Energy 368 (2026) 142993



Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy



Towards automation of fault detection and diagnosis for district heating substations: A three-layer hybrid approach

Laura Block *, Martin Rätz, Dirk Müller

RWTH Aachen University, E.ON Energy Research Center, Institute for Energy Efficient Buildings and Indoor Climate, Mathieustraße 10, Aachen, 52074, Germany

ARTICLE INFO

Keywords:
Fault detection
Fault diagnosis
District heating
Machine learning
Novelty detection
Energy efficiency

ABSTRACT

District heating (DH) networks play a crucial role in achieving climate neutrality goals, but faults in these systems significantly impact energy efficiency. Traditional rule-based fault detection lacks generalization capabilities. Emerging machine learning (ML) techniques frequently suffer from a scarcity of labeled training data. This paper presents a fault detection and diagnosis (FDD) methodology for DH substations that addresses these challenges by significantly automating the setup of the FDD for an individual substation. The focus is on detecting hard faults at the substation level rather than suboptimality in the DH network. The proposed three-layer approach combines ML models, mode-based rules, and rule-based system boundary monitoring. The methodology is validated using measurement data from real-world district heating network substations. A novel kNN-based selection algorithm is introduced to automatically isolate representative training data, addressing a major bottleneck in regression-based FDD. The ML-based FDD then employs regression models to identify deviations from normal operation. For the mode-based FDD, rules derived from the substation control strategy detect faulty operating modes. System boundaries are monitored using defined thresholds. Validation with historical real-world measurements demonstrates the effectiveness of this three-layer approach in detecting various fault types. In two detailed case studies, the FDD framework detected the faults 8.5 h before the presumed technician intervention and 104 h before the customer report, and maintains a low false alarm rate of 0.27% and 0.01%. These findings indicate that the proposed methodology supports scalable, proactive maintenance for district heating substations.

1. Introduction

District heating (DH) networks enable efficient heat supply and are suitable for the integration of renewable energies [1]. Faults in district heating networks have a significant impact on energy consumption. Previous studies reveal that 43 %, 50–60 % and 75 % of investigated substations operate suboptimally [2–4]. The detection and handling of faults in substations is essential to secure the heat supply and enable a faster decarbonization of the heating sector [3,5]. Through these measures, district heating customers can reduce energy consumption by an average of 14 % [6]. Additionally, early detection can reduce costs through more flexible deployment planning of technicians. The company Samson highlights a 30–80 % drop in service trips due to smart detection [7]. In general, fault detection and diagnosis (FDD) can contribute to the cost-effectiveness, increased efficiency, and security of heat supply [5]. The digitalization of district heating networks has led to a significant increase in available sensor data, bringing FDD for district heating systems into the focus of research [8]. Van Dreven et al. [8] provide a comprehensive overview of intelligent FDD approaches for district heating systems, highlighting both challenges and opportunities. The authors emphasize that artificial intelligence and machine learning (ML) in predictive maintenance can reduce downtime. They advocate for the development of automated frameworks, as manual FDD methods are considered time-consuming, costly, and error-prone. Critically, the authors identify a research gap regarding real-world scenario applications, as many studies rely exclusively on synthetic data.

Traditionally, rule-based methods have been widely used for FDD in district heating substations. Yang et al. [9] investigated four fault types in buildings connected to DH with rule-based FDD, utilizing, for instance, the heating curve of the substation. Leiria et al. [6] analyzed 158 fault reports from 90 buildings using district heating in Denmark, comparing five rule-based fault indicators before and after the intervention of a technician. Sandin et al. [10] carried out a correlation analysis of primary supply temperatures of different substations

* Corresponding author.
E-mail addresses: laura.block@eonerc.rwth-aachen.de (L. Block), martin.raetz@eonerc.rwth-aachen.de (M. Rätz), dmueller@eonerc.rwth-aachen.de (D. Müller).

https://doi.org/10.1016/j.energy.2026.142993
Received 11 June 2026; Received in revised form 5 October 2026; Accepted 29 October 2026
Available online 5 November 2026
0360-5442/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Open access funding enabled and organized by Projekt DEAL.

and limit checking combined with cluster analysis of thermal power usage. While rule-based methods are straightforward to implement and interpret, they often lack adaptability to varying operational conditions and may not capture complex fault patterns effectively. This motivates data-driven methods that learn the behavior of a substation from its own measurements.

The application of ML methods for FDD in district heating substations has become more feasible with the increasing availability of sensor data. ML can be categorized into supervised and unsupervised methods [11]. Unsupervised methods focus on identifying anomalies in the data without the need for labeled fault data. In the context of district heating substations, unsupervised methods are mostly used to identify suboptimal substations or to group similar substations and detect deviations from the group behavior. Calikus et al. [12] employ clustering techniques on a dataset of 1385 Swedish buildings to identify suboptimal stations. Van Dreven et al. [13] detect anomalies related to supply temperature and substation performance using an advanced shared nearest neighbor algorithm. Leiria et al. [14] analyze 50 reported faults by grouping fault symptoms with self-organizing maps with k-means clustering to advance towards automated fault diagnosis. Farouq et al. [15] used k-nearest neighbor (kNN) to identify the reference group of a substation and identified faults with isolation trees based on deviation of the substation from its reference group. Van Dreven et al. [16] introduce a novel approach to group substations based on temperature profiles to improve fault detection performance. These approaches identify persistently suboptimal substations by comparison with their peers but are not designed to detect abrupt failures at a single substation in time.

Alongside these unsupervised approaches, supervised ML methods are the subject of recent research. Supervised ML methods can be divided into two different tasks: regression and classification. Due to the lack of data labeled by fault type, classification is not used for FDD in DH substations [17]. Regression models are used to learn the normal operation of the DH substations. During FDD, a deviation from the learned behavior is detected as a fault. Linear regression is used by Gadd and Werner [4], Månsson et al. [2] and Theusch et al. [18]. Månsson et al. [19] predict substation volume flow using a Gradient Boosting Regressor based on outdoor temperature, primary supply temperature, and time of day. Their approach shows promise for detecting changes in fault-induced datasets. Beyond substations, Pourbozorgi Langroudi et al. [20] evaluate multiple regression algorithms for predictive maintenance of district heating pipe aging, finding Random Forest to achieve the highest accuracy. Their work demonstrates that features such as relative humidity, sunshine duration, month, and hour improve prediction quality. Van Dreven et al. [21] achieved an accuracy of 86.21 % using Isolation Forest for fault detection and Random Forest for fault diagnosis on a fault-induced lab dataset. They state that misclassifications of normal operation are a significant challenge of their methodology. Kim et al. [22] developed a method using a virtual sensor measurement as an additional input variable for Multilayer Perceptron models to detect fouling in the heat exchanger of the substation. Hermans et al. [17] focus on the underlying distribution and replace a Gaussian error model with a quantile regression model. They conclude that the quantile regression model better reflects the actual error distribution. Furthermore, the authors state that ensemble models trained on different time periods improve robustness. More complex architectures are also used, such as convolutional neural networks (CNN) with principal component analysis [23], a Long Short-Term Memory network (LSTM) [24], a CNN-LSTM with a novel data augmentation method [25], an autoencoder [26] or combining a simplified physical model with an LSTM-based variational autoencoder [27]. Common to regression approaches is the assumption that the training dataset is fault-free.

To achieve the widespread adoption of FDD and prognosis technologies, the limitations of single-model concepts must be overcome [28]. According to Guevara Bastidas et al. [5] different methods are required to detect the whole spectrum of faults in DH substations. Parallel combinations have been proposed outside of district heating. In FDD for Heating, Ventilation, and Air Conditioning (HVAC) systems, ML methods and rule-based systems have been applied in parallel. For example, Dietz et al. [29] filtered the historical data of an air handling unit using rules to ensure the data for training of the ML algorithm is fault-free. During the operation of the FDD, the rule-based filter and ML-based method are both applied, reaching a false alarm rate below 1 %. In the approaches for district heating applications, the combined methods affect a common fault indicator or are applied in series for detection and subsequent classification. In contrast, the three layers proposed in this work are complementary, each covering different fault classes, and their indicators feed a hierarchical fault structure that assigns a root cause.

For the application of ML-based FDD algorithms it is essential to know which data is fault-free and which data contains faults [30]. For district heating substations there is a lack of labeled monitoring data [17,31]. The lack of labeled data for ML-based FDD in district heating substations is not only a prohibitive barrier to classification but also a critical challenge for regression. The aim of regression models for FDD is to predict the normal system behavior. To achieve this, the selected training dataset must be representative of all operational states and fault-free. Thus, the selection of training datasets is crucial for the performance of ML-based FDD methodologies. Neumayer et al. [32] analyze 20 real-world datasets used in FDD in district heating networks. They identify datasets as a critical part of the development of FDD methodologies and state that methodology and dataset can only be evaluated together. However, the survey does not address how training data was selected for the reviewed datasets or provide a recommendation on how training data should be selected within a dataset. Theusch et al. [18] implement a non-manual approach, using kNN classification with Euclidean distance to identify and eliminate outliers in outdoor temperature and power consumption data. Dietz et al. [29] explicitly stated the application of rule-based fault detection to ensure the training data of the ML algorithm is fault-free. Nevertheless, they investigated an air handling unit, not a DH substation. Most studies for DH substations rely on manually selected time periods without considering data distribution, balance, or representation of all system states. This leads to imbalanced datasets, with unequal representation of different operational states potentially leading to bias towards overrepresented conditions. Imbalanced datasets are a fundamental problem in district heating networks [8]. The literature reveals significant limitations in data selection methodologies for FDD in district heating. To address the limitations of manual data selection, automated methods for selecting representative training datasets are needed.

In this work, novelty detection is identified as a suitable approach to address the challenge of automated selection of a representative training dataset. The technique aims to identify how a data sample differs from the training data of the novelty detector in some respect [33]. In general, novelty detection methods determine the epistemic uncertainty represented by a novelty score [34]. Epistemic uncertainty arises from the model's lack of knowledge about the system. In the context of this work, novelty detection is used to identify the most unknown data points in the available data based on an initial dataset. In this way, unknown operational states can be identified and added iteratively to the training dataset. Unlike clustering methods, which partition a fixed dataset, a novelty detector scores each sample relative to a reference set and therefore supports the iterative extension of that set required here. Rätz et al. [34] demonstrate the application of novelty detection for identifying validity domains of ML models in building energy systems. Their comprehensive evaluation of 15 novelty detection algorithms identifies kNN, Kernel-based Density Estimation, Gaussian Process Regression, One-Class Support Vector Machine, and Isolation Forest as particularly suitable for building energy applications. Rätz et al. [34] emphasize that kNN performs well in both two-dimensional and multi-dimensional cases while being computationally efficient and

robust to hyperparameter settings. Consequently, a kNN-based novelty detector is selected for this work to automate the selection of a representative training dataset for the regression models used in the FDD methodology.

