysed for each of them using the described methodology and, in the event of a fault, the affected part of the system can be diagnosed. The algorithm is also very robust against the transmission failures between gateway and MDM described in Sect. 2.3, as it does not require a consistent recording of the sensor values to train the ML model, unlike representatives of time series analysis such as Matrix Profile which assume an equally sampled, complete time series (Yeh et al., 2016). The cluster consistency can be used as an indicator for a suitable *hyperparameterization* and the estimation of the *analysis accuracy*. This allows the identification of DHSs in advance where the algorithm is not suitable or where the length of the fault intervals must be significantly longer to reliably detect a fault. On the other hand, DHSs with high cluster consistency should be given stricter thresh-

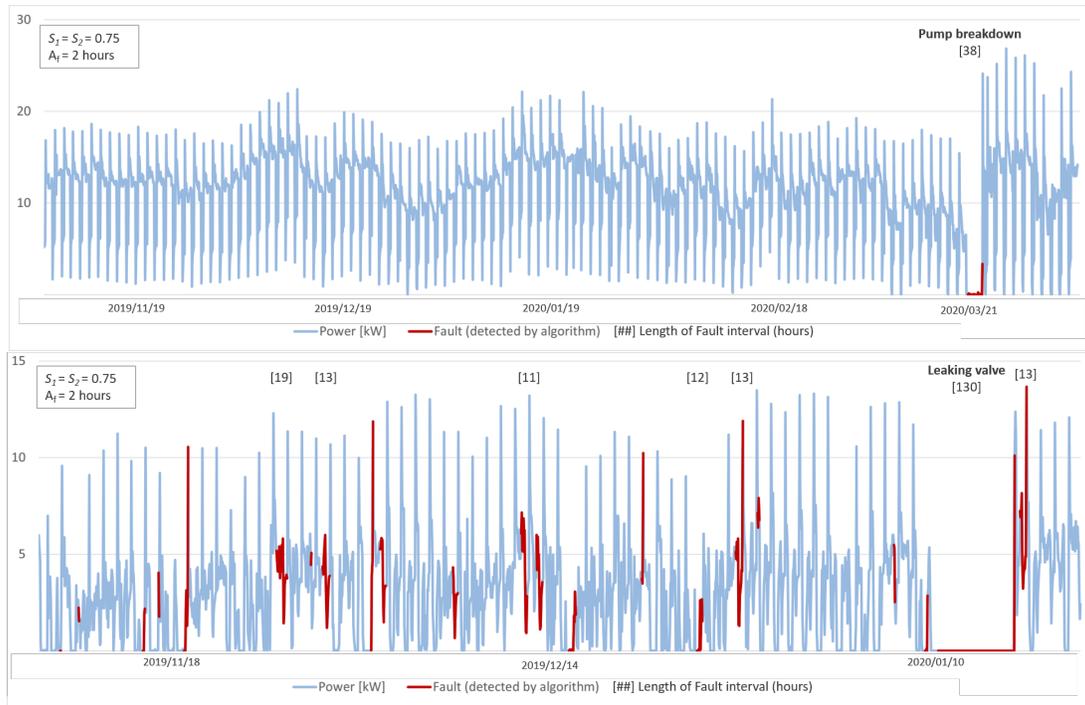


Figure 6. Comparison of the detected fault intervals for DHS A (top) and DHS B (bottom)

olds S_1 and S_2 to react to faults that have a lower impact on the sensor values. Nevertheless, one focus of the further development of the approach should be an efficient and, in the best case, an automated hyperparameter optimization (Feurer & Hutter, 2019). In this way, an optimal balance between *recall* and *precision* can be found.

As the example of DHS B shows, our model currently does not consider the influence of additional heat sources, such as solar thermal energy. In this case, the regression analysis could, for example, be supplemented by the independent variable of solar radiation (multivariate regression). Moreover, our case study also revealed the limitations of our MDM data, e.g. for predicting the remaining useful life (RUL). For example, we could not detect any patterns in the data that could be used as indicator of minor leaks in advance of the failure at DHS B. Thus, additional sensors (e.g. humidity) would have to be installed and made available to the MDM data in order to be able to act in advance. Especially for the case of DHS A, measuring current or vibration of the pump could be helpful to detect wear in advance (Fausing Olesen & Shaker, 2020).

