Energy 276 (2023) 127569

Contents lists available at ScienceDirect

Energy

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

Fault and anomaly detection in district heating substations: A survey on methodology and data sets

Martin Neumayer [a], Dominik Stecher [b], Sebastian Grimm [c], Andreas Maier [d], Dominikus Bücker [e,a], Jochen Schmidt [b,*]

[a] Institut für nachhaltige Energieversorgung GmbH (INEV), Eduard-Rüber-Str. 7, 83022 Rosenheim, Germany
[b] Rosenheim Technical University of Applied Sciences, Dep. of Computer Science, Hochschulstr. 1, 83024 Rosenheim, Germany
[c] AGFW-Projektgesellschaft für Rationalisierung, Information und Standardisierung mbH, Stresemannallee 30, 60596 Frankfurt/Main, Germany
[d] Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nürnberg, Martensstr. 3, 91058 Erlangen, Germany
[e] Rosenheim Technical University of Applied Sciences, Dep. of Engineering Sciences, Hochschulstr. 1, 83024 Rosenheim, Germany

ARTICLE INFO

Keywords:
District heating systems
Fault/anomaly detection
Machine learning

ABSTRACT

District heating systems are essential building blocks for affordable, low-carbon heat supply. Early detection and elimination of faults is crucial for the efficiency of these systems and necessary to achieve the low temperatures targeted for 4th generation district heating systems. Especially methods for fault and anomaly detection in district heating substations are currently of high interest, as faults in substations can be repaired quickly and inexpensively, and smart meter data are becoming widely available. In this paper, we review recent scientific publications presenting data-driven approaches for fault and anomaly detection in district heating substations with a focus on methods and data sets. Our review indicates that researchers use a wide variety of methods, mostly focusing on unsupervised anomaly detection rather than fault detection. This is due to a lack of labeled data sets, preventing the use of supervised learning methods and quantitative analysis. Together with the lack of publicly available data sets, this impedes the accurate comparison of individual methods. To overcome this impediment, increase the comparability of different methods and foster competition, future research should focus on establishing publicly available data sets, and industry-relevant metrics as benchmarks.

1. Motivation

About half of the EU's energy consumption accounts for heating and cooling of buildings and industries. District heating represented only 12% of the supply in Europe [1], however, it is seen as a crucial element to achieve the decarbonization of the heating and cooling sector. Several studies underline the key role of district heating expansion in the energy or heat transition to achieve the climate targets, both on a national level, e. g., in Germany [2,3], Spain [4], Denmark [5,6], and EU wide [7,8]. This can be done by upgrading to 4th and 5th generation district heating networks and increasing the share of renewable energies in combined heat and power plants.

The history of DHS starts in the early 19th century with the idea to simply connect heat sources with suitable heat demands by a distribution network or grid. In this way, heat can be distributed efficiently in urban areas and is directly used by customers [9]. Therefore, three steps are necessary along the district heating (DH) supply chain: heat generation, heat transition and heat consumption. While the main idea of DHS is still the same today, there have been massive developments in each of these steps during the past century. Lund et al. [10] defined four generations of DHS clustered by similar properties and significant technological advancement in all three sub-systems, leading to increased energy efficiency of the total DHS while the temperature level of the transport medium decreases. In general, utilities constantly strive for lower temperature levels, as the reduction of DHS temperature level improves almost all efficiencies of the entire DHS, reducing heat losses in the distribution system and increasing the number of possible usable heat sources [11–13].

The fourth generation of district heating systems is characterized, among other aspects, by the ability to use heat from renewable sources and distribute it in a low-temperature grid [10]. To reach the targeted temperatures, an accurate forecast of the heating demand and efficient operation of the whole system is required [10]. In contrast, previous studies have identified substations as a frequent source of faults and inefficiencies [11]. While faults at substations can often be fixed quickly and inexpensively [12], the identification of these faults remains difficult even if the sub-optimal operation is known [15]. However, more

* Corresponding author.
E-mail address: jochen.schmidt@th-rosenheim.de (J. Schmidt).

https://doi.org/10.1016/j.energy.2023.127569
Received 19 December 2022; Received in revised form 27 February 2023; Accepted 17 April 2023
Available online 20 April 2023

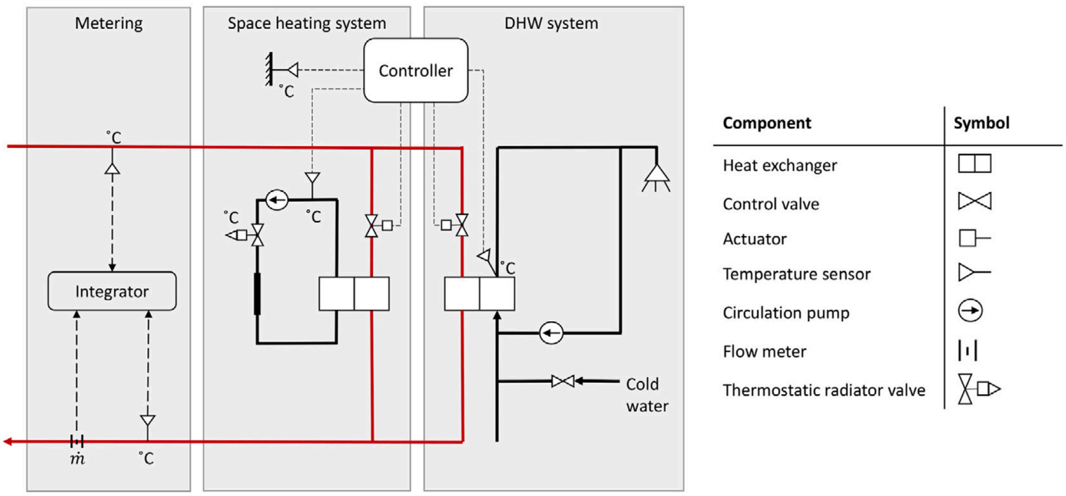

Fig. 1. Schematic diagram of a parallel-connected substation. Reprinted from [14] with permission from Elsevier.

and more substation data are becoming available as digitalization and legal requirements advance, e. g., [16] in Germany. Lastly, the current trend of massively increasing prices for electricity and primary energy sources will increase the pressure to identify inefficiently operating substations in any district heating system.

Data-driven methods show promising results for user demand prediction in other utility sectors such as electricity [17,18], water [19], and district heating [20,21]. Using them for detecting faults and anomalies automatically, or assisting humans to do so, can bridge the gap between hard identification and easy-to-fix and have gained interest in the scientific community as well as with practitioners. Digital support promises two main advantages: Faults can be detected and fixed early on, keeping the effects on the system to a minimum. Further, the personnel needed to identify faults can be minimized. While interest in the topic increases, to the authors' knowledge a comprehensive review focused on the methods and data sets used for fault and anomaly detection in district heating substations does not exist. Therefore, we conduct a structured literature review in this paper to address the following research questions:

1. What methods are used to detect faults and anomalies in district heating substations?
2. What data sets are used to evaluate (and possibly train) the methods?

The remainder of this paper is structured as follows: Section 2 introduces the most important terms as well as the scope of our review. In Section 3, we elaborate on our methodology and report our findings on the used methods and data sets. Lastly, Section 4 summarizes our findings, discusses challenges that still need to be overcome, and identifies how further research can help address these challenges.