While different studies have proposed methodologies for FDD in district heating substations, several gaps remain unaddressed. The literature focuses predominantly on identifying suboptimal operation rather than system failures, and often employs only single regression models [8]. However, the application of different methods is needed to cover the variety of fault types [5]. Due to the lack of labeled datasets, the use of supervised learning methods is limited, and quantitative analysis is often not possible. In particular, the selection of training datasets for FDD in district heating substations and its automation is not the focus of the literature. Moreover, there is a notable lack of automated methodologies evaluated on real-world applications, with many studies relying exclusively on synthetic data. HVAC system ML-based FDD solutions are rarely implemented in practice [29]. Andersen et al. [35] state that FDD methods need to be applied on real-life systems to show their efficacy and utilize their benefits. They identify the implementation of FDD demonstrators as a key driver for FDD interest. This work builds on the proof-of-concept study of Rätz et al. [36], which showed that regression models trained on manually selected data can detect severe substation faults. The aim of the present paper is to advance the automation of FDD for individual district heating substations, so that the effort of adapting the method to each substation is reduced. The following aspects are addressed: we (i) introduce a kNN-based novelty detection method that automatically selects a representative training dataset and reduces manual verification effort, (ii) combine regression-based, mode-based, and boundary-based detection in a three-layer framework that covers a broader spectrum of faults than a single approach, and (iii) evaluate the framework over 20 months of operation of two substations in the Aachen district heating network against faults documented by the operator.

This paper is structured as follows: Section 2 describes the methodology used in this study, including the case study of the Aachen district heating network and the automated three-layer FDD methodology. Section 3 presents the results of the dataset selection, the results of the station mode detection, and the performance of the FDD methodology evaluated for different data selection methods and in detail for two faults in different substations for the proposed data selection. Section 4 discusses the findings and addresses limitations. Finally, Section 5 summarizes the main contributions and provides an outlook on future research directions.

2. Methodology

Section 2.1 provides a brief technical overview of the investigated substations, including available sensor data and the control strategy. Section 2.2 presents the three-layer FDD framework, comprising regression model-based fault detection, station-mode-based fault detection, rule-based system boundary monitoring, and fault diagnosis. For the regression model-based fault detection, the automated selection of a representative training dataset using novelty detection is described.

2.1. Case Study: Substations of the Aachen district heating network

The Aachen district heating network is used as a case study because, in cooperation with the network operator, ~300 substations were equipped with monitoring equipment. The two analyzed substations were chosen as they were the only two substations with useful fault documentation during the monitored period. Given the heterogeneity in substation designs, this study focuses on the most prevalent configuration within the Aachen district heating network. The technical details provided in the following apply to this specific subset. Fig. 1 shows a schematic representation of this substation type. For this substation type, space heating and domestic hot water (DHW) are

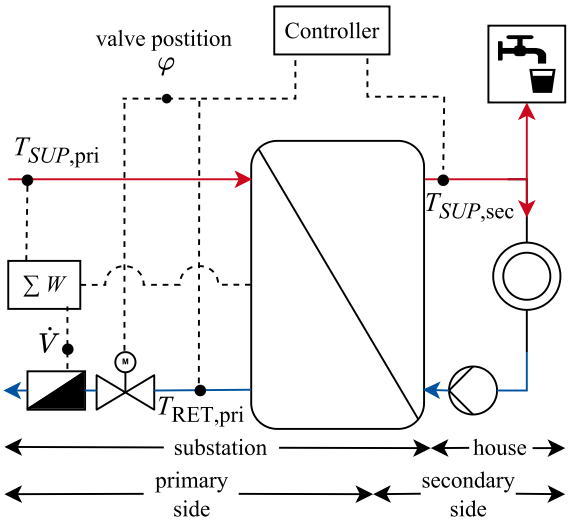


Fig. 1. Schematic representation of a substation and the positioning of the sensors of the monitored variables.

supplied through a single heat exchanger. The DHW preparation is part of the customer's installation and it branches off the secondary supply outside the monitored substation boundary. The controller serves both demands sequentially by switching the secondary supply temperature setpoint between the operating modes as described below.

The substations are indirectly connected to the district heating network, with a maximum design heat load of up to 200 kW. The maximum design heat load is determined for each substation individually according to DIN EN 12831 [37] and DIN 4708 [38] for a design outdoor temperature of −8 °C. The maximum primary volume flow rate is derived from the design heat load. Further dimensioning details can be found in the technical design conditions of the district heating operator [39]. All substations are equipped with similar controllers, specifically different versions of the SAMSON Trovis series. A PID controller regulates the secondary supply temperature by modulating the valve position. The setpoint for the secondary supply temperature depends on the substation's operating mode. These modes are based on rules implemented in the controller, as described in the manufacturer's manual [40]. The three operating modes relevant to this study are described below. In weather-compensated operating mode, the secondary supply temperature setpoint is determined by the outdoor temperature measured in real time at the individual substation. In many substations, a night setback is also implemented, which lowers the setpoint during the night. In summer mode, the substation only provides DHW. The substation switches from weather-compensated to summer mode when the daily average outdoor temperature (7 am–10 pm) exceeds a defined threshold for n consecutive days. It returns to weather-compensated mode after m consecutive days below this threshold. By default, n is set to 2 days, m to 1 day, and the threshold to 18 °C. In DHW mode, the secondary supply temperature setpoint is determined by the DHW temperature setpoint. The mode is activated when the DHW temperature falls 0.1 °C below the setpoint. The factory default for the DHW temperature setpoint is 60 °C. The measurement data is collected by the controller and transmitted to the cloud via LoRaWAN. It is stored in a time-series database and can be accessed via an API. The monitored variables shown in Table 1 are transmitted every 15 min and have been available since February or September 2023 until October 2024, depending on the installation date of the monitoring equipment at each substation. Additional variables are available on demand. In this work, the following on-demand variables are used during the calibration of the FDD methodology: the setpoint of the secondary supply temperature ($T_{\mathrm{SUP,sec,set}}$ in °C) and the DHW storage charging

Table 1
Monitored variables of the substations.

Name	Symbol	Unit	Data point type
Error code controller	–	–	–
Cumulative heat	W	kWh	Balance heat meter
Primary supply temperature	$T_{\mathrm{SUP,pri}}$	°C	Measurement
Primary return temperature	$T_{\mathrm{RET,pri}}$	°C	Measurement
Secondary supply temperature	$T_{\mathrm{SUP,sec}}$	°C	Measurement
Primary volume flow rate	$\dot{V}$	m^3/h	Measurement
Valve position	φ	%	Setpoint

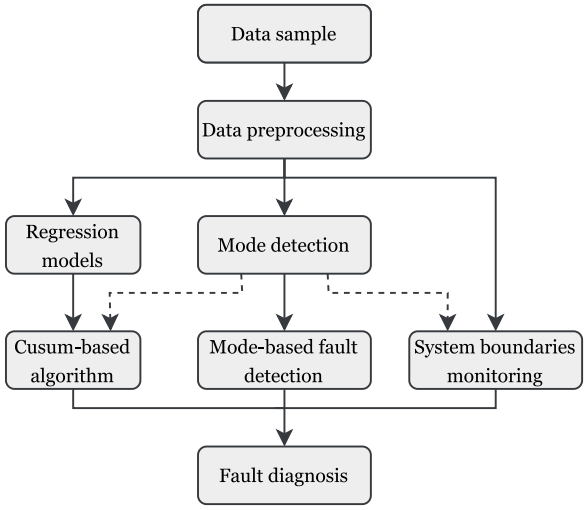


Fig. 2. Schematic overview of the three-layer FDD methodology. The logical connection between the different steps of the methodology, as well as the data flow within the three detection layers, is illustrated. Necessary steps for initialization of the three fault detection layers are not shown.

temperature (T_{DHW} in °C). Additionally, the outdoor temperature (T_{ODA} in °C) and global radiation (G_{Hor} in W/m^2) from the E.ON ERC main building weather station are used to estimate the values at the location of the substation.

The district heating operator provides limited information on past substation faults. This information consists primarily of unstructured, qualitative descriptions recorded by technicians during service visits, which vary in quality and completeness. In the results section, this limited information about specific faults is used to evaluate the performance of the FDD methodology.

2.2. FDD framework

This section describes the proposed three-layer FDD framework. A graphical overview of the framework is provided in Fig. 2. Each layer is presented in a dedicated subsection, starting with regression model-based fault detection in Section 2.2.1, followed by station-mode-based fault detection in Section 2.2.2, and rule-based system boundary monitoring in Section 2.2.3. The subsequent fault diagnosis is then described in Section 2.2.4. The necessary steps for initialization are detailed within each respective subsection. The dataset selection for the regression models is described in Section 2.2.1.

2.2.1. Regression model-based fault detection

For the regression model-based fault detection, additional features are engineered to support the data selection and the training of the regression models. The following temperature differences are calculated from the monitored variables:

$$\Delta T_{\mathrm{SUP}} = T_{\mathrm{SUP,pri}} - T_{\mathrm{SUP,sec}} \tag{1}$$

$$\Delta T_{\mathrm{pri}} = T_{\mathrm{SUP,pri}} - T_{\mathrm{RET,pri}}. \tag{2}$$

Moreover, temporal features are derived for each monitored variable and calculated temperature difference. These features are the one-step time lag (V_{t-1}) and the rate of change (ΔV_t), defined as:

$$\Delta V_t = V_t - V_{t-1}. \tag{3}$$

Prior to training the regression models, a representative training dataset is selected using novelty detection. As described in Section 1, kNN-based novelty detection is used for this task. The novelty score of each sample is the distance to its kth nearest neighbor. This is calculated using the Euclidean distance, which determines the distance between two points $X = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and $Y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ as follows:

$$D(X,Y) = \left(\sum_{i=1}^{n} |x_i - y_i|^2\right)^{1/2}. \tag{4}$$

The novelty detector is trained with the following features and their lag and delta features: $T_{\mathrm{RET,pri}}$, $T_{\mathrm{SUP,pri}}$, $T_{\mathrm{SUP,sec}}$, φ, $\dot{V}$, T_{ODA}, G_{Hor}, W, ΔW, ΔT_{pri} and ΔT_{SUP}. Each feature is normalized to the range $[0, 1]$ using min–max normalization. The neighborhood size is set to $k = 5$. Since the selection criterion is the daily average of the novelty scores rather than the score of an individual sample, the influence of k on the selected days is limited, which is examined in Section 3.1. The selection of the representative dataset begins by excluding all faulty periods known to the network operator from the available data. Subsequently, the remaining data is split into daily segments. Segments with more than 33 % missing data are excluded. The novelty detector is initially trained using the first available day of historical data. The novelty score is calculated for each sample in the remaining data. For each daily segment, the average novelty score is computed. The day with the highest average novelty score is manually verified to be fault-free and added to the representative dataset. A new novelty detector is trained with the updated representative dataset. This procedure is repeated iteratively until seven days are included in the representative dataset. Seven days are chosen as a compromise between representativeness and the effort required for manual verification. The selection of the representative dataset can be followed in detail for an exemplary substation in Fig. 4 in the results (Section 3.1).