3.2. Condition Monitoring

The objective of the machine learning pipeline presented in Sect. 3.1 is to detect unexpected failures of the substations and to initiate corrective maintenance as soon as possible.

These failures typically lead to an insufficient heat supply, which is noticed by the customer after some time due to the drop in room temperature and a reduced tap water temperature.

From the DHN operator’s point of view, however, it is also of great interest to work with low flow and temperature levels and to identify substations that cause a significant drop in overall grid efficiency and thus have a negative impact on its profitability. Up to now, a *time-based maintenance* strategy (Selcuk, 2017) is applied in the examined DHN, in which all DHS are maintained once a year according to a fixed scheme. This includes, among other things, a visual or functional check of the pipes, valves and the heat meter as well as the cleaning of the strainers. The disadvantage of this strategy is that maintenance is done regardless of the condition of the DHS and thus the remaining useful life cannot be optimally utilised. In modern DHNs, the permanent and centralised monitoring of the DHS offers new possibilities for *condition-based maintenance*. The following subsections describe two key performance indicators (KPI) for assessing the need of maintenance of DHSs using condition monitoring.

3.2.1. High Return Temperature

Increased return temperatures often have a negative impact on the overall efficiency of a DHN, especially when geothermal energy is used. Analogous to Eq. 1, the output of a geother-

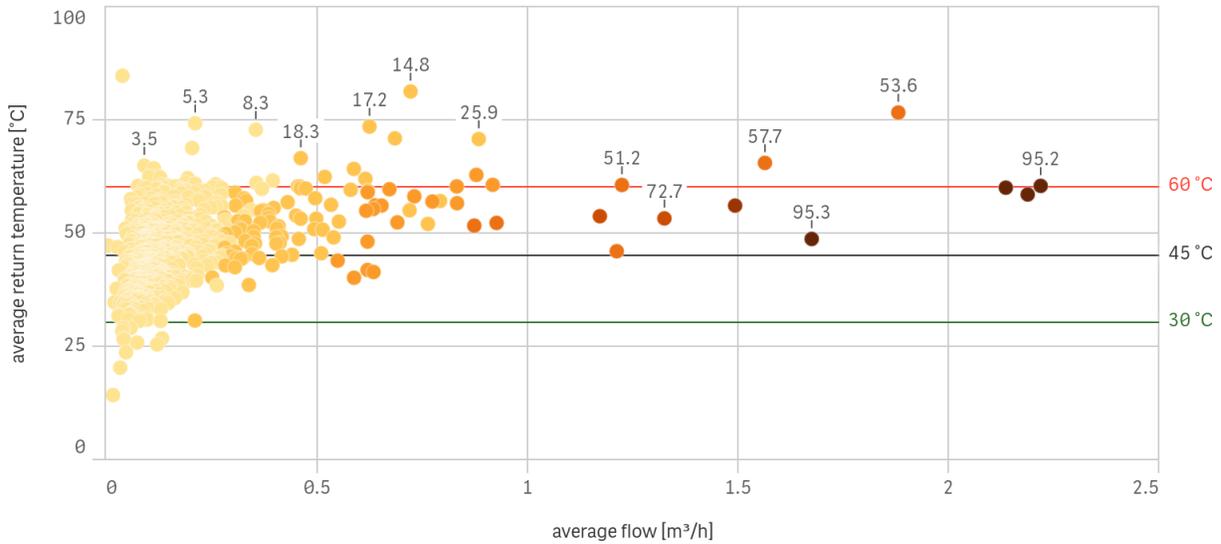


Figure 7. Return temperature in relation to the flow rate (dark points imply a higher power, which is partly indicated)

mal plant is the product of the volume of circulated thermal water and the temperature difference between the thermal water pumped into the grid and the thermal water returned from the grid (p and c are neglected for simplicity). Nevertheless, the advantages of low return temperatures apply not only to geothermal energy, but to all types of district heating networks (Gadd & Werner, 2014).

In geothermal DHNs, the amount of thermal water extracted from the ground per unit of time and its temperature are fixed parameters. The return temperature of the DHN is therefore the key parameter to increase the overall performance and to reduce heat losses in the pipelines. For this reason, heat supplier and customers agree on a maximum permissible average return temperature in many DHNs. In the analysed data set, 48 out of 896 buildings have an average return temperature above the agreed limit of 60 °C. But even the mean value of all substations, at 52.3 °C, is still well above the return temperatures targeted by the 4GDH. Table 3 outlines the positive influence of low return temperatures based on a real example of a DHS with $P = 14.8$ kW. If T_{pr} can be reduced to the maximum permissible value of 60 °C, V is reduced by more than half and thus also the corresponding energy losses in the DHN.