2. Scope and related work

2.1. Fault and anomaly detection

In this section, we clarify our understanding of the fundamental concepts, i. e., fault and anomaly detection. Following Chandola et al. [22], we define anomalies as patterns in data that do not conform to a previously defined notion of expected and normal behavior. Similar or synonymous terms found in our survey are outliers, deviations, atypical behavior, or abnormal patterns. While anomalies may indicate faults or inefficient usage, there may also be other reasons, such as changes in user behavior. Also, the boundary between normal and anomalous behavior is often fuzzy and hard to define. Therefore, anomalous observations close to the boundary can actually be normal and vice-versa [22]. Anomaly detection can be defined as the task of finding these non-conforming, atypical patterns in data [22]. Within the scope of our work, anomaly detection is often achieved using unsupervised learning (e. g., [23]). Unlike fault detection, this makes the process of anomaly detection less time and labor intensive as no annotated data set is required.

In contrast, a fault is defined as a change in a system so that it can no longer operate satisfactorily and meet the requirements of its user [24]. Typically, to fix a fault, corrective action is required, such as replacement or maintenance. Based on this notion, we can define fault detection as the problem of finding changes in a system so that it can no longer operate in accordance with the requirements of its user. Fault detection typically requires some form of a labeled data set, either providing annotated [25,26] or synthetically created [27] fault examples, or providing reference substations to compare against. Using labeled data, fault detection can directly return detailed information about the nature of the fault.

Overall, fault detection (relying on labeled data sets) is typically done using classification (Section 3.2.5) or visualization (Section 3.2.1), whereas anomaly detection often uses clustering (Section 3.2.6), regression (Section 3.2.4), and thresholds (Section 3.2.2).

2.2. Scope: District heating substations

In this article, we focus on faults and anomalies in DH substations. A substation hydraulically separates the water in the DHS from that in the customer installation. Fig. 1 shows a schematic diagram of a parallel-connected substation by Månsson et al. [14], who give a detailed description of its components. The work also includes a compendium of typical faults: heat exchanger, control system and controller, actuators, control valves, and internal heating system/domestic hot water system (DHW) on the customer side. Their reported findings from Swedish utilities show that almost two-thirds of all faults are leakages (33%) and faults in the internal heating system (31%); 13% affect control valves, 10% actuators. The remaining faults are distributed between inferior gaskets, heat exchangers, and control system and controller. In a workshop and bilateral interviews, Austrian DHS operators reported 113 faults, which are distributed among various components as follows: substation valve (24%), substation control system (15%), secondary-side distribution line (14%), substation (12%), domestic hot water system (12%), space heating system (8%), others (15%) [28]. Sandin et al. [29,30] add that, besides defective components, faults and anomalies can also be caused by incorrect installation or configuration, fraud, and communication issues between components.

2.3. Related work

Fundamental organizational aspects of fault and anomaly detection are addressed in [31,32]: In collaboration with district heating utilities, the authors put forward a taxonomy for labeling deviations in district heating customer data [31] and suggest a fault handling process [32]. Dealing with faults, fouling, and inefficiency in substations is also proposed as part of strategies for reducing the supply temperature [33]. Leoni et al. discuss the development of new business models for reducing return temperatures, e. g., by increasing customer engagement in fault detection in [28]. While these papers focus on organizational aspects, we consider technical and methodological aspects.

Other surveys address the current state of [34] or trends in [35] district heating research, without going into details on fault and anomaly detection due to the broad range of different topics. A review considering control strategies and fault detection in district heating and cooling systems is presented by [15]; software- and hardware-based solutions are addressed. Yet, it lacks a detailed analysis of data sets and focuses on general methods and detecting specific faults, not on specific classes of algorithms. An extensive survey of machine learning methods applied in DHS can be found in [36,37]. Both articles feature sections about fault and anomaly detection in DHS, including the distribution network or heat supply. However, due to their focus on machine learning, other methods apart from machine learning are not covered.

In contrast to these papers, we concentrate on fault and anomaly detection in DH substations and consider all kinds of algorithms without restrictions. We differentiate individual classes of algorithms from the viewpoint of a machine learning expert and describe the characteristics of data sets used. This is in contrast to the district heating expert point of view taken by most other review papers and can improve interdisciplinary cooperation on the topic.

3. Survey: Methods and data sets

3.1. Methodology

In this section, we document our methodology used to find, filter, and categorize suitable articles. To find suitable articles, we used the Google Scholar search engine. We collected the found literature, then excluded duplicates and publications without peer review. Other surveys are discussed in Section 2.3 and were also excluded from the review. We further excluded articles that do not present a data-driven method and surveyed the remaining 25 articles in detail regarding the methods and data sets used. To do so, we categorize the articles according to the methods used and report our findings in Section 3.2. An overview is shown in Table 1. Our findings on the used data sets are presented in Section 3.3.

3.2. Methods for fault and anomaly detection

The methods used for fault and anomaly detection and their application are described below. Our categorization includes typical machine learning methods, such as classification, regression, and clustering, as well as more traditional techniques, such as manual analysis, thresholds, and physical models. While we discuss the methods or categories individually, it should be noted that the methods presented can be used both separately or in combination. Also note that we use the terminology applied in the cited articles, even if it does not match our definitions (see Section 4 for a discussion). We assign at least one of the following categories to each article:

Visualization and Manual Analysis In manual analysis, human experts examine substation measurements or derived metrics with the help of visualizations. Due to the use of human experts, manual analysis offers great flexibility, although it is also subjective. In general, manual analysis can be used to find anomalies and faults or to further investigate anomalies and faults already detected.

Threshold-based Methods Threshold-based methods use a threshold to decide the state of a DHS. Thresholds can be set manually by a human expert, according to assumptions about the error rate (e. g., 5% error rate) or after statistical analysis (e. g., three standard deviations).

Physical Models Physical models simulate the hydraulic network or substation parts, e. g., to detect deviations from the simulated normal conditions. This can be done without personalized user information but at the cost of accuracy and model fidelity. Models that determine physical relationships through regression are not categorized as physical models in this paper but as regression models.

Regression Regression aims at estimating the relationship between a dependent variable and one or more independent variables in a given data set. Given the independent variables as input data, a regressor predicts the dependent variable as a real-valued output. In the context of DHS, regression is typically used to forecast a certain variable and evaluate the deviation between predicted and measured values.

Classification Classification refers to the machine learning task of assigning an item to one of several given classes. While many other methods have the same goal, we limit this category to approaches that employ a classification algorithm to directly determine whether or which fault has occurred. Unlike a regressor, which outputs a real number, the result of a classifier is the assignment of an item to one of a finite set of classes, typically represented by an integer number. In contrast to clustering, classification algorithms learn from annotated or labeled examples of faults. Typical classification algorithms include neural networks, decision trees, and support vector machines.

Clustering Clustering is also concerned with assigning items to groups or clusters. However, instead of learning from annotated examples, clustering organizes the items in a way that they share more similar properties with items in their cluster than with those in other clusters. Therefore, no annotated examples are needed. Clustering algorithms are used to either group DHS according to certain performance indicators or their consumption patterns. Outliers or marginal clusters can then be identified and further analyzed. Typically used algorithms are k-means, k-shape, and k-nearest-neighbors.

Other Methods Some fault and anomaly detection methods, e. g., isolation forest or autoencoder, do not fit into any of the previously mentioned categories. These methods are subsumed here.