The representative dataset is used for the training of the regression models. For each substation, five regression models are trained to represent different aspects of the system behavior. The specific features and targets used, as well as the faults detectable by each model, are described in Section 2.2.4. Random Forests are used as regression models, as they achieve solid results with default hyperparameter settings [41]. This is essential here, since tuning each model individually would counteract the automated setup. Moreover, the representative dataset comprises seven days, i.e., 672 samples, which excludes model classes that require extensive training data. Ten decision trees are used for each Random Forest. The models are trained using bootstrapping. The representative dataset is split into subsets: 80 % is used for training and validation, while the remaining 20 % serves as a test dataset to evaluate the performance of the regression models. The split is performed randomly on sample level. Since the lag and delta features link neighboring samples, the test dataset is not fully independent of the training dataset and the reported test metrics are therefore optimistic. The calibration of the fault detection is not affected, as it is based on the prediction errors of the entire representative dataset, as described below.

In the fault detection phase, the regression models predict the target values based on the input features. For each time step, the prediction error is calculated as the difference between the predicted and the measured value. For robust fault detection, the prediction error is passed to a CUSUM-based algorithm. Simplified, the CUSUM value is the bounded recursive sum of the prediction error with a decay term. The CUSUM values are calculated separately for positive and

negative prediction errors. This enables the detection of both positive and negative drifts in the prediction error. The exact algorithm for positive prediction errors is presented in Algorithm 1. If the CUSUM value ($\text{cusum}_{+,t}$) exceeds a predefined threshold, a fault is indicated.

Algorithm 1 Pseudocode of the CUSUM-based algorithm for positive prediction errors

$residual_t \leftarrow prediction_t - actual_t$
$\delta_t \leftarrow max(0, residual_t - \text{cusum-limit}_+)$
if $\delta_t > 0$ **then**
  if $target \neq \dot{V} \wedge station_mode_t = \text{off}$ **then**
    $cusum_{+,t} \leftarrow max(0, cusum_{+,t-1} - \text{cusum-reducer}_+)$
  else if $cusum_{+,t-1} = \text{cusum-max}_+$ **then**
    $cusum_{+,t} = cusum_{+,t-1}$
  else
    $cusum_{+,t} \leftarrow cusum_{+,t-1} + \delta_t$
    $cusum_{+,t} \leftarrow min(cusum_{+,t}, \text{cusum-max}_+)$
  end if
else
  $cusum_{+,t} \leftarrow max(0, cusum_{+,t-1} - \text{cusum-reducer}_+)$
end if

The parameters of the CUSUM-based algorithm, cusum-limit_+, cusum-reducer_+, cusum-max_+, and threshold_+, are determined individually for each regression model. Since the representative dataset is assumed to be fault-free, the parameters are calibrated to ensure that no faults are indicated within this dataset. Consequently, the maximum prediction error $E_{\max}$ of the representative dataset is used to derive the parameters as follows:

$$\text{cusum-limit}_+ = E_{\max} \cdot 0.8, \tag{5}$$

$$\text{cusum-reducer}_+ = E_{\max}, \tag{6}$$

$$\text{cusum-max}_+ = \text{cusum-limit}_+ + \text{cusum-reducer}_+ \cdot 5, \tag{7}$$

$$\text{threshold}_+ = \text{cusum-limit}_+ \cdot 2. \tag{8}$$

The determination of the parameters for negative prediction errors is performed analogously.

2.2.2. Station-mode-based fault detection

For each data sample, the station mode is determined using a rule-based approach. The rules are derived from the control strategy described in Section 2.1. Three distinct modes are detected: heating mode, DHW mode and off mode. Any sample that does not match these criteria is classified as undefined. Based on the detected mode, mode-based fault detection strategies are applied. The following first describes the calibration and execution of the mode detection. Thereafter, the mode-based fault detection strategies are explained.

The **mode detection** of the heating mode is calibrated using a seven-day period during winter. The off mode and the DHW mode are calibrated without a specific calibration period. The off mode is calibrated based on the maximum volume flow rate of the substation, as provided by the district heating operator. If the volume flow rate is below 3 % of this maximum, the substation is considered to be in off mode. For the calibration of the DHW mode, the DHW setpoint (T_{DHW}) is queried. If the data point is not available, a standard setpoint of 80 °C is assumed. During mode detection, a sample is in DHW mode if the deviation from the DHW storage charging temperature is less than 5 K. For the heating mode, a substation can have different control strategies to set the secondary supply temperature setpoint. The control strategy can be weather-compensated control with and without night setback, or a fixed setpoint. The control strategy is fixed for each substation over its operating time. This strategy is identified for each substation during the calibration. Therefore, all samples in DHW mode and off mode are removed from the calibration dataset. Samples with a deviation of the secondary supply temperature of more than 3.5 K from the previous time step are removed, as these are considered to be in a transition between modes. Firstly, it is checked if the substation has a fixed setpoint control strategy. The secondary supply temperature setpoint is queried and if all samples of the calibration dataset are within ±5 K of this setpoint, the substation is considered to have a fixed setpoint control strategy. During the mode detection phase, the heating mode is detected if the secondary supply temperature is within the same range. If the fixed setpoint control strategy is not detected, the weather-compensated control strategy is assumed. To determine if night setback is implemented, the continuous secondary supply temperature is approximated by a heating curve with a slope of 1.8, derived from the controller's datasheet [40]. For the calibration dataset, the average difference between the measured secondary supply temperature and the approximated secondary supply temperature is calculated. Using the temperature difference between night and day, an adjusted average is calculated under the assumption that 30 % of the samples are in the night setback period. Based on the adjusted average, the dataset is split into two subsets. If the average secondary supply temperature of these two subsets differs by more than 10 K, a night setback is detected. For mode detection during the fault detection phase, the averages and standard deviations of the two subsets are determined. A sample is classified as being in heating mode if its secondary supply temperature is within two standard deviations of the respective average and the difference from the previous sample is less than ±3.5 K. If a night setback is not detected, the calibration dataset is not split, and the average and standard deviation of the entire dataset are calculated. The mode detection during the fault detection phase is analogous to the case with night setback.

In the **fault detection phase**, after mode detection is complete, the mode-based fault detection strategies are applied. The following faults can be detected: too frequently in DHW mode, too infrequently in heating mode and too frequently in undefined mode. A moving window of 24 h is used to detect these faults. If more than 20 % of the data within the moving window is not available, the mode-based fault detection is skipped for that time step. For each time step, the number of samples in each mode and the average outdoor temperature within the moving window are calculated. Using the total number of available samples in the moving window (n_{total}), the minimum required number of samples in heating mode ($n_{\text{heating,min}}$) and the limit for samples in an undefined mode ($n_{\text{undefined,limit}}$) are determined as follows:

$$n_{\text{undefined,limit}} = 0.5 \cdot n_{\text{total}}, \tag{9}$$

$$n_{\text{heating,min}} = 0.3 \cdot n_{\text{total}} - n_{\text{undefined}}. \tag{10}$$

The maximum number of samples in DHW mode ($n_{\text{DHW,max}}$) is set to a fixed value of 30, and the outdoor temperature threshold for heating mode ($T_{\text{ODA,limit}}$) is set manually for each substation. The fault too frequently in DHW mode is indicated if the number of samples in DHW mode exceeds this maximum. The fault too frequently in undefined mode is indicated if the number of samples in undefined mode is above the threshold. The fault too infrequently in heating mode is evaluated if the current sample falls between October and March and the average outdoor temperature is below its threshold. This fault is indicated if the combined number of samples in heating and DHW mode is below the minimum required value ($n_{\text{heating,min}}$). Samples in DHW mode are included in this count because the secondary supply temperature in DHW mode is also high enough to satisfy the heating demand. The fault indicators for each sample are used in the fault diagnosis stage to isolate the root cause, as described in Section 2.2.4.

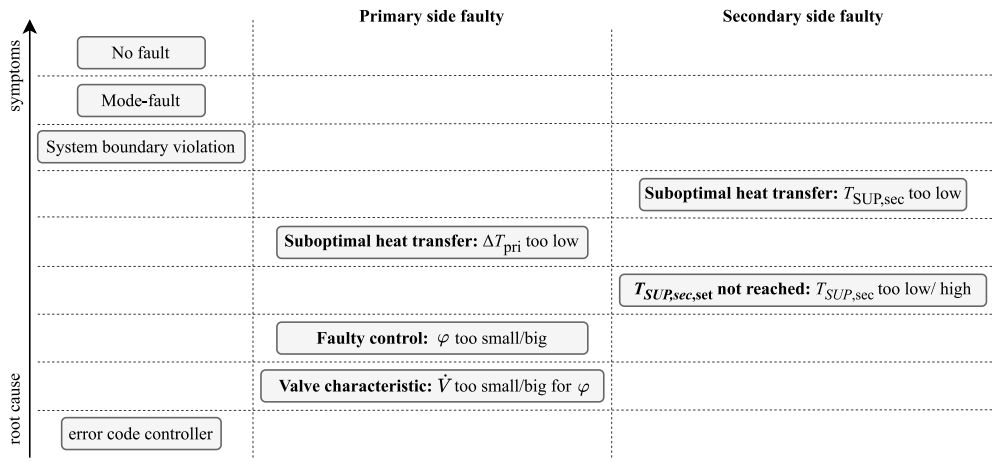


Fig. 3. Fault tree diagram showing the systematic approach to fault diagnosis. The tree illustrates the hierarchical structure of the fault types from root causes to symptoms. The diagram separates the regression model-based faults into primary and secondary side faults.

Table 2
System boundaries for the monitored variables, based on the DH network operator's expertise and the TAB [39], used for fault indication.

Data point	Lower threshold	Upper threshold
$T_{RET,pri}$	20 °C	80 °C
$T_{SUP,pri}$	60 °C	130 °C
$T_{SUP,sec}$	20 °C	100 °C
φ	0 %	100 %
$\dot{V}$	0 m^3/h	–
ΔT_{pri}	0 K	–
ΔT_{SUP}	0 K	–

Table 3
Target variables and features used in the regression models during fault detection.

Regression model	Target	Features
Deviation valve characteristic	$\dot{V}$	φ
Faulty control	φ	$T_{SUP,pri} + T_{ODA} + T_{SUP,sec}$
$T_{SUP,sec,set}$ not reached	$T_{SUP,sec}$	$T_{ODA} + h + station_mode$
Suboptimal heat transfer	ΔT_{pri}	$T_{SUP,pri} + T_{SUP,sec} + \dot{V}$
Suboptimal heat transfer	$T_{SUP,sec}$	$\Delta T_{pri} + \dot{V} + T_{ODA}$

2.2.3. Rule-based system boundary monitoring

The system boundaries for the monitored variables are derived from the TAB [39] and shown in Table 2. Each sample is checked to determine if its monitored variables are within these boundaries. A data point outside the system boundaries indicates a fault. When the substation is in off mode, the volume flow of the substation is zero. Thus, the system boundaries for temperatures and temperature differences are not checked, as temperatures converge to the ambient level. The fault indicators for each sample are then passed to the fault diagnosis.