Table 3. Comparison of V for different T_{out}

ϕP	ϕT_{ps}	ϕT_{pr}	ϕV^*
14.8 kW	100.2 °C	81.4 °C	0.689 m³/h
14.8 kW	100.2 °C	60.0 °C	0.322 m³/h
14.8 kW	100.2 °C	45.0 °C	0.235 m³/h
14.8 kW	100.2 °C	30.0 °C	0.184 m³/h

* calculated with $c = 4.185$ kJ/(kg K)

The scatter plot in Fig. 7 shows the distribution of the return temperature in relation to the flow rate of all DHSs. It can be seen that DHS with large power consumption tend to have high return temperatures and flow rates. In the context of maintenance planning, those DHSs that fulfil these three characteristics (large P , large V , large T_{pr}) should therefore be maintained primarily. When considering low return temperature, the overall situation of the DHS must also be taken into account, e.g. if domestic hot water heating takes place and the return temperature may not fall below a certain threshold for hygiene reasons. Possible causes for high return temperatures are a missing or incorrect hydronic balancing of the radiators as well as faulty valves that lead to a permanent flow through the system or an ineffective outdoor temperature-dependent control (Wirths, 2008). Especially with hydronic balancing of the individual radiators, the return temperature can often be significantly reduced and thermal comfort improved at the same time. In order to meet the requirements of 4GDH, DHSs with high return temperatures should therefore be assessed and, for example, the use of flow valve limiters at the radiators should be examined (Averfalk & Werner, 2017).

3.2.2. Low Average Temperature Difference

The maximum supply temperature of a DHS is determined by its position in the DHN. The greater the distance to the power plant, the higher the heat losses in the pipes and the associated reduction in T_{ps} . In the data set analysed, the range of the average supply temperature is approximately between 100 °C and 85 °C, which means that near buildings receive up to 15 °C warmer supply temperatures. (Gadd & Werner, 2014) therefore suggest investigating the temperature difference ΔT between supply and return in relation to the outside temperature as a performance indicator.

The achievable temperature difference of a DHS depends very much on the conditions of the heated building. The average ΔT is 44 °C, with the best DHS reaching 67 °C and the worst only 15.2 °C. However, since no precise information is available about the buildings (e.g. year of construction or heated living space) and the heating system, no target values for well-functioning DHS can be given. (Gadd & Werner, 2015) therefore propose a target value of 45 °C as a threshold value for identifying installations to be inspected.

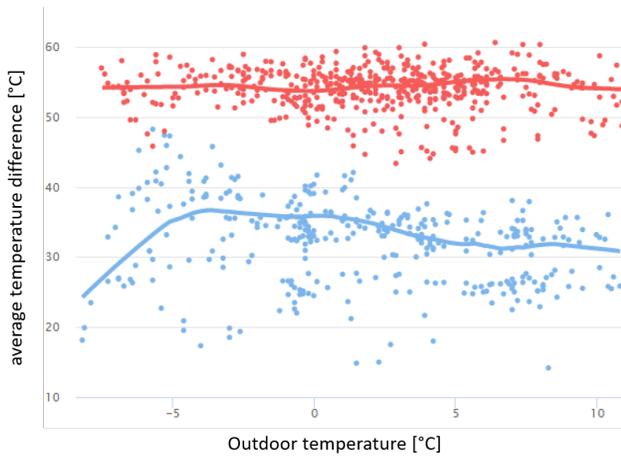


Figure 8. ΔT before (blue) and after maintenance (red)

For demonstration purposes, Fig. 8 shows the positive effects after a maintenance intervention where the strainer (primary side) was cleaned. The hourly ΔT values are plotted and a locally weighted regression line is calculated using the LOESS (locally estimated scatterplot smoothing) algorithm. As a result of the maintenance, the DHS's performance improves remarkably, as ΔT increases by an average of 21 degrees. The system is subsequently much more efficient, as significantly more energy can be extracted from the same quantity of water.