3.2.1. Visualization and manual analysis

Manual analysis is used to detect anomalies and faults in visualizations of raw substation measurements or metrics calculated from those measurements. In [11], hourly meter readings of Swedish substations are analyzed manually and using thresholds to detect faults and their symptoms. The authors identify unsuitable heat load patterns and poor substation control by visualization and manual analysis. Unsuitable heat load patterns are defined as heat load patterns that do not fit the substation's customer category, e. g., an office building should have an increased heat demand on a working day. Poor substation control is detected by irregular oscillations and a bad correlation between the substation's heat demand and the outdoor temperature. Gadd and Werner [42] introduce typical heat load patterns for different DHS customer groups as well as two descriptive parameters to identify outliers in visual analysis. The authors put forward the annual relative daily variation and the annual relative seasonal variation as parameters that can be used to manually identify outliers

Table 1
Overview — paper categorization.

Paper	Manual analysis	Thresholds	Physical model	Regression	Classification	Clustering	Other methods
[11]	•	•					
[23]	•	•				•	
[25]					•		
[26]		•		•			•
[27]	•			•			
[38]	•					•	•
[39]	•					•	
[40]	•					•	
[41]	•		•				
[42]	•						
[43]	•	•		•			
[44]	•						
[45]	•						
[46]		•	•		•		
[47]		•		•			
[29]	•	•				•	
[48]	•					•	
[49]		•		•		•	
[50]		•					
[51]			•		•		
[52]		•				•	•
[53]				•		•	
[54]				•		•	
[55]		•		•			
[56]						•	•

in corresponding visualizations. Gadd and Werner present a temperature difference signature-based visualization method to identify faults manually in [43]. The authors refer to the temperature difference signature as a diagram, where the daily average temperature difference is visualized as a function of the daily average outdoor temperature. The diagram is supplemented with a regression line and threshold lines to aid manual analysis. Article [27] uses manual analysis to detect deviations by visually comparing predicted values from a gradient boosting regressor to the measured values. A method to assist operators in detecting fouling in substations' heat exchangers is presented in [44]. To detect fouling the behavior of a given heat exchanger in a clean state, i. e., directly after cleaning, is compared to the behavior of the same heat exchanger in its current state. The results of the comparison are then visualized in different ways.

Other approaches focus on creating more complex visualizations, e. g., based on pattern mining or clustering analysis performed on substation measurements: Abghari et al. [38] propose visualization techniques based on pattern mining and clustering analysis to aid domain experts in interpreting substations' behavior and fault detection. In addition, [40] proposes an approach to support operators in manually analyzing DHS performance and fault detection. An operator can use the approach to visualize clusters of substations stepwise by their features, i. e., after grouping the substations by one feature, the operator can group the remaining substations by another feature. In a comparable approach, Abghari et al. [39] model a substation's operational behavior by extracting weekly patterns and clustering the resulting patterns into similar groups. The substation behavior models are then linked to performance indicators, such as substation effectiveness or the difference between primary and secondary return temperatures. The authors present visualization results that may aid practitioners in analyzing substation behavior and detecting deviations. In [45], the operational behavior of three substations is studied using different visualization techniques and performance metrics, such as the Pearson coefficient. The authors propose to examine the substation's functionality visually with contour mapping, parallel coordinates, and scatter plot matrices. Outliers and relationships between the substations' measurements are identified using the aforementioned techniques.

Lastly, manual analysis is used to examine suspicious patterns or anomalies found by other means in greater detail. Often the aim is to identify the cause of the anomaly. In [23], abnormal heat load patterns from a clustering-based analysis are manually investigated to detect the underlying fault. To this end, human experts use visualizations and further information from the corresponding customers. In [29], substations found to have higher variations in their measurements are investigated manually in greater detail. Sun et al. [48] manually examine substation measurements categorized as anomalies by the previous clustering. In this way, the authors aim to identify and explain the cause of the anomaly. Bergstraesser et al. [41] visit and manually inspect substations with high excess flow on-site to document faults and recommend measures to reduce the substation's return temperature.

3.2.2. Threshold-based methods

Thresholds can be used to aid manual analysis, e. g., by providing visual indicators: In [43], Gadd and Werner plot two threshold lines at ± 3 standard deviations from the average line found by linear regression to support manual analysis.

In contrast, other authors use thresholds or limit checking to detect faults or anomalies automatically. If one or more measurements are beyond the threshold value, a fault or anomaly is detected and an appropriate action can be initiated: Gadd and Werner detect low average annual temperature differences in substations by comparing to a fixed, manually set threshold [11]. However, the authors argue that a low temperature difference is a symptom of faults rather than a fault itself. Månsson et al. [47] count measurements outside a threshold of ± 3 standard deviations to detect and rank poorly performing substations automatically. In [50], the authors calculate the variance of temperature sensor measurements and use a threshold of ± 2 standard deviations on the variance to estimate whether a sensor fault occurred. In [29], the authors propose limit-checking substation measurements with thresholds. Their methods can be improved by checking the moving average and standard deviation instead of the raw measurements. Further, comparing measurements of one substation with other substations that are geographically nearby or show a high correlation in their supply temperature can help to reveal faults. Theusch et al. [49] monitor the residual of the measured and the predicted energy consumption of substations. The residual is set in relation to the expected value and is then averaged over multiple time steps to be less prone to short-term fluctuations. Finally, the resulting metric is monitored using thresholds: If it exceeds the upper threshold or falls below the lower threshold, the substation is considered faulty. Calikus et al. [23] apply a threshold in a clustering-based analysis to separate abnormal from regular heat load profiles. A given substation should not have a greater distance to its cluster centroid than the mean distance between the cluster members and the cluster centroid plus three times the standard deviation. Zhang and Fleyeh [26] use thresholds of 93% and 99.5% on the reconstruction error of an autoencoder to detect anomalies. To detect fouling in substations, the authors of [55] compare the difference between the modeled and measured values to a given threshold. In [52], the authors compare their reference-group-based approach to detect outliers against using a fixed threshold, concluding that the reference-group-based approach might find additional outliers, not detected by a fixed threshold, as it adds local context.

Finally, thresholds are used as a trigger for other methods: To reduce false alerts Li et al. [46] apply a threshold to trigger a fault detection and isolation scheme based on convolutional neural networks. The fault detection and isolation scheme is only called if measured values deviate from the observed values in a healthy state.

3.2.3. Physical models

Bergstraesser et al. [41] detect and rank poorly performing substations of three DHS in Germany using excess flow analysis. Excess flow is calculated as the additional flow needed to compensate for a low difference in supply and return temperature. To calculate excess flow, yearly averages of the heat demand, return and supply temperatures plus a target return temperature are used. In [51], a Swedish DHS is modeled in the Modelica language using differential equations. The model is used to predict pressure values which are then compared to measured values and analyzed using a Bayesian network where appropriate. Li et al. [46] use an integrated energy-based district heating system model, including heat production, heat distribution networks, and buildings with substations, to evaluate their proposed fault detection and isolation scheme. The system model is taken from previous work [57].

3.2.4. Regression

Regression can be used to support the manual analysis, for example, by adding the result of regression as a reference to an existing visualization. Gadd and Werner perform a linear regression on a set of well-performing substations to plot the daily temperature difference in dependence on the outdoor temperature. The resulting regression line is used to support manual analysis [43]. In a similar approach, Månsson et al. [47] use a piecewise linear regression on so-called reference cases, i. e., well-performing installations, to model the relationship between three meter readings (temperature difference, return temperature, and energy) and the outside temperature. The authors of this approach argue that besides manual analysis, the number of measurements outside statistical thresholds can be counted and used to detect and rank substations.

