Virtual Sensor-Based Fault Detection and Diagnosis Framework for District Heating Systems: A Top-Down Approach for Quick Fault Localisation

Theis Bank, Frederik Wagner Madsen, Lasse Kappel Mortensen(✉), Henrik Alexander Nissen Søndergaard, and Hamid Reza Shaker

University of Southern Denmark, 5230 Odense, Denmark
lkmo@mmmi.sdu.dk

Abstract. For district heating systems (DHS) to operate cost-effectively, avoid disturbances of loads, and increase overall energy efficiency, faults in DHSs must be detected, located, and rectified quickly. For this purpose, a novel digital twin-based fault detection and diagnosis framework with virtual sensor employment have been developed. The framework defines virtual sensors measuring the mass flow rate in points in the DHS where sensors are absent by using the existing sensors in the system. Faults in the virtual sensors are detected when deviations occur between the calculated and digital twin-simulated mass flow rate using a bound of normal operation, allowing some degree of modelling error. To define which virtual sensors are of interest, a novel Specialised Agglomerative Hierarchical Clustering algorithm will be used. A case study on a DHS of a suburb in Odense showed how the framework was able to locate faults with a top-down approach and could indicate whether the fault was local or due to upstream faults. The framework has the potential to be implemented in real-time monitoring of a DHS.

Keywords: Fault detection and diagnosis · District heating systems · Digital twin · Virtual sensor · Machine learning

1 Introduction

In Denmark, a large share of households uses district heating (DH) for hot water use and space heating. The objective of a district heating system (DHS) is to distribute cost-effective heat to consumers efficiently [2]. As hot water use and space heating are a large proportion of total energy use in the EU [5], the DHSs are critical infrastructure, which means it is a relevant domain to improve energy efficiency and thereby lowering CO_2 emissions. The vast amount of functionalities required for the operation of the DHS makes it complex and failure prone. To guarantee the DHS works efficiently, these faults need to be detected and corrected quickly, as the faults result in suboptimal operating conditions of the DHS [2,9,12].

B. N. Jørgensen et al. (Eds.): EI.A 2023, LNCS 14468, pp. 292–307, 2024.
https://doi.org/10.1007/978-3-031-48652-4_19

Currently, most DH companies perform reactive- and preventive maintenance, whereas proactive maintenance is rarely used. With a reactive maintenance strategy, it is difficult to detect faults quickly. It may also lead to faults never being detected, e.g., if the DHS can compensate for the fault and still deliver heat, but this will still waste energy. Preventive maintenance often leads to a waste of resources due to redundant maintenance. To utilise the still ongoing digitalisation of the DH sector by implementing proactive maintenance, diagnostic tools using automatic fault detection and diagnosis (FDD) methods are of great interest. Developing such tools and methods to detect faults expeditiously and affordably in a DHS would reduce the energy and time waste and furthermore reduce CO_2 emissions.

In [6], three sub-types are used to categorise different FDD methods: quantitative model-based, qualitative model-based, and process history-based, where all three have certain advantages and disadvantages. An advantage of the quantitative model-based method is the utilisation of a very precise model, which is built on the basis of thorough physical or engineering principles. On the other hand, the precise model is also a disadvantage of the method because of the level of complexity and required amount of input needed to build the model, which can lead to reducing the scalability of the approach. [19] models a DHS in OpenModelica, and uses the output of the model together with pressure sensors to calculate residuals. A Bayesian Network is then used to compare these residuals, calculate fault probability, and evaluate the system. [1] also uses a qualitative model-based approach, where a DH pipe network is simulated together with an optimisation problem, used to detect both thermal and hydraulic faults.

While the difference between quantitative- and qualitative model-based methods can be vague due to overlapping features, the process history-based method differs a lot. This category covers data-driven approaches, that do not necessarily require knowledge about the complex physics of the DHS system, and offers great scalability. On the other hand, these methods may require vast amounts of correct data, which can be hard to obtain. Furthermore, they can have high computational demands.