3.2.3. Impact on Maintenance Strategy

In a sample of 30 maintenance operations, no noticeable effect on parameters ΔT and T_{pr} was found. This means that their performance did not change considerably after the maintenance and is questionable if the on-site inspection by a technician was really necessary. Thus, the metrics illustrate the weaknesses of the current time-dependent maintenance strategy: A large part of the DHSs that are regularly inspected on-site do not yet need to be maintained due to their current state of wear. On the other hand, a significant number do not work optimally, for example because their return temperature is permanently too high (c.f. Fig. 7). For this reason, we argue that smart meter data should become a much stronger focus of maintenance planning.

In discussions with domain experts, it was confirmed that it

is in general possible to customise the maintenance rhythm of a DHS. Therefore, a *condition-based maintenance strategy* (Jardine, Lin, & Banjevic, 2006) should be applied in the future, prioritising those DHS whose performance indicators are particularly poor or where these values have noticeably deteriorated over time. In addition to the existing sensors, which are mainly used for control purposes, extra sensors such as humidity or power consumption (e.g. of the circulator pump) could enable to detect faults at a relatively early stage in order to enable predictive maintenance in the future.

4. CONCLUSION AND FUTURE WORK

In this paper, various data-driven applications for failure detection and condition monitoring using smart meter data in the field of district heating were demonstrated. First, an ML pipeline for automatic failure detection was designed and discussed on the basis of two selected failure cases. It was shown that learning regular load patterns can be used for sudden failure detection, but the accuracy of the algorithm depends significantly on the regularity of the substation's behaviour. Future failure detection models should therefore be supplemented by the influence of exogenous factors such as additional solar thermal heat use. In addition to failure detection using cluster and regression analysis, two metrics (KPIs) for improving the maintenance strategy and the overall efficiency of DHN were also discussed. We therefore recommend the permanent assessment of the DHSs' condition via suitable KPIs and their consideration for a condition-oriented maintenance strategy. Finally, all our results as well as their usefulness are verified and confirmed by domain experts of the DHN operator.

Future work can focus on incorporating of gained experience from previously conducted maintenance actions using the methodology of case-based reasoning (Bergmann, 2003) for diagnosis in addition to fault detection. In this regard, we see one major challenge in transferring knowledge between DHS with different load patterns, which requires the application of transfer learning (Moradi & Groth, 2020). Furthermore, it can be explored how semantic technologies can be used to solve interoperability problems by relating domain knowledge in form of ontologies and describing data streams from heterogeneous sources in order to leverage them for condition monitoring (Al-Shdifat, Emmanouilidis, Khan, & Starr, 2021). In addition to the use cases outlined in this paper, data from smart metering offers further potential, especially with regard to the increasing complexity within future *smart grids* and the growing interconnection of different sectors such as electricity, heat supply, industry and transport according to the principle of *sector coupling*. In contrast to conventional district heating systems, DHSs in the future will be controlled adaptively and in real time as part of a cooperating network via the Internet of Things.