2.2.4. Fault diagnosis

The fault indicators from the three-layer fault detection are passed to the fault diagnosis, which aims to identify the root cause. To achieve this, the faults are structured hierarchically, as shown in Fig. 3. This structure is based on technical system understanding and expert knowledge. The fault indicators are mapped to this hierarchical fault structure. During diagnosis, the fault indicators are evaluated from the bottom to the top of the hierarchy, i.e., from root causes to symptoms. The first fault indicated in the hierarchical structure is considered the root cause. The diagnosed system state is set to this fault if it persists as the root cause for the last three time steps. This serves as a filter to avoid false alarms. The diagnosis proceeds in the following order: First, the controller's error code is checked. Then, the regression model-based fault indicators are evaluated. These models are designed to detect specific faults, as described in Table 3. Afterwards, the rule-based system boundary monitoring fault indicators are evaluated. Finally, the mode-based fault detection indicators are checked. If no fault is detected, the system state is set to normal operation.

Regression Models

In Fig. 3, the regression model-based faults are separated from the other faults and split into primary and secondary side faults. This separation is made because it is important for the DH network operator's maintenance team to know whether the fault lies within their responsibility. The target variables and features for the regression models are shown in Table 3. For each feature, corresponding lag and delta features are also used. Here, h represents the hour of the day and $station_mode$ is the detected station mode, as described in Section 2.2.2. The regression models are designed to represent different aspects of the system behavior, making it possible to isolate different faults. In the following, each regression model and its corresponding system behavior is described. Unless stated otherwise, deviations in both positive and negative directions are considered.

The fault **deviation valve characteristic** is detected using a regression model with the target variable volume flow rate ($\dot{V}$). This model represents the relationship between the valve position (φ) and the volume flow rate. If the valve characteristic deviates, the volume flow rate is not as expected for a given valve position. The fault **faulty control** is detected by a regression model with the target valve position (φ). The valve position is set by a PID controller to reach the secondary supply temperature setpoint. The outdoor temperature also influences the valve position, as the secondary supply temperature setpoint is typically weather-compensated. Furthermore, the primary supply temperature affects the valve position because it limits the achievable secondary supply temperature. If the control is faulty, the valve position deviates from the expected value for a given primary supply temperature, outdoor temperature, and secondary supply temperature. The fault $T_{SUP,sec,set}$ **not reached** is detected by a regression model with the secondary supply temperature as the target variable. This model investigates whether the secondary supply temperature is as expected for a given outdoor temperature, hour of the day, and station mode. Both the night setback and the station mode influence the secondary supply temperature setpoint. The fault **suboptimal heat transfer** is detected by two regression models. The first model uses the primary temperature difference (ΔT_{pri}) as the target variable. This difference is influenced by the primary supply temperature, the secondary supply temperature, and the volume flow rate. These features approximate the heat transfer in the heat exchanger. A fault is indicated only if

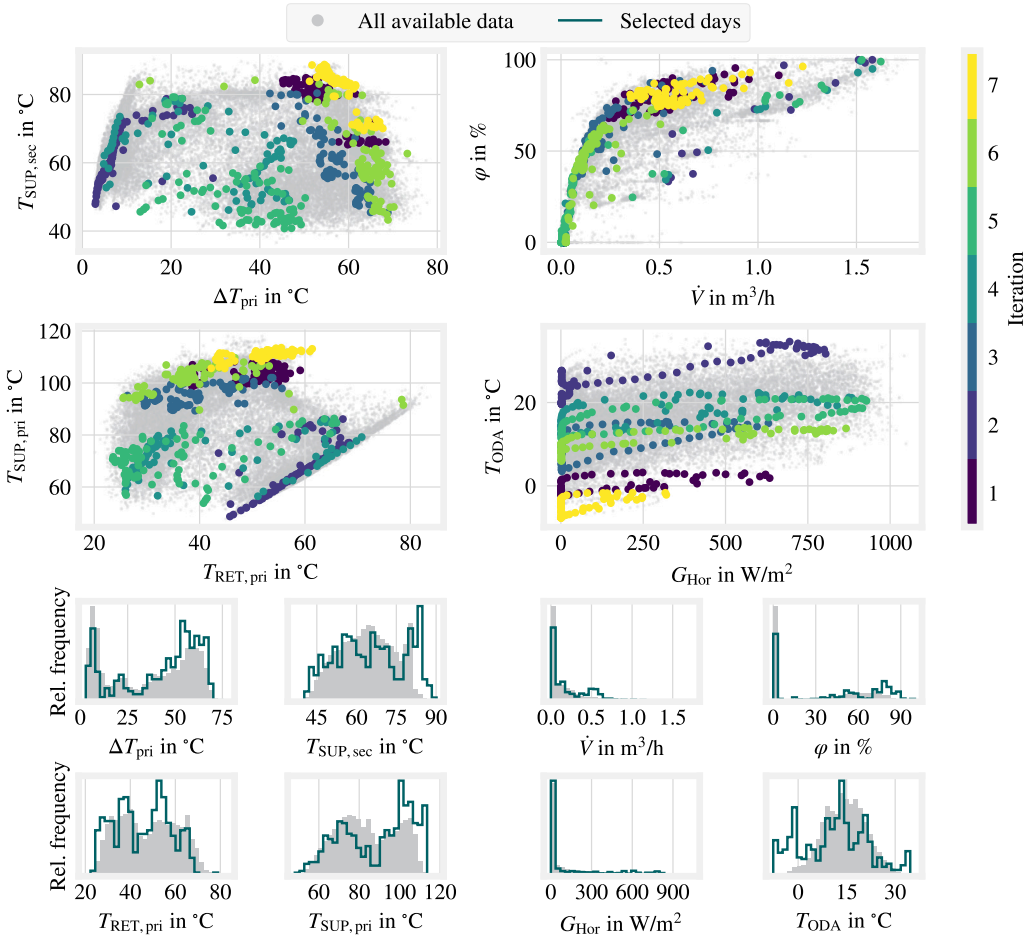


Fig. 4. Data selection for substation 40019356 demonstrated for all iterations of the selection algorithm. Histograms of the selected dataset versus all available data are shown for the different variables.

the primary temperature difference is lower than expected, as a high difference is desired for efficient district heating network operation. The second regression model has the secondary supply temperature ($T_{SUP,sec}$) as the target variable. The primary temperature difference and the volume flow rate approximate the heat transfer. The outdoor temperature is also used as a feature, as it determines the setpoint of the secondary supply temperature. A fault is indicated only if the secondary supply temperature is lower than expected, as this ensures customer comfort.

2.2.5. Evaluation metrics

A sample is diagnosable if sensor data is available. Diagnosable samples outside the fault periods documented by the network operator or found through the analysis are negatives (N_{neg}), and the false alarm rate follows as

$$\mathrm{FAR} = \frac{N_{FA}}{N_{neg}}, \tag{11}$$

where N_{FA} is the number of negatives in which a fault is detected.

3. Results

Section 3.1 demonstrates the data selection in detail for a single substation. Moreover, the data selection of both substations is compared against manual and random data selection and different neighborhood sizes k. Subsequently, Section 3.2 presents the results of the station mode detection for one substation. Finally, the results of the FDD for the kNN-based data selection are contrasted with the results for random and manual data selection in Section 3.3 for both substations. For the kNN-based data selection, two faults of different substations are evaluated in detail.

3.1. Data selection

This subsection demonstrates the data selection using kNN-based novelty detection for substation 40019356 in detail. The kNN-based data selection is compared against manual and random data selection for both substations. The data selection is critical for the performance of the regression models, as the models learn the normal system behavior from the selected data. Fig. 4 illustrates the iteration process of the kNN-based novelty detection. Eight features of the novelty detector are plotted in pairs, and the days added in each iteration are highlighted in different colors. The histograms compare the distribution of the selected days with that of all available data. Two metrics quantify the coverage per feature. The range share is the ratio of the value range of the selected days to the value range of all available data. The bin coverage is the share of histogram bins occupied by all available data that also contain selected samples, using 20 bins per feature. Averaged over the eight features, the seven selected days reach a range share of 93 % and a bin coverage of 94 %. The selected days therefore cover nearly the entire operating range of the substation. The mean two-sample Kolmogorov–Smirnov distance between the selected days and all available data is 0.15. It is largest for the outdoor temperature (0.22), because, as intended by novelty-based selection, rare operating conditions are selected, which are thus overrepresented relative to their frequency.

To assess the sensitivity to the neighborhood size, the selection was repeated for $k = 3$ (K1), $k = 7$ (K2) and $k = 10$ (K3), with $k = 5$ (K0) as

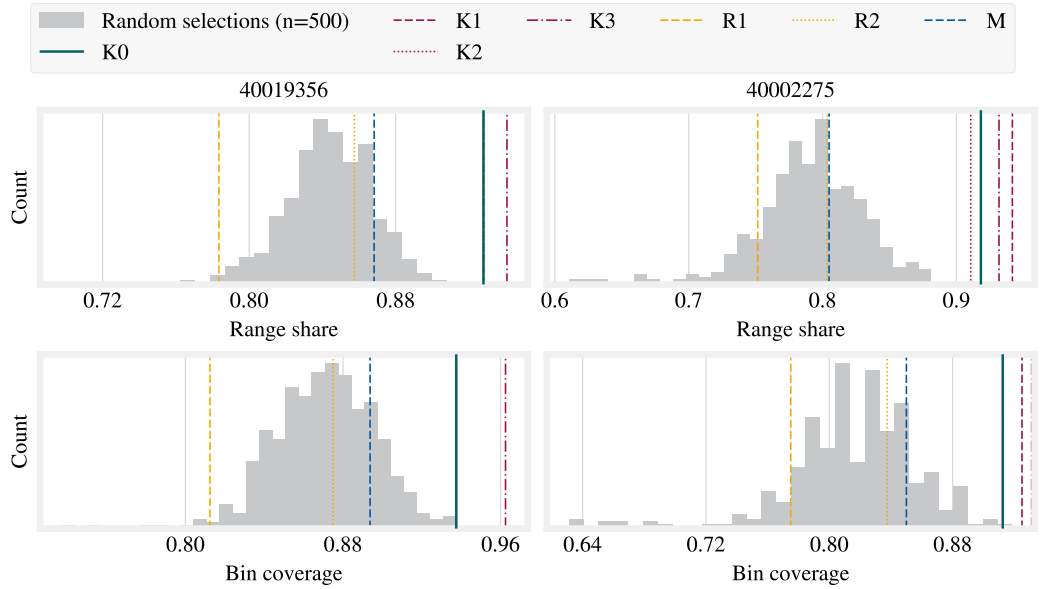


Fig. 5. Range share and bin coverage of the seven selected days relative to all available data for the kNN configurations (K0 to K3), the random selections (R1, R2), the manual selection (M) and 500 random draws of seven days.

the reference configuration. Tables A.4 and A.5 list the selected days. For substation 40019356, K1 and K2 select the same seven days as K0. K3 replaces three days, two by the directly adjacent day and one by a day in the same month. For substation 40002275, K2 differs from K0 in one day and K1 in two days, whereas K3 replaces four of the seven days. Fig. 5 compares the range share and bin coverage of these configurations with 500 random draws of seven days, with two random selections (R1, R2) and with a manual selection (M). R1 and R2 are the random selections with the lowest and the highest coverage out of five random runs. For M, days from all seasons were selected manually. For both substations, all kNN configurations exceed every random draw in range share and lie at the upper edge of or above the random distribution in bin coverage. The manual and random selections lie within the random distribution. Although K3 selects different days for substation 40002275, its coverage remains at the level of K0. The coverage achieved by the novelty-based selection is therefore robust to the choice of k, even where individual days change.