However, regression can also be used to model and predict a substation's behavior. In these approaches, a regressor is often used to predict a substation's future measurements. These predictions are then compared to the actual measurements in a residual analysis to detect inconsistencies between the learned model and actual behavior. In [27], the authors use the tree-based pipeline optimization tool to find a suitable regressor to model the behavior of one substation. The resulting gradient boosting regressor predicts a substation's mass flow from recent and historical outdoor temperature, supply temperature, and the hour of the day. In a residual analysis the results of the predicted values from the gradient boosting regressor can be compared to the measured values and analyzed visually. After separating substation measurements into two clusters (high and low energy consumption), Theusch et al. [49] perform linear regressions to predict the energy consumption (dependent variable) as a function of the outdoor temperature (independent variable) for each cluster. The resulting regression models are then used to predict future energy consumption. In [53,54], a step-wise regression is used to model the relation between outdoor temperature and energy consumption of a substation. The authors of [55] use a multilayer perceptron neural network to predict the temperature control valves opening rate or the return temperature from other substation measurements. If the difference between the modeled and measured values is larger than a given threshold, fouling conditions are detected.

Lastly, Zhang and Fleyeh [26] use a piecewise linear regression to separate the demand for space heating from the demand for hot water preparation and other processes in substations. The authors argue that the demand for space heating is regular and dependent on the outside temperature, while the demand for hot water preparation and other processes is random or semi-random. Therefore, a piecewise linear regression can be used to extract the demand for hot water preparation and other processes from the original measurements.

3.2.5. Classification

Li et al. present a two-level fault detection and isolation scheme based on convolutional neural network classifiers in [46]. The authors use an upper-level classifier to detect and localize faults, such as sensor, actor, or component failure, and a lower-level classifier to distinguish between different subfaults, e. g., constant bias or drift in sensor values. To reduce false alerts the classifiers are only used if a given threshold is exceeded. The convolutional neural network-based detection and isolation scheme is compared to other classifiers (k-nearest neighbor, random forest, backpropagation neural network) and found to be superior in terms of F_1 score. An explainable anomaly detection scheme based on a random forest classifier is presented in [25]. The authors train the random forest classifier to detect anomalies in differential pressure control valve measurements. Further, the authors use Shapley Additive Explanations to quantify the influence of the input variables on the classifier's decision and, therefore, make the result interpretable for human operators. In [51], the authors use a Bayesian network to detect faults (jumping and drifting values in pressure sensors) and attribute them to an effect (leak and sensor fault) in a simulation model of a DHS. If the readings of one sensor lie outside a given tolerance range repeatedly, the Bayesian network differentiates between two cases: If the next sensor nearby behaves consistently, i. e., its readings are also outside the tolerance range, a leak is detected. However, if the next sensor does not show similar behavior, i. e., its readings are inside the tolerance range, then the first sensor is faulty. Both scenarios are validated on a model based on a real DHS in Sweden, though scenarios for sensor faults and leakage are simulated.

3.2.6. Clustering

In the context of fault and anomaly detection, clustering finds application in different ways, e. g., clustering algorithms are used to separate anomalies from normal behavior: For instance, Calikus et al. [23] present a large-scale clustering-based analysis of heat load patterns in two Swedish DHS. The authors use the k-shape algorithm to cluster substations according to their heat load and extract representative weekly heat load profiles for each cluster and season. Abnormal heat load patterns are detected by applying a threshold to the distance between the cluster members and the cluster's centroid. Article [48] compares different clustering algorithms regarding how well they separate anomalies from normal behavior in Chinese substations. Therefore, hourly data from six substations serving 17,000 apartments over three heating periods is divided into eight clusters. Four out of these eight clusters are identified as anomalous, described by certain features, such as high differential temperature and high heating power. The authors then compare the k-means, a kernel Gaussian mixture cluster, and a Gaussian mixture model algorithm in assigning records to the aforementioned clusters, concluding that the kernel Gaussian mixture cluster algorithm outperforms the others in terms of detection rate and false positive rate. Outliers in terms of energy consumption are found and extracted using the k-nearest neighbors algorithm in [53,54].

However, the results of clustering can also be used to create visualizations to support manual analysis. In their approach, Abghari et al. [38] model the substations' behavior by discretizing measurements from the primary and secondary side, extracting weekly patterns, and clustering the resulting patterns using affinity propagation and consensus clustering. If a significant difference between the model and the actual behavior is observed, further analysis is done by building a minimum spanning tree and manual analysis by experts. In [40], Abghari et al. introduce a multi-view clustering approach to substation measurements from 70 Swedish substations. Substations are clustered according to primary and secondary side measurements alongside computed performance indicators using minimum spanning tree clustering and affinity propagation. The results from clustering are then used for manual monitoring and fault detection. In a comparable approach, Abghari et al. [39] model a substation's operational behavior by extracting weekly patterns and clustering the resulting patterns into similar groups. The authors use affinity propagation and consensus clustering to do so.

Lastly, some authors use clustering to divide measurements or substations into groups of similar behavior to create more accurate models or find better threshold values for the found groups. Theusch et al. [49] use k-means clustering to separate typical substation measurements into two clusters (high and low energy consumption) before performing a regression for each cluster. Similarly, the authors of [29] propose to use clustering analysis to distinguish between consumption profiles.

Distinguishing between high and low energy consumption profiles can help to find meaningful thresholds, especially in buildings with time-varying thermal energy consumption, such as offices. In [56], the authors use the following clustering algorithms to divide substation measurements into several subsets based on different operating patterns: k-means, partitioning around medoids (PAM), and agglomerative hierarchical clustering. In [52], Farouq et al. present a reference-group-based approach towards monitoring the return temperature of substations to detect atypical and faulty behavior. They argue that modeling every substation might become costly in large networks, while creating one global model might be imprecise due to the substations' diversity. To overcome these limitations, the authors assign each substation to a reference-group, i. e., a group of similar behavior based on one month of data, using the k-nearest neighbors algorithm. As the reference groups highly depend on the used similarity measure, the authors test different similarity measures for their stability, i. e., substations will stay in the same reference group when focusing on another time interval.

3.2.7. Other methods

In [52], Farouq et al. present a reference-group-based approach towards detecting atypical and faulty behavior. Reference groups are revisited every month and outliers, i. e., substations whose behavior has changed significantly, are found using the isolation forest algorithm. In [26], a long short-term memory (LSTM) based variational autoencoder is used to detect labeled anomalies in the energy consumption of a Swedish substation. The authors use a piece-wise linear regression against the outdoor temperature to separate the substation's energy consumption into energy used for space heating and energy used for other processes, e. g., hot water preparation. An LSTM-based variational autoencoder is then trained to reconstruct the time series describing the energy used for other processes. If the reconstruction error exceeds a given threshold, an anomaly is detected. To evaluate their approach, the authors compare it against an LSTM-based autoencoder and an LSTM neural network using different thresholds to conclude that their approach outperforms the others in terms of area under receiver operating characteristic curve (AUC) and F_1 score. Abghari et al. [38] construct a minimum spanning tree on the basis of substations' behavior patterns derived from clustering. The longest edges of the minimum spanning tree are then removed as outliers with deviating behavior. Xue et al. [56] use association analysis to uncover frequent correlations in DH substation operating patterns. These correlations are represented in the form of association rules. As only a fraction of the association rules contains valuable knowledge, further examination by a human expert might be required.