REFERENCES

- Al-Shdifat, A., Emmanouilidis, C., Khan, M., & Starr, A. G. (2021). Ontology-based context modeling in physical asset integrity management. *Data-Driven Cognitive Manufacturing-Applications in Predictive Maintenance and Zero Defect Manufacturing*.
- Averfalk, H., & Werner, S. (2017). Essential improvements in future district heating systems. *Energy Procedia*, 116, 217–225.
- Bergmann, R. (2003). *Experience management: foundations, development methodology, and internet-based applications* (Vol. 2432). Springer.
- Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE transactions on Pattern Analysis and Machine Intelligence*(2), 224–227.
- European Union. (2018). *Directive 2018/2002 of the European Parliament and of the Council of 11 December 2018 amending Directive 2012/27/EU on energy efficiency*. Retrieved 2021-03-22, from <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018L2002>
- Fausing Olesen, J., & Shaker, H. R. (2020). Predictive maintenance for pump systems and thermal power plants: State-of-the-art review, trends and challenges. *Sensors*, 20(8).
- Feurer, M., & Hutter, F. (2019). Hyperparameter optimization. In *Automated machine learning* (pp. 3–33). Springer, Cham.
- Gadd, H., & Werner, S. (2014). Achieving low return temperatures from district heating substations. *Applied Energy*, 136, 59–67.
- Gadd, H., & Werner, S. (2015). Fault detection in district heating substations. *Applied Energy*, 157, 51–59.
- Gianniou, P., Liu, X., Heller, A., Nielsen, P. S., & Rode, C. (2018). Clustering-based analysis for residential district heating data. *Energy conversion and management*, 165, 840–850.
- Jahn, M., Jentsch, M., Prause, C. R., Pramudianto, F., Al-Akkad, A., & Reiners, R. (2010). The energy aware smart home. In *5th International Conference on Future Information Technology* (pp. 1–8).
- Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7), 1483–1510.
- Kiluk, S. (2012). Algorithmic acquisition of diagnostic patterns in district heating billing system. *Applied Energy*, 91(1), 146–155.
- Li, X., Bowers, C. P., & Schnier, T. (2009). Classification of energy consumption in buildings with outlier detection. *IEEE Transactions on Industrial Electronics*, 57(11), 3639–3644.
- Lund, H., Werner, S., Wiltshire, R., Svendsen, S., Thorsen, J. E., Hvelplund, F., & Mathiesen, B. V. (2014). 4th Generation District Heating (4GDH): Integrating smart thermal grids into future sustainable energy systems. *Energy*, 68, 1–11.
- Månsson, S., Kallioniemi, P.-O. J., Sernhed, K., & Thern, M. (2018). A machine learning approach to fault detection in district heating substations. *Energy Procedia*, 149, 226–235.
- Moradi, R., & Groth, K. (2020). On the application of transfer learning in prognostics and health management. *Annual Conference of the PHM Society*, 12(1).
- Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., ... others (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Intergovernmental Panel on Climate Change (IPCC).
- Rybach, L. (2003). Geothermal energy: sustainability and the environment. *Geothermics*, 32(4-6), 463–470.
- Sandin, F., Gustafsson, J., & Delsing, J. (2013). *Fault detection with hourly district energy data: Probabilistic methods and heuristics for automated detection and ranking of anomalies*. Svensk Fjärrvärme.
- Sandin, F., Gustafsson, J., Delsing, J., & Eklund, R. (2012). Basic methods for automated fault detection and energy data validation in existing district heating systems. In *International Symposium on District Heating and Cooling*.
- Sayegh, M., Danielewicz, J., Nannou, T., Miniewicz, M., Jadwiszczak, P., Piekarska, K., & Jouhara, H. (2017). Trends of european research and development in district heating technologies. *Renewable and Sustainable Energy Reviews*, 68, 1183–1192.
- Seem, J. E. (2005). Pattern recognition algorithm for determining days of the week with similar energy consumption profiles. *Energy and Buildings*, 37(2), 127–139.
- Seem, J. E. (2007). Using intelligent data analysis to detect abnormal energy consumption in buildings. *Energy and Buildings*, 39(1), 52–58.
- Selcuk, S. (2017). Predictive maintenance, its implementation and latest trends. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 231(9), 1670–1679.
- Skagestad, B., & Mildenstein, P. (2002). *District heating and cooling connection handbook*. NOVEM, Netherlands Agency for Energy and the Environment.
- Sun, Q., Li, H., Ma, Z., Wang, C., Campillo, J., Zhang, Q., ... Guo, J. (2015). A comprehensive review of smart energy meters in intelligent energy networks. *IEEE Internet of Things Journal*, 3(4), 464–479.
- Wang, Y., Chen, Q., Hong, T., & Kang, C. (2018). Review of smart meter data analytics: Applications, methodologies, and challenges. *IEEE Transactions on Smart Grid*, 10(3), 3125–3148.

- Wirths, A. (2008). Einfluss der Netzurücklaufemperatur auf die Effizienz von Fernwärmesystemen. In *13. Dresdner Fernwärme-Kolloquium*.
- Xue, P., Zhou, Z., Fang, X., Chen, X., Liu, L., Liu, Y., & Liu, J. (2017). Fault detection and operation optimization in district heating substations based on data mining techniques. *Applied Energy*, *205*, 926–940.
- Yeh, C.-C. M., Zhu, Y., Ulanova, L., Begum, N., Ding, Y., Dau, H. A., ... Keogh, E. (2016). Matrix profile i: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In *2016 IEEE 16th International Conference on Data Mining (ICDM)* (pp. 1317–1322).
- Zheng, J., Gao, D. W., & Lin, L. (2013). Smart meters in smart grid: An overview. In *2013 IEEE Green Technologies Conference (GreenTech)* (pp. 57–64).