Generally, for the domain of DHS, data-driven FDD is more researched compared to model-based approaches. A data-driven method is seen in [14], which utilises three FDD methods: Hotelling's T^2 and Q statistics, contextual Shewhart chart, and linear regression, enabling to check if all methods agree on detected faults. More examples of data-driven approaches are seen in [8,10,15,16]. [15] identifies operations patterns by three clustering methods, [8] predicts normal operation of DH substations with the use of gradient boosting regressor, and [16] finds rule patterns for the operation of a DH substation using cluster- and association analysis.

This paper utilises a quantitative model as DH companies often have thorough knowledge about their pipe network regarding configuration, dimensions, and heat transfer coefficients, which can be used to set up a digital twin (DT) in industry-specialised software. Software, that supports geographical information systems also allows the operator to locate and correlate faults quickly on a map.

[9,12] argues that previously FDD frameworks for DHS were old and not sufficiently advanced for the move towards Industry 4.0. Though more methods are being developed, as we have highlighted, there is still a general lack of research on FDD in the DHS domain, as is also emphasized by [2]. For these reasons and to support the move towards Industry 4.0, the framework proposed in this paper integrates data from modern residential smart heat meters alongside substations' sensor data to operationalise a digital twin of a real DH network for improved monitoring and proactive maintenance.

One way of utilising a digital twin of a DHS is through the ability to calculate certain properties everywhere in the system. This can be used together with virtual sensors and thereby enabling fault detection by comparing the results from the digital twin and the virtual sensors. For the energy systems domain, virtual sensors have been developed in different applications, such as heat, ventilation, and air conditioning (HVAC) systems of buildings and in residential buildings in a DHS. [7] uses physical relations inside the HVAC to establish virtual sensors, which enables FDD and introduces cost-efficient redundancy. In [17], virtual sensors are established by grey-box modelling, which estimates heating loads in residential buildings in the absence of sensors. To the best of the authors' knowledge, using virtual sensors and a digital twin to locate faults in the pipe network of a DHS has not been done before.

The contribution of this paper to the apparent research gap in the literature for FDD in DHSs is two-fold. Firstly, a digital twin-based FDD framework utilising virtual sensors is developed, which can detect and locate faults in a DHS. To the best of our knowledge, this has not been done before. Furthermore, the FDD method makes use of a novel Specialised Agglomerative Hierarchical Clustering algorithm that validates discovered clusters with information about virtual sensors. Secondly, the framework is tested by implementing it on a case study with real-world sensor data and a digital twin of a DHS.

In Sect. 2, the different steps of the framework are presented, and in Sect. 3, the framework is implemented on a case study, and results are shown.

2 Methodology

This section will present a novel digital twin-based FDD framework using virtual sensors for DHSs.

The framework will utilise sensor measurements and a digital twin to detect faults in virtual sensors in DHSs. Faults in this context are abnormalities and deviations between measured- and digital twin simulated data. Faults can be located in the DHS to correlate possible multiple faulty consumers in an area. Locating and correlating faults in a specific area will enable a maintenance crew to focus resources on local related faults instead of individual faults at consumers. This may lead to the use of fewer resources and achieve better system performance. The unsupervised nature of the framework means it does not require historical maintenance records or fault characteristics. This both strengthens the framework's applicability as it has fewer requirements for implementation than would otherwise be the case, but it also comes at the cost of weakened diagnosis capabilities.

The framework, seen in Fig 1, is divided into eight steps and shows how sensor measurements in the DHS are simulated in a digital twin, how data is processed, and how virtual sensors and consumer clusters are used to perform FDD on pipes in the system. For the purpose of this work, we adopt the general definition of a digital twin stated by Yu et al. [18] that "a DT is a digital (or virtual) representation that looks-like, behaves-like, and connects-to a physical part or system with the goal of improving or optimising decision making for any time horizon".