The selection algorithm does not guarantee that the selected days are fault-free. Each selected day requires manual verification before inclusion in the dataset. Data selection stops after seven iterations since the effort for manual verification increases, while the benefit of adding more data decreases. The regression models are trained, validated, and tested using the representative dataset, as described in Section 2.2.1. The MAE, R^2, and RMSE of the regression models on the test dataset for configuration K0 for substation 40019356 and substation 40002275 are given in Tables B.6 and B.7, respectively.

3.2. Mode detection results

In this subsection, the mode detection results are evaluated as an example for substation 40019356. The mode detection showed similar performance for other substations of the Aachen district heating network. Fig. 6 shows the determined station modes and the secondary supply temperature. The data covers two nine-day periods: one in summer and one in winter. The heating mode is predominant during the winter period, as expected due to low outdoor temperatures. The secondary supply temperature shows higher temperatures during the day and lower temperatures at night, which implies weather-compensated behavior with night setback. The algorithm correctly detects heating mode for this behavior. DHW mode events interrupt the weather-compensated behavior. The mode detection clearly identifies these events. Undefined and off modes are detected for some samples. This is expected due to measurement noise and transitions between modes. Adjusting the mode detection parameters could reduce these effects. However, parameter optimization was not the focus of this study. The general performance is considered sufficient for the subsequent mode-based fault detection. In the summer period, the secondary supply temperature follows a typical pattern. Peaks in the secondary supply temperature characterize this pattern. These peaks are caused by DHW load events. After each DHW load event, the temperature slowly decreases towards ambient temperature until the next event. The algorithm mostly detects DHW mode during the peaks and off mode thereafter. This is expected for typical summer operation. Occasionally, the heating mode or undefined mode is detected during DHW load events. Adjusting the mode detection parameters for the specific substation could reduce these misclassifications. However, this would reduce the generalizability of the methodology. Overall, the mode detection shows suitable performance for the FDD methodology.

3.3. Fault detection and diagnosis

The FDD performance is evaluated for two substations in the Aachen district heating network. Both substations are evaluated over the same period of 20 months, corresponding to 58,368 samples each. The data availability is 76.9 % for substation 40019356 and 62.5 % for substation 40002275. Four fault periods are labeled, one for substation 40019356 and three for substation 40002275. They were either documented by the network operator or verified through the analysis of the measurement data. Excluding the samples within these periods leaves 44,729 and 35,498 negatives. All four labeled faults are detected for every selection, and the false alarm rate of K0 is 0.27 % and 0.01 %, respectively. No detection rate is reported, as the documentation of the network operator is based on unstructured technician notes and cannot be assumed to cover all faults that occurred. Fig. 7 compares the FDD results of K0 with those of the random selections (R1, R2) and the manual selection (M). The alarms of the system boundary monitoring and of the mode-based fault detection are identical for all selections, as neither layer depends on the training data. For substation 40019356, these alarms account for all false alarms of K0 and M. The differences between the selections therefore result from the regression models. R1, the random selection with the lowest coverage, contains no winter days

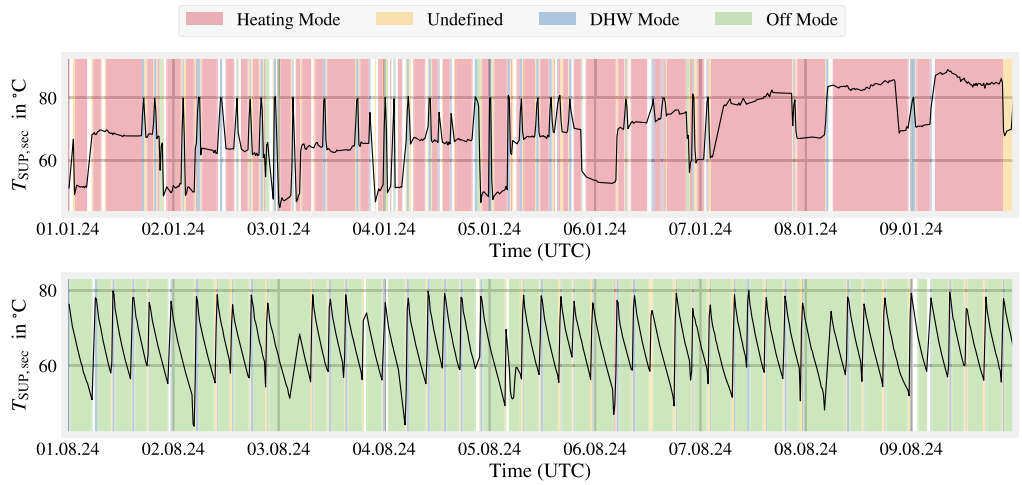


Fig. 6. Mode detection results for substation 40019356 showing determined station modes and secondary supply temperature over two nine-day periods (summer and winter).

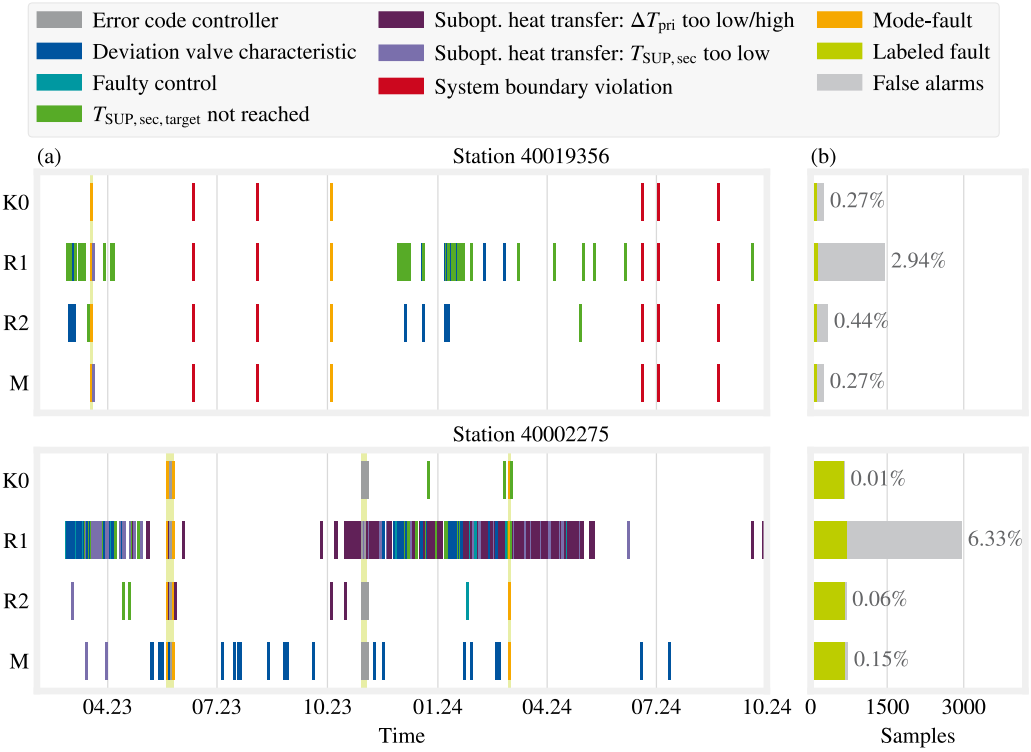


Fig. 7. FDD results for both substations over the entire evaluation period for the kNN-based selection (K0), two random selections (R1, R2) and the manual selection (M). (a) Alarms of each selection over time, colored by fault type. (b) Number of alarm samples per selection, split into alarms within a labeled fault period (green) and false alarms (gray), with the false alarm rate (FAR) next to each bar.

for either substation. Its regression models consequently raise frequent false alarms during the heating periods, resulting in false alarm rates of 2.94 % and 6.33 %. For substation 40019356, M reaches the same false alarm rate as K0, while R2 is slightly higher with 0.44 %. For substation 40002275, R2 (0.06 %) and M (0.15 %) exceed K0, the latter mainly due to repeated alarms for deviation valve characteristic. K0 thus matches or outperforms the manual selection, while automating the search for suitable days, so that expert knowledge is only required to verify fault-freeness. A random selection, in contrast, does not ensure coverage of the operating conditions. As the difference between R1 and R2 shows, its performance varies strongly between draws and deteriorates significantly if relevant conditions are missing. The kNN-based selection therefore achieves the performance of a manual selection with reduced effort and avoids the risk of an insufficient random selection.