3.3. Data sets for fault and anomaly detection

Data sets are a critical part of the development process, be that as a proof of concept for visualization, manual analysis, and statistical methods, or as training data for machine learning methods such as regression, clustering, and classification. As such, one cannot consider either data sets or methods alone. We examine various aspects of data set size and also answer questions about where the data was recorded, whether it is synthetic or real, labeled or not. This can serve as a reference for other researchers to compare their own data set to as well as design their methods to be compatible with the most widespread sensor types, sampling resolution, and so on. Additionally, it can serve as a guide for utility companies who are planning their smart meter setup or designing a database for long-term data storage. Table 2 summarizes all relevant attributes. None of the data sets is public.

For some data sets, no definitive numbers could be found and others miss critical information such as the actual number of substations considered, sampling rate, or the exact time frame. Synthetic data sets from physical models were also excluded from the data set analysis. The following values should thus be taken as trend indicators. This limits the number of publications using real-world measurements to 22.

Analyzing the origin of the data sets used, we find that 14 out of 25 data sets were recorded in Sweden with a 15th data set not specified but provided by a Swedish company. In second place are China, Germany, and South Korea with two data sets each. Italy is in third place with only one data set. There are also three synthetic data sets from simulations, one based on the DHS of a Swedish city. This matches the fact that Sweden has the highest share of DH for residential heating in the European Union [58], but also highlights the need for more data from other countries with a significant share of DH. However, this also increases the risk of methods that suit only one climate region or are influenced by country-specific regulatory factors that impact the design or operation of DH networks.

The number of substations included in the evaluation varies from 1 to 3000 with a median of 140 and a mean of approximately 480. Typically, the authors have limited their data set to substations with complete measurement records [54], to a specific building type [43], geographic location [44], or consumption — e. g., the n largest consumers [43,47]. While this makes the task at hand easier thanks to a more homogeneous subset, it can also result in more specialized solutions. For example, [42] describes a data set of 13,000 substations in two DH networks, but only 141 were selected for the study. While this is done to maximize the impact of the resulting maintenance work, excluding 10,000 one- and two-dwelling buildings due to their low heat consumption may have an influence on this method's transferability.

Similarly, we can look at the time frame during which the data were recorded. In general, data are either recorded yearly or only during the heating season in eight cases. 16 publications looked at 1 year of data or less with [41] using mixed observation periods between 1 and 2 years in three different DH systems. In [53], data from the 2002/2003 heating period is mentioned and shown in graphs, but the evaluation is done on data from the 2012/2013 heating period. The shortest duration is found in [52] at 2 months, the longest in [40] at 3 years. While the focus on the heating period is directing the effort at the time when improvements have the most significant impact, in systems with domestic hot-water supply, a year-round data set should yield a more generalized fault detector.

The de-facto default sampling rate is 1 h — used in 15 publications. 3-minute resolution is found in [49], but downsampled to hourly values prior to using it. 5-minute resolution is used in [44] and 10-minute resolution in [56]. In [55], consumption data are recorded at 1 h with system parameters at 1 min resolution which is then downsampled and smoothed to 10 min values. Daily values are used once — downsampled from hourly values [47] — and [38] analyzes weekly patterns derived from hourly values. A notable exception is [41] using yearly billing data. While a higher temporal resolution may yield better results, using shorter than 1 h intervals to develop a new fault detection algorithm might limit its transferability to other DHS in the short term due to a lack of widespread deployment of high-frequency meters. Contrary, higher sampling frequencies may be necessary to detect certain faults, and work in this area should be continued. Further effort to determine the impact of sampling frequency on fault detection is necessary.

As far as sensor types are concerned, we find a collection of supply and return temperature as well as flow rate to be the current standard. Typically, the transferred heat energy — calculated from these three values — is also included, either directly or as billing information [53]. Other data from the primary side such as pumps and valves are rarely measured and more often found in simulations, as are values from the secondary side. The same goes for pressure sensors. Beyond the scope of the substation, weather data is often included [25] as heat demand and outside temperature are directly correlated during the heating period [49]. Some approaches also use calendrical information [23,25] as customer behavior can change depending on working days, weekends, school holidays, and such. It should be noted that not every publication provides a complete list of sensors used. To mitigate

Table 2
Data set overview. Frequency only applies to consumption data.

Paper	Data Source	Labeled	# Substations	Recorded Time	Frequency	Country
[11]	Real	No	135	1 year	1 h	Sweden
[23]	Real	No	1385	1 year	1 h	Sweden
[25]	Real	Yes	unclear	8 months	unclear	South Korea
[26]	Real	Yes	1	1 year	1 h	Sweden
[27]	Synthetic Faults	Synthetic	1	1 year	1 h	Sweden
[38]	Real	No	10	2 years	1 h	Sweden
[39]	Real	No	10	1 year	1 h	Sweden
[40]	Real	No	70	3 years	1 h	Sweden
[41]	Real	No	1470	1-2 years	1 year[a]	Germany
[42]	Real	No	141	1 year	1 h	Sweden
[43]	Real	No	140	1 year	1 h	Sweden
[44]	Real	No	325	>1 year	5 min	Italy
[45]	Real	No	3	–	1 h	Sweden[b]
[46]	Simulation	Sim	–	–	–	Simulation
[47]	Real	No	3000	1 year	1 h	Sweden
[29]	Real	No	200	1 year	1 h	Sweden
[48]	Real	No	6	3 heating periods	1 h	China
[49]	Real	No	896	1 heating period	3 min	Germany
[50]	Simulation	Sim	–	–	–	Simulation
[51]	Simulation	Sim	–	–	–	Simulation
[52]	Real	No	778	2 months	1 h	Sweden
[53]	Real	No	>1000	2 heating periods	1 h	Sweden
[54]	Real	No	>1000	4 months	1 h	Sweden
[55]	Synthetic Faults	Synthetic	1	4 months	1 h	South Korea
[56]	Real	No	2	2 heating periods	10 min	China

Note: heating periods vary in length due to different climates.
[a]Yearly billing data used.
[b]Data provided by Swedish data science company.

inconsistent sensor coverage between individual substations or unstable data channels, [55] demonstrates using virtual sensors as replacements.

Aside from three simulation data sets, 20 publications use real-world data sets. However, out of these 20, only the sets of [25,26] are labeled. Real-world data with synthetic — and thus labeled — faults are utilized in [27,55]. All other works use unlabeled data and perform additional analysis on the result to evaluate their methods. Labeling in this context is done in a variety of ways, e. g., with a boolean [26] or on a scale of 1–5 with 4 and 5 signaling faults [25]. While those labels work for fault detection, they do not allow for more advanced methods such as fault classification nor do they work for targeted approaches to certain fault types unless the data set was labeled for those specific fault types in the first place.