3.3.1. Case Study 1: Substation 40019356

Fig. 8 shows the measured secondary supply temperature ($T_{SUP,sec}$), primary volume flow rate ($\dot{V}$) and valve position (φ) of substation 40019356 and the outdoor temperature (T_{ODA}) of the weather station from 15.03.2023 to 19.03.2023. In addition, it contains the diagnosed system state, the residual ($r_{\dot{V}}$) and CUSUM values ($C_{\dot{V}}$) of the deviation valve characteristic model, the 24 h calculation period of the mode-based fault and the fault report documented by the network operator. Until 17.03.2023, $T_{SUP,sec}$ follows a weather-compensated pattern with peaks for DHW demand. On the morning of 17.03.2023, the valve opens fully to 100 % while $\dot{V}$ drops to zero, and $T_{SUP,sec}$ decreases to about 25 °C. For a fully open valve, the regression model predicts a volume flow rate of about 1.2 m^3/h above the measured value. The corresponding CUSUM value exceeds its threshold within a few time steps, and the fault is diagnosed as deviation valve characteristic. At

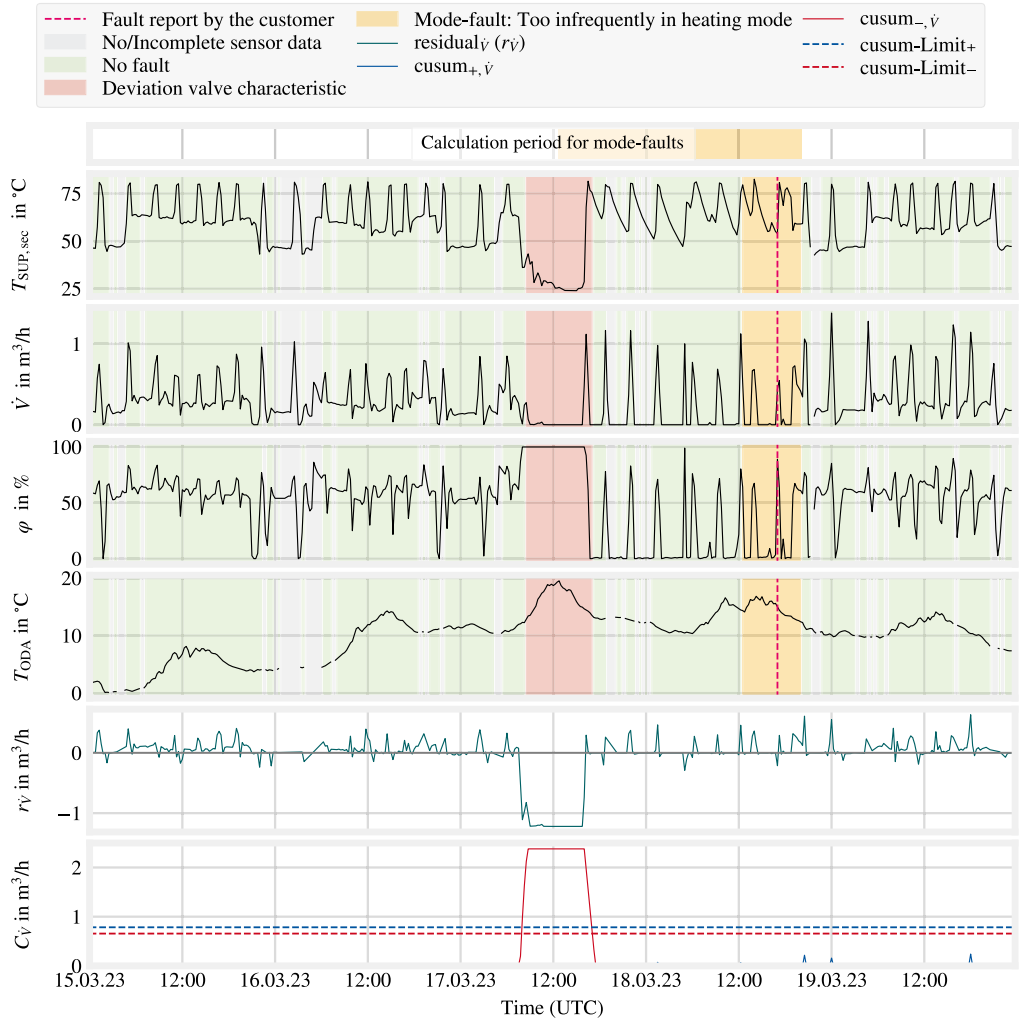


Fig. 8. FDD results, measurements, and residual and CUSUM values of the deviation valve characteristic model for substation 40019356 from 15.03.2023 to 19.03.2023.

17:00, the volume flow resumes and the residual returns to zero. However, the substation now operates in a summer pattern, with short DHW peaks followed by periods in which φ and $\dot{V}$ are zero, although the average outdoor temperature remains below the summer mode threshold of 18 °C. On 18.03.2023 at about 12:30, the number of samples in heating mode within the 24 h window falls below the minimum, and the fault too infrequently in heating mode is indicated. The delay results from the length of this window. In the evening of 18.03.2023, the substation returns to weather-compensated operation. According to the documentation of the network operator, the customer reported the fault on 18.03.2023 at 17:00, and a technician visited the substation on the same day, found no fault and assumed it to be on the customer side. The measurements instead indicate an intervention on 17.03.2023, as the abrupt end of the first fault at 17:00 has no other apparent cause and is consistent with a manual restart. Since the documented date is therefore unreliable, the lead time is measured against the end of the first fault and amounts to 8.5 h. Both faults are detected correctly.

3.3.2. Case Study 2: Substation 40002275

Fig. 9 shows the measured $T_{SUP,sec}$, $\dot{V}$ and φ of substation 40002275 and the outdoor temperature of the weather station from 19.05.2023 to 25.05.2023. In addition, it contains the diagnosed system state, the number of samples in DHW mode within the 24 h window with its threshold, the 24 h calculation period of the mode-based fault and the fault report documented by the network operator. On 19.05.2023, $T_{SUP,sec}$ shows short peaks for DHW demand. From 20.05.2023, it remains at about 65 °C for many consecutive time steps while $\dot{V}$ stays low. The number of samples in DHW mode within 24 h rises to 63, corresponding to almost 16 h, and exceeds the threshold of 30 samples on 20.05.2023. The fault too frequently in DHW mode is therefore indicated until 21.05.2023, when the count falls below the threshold again. From 22.05.2023, the controller repeatedly reports a sensor failure via its error code, while $T_{SUP,sec}$ remains between 30 and 55 °C and $\dot{V}$ is close to zero. According to the documentation of the network operator, the customer reported a fault on 24.05.2023 at 14:00, stating that the DHW temperature had dropped to ambient level. A technician found a broken temperature sensor in the DHW storage tank and replaced it. The broken sensor explains both detected fault periods. It first indicated a permanent charging demand, which kept the substation in DHW mode, and its subsequent failure was reported by the controller. After the replacement, the storage tank is recharged and the substation returns to summer operation with regular DHW peaks. As the recharging falls within the 24 h window, the DHW mode count exceeds the threshold again on 25.05.2023 for a few hours. The fault too frequently in DHW mode is detected 4 days and 8 h before the customer report and 2 days and 4 h before the first error code of the controller. Since the methodology only forwards the error code, this lead time demonstrates the added value of the mode-based fault detection.

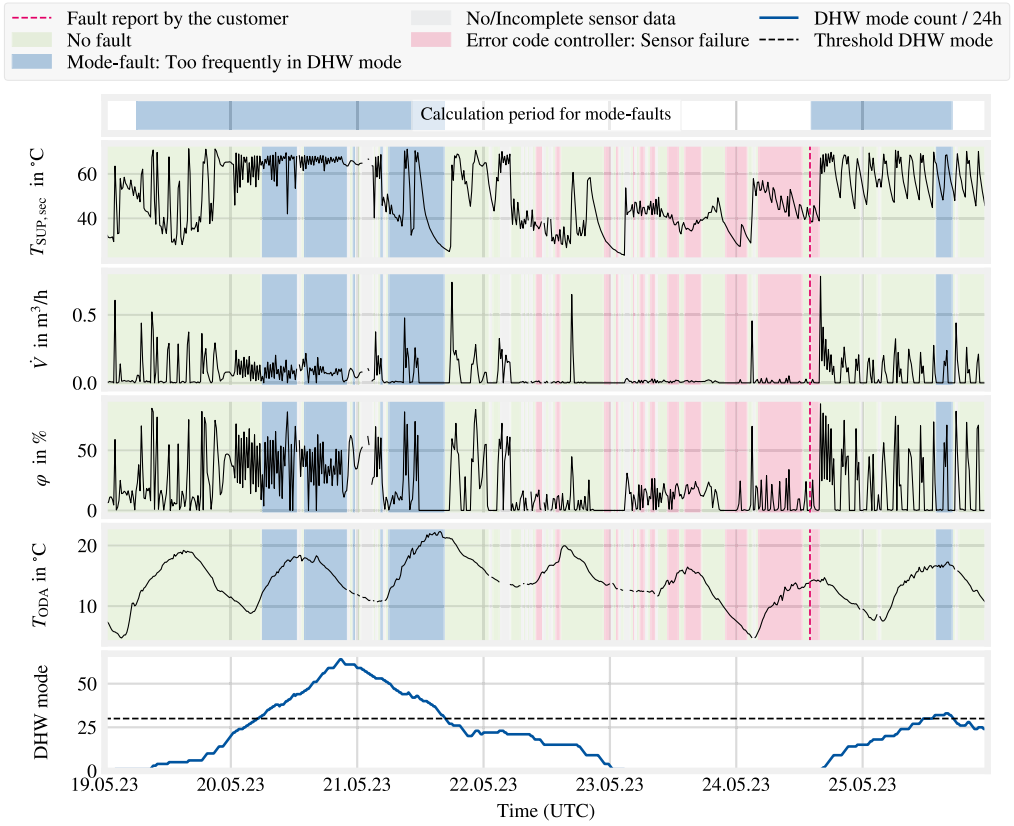


Fig. 9. FDD results, measurements and number of samples in DHW mode within 24 h for substation 40002275 from 19.05.2023 to 25.05.2023.

4. Discussion

The results confirm the capability of the proposed FDD framework to detect and diagnose substation faults. The following discussion examines the data selection process, the trade-offs between detection speed and false alarm rates, the contribution of the individual layers and the transferability of the approach. The proposed kNN-based data selection creates a representative dataset for training the regression models. The selected days cover nearly the entire value range of the available data, exceed all random draws in range share and are robust to the choice of k. The FDD results show that this coverage translates into a low false alarm rate, whereas training data that misses relevant operating conditions, such as the heating period in R1, leads to frequent false alarms. As the test metrics of the regression models are optimistic due to the random split, the selection is assessed by the FDD results rather than by these metrics. However, the selected data requires manual verification to ensure that only fault-free data is included. This manual verification is time-consuming and limits the scalability of the methodology. Nevertheless, the proposed data selection significantly reduces the volume of data requiring manual verification. The manual effort for selecting training data is lower than for methods in the literature, where experts typically select fault-free data manually. Alternatively, a human-in-the-loop approach could enable iterative data cleaning. If a fault remains undetected by the automated FDD, technicians can manually flag the anomalous interval. Subsequently, a kNN-based algorithm can identify and remove similar instances from the training data, facilitating continuous data refinement. The kNN-based data selection allows the methodology to be adapted to different substations with varying operating conditions. Consequently, the kNN-based data selection is applicable for selecting representative datasets for training regression models in district heating substations, assuming the data is fault-free.

In both case studies, the FDD methodology detects the faults correctly, 8.5 h before the presumed technician intervention and 4 days and 8 h before the customer report, respectively. This allows the network operator to schedule maintenance more efficiently (e.g., during the day instead of the night shift) and increases customer satisfaction by resolving faults faster. The two further labeled faults of substation 40002275 are also detected, one by the controller error code (sensor failure) and one by the deviation valve characteristic regression model and the mode-based fault detection. Detections outside the labeled fault periods did not affect the operation of the substations and are counted as false alarms, resulting in a false alarm rate of 0.16 % across both substations for K0. For substation 40019356, all false alarms of K0 originate from the system boundary monitoring and the mode-based fault detection. As these layers use generic thresholds, adjusting them to the individual substation could reduce these false alarms at the cost of additional manual effort. Nevertheless, false alarms require critical consideration. They could lead to unnecessary maintenance actions and reduce the network operator's trust in the methodology. False alarms could be further filtered by increasing the required number of consecutive time steps for setting a fault state. However, this would increase the detection time for real faults and could also lead to missing short faults. Short faults typically do not significantly affect substation operation. Thus, missing these faults is considered acceptable. Due to the low false alarm rate, the current setting is considered a good trade-off between false alarm rate and detection time. In addition, the 24 h window of the mode-based fault detection delays the detection, as observed in both case studies, and can extend an alarm beyond the end of a fault, as after the sensor replacement in Case Study 2. A possible solution to reduce the cost of unnecessary maintenance is to first check the substation data remotely and only schedule a maintenance action if an expert confirms the fault.

As the fault diagnosis is primarily correct, the implemented three-layer fault detection provides useful information for isolating the root cause. However, not all fault types implemented in the hierarchical structure were present in the two analyzed substations. Moreover, Fig.

7 shows the diagnosed root cause after the hierarchical evaluation, in which the first confirmed fault determines the reported state, so a layer can be active in a time step without appearing in the figure. The sparse appearance of individual fault types therefore reflects this prioritization and the limited number of documented faults rather than a limitation of the respective layer. To further evaluate the diagnosis performance, additional fault types must be investigated. The mode-based fault detection effectively supplements the regression model-based detection. It detects faults that are not covered by the regression models, as shown in both case studies. To ensure this enhancement, the mode detection must be sufficiently accurate. This is the case for the currently implemented mode-based faults. Adding more mode-based faults requires further improvements to the mode detection. Examples include too frequently in heating mode during summer or too frequently in off mode during winter. Additional mode-based faults could further enhance the fault diagnosis performance. If additional continuous sensor data becomes available, other regression model-based faults could be implemented.

Further adaptation of the methodology to the Aachen district heating network makes it less generalizable to other networks. The current methodology is already tailored to the specific control strategies and operating conditions of the Aachen network. The regression models are not transferred between substations but retrained for each one, which the automated data selection makes feasible, so that differing load characteristics and control strategies are captured by construction. A differing hydraulic configuration, in contrast, requires adapting the mode-based rules and the fault structure. Applying the methodology to other district heating networks requires adaptation to the specific control strategies, operating conditions, and available sensor data of the respective network. Scaling to a large number of substations is limited by the manual verification of the selected days rather than by computational requirements (see Appendix C for measured timings). The evaluation is based on real-world measurements that include noise and missing data. A systematic sensitivity analysis to sensor errors was not performed, as synthetic perturbations of single variables would not reproduce their physical coupling to the other variables. Nevertheless, the general approach can be transferred to other district heating networks.

5. Conclusion and future work

The evaluation of the proposed FDD methodology shows promising results. It was tested on two substations within the Aachen district heating network. The kNN-based data selection reduces the manual effort for selecting training data. The selected days cover nearly the entire operating range of the substations, and the resulting false alarm rate matches or improves on a manual selection, whereas random selections can perform significantly worse. The methodology combines regression model-based fault detection, mode-based fault detection, and rule-based system boundary monitoring. This combination detects all labeled faults, in the case studies 8.5 h before the presumed technician intervention and 4 days and 8 h before the customer report. The fault diagnosis provides useful information regarding the root cause. The false alarm rate is 0.16 % across the two evaluated substations for the kNN-based selection. This represents a good trade-off between the false alarm rate and detection time. Overall, the proposed methodology demonstrates high potential for improving FDD in district heating substations. It enables more efficient maintenance scheduling and increased customer satisfaction. Additionally, resolving faults earlier can lead to energy savings. The methodology requires adaptation to specific district heating networks. However, the general concept provides a scalable solution for operators to monitor a large number of substations with limited manual effort, as the computational requirements are low and the manual effort is reduced to verifying seven selected days per substation.

Future work should prioritize the investigation of generalizability and scalability. To comprehensively evaluate performance, the methodology should be tested in a larger number of substations and adapted to different district heating networks. Structured fault documentation by the network operator would additionally enable the calculation of detection rates. A key challenge is developing a strategy that minimizes the manual effort required for such adaptation. In parallel, the data selection process should be improved towards automation of the verification of fault-freeness of the selected representative data. Strategies such as unsupervised or active learning approaches represent promising candidates for this purpose. Since substation behavior changes over time, for instance after component replacement or changes to the control settings, procedures for periodically repeating the data selection and retraining the regression models should be developed. The applicability to gradual degradation, sensor drift, and intermittent faults, which lie outside the present focus on hard faults, is a further direction, with the central question being how often retraining is required without absorbing slowly developing faults into the learned normal behavior. Furthermore, investigations should focus on further reducing false alarms without compromising the detection time, for instance, through advanced post-processing of alarm signals. Adapting the generic thresholds of the system boundary monitoring and the mode-based fault detection to the individual substation is a further option. Finally, the scope of application should be expanded by increasing the number of detectable fault types. This could be achieved by utilizing enhanced mode detection or additional sensor data. The broader application could be validated by quantifying the economic benefits for the network operator through a cost–benefit analysis.

CRediT authorship contribution statement

Laura Block: Writing – original draft, Visualization, Software, Methodology, Investigation, Conceptualization. **Martin Rätz:** Writing – review & editing, Supervision, Software, Project administration, Methodology, Funding acquisition, Conceptualization. **Dirk Müller:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used Google Gemini, Cline, Claude, and GitHub Copilot to refine the language, improve readability, and assist with code generation for data visualization. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We gratefully acknowledge the financial support by the Federal Ministry for Economic Affairs and Energy (BMWE), promotional reference 03EN3060D and 03EI4081A.

This research was funded by CETPartnership, the Clean Energy Transition Partnership under the 2022 joint call for research proposals, co-funded by the European Commission (GA No. 101069750) and with the funding organizations detailed on https://cetpartnership.eu/funding-agencies-and-call-modules.

Table A.4
Selected days per configuration for substation 40019356. Days are listed in chronological order. Selected days for different neighborhood sizes (K1, K2, K3) not contained in the dataset selected by configuration K0 are marked in red.

Days	K0	K1	K2	K3	M	R1	R2
1	26.02.23	26.02.23	26.02.23	26.02.23	26.02.23	16.05.23	06.04.23
2	19.04.23	19.04.23	19.04.23	28.04.23	01.05.23	03.07.23	29.04.23
3	09.08.23	09.08.23	09.08.23	09.08.23	01.07.23	04.07.23	25.07.23
4	10.01.24	10.01.24	10.01.24	11.01.24	01.08.23	02.08.23	12.08.23
5	08.06.24	08.06.24	08.06.24	07.06.24	01.01.24	26.06.24	03.03.24
6	13.08.24	13.08.24	13.08.24	13.08.24	01.02.24	27.06.24	21.04.24
7	05.10.24	05.10.24	05.10.24	05.10.24	29.07.24	01.08.24	19.08.24

Table A.5
Selected days per configuration for substation 40002275. Days are listed in chronological order. Selected days for different neighborhood sizes (K1, K2, K3) not contained in the dataset selected by configuration K0 are marked in red.

Days	K0	K1	K2	K3	M	R1	R2
1	26.02.23	26.02.23	26.02.23	26.02.23	26.02.23	09.04.23	04.03.23
2	08.04.23	06.05.23	08.04.23	06.05.23	01.05.23	08.06.23	25.03.23
3	06.05.23	25.06.23	06.05.23	25.06.23	01.07.23	10.06.23	29.06.23
4	25.06.23	28.12.23	25.06.23	28.12.23	01.08.23	07.07.23	17.07.23
5	18.01.24	10.01.24	20.01.24	10.01.24	01.01.24	13.06.24	28.01.24
6	26.04.24	26.04.24	26.04.24	30.07.24	01.02.24	14.06.24	01.04.24
7	12.08.24	12.08.24	12.08.24	05.10.24	29.07.24	20.07.24	07.08.24

Table B.6
MAE, R^2, and RMSE of the regression models on the respective test dataset for the configuration K0 for substation 40019356.

Regression model	MAE	R^2	RMSE
Deviation valve characteristic	0.05 m^3/h	0.9	0.11 m^3/h
Faulty control	5.26 %	0.92	9.4 %
$T_{SUP,sec,set}$ not reached	3.34 K	0.82	5.45 K
Suboptimal heat transfer ΔT_{pri}	2.62 K	0.94	5.09 K
Suboptimal heat transfer $T_{SUP,sec}$	1.48 K	0.96	2.65 K

Table B.7
MAE, R^2, and RMSE of the regression models on the respective test dataset for the configuration K0 for substation 40002275.

Regression model	MAE	R^2	RMSE
Deviation valve characteristic	0.02 m^3/h	0.87	0.05 m^3/h
Faulty control	6.69 %	0.9	10.4 %
$T_{SUP,sec,set}$ not reached	3.26 K	0.87	4.6 K
Suboptimal heat transfer ΔT_{pri}	3.25 K	0.94	5.36 K
Suboptimal heat transfer $T_{SUP,sec}$	2.55 K	0.91	3.98 K

Appendix A. Selected training days per configuration

See Tables A.4 and A.5.

Appendix B. Test performance of the regression models

See Tables B.6 and B.7.

Appendix C. Computational requirements

See Table C.8.

Data availability

The authors do not have permission to share data.

Table C.8
Measured computational requirements of the proposed method, per substation.

	Value
Once per substation	
Training day selection	263 s
Model training and threshold calibration	8.3 s
Total computation	271 s
Online operation	
Inference, five models, one 15 min time step	20.6 ms
Model storage per substation	1.5–2.0 MB

Measured on a single machine (Intel(R) Core(TM) Ultra 7 165H, 22 logical cores, Python 3.12.5, scikit-learn 1.4.2) for a substation with 7 training days at a 15 min resolution.

References

[1] Lund H. Renewable heating strategies and their consequences for storage and grid infrastructures comparing a smart grid to a smart energy systems approach. Energy 2018;151:94–102. http://dx.doi.org/10.1016/j.energy.2018.03.010.

[2] Månsson S, Davidsson K, Lauenburg P, Thern M. Automated statistical methods for fault detection in district heating customer installations. Energies 2019;12(1):113. http://dx.doi.org/10.3390/en12010113.

[3] Marszal-Pomianowska A, Leiria D, Johra H, Pomianowski M, Praulins I, Anoruo JCA. Fault detection in district heating substations: Overview of real-life faults in residential heating installations. In: Jørgensen BN, Ma ZG, Wijaya FD, Irnawan R, Sarjiya S, editors. Energy informatics. Cham: Springer Nature Switzerland; 2025, p. 357–64. http://dx.doi.org/10.1007/978-3-031-74738-0_23.

[4] Gadd H, Werner S. Achieving low return temperatures from district heating substations. Appl Energy 2014;136:59–67. http://dx.doi.org/10.1016/j.apenergy.2014.09.022.