3.4. Current state of the art, metrics, and comparison

To determine the quality of a method, a well-labeled ground truth is necessary to calculate the metrics used in the field of machine learning. This poses the first issue, as only two data sets are labeled and contain real-world data. Another two data sets consist of real-world data with synthetically generated faults. Three more data sets are completely simulated using software such as Simulink [50]. The advantage of synthetic faults or completely synthetic data sets is that they are automatically labeled, thus avoiding a time-consuming labeling process. Publications lacking such labeled data sets typically feature a qualitative analysis of the result [27,47], graphs or other visual representations to demonstrate the result [23,48], or the conclusion that verifying the result would require a manual review of every fault instance detected [23]. The latter equates to labeling only part of the data set but would only find correct instances and false positives — i. e., predicted faults during normal behavior — while missing false negatives — i. e., faults that the method missed and were thus not looked at. This skews the result of the analysis. A special case is [49], where the authors verified their method using one fault each for two real-world substations representing a best and worst case scenario in terms of being easy to detect.

Among the simulated data sets, [50,51] did not provide performance metrics, using qualitative analysis and theoretical mathematical proofs instead. Contrary, [46] provides accuracy, precision, recall, and the F1 score for multiple classifiers and conditions such as artificial sensor noise, fault type, and the duration of the data segment the classifiers used. Accuracy is the percentage of samples correctly classified as fault or no fault. Precision describes the purity of the detected faults and whether there are false positives present. Recall describes the percentage of existing faults detected by the classifier. The F1 score is the harmonic mean of precision and recall. The F1 score is significantly more meaningful for unbalanced data sets, where one class, e. g. faults, is much rarer than the other class [17]. If 1% of samples are faults, a classifier always returning “No Fault” would score an accuracy of 0.99 — close to a perfect result of 1.00 — whereas the F1 score would be 0.00. As an example, good classifiers using a convolutional neural network (CNN) yield F1 scores above 0.95 with several cases reaching a perfect 1.00 score [46].

For synthetic fault data sets, [27] provides extensive qualitative analysis of the obtained results. A correct and false alarm rate similar to the precision metric are given by [55]. Correct alarm rate can vary from 6% to 100% depending on the conditions and test data with false alarm rates ranging from 0.2% up to 32.9%. In addition, the authors provide a table with the time delay for fault detection. Faults are generally detected within 5 min but two worst-case scenarios take 69 and 180 min respectively.

The two publications using labeled real-world data sets both use the F1 score. In addition, [25] uses accuracy, precision, and recall like [46], but split into two categories for normal and abnormal samples (i. e., faults). This yields F1 scores around 0.98 up to 1.00 for normal samples and 0.78 up to 0.95 for abnormal samples. Due to this split, these values are not directly comparable to the values shown in [46]. The area under the curve (AUC) obtained from the receiver operating characteristics (ROC) curve is used by [26]. The ROC curve describes how the ratio between true positives and false positives — i. e., the purity of the result — shifts. Typically, one starts with one or more true positives yet missing the majority of faults, but as the threshold for classification shifts and becomes more generalizing, more false positives are generated by the classifier. This is due to real-world problems not being perfectly separable and outliers getting misclassified as faults. The larger the area under the ROC curve, the better a classifier is. Inversely, one can choose a point on the curve as a balance between finding the most faults yet getting few false positives. The authors of [26] use this to provide several different classifiers and show the relation between ROC, AUC, and F1 score. Due to this, the values are not directly comparable to the F1 scores of either [46] or [25]. However, near-perfect values around 1.00 are achieved by [26] as well.

Overall, it is very difficult to make comparisons. Publications using unlabeled data sets must rely on qualitative evaluation, interpretation, and graphs. Publications using labeled data sets tend to use a variety of metrics or the same metrics in different ways. In the end, even if metrics were used consistently, comparisons would still be done on different data sets, which can significantly change the result. There is no guarantee that a method with a higher score would actually perform better on a different data set.

4. Discussion and conclusion

While we see many successful applications of fault and anomaly detection in Section 3, we argue that there are still some challenges that hinder a wider implementation and adoption in practice. In this section, we want to summarize our findings, discuss these challenges and conclude with possible solutions to overcome these challenges in further research.

4.1. Summary

Our review of recent scientific literature indicates rising interest in data-driven fault and anomaly detection in district heating substations as the number of publications grows. Many publications rely on manual analysis with visualizations and thresholds to detect anomalies. Although some authors use machine learning techniques to detect anomalies, we share the conclusion of [36] that we currently see an initial phase of machine learning adoption. Regression is typically used to model a substation's behavior, e. g., its heat demand. Future deviating behavior can then be compared to the model and detected as anomalous. Physical models beyond regression are rarely used but could be an interesting addition to machine learning models in the future. Further, authors use clustering, e. g., to divide measurements and substations into groups of similar behavior or detect anomalous substations. Anomalous substations are detected by measuring the distance of a substation to its cluster centroid or by inspecting the smallest clusters. However, only a few articles showcase the potential of fault detection in the sense of classification. One reason for the focus on anomaly detection and the lack of classification approaches could be the absence of suitable data sets: While real-world data are used in most articles, publicly available and labeled data sets required for classification are lacking. Existing data sets typically have a temporal resolution of 1 h, extend from at least 2 months up to several years, and cover from 1 up to several thousand substations. We further observe that some articles do not clearly distinguish between faults, anomalies, and poor-performing substations, or share a common understanding, i. e., clear terminology is missing.

4.2. Terminological challenges

One challenge arises from the heterogeneous terminology used in the reviewed articles. As discussed in Section 2.1, different notions of faults and anomalies exist, and not all articles refer to related work or clarify their understanding. This complicates research and can cause confusion for practitioners and scientists alike. Further research should seek consensus on the terminology used, referring to established definitions if possible, adapting and expanding them if necessary.

4.3. Challenges related to data sets

The current situation with regard to DH data sets is as follows: Most authors get access to only a single data set from a cooperating utility company. This data set is generally not labeled, nor is it publicly available. This makes it time-consuming to compare between methods, introduces an additional source of error, and limits progress. We have identified three major issues and propose the following solutions. The primary issue is simply the lack of publicly available data sets. For instance, in image classification, there are a number of reference or benchmark data sets such as MNIST [59], ImageNet [60], CIFAR-10/CIFAR-100 [61]. This allows for direct comparison using exact metrics and fosters competition, accelerating the progression in the field. A district heating fault detection equivalent should contain data from multiple DH networks from various countries. Only by covering a wide gamut of options can such a data set become a meaningful reference. Second, in order to create such a benchmark data set for district heating, a clear, strict, identical, and complete taxonomy is necessary. For faults, [31] provides a thorough taxonomy for DH substations. There is also need for identical categories for the consumer type, e. g., single-family, multi-family, office, public, and industry. This unified taxonomy must then be used to create the individual subsets of the reference data set. An open question is whether, and if so, how, information about the network architecture can be included in such a data set. Third, this reference data set will contain personal information and suitable measures to protect the privacy of individual customers while maintaining the usefulness of the data set. While there exist a variety of methods for data set anonymization, choosing the right ones for this task is not easy, especially when it comes to including the DH network architecture.

In order to advance the field of fault detection in district heating, such a reference data set is — in our opinion — inevitable and essential. We ask to consider the impact of either the MNIST [59] or ImageNet [60] data sets on the machine learning and image recognition community as an example.

4.4. Challenges related to methodology

The focus on anomaly detection rather than fault detection also poses a challenge to adoption in the field. In our experience, practitioners see great value in the early detection of faults in substations to better plan maintenance and repair work. In contrast, anomalous behavior can sometimes not be followed up due to lack of time and personnel, especially as the manual identification of faults on site remains difficult and costly even if anomalous behavior is known [15]. To avoid a mismatch between research and practice future research should therefore focus on enabling and proposing methods for fault detection. This includes the creation of publicly available labeled data sets as well as a set of evaluation metrics focusing on targets set by DH companies to represent their practical use cases.

CRediT authorship contribution statement

Martin Neumayer: Investigation, Writing – original draft, Writing – review & editing. **Dominik Stecher:** Investigation, Writing – original draft, Writing – review & editing. **Sebastian Grimm:** Writing – original draft. **Andreas Maier:** Writing – review & editing. **Dominikus Bücker:** Writing – review & editing. **Jochen Schmidt:** Writing – original draft, Writing – review & editing.

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.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work was funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK), grant 03EN3033, and supported by the Bayerische Wissenschaftsforum (BayWISS).

References

[1] Volt S, Roca Reina J, Carlsson J, Georgakaki A, Letout S, Kuokkanen A, et al. Clean energy technology observatory: District heat and cold management in the European Union – 2022 Status report on technology development, trends, value chains and markets. Luxembourg (Luxembourg): Publications Office of the European Union; 2022, http://dx.doi.org/10.2760/168004.

[2] Blesel M, Koziol M, Ludwig C, Rapp H, Tenberg B, Vautz S, et al. 40/40 Strategie. Unser Konzept für die Wärmewende/40/40 strategy. Our concept for the change in heating. 2018, https://www.agfw.de/strategien-der-waermewende/perspektive-der-fw-7070-4040. (visited on 6 Dec 2022). AGFW, Frankfurt am Main.

[3] Bründlinger T, König JE, Frank O, Gründig D, Jugel C, Kraft P, et al. dena-Leitstudie Integrierte Energiewende/dena(German energy agency GmbH) pilot study: integrated energy transition. 2018, https://www.dena.de/fileadmin/dena/Dokumente/Pdf/9262_dena-Leitstudie_Integrierte_Energiewende_Ergebnisbericht.pdf. (visited on 6 Dec 2022). Deutsche Energie-Agentur GmbH (dena).

[4] Soltero V, Chacartegui R, Ortiz C, Velázquez R. Evaluation of the potential of natural gas district heating cogeneration in Spain as a tool for decarbonisation of the economy. Energy 2016;115:1513–32. http://dx.doi.org/10.1016/j.energy.2016.06.038, URL https://www.sciencedirect.com/science/article/pii/S036054421630812X. Sustainable Development of Energy, Water and Environment Systems.

[5] Johansen K, Werner S. Something is sustainable in the state of Denmark: A review of the Danish district heating sector. Renew Sustain Energy Rev 2022;158:112117. http://dx.doi.org/10.1016/j.rser.2022.112117, URL https://www.sciencedirect.com/science/article/pii/S1364032122000466.

[6] Lund H, Möller B, Mathiesen B, Dyrelund A. The role of district heating in future renewable energy systems. Energy 2010;35(3):1381–90. http://dx.doi.org/10.1016/j.energy.2009.11.023, URL https://www.sciencedirect.com/science/article/pii/S036054420900512X.

[7] Connolly D, Lund H, Mathiesen B, Werner S, Möller B, Persson U, et al. Heat Roadmap Europe: Combining district heating with heat savings to decarbonise the EU energy system. Energy Policy 2014;65:475–89. http://dx.doi.org/10.1016/j.enpol.2013.10.035, URL https://www.sciencedirect.com/science/article/pii/S0301421513010574.

[8] Jimenez-Navarro J-P, Kavvadias K, Filippidou F, Pavičević M, Quoilin S. Coupling the heating and power sectors: The role of centralised combined heat and power plants and district heat in a European decarbonised power system. Appl Energy 2020;270:115134. http://dx.doi.org/10.1016/j.apenergy.2020.115134, URL https://www.sciencedirect.com/science/article/pii/S0306261920306462.

[9] Frederiksen S, Werner S. District heating and cooling. Lund: Studentlitteratur AB; 2013.

[10] Lund H, Werner S, Wiltshire R, Svendsen S, Thorsen JE, Hvelplund F, et al. 4Th Generation District Heating (4GDH): Integrating smart thermal grids into future sustainable energy systems. Energy 2014;68:1–11. http://dx.doi.org/10.1016/j.energy.2014.02.089.

[11] Gadd H, Werner S. Fault detection in district heating substations. Appl Energy 2015;157:51–9. http://dx.doi.org/10.1016/j.apenergy.2015.07.061.

[12] Østergaard DS, Smith KM, Tunzi M, Svendsen S. Low-temperature operation of heating systems to enable 4th generation district heating: A review. Energy 2022;248:123529. http://dx.doi.org/10.1016/j.energy.2022.123529.

[13] Averfalk H, Werner S. Essential improvements in future district heating systems. Energy Procedia 2017;116:217–25. http://dx.doi.org/10.1016/j.egypro.2017.05.069, 15th International Symposium on District Heating and Cooling, DHC15-2016, 4-7 September 2016, Seoul, South Korea.

[14] Månsson S, Johansson Kallioniemi P-O, Thern M, Van Oevelen T, Sernhed K. Faults in district heating customer installations and ways to approach them: Experiences from Swedish utilities. Energy 2019;180:163–74. http://dx.doi.org/10.1016/j.energy.2019.04.220.

[15] Buffa S, Fouladfar MH, Franchini G, Lozano Gabarre I, Andrés Chicote M. Advanced control and fault detection strategies for district heating and cooling systems — A review. Appl Sci 2021;11(1):455.

[16] Verordnung über die Verbrauchserfassung und Abrechnung bei der Versorgung mit Fernwärme oder Fernkälte (FFVAV)/Ordinance on metering and billing for the supply of district heating or district cooling. 2021, https://www.gesetze-im-internet.de/ffvav/BJNR459110021.html. (visited on 9 Dec 2022).

[17] Wenninger M, Schmidt J, Goeller T. Appliance usage prediction for the smart home with an application to energy demand side management-and why accuracy is not a good performance metric for this problem. In: International conference on smart cities and green ICT systems. Vol. 2. SCITEPRESS; 2017, p. 143–50.

[18] Wenninger M, Maier A, Schmidt J. DEDDIAG, a domestic electricity demand dataset of individual appliances in Germany. Sci Data 2021;8(1):1–15. http://dx.doi.org/10.1038/s41597-021-00963-2.

[19] Gourmelon N, Bayer S, Mayle M, Bach G, Bebber C, Munck C, et al. Implications of experiment set-ups for residential water end-use classification. Water 2021;13(2):236.

[20] Chatterjee S, Bayer S, Maier A. Prediction of household-level heat-consumption using PSO enhanced SVR model. 2021, arXiv preprint arXiv:2112.01908.

[21] Ramachandran A, Chatterjee S, Bayer S, Maier A, Flensmark T. Heat demand forecasting with multi-resolutional representation of heterogeneous temporal ensemble. 2022, arXiv preprint arXiv:2210.13108.

[22] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey. ACM Comput Surv 2009;41(3). http://dx.doi.org/10.1145/1541880.1541882.

[23] 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.

[24] Worden K, Dulieu-Barton JM. An overview of intelligent fault detection in systems and structures. Struct Health Monit 2004;3(1):85–98. http://dx.doi.org/10.1177/1475921704041866.

[25] Park S, Moon J, Hwang E. Explainable anomaly detection for district heating based on Shapley additive explanations. In: 2020 International conference on data mining workshops. 2020, p. 762–5. http://dx.doi.org/10.1109/ICDMW51313.2020.00111.