[5] Bastidas E Guevara, Faulstich S, Dittmer H, Neumayer M, Mohan GS, Sercan-Calismaz K, Hosenfelder F, Glenewinkel T, Fischer-Florschütz K, Cadenbach A. Prioritisation of faults in district heating substations: Towards predictive maintenance and optimised operation. Energy 2025;333:137210. http://dx.doi.org/10.1016/j.energy.2025.137210.

[6] Leiria D, Andersen KH, Melgaard SP, Johra H, Marszal-Pomianowska A, Piscitelli MS, Capozzoli A, Pomianowski MZ. Towards automated fault detection and diagnosis in district heating customers: Generation and analysis of a labeled dataset with ground truth. In: Proceedings of building simulation 2023: 18th conference of IBPSA. IBPSA, Shanghai, China, 2023, p. 3615–23. http://dx.doi.org/10.26868/25222708.2023.1576.

[7] Kleppmann N. From the substation outwards – anomalies, prediction and optimization. In: Proceedings of the IEA DHC annex TS4 conference. Kassel, Germany: Fraunhofer Institute for Energy Economics and Energy System Technology (IEE); 2023, p. 41, [Accessed 30 March 2026]. URL https://www.iea-dhc.org/fileadmin/documents/Annex_TS4/20231121_IEA_DHC_Annex_TS4_Conference_Proceedings.pdf.

[8] van Dreven J, Boeva V, Abghari S, Grahn H, Koussa J Al, Motoasca E. Intelligent approaches to fault detection and diagnosis in district heating: Current trends, challenges, and opportunities. Electronics 2023;12(6):1448. http://dx.doi.org/10.3390/electronics12061448.

[9] Yang Q, Salenbien R, Smith KM, Tunzi M. Identifying untraced faults associated with high return temperatures from heating systems in buildings connected to district heating networks. Energy 2024;309:133097. http://dx.doi.org/10.1016/j.energy.2024.133097.

[10] Sandin F, Gustafsson J, Delsing J, Eklund R. Basic methods for automated fault detection and energy data validation in existing district heating systems. In: Proceedings of the 13th international symposium on district heating and cooling, 3–4 September 2012. Copenhagen, Denmark: District Energy Development Center; 2012, URL https://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-40519.

[11] Bishop CM. Pattern recognition and machine learning, information science and statistics. New York: Springer; 2006.

[12] Calikus E, Nowaczyk S, Sant'Anna A, Gadd H, Werner S. A data-driven approach for discovering heat load patterns in district heating. Appl Energy 2019;252:113409. http://dx.doi.org/10.1016/j.apenergy.2019.113409.

[13] van Dreven J, Cheddad A, Alawadi S, Ghazi AN, Koussa J Al, Vanhoudt D. SHEDAD: SNN-enhanced district heating anomaly detection for urban substations. In: 2024 9th international conference on fog and mobile edge computing. FMEC, 2024, p. 130–7. http://dx.doi.org/10.1109/FMEC62297.2024.10710205.

[14] Leiria D, Johra H, Anoruo J, Praulins I, Piscitelli MS, Capozzoli A, Marszal-Pomianowska A, Pomianowski MZ. Is it returning too hot? Time series segmentation and feature clustering of end-user substation faults in district heating systems. Appl Energy 2025;381:125122. http://dx.doi.org/10.1016/j.apenergy.2024.125122.

[15] Farouq S, Byttner S, Bouguelia M-R, Nord N, Gadd H. Large-scale monitoring of operationally diverse district heating substations: A reference-group based approach. Eng Appl Artif Intell 2020;90:103492. http://dx.doi.org/10.1016/j.engappai.2020.103492.

[16] van Dreven J, Cheddad A, Ghazi AN, Alawadi S, Koussa J Al, Vanhoudt D. HEAT: Hierarchical-constrained encoder-assisted time series clustering for fault detection in district heating substations. Energy AI 2025;21:100548. http://dx.doi.org/10.1016/j.egyai.2025.100548.

[17] Hermans C, Koussa J Al, Oevelen T Van, Vanhoudt D. Fault detection for district heating substations: Beyond three-sigma approaches. Smart Energy 2024;16:100159. http://dx.doi.org/10.1016/j.segy.2024.100159.

[18] Theusch F, Klein P, Bergmann R, Wilke W, Bock W, Weber A. Fault detection and condition monitoring in district heating using smart meter data. PHM Soc Eur Conf 2021;6(1). http://dx.doi.org/10.36001/phme.2021.v6i1.2786, 11–11.

[19] Månsson S, Kallioniemi P-OJ, Sernhed K, Thern M. A machine learning approach to fault detection in district heating substations. Energy Procedia 2018;149:226–35. http://dx.doi.org/10.1016/j.egypro.2018.08.187.

[20] Langroudi P Pourbozorgi, Weidlich I, Hay S. Backward simulation of temperature changes of district heating networks for enabling loading history in predictive maintenance. Energy Rep 2021;7:119–27. http://dx.doi.org/10.1016/j.egyr.2021.09.031.

[21] van Dreven J, Boeva V, Abghari S, Grahn H, Koussa J Al. A systematic approach for data generation for intelligent fault detection and diagnosis in district heating. Energy 2024;307:132711. http://dx.doi.org/10.1016/j.energy.2024.132711.

[22] Kim R, Hong Y, Choi Y, Yoon S. System-level fouling detection of district heating substations using virtual-sensor-assisted building automation system. Energy 2021;227:120515. http://dx.doi.org/10.1016/j.energy.2021.120515.

[23] Jallal MA, Vallée M, Lamaison N. Fouling fault detection and diagnosis in district heating substations: Validation of a hybrid CNN-based PCA model with uncertainty quantification on virtual replica synthesis and real data. Energy 2024;312:133590. http://dx.doi.org/10.1016/j.energy.2024.133590.

[24] Xue P, Shi L, Zhou Z, Liu J, Chen X. An online fault detection and diagnosis method of sensors in district heating substations based on long short-term memory network and adaptive threshold selection algorithm. Energy Build 2024;308:114009. http://dx.doi.org/10.1016/j.enbuild.2024.114009.

[25] van Dreven J, Cheddad A, Alawadi S, Ghazi AN, Koussa J Al, Vanhoudt D. From data scarcity to diagnostic precision: A novel data augmentation and fault diagnosis framework for district heating substations. Eng Appl Artif Intell 2025;151:110662. http://dx.doi.org/10.1016/j.engappai.2025.110662.

[26] van Dreven J, Cheddad A, Alawadi S, Ghazi AN, Koussa J Al, Vanhoudt D. From bearings to substations: Transfer Learning for fault detection in district heating. Energy 2025;335:138016. http://dx.doi.org/10.1016/j.energy.2025.138016.

[27] Zhang F, Fleyeh H. Anomaly detection of heat energy usage in district heating substations using LSTM based variational autoencoder combined with physical model. In: 2020 15th IEEE conference on industrial electronics and applications. ICIEA, 2020, p. 153–8. http://dx.doi.org/10.1109/ICIEA48937.2020.9248108.

[28] Wang Z, Qin Y, Kong Y, Wang L, Leng Q, Zhang C. Advanced fault detection, diagnosis and prognosis in HVAC systems: Lifecycle insight, key challenges, and promising approaches. Renew Sustain Energy Rev 2025;219:115867. http://dx.doi.org/10.1016/j.rser.2025.115867.

[29] Dietz S, Réhault N, Frison L, Rist T, Scholzen F. Increasing transferability for automated fault detection and diagnosis in HVAC systems through a hybrid AI methodology. Open Res Eur 2025;5:91. http://dx.doi.org/10.12688/openreseurope.19029.1, version 1; peer review: awaiting peer review.

[30] Mbiydzenyuy G, Nowaczyk S, Knutsson H, Vanhoudt D, Brage J, Calikus E. Opportunities for machine learning in district heating. Appl Sci 2021;11(13):6112. http://dx.doi.org/10.3390/app11136112.

[31] Stecher D, Ziegltrum L, Reiprich P, Fuchs C, Maier A, Schmidt J. Neural network synthetic dataset generation for fault detection in district heating substations. Smart Energy 2025;20:100206. http://dx.doi.org/10.1016/j.segy.2025.100206.

[32] Neumayer M, Stecher D, Grimm S, Maier A, Bücker D, Schmidt J. Fault and anomaly detection in district heating substations: A survey on methodology and data sets. Energy 2023;276:127569. http://dx.doi.org/10.1016/j.energy.2023.127569.

[33] Pimentel MAF, Clifton DA, Clifton L, Tarassenko L. A review of novelty detection. Signal Process 2014;99:215–49. http://dx.doi.org/10.1016/j.sigpro.2013.12.026.

[34] Rätz M, Henkel P, Stoffel P, Streblow R, Müller D. Identifying the validity domain of machine learning models in building energy systems. Energy AI 2024;15:100324. http://dx.doi.org/10.1016/j.egyai.2023.100324.

[35] Andersen KH, Melgaard S Pommerencke, Johra H, Marszal-Pomianowska A, Jensen R Lund, Heiselberg P Kvols. Barriers and drivers for implementation of automatic fault detection and diagnosis in buildings and HVAC systems: An outlook from industry experts. Energy Build 2024;303:113801. http://dx.doi.org/10.1016/j.enbuild.2023.113801.

[36] Rätz M, Streblow R, Müller D. Diagnosis in digitized district heating network substations: A machine learning regression approach. In: Proceedings of the 15th ACM international conference on future and sustainable energy systems. e-Energy'24, New York, NY, USA: Association for Computing Machinery; 2024, p. 472–3. http://dx.doi.org/10.1145/3632775.3661971.

[37] Deutsches Institut für Normung e V. DIN EN 12831-1:2017-09: Energetische bewertung von gebäuden – verfahren zur berechnung der norm-heizlast – Teil 1: Raumheizlast, modul M3-3; deutsche fassung EN 12831-1:2017. Berlin: DIN Media; 2017, http://dx.doi.org/10.31030/2571775.

[38] Deutsches Institut für Normung e V. DIN 4708-1:1994-04: Zentrale wassererwärmungsanlagen – begriffe und berechnungsgrundlagen. Berlin: DIN Media; 1994, http://dx.doi.org/10.31030/2480131.

[39] STAWAG Stadt- und Städteregionswerke Aachen AG. TAB fernwärme – technische anschlussbedingungen für die versorgung mittels heizwasser aus den fernwärmenetzen der STAWAG. 2024, [Accessed 11 October 2024]. URL https://www.stawag.de/fileadmin/stawag/content/Dokumente/Waerme/24-10-01_TAB_neu.pdf.

[40] SAMSON AG. Einbau- und bedienungsanleitung EB5579: Heizungs- und fernheizungsregler TROVIS 5579. 2023, Frankfurt am Main, [Accessed 01 October 2024]. URL https://www.samsongroup.com/document/e55790de_2.6x.pdf.

[41] Probst P, Wright MN, Boulesteix A-L. Hyperparameters and tuning strategies for random forest. WIREs Data Min Knowl Discov 2019;9(3):e1301. http://dx.doi.org/10.1002/widm.1301.