[26] 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. 2020, p. 153–8. http://dx.doi.org/10.1109/ICIEA48937.2020.9248108.

[27] 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, 16th International Symposium on District Heating and Cooling, DHC2018, 9–12 September 2018, Hamburg, Germany.

[28] Leoni P, Geyer R, Schmidt R-R. Developing innovative business models for reducing return temperatures in district heating systems: Approach and first results. Energy 2020;195:116963. http://dx.doi.org/10.1016/j.energy.2020.116963, URL https://www.sciencedirect.com/science/article/pii/S0360544220300700.

[29] Sandin F, Gustafsson J, Delsing J, Eklund R. Basic methods for automated fault detection and energy data validation in existing district heating systems. In: International symposium on district heating and cooling. District Energy Development Center; 2012.

[30] Sandin F, Gustafsson J, Delsing J. Fault detection with hourly district energy data : Probabilistic methods and heuristics for automated detection and ranking of anomalies. Tech. rep, Svensk Fjärrvärme; 2013, p. 120.

[31] Månsson S, Lundholm Benzi I, Thern M, Salenbien R, Sernhed K, Johansson Kallioniemi P-O. A taxonomy for labeling deviations in district heating customer data. Smart Energy 2021;2:100020. http://dx.doi.org/10.1016/j.segy.2021.100020.

[32] Månsson S, Thern M, Johansson Kallioniemi P-O, Sernhed K. A fault handling process for faults in district heating customer installations. Energies 2021;14(11). http://dx.doi.org/10.3390/en14113169.

[33] Guelpa E, Capone M, Sciacovelli A, Vasset N, Baviere R, Verda V. Reduction of supply temperature in existing district heating: A review of strategies and implementations. Energy 2023;262:125363. http://dx.doi.org/10.1016/j.energy.2022.125363, URL https://www.sciencedirect.com/science/article/pii/S0360544222022459.

[34] Werner S. International review of district heating and cooling. Energy 2017;137:617–31.

[35] Sayegh M, Danielewicz J, Nannou T, Miniewicz M, Jadwiszczak P, Piekarska K, et al. Trends of European research and development in district heating technologies. Renew Sustain Energy Rev 2017;68:1183–92.

[36] Ntakolia C, Anagnostis A, Moustakidis S, Karcanias N. Machine learning applied on the district heating and cooling sector: A review. Energy Syst 2021;1–30.

[37] 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.

[38] Abghari S, Boeva V, Brage J, Johansson C, Grahn H, Lavesson N. Higher order mining for monitoring district heating substations. In: 2019 IEEE international conference on data science and advanced analytics. 2019, p. 382–91. http://dx.doi.org/10.1109/DSAA.2019.00053.

[39] Abghari S, Boeva V, Brage J, Johansson C. District heating substation behaviour modelling for annotating the performance. In: 19th Joint European conference on machine learning and principles and practice of knowledge discovery in databases, ECML PKDD 2019; Würzburg; Germany; 16–20 September 2019. Vol. 1168. Springer; 2020, p. 3–11.

[40] Abghari S, Boeva V, Brage J, Grahn H. Multi-view clustering analyses for district heating substations. In: 9th International conference on data science, technology and applications, DATA 2020, Virtual, Online; France, 7–9 July 2020. SciTePress; 2020, p. 158–68.
[41] Bergstraesser W, Hinz A, Braas H, Orozaliev J, Vajen K. Lessons learned from excess flow analyses for various district heating systems. Smart Energy 2021;1:100005. http://dx.doi.org/10.1016/j.segy.2021.100005.
[42] Gadd H, Werner S. Heat load patterns in district heating substations. Appl Energy 2013;108:176–83. http://dx.doi.org/10.1016/j.apenergy.2013.02.062.
[43] 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.
[44] Guelpa E, Verda V. Automatic fouling detection in district heating substations: Methodology and tests. Appl Energy 2020;258:114059. http://dx.doi.org/10.1016/j.apenergy.2019.114059.
[45] Johansson C, Wernstedt F. N-dimensional fault detection and operational analysis with performance metrics. In: The 13th international symposium on district heating and cooling. 2012.
[46] Li M, Deng W, Xiahou K, Ji T, Wu Q. A data-driven method for fault detection and isolation of the integrated energy-based district heating system. IEEE Access 2020;8:23787–801. http://dx.doi.org/10.1109/ACCESS.2020.2970273.
[47] Månsson S, Davidsson K, Lauenburg P, Thern M. Automated statistical methods for fault detection in district heating customer installations. Energies 2019;12(1). http://dx.doi.org/10.3390/en12010113.
[48] Sun W, Cheng D, Peng W. Anomaly detection analysis for district heating apartments. J Appl Sci Eng 2018;21(1):33–44. http://dx.doi.org/10.6180/jase.201803_21(1).0005.
[49] Theusch F, Klein P, Bergmann R, Wilke W, Bock W, Weber A. Fault detection and condition monitoring in district heating using smart meter data. In: PHM society European conference. Vol. 6. No. 1. 2021, p. 11.
[50] Yliniemi K, Van Deventer J, Delsing J. Sensor fault detection in a district heating substation. In: IMEKO TC10 international conference on technical diagnostics. 2005.
[51] Zimmerman N, Dahlquist E, Kyprianidis K. Towards on-line fault detection and diagnostics in district heating systems. Energy Procedia 2017;105:1960–6. http://dx.doi.org/10.1016/j.egypro.2017.03.567, 8th International Conference on Applied Energy, ICAE2016, 8-11 October 2016, Beijing, China.
[52] 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.
[53] Kiluk S. Algorithmic acquisition of diagnostic patterns in district heating billing system. Appl Energy 2012;91(1):146–55. http://dx.doi.org/10.1016/j.apenergy.2011.09.023.
[54] Kiluk S. Diagnostic information system dynamics in the evaluation of machine learning algorithms for the supervision of energy efficiency of districtheating-supplied buildings. Energy Convers Manage 2017;150:904–13. http://dx.doi.org/10.1016/j.enconman.2017.05.006.
[55] 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.
[56] Xue P, Zhou Z, Fang X, Chen X, Liu L, Liu Y, et al. Fault detection and operation optimization in district heating substations based on data mining techniques. Appl Energy 2017;205:926–40. http://dx.doi.org/10.1016/j.apenergy.2017.08.035.
[57] Jiang X, Jing Z, Li Y, Wu Q, Tang W. Modelling and operation optimization of an integrated energy based direct district water-heating system. Energy 2014;64:375–88. http://dx.doi.org/10.1016/j.energy.2013.10.067.
[58] Bertelsen N, Vad Mathiesen B. EU-28 residential heat supply and consumption: Historical development and status. Energies 2020;13(8). http://dx.doi.org/10.3390/en13081894.
[59] LeCun Y, Cortes C, Burges C. MNIST handwritten digit database. 1998, http://yann.lecun.com/exdb/mnist/. (visited on 16 Dec 2022).
[60] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. 2009, p. 248–55. http://dx.doi.org/10.1109/CVPR.2009.5206848.
[61] Krizhevsky A, Nair V, Hinton G. CIFAR-10/CIFAR-100 (Canadian Institute for Advanced Research). 2009, http://www.cs.toronto.edu/~kriz/cifar.html. (visited on 16 Dec 2022).