In the first step of the framework, data is collected from sensors at the individual consumers and at sub-stations. In step two, this data is processed and formatted together with a thorough investigation of the data, finding abnormalities and physics-violations, to validate it. Invalidated data is reconstructed in step 2. In the third step, some of the sensor data is imported as time series to a digital twin as boundary conditions, which can run quasi-dynamic simulations showing the state of the DHS. The proposed FDD framework exploits the increasing modelling error in the digital twin, which occurs due to faults or when the boundary conditions indicate abnormal operation at an individual consumer, to detect deviations between the sensor- and digital twin data The

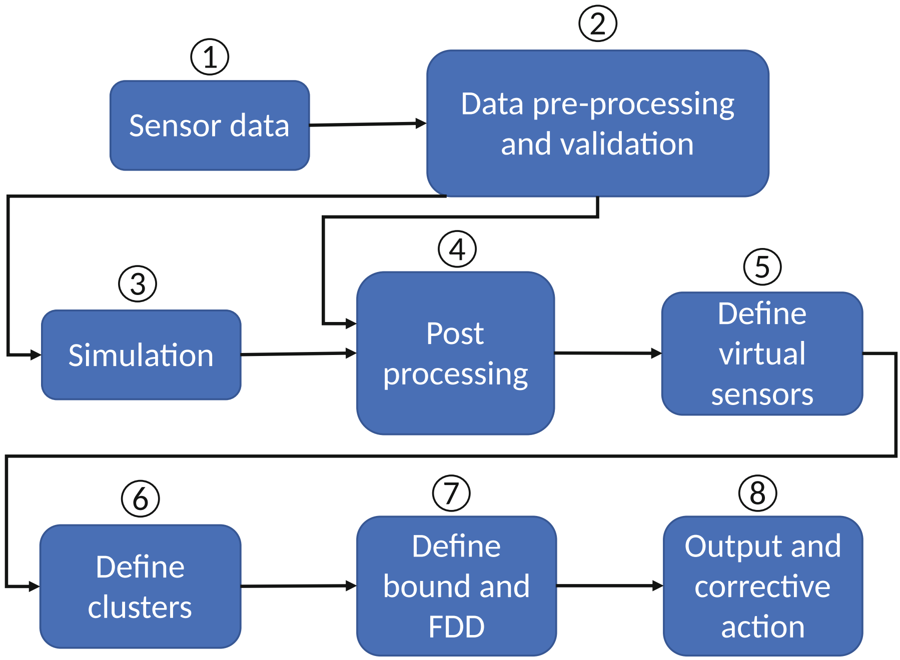

Fig. 1. Flow diagram of digital twin-based FDD framework with virtual sensor employments.

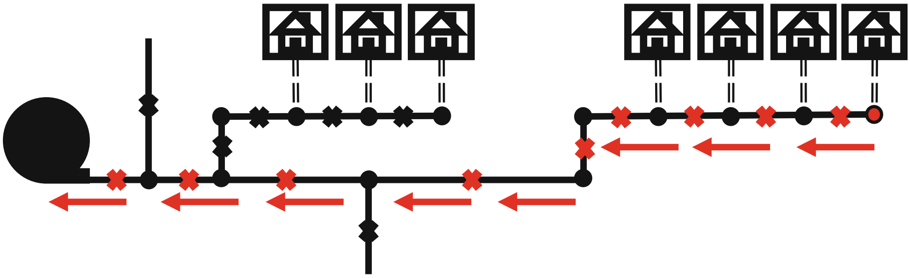

Fig. 2. Virtual sensors are represented as crosses, and nodes connecting pipes and consumers are represented as dots. The arrows show the shortest route from the red end node to the sub-station. (Color figure online)

desired output from the digital twin is consumer properties and the mass flow rate in the pipes used later for the FDD and fault location. The results from the digital twin are formatted and validated in step four for the further use of the data.

The virtual sensors employed in the DHS, in step five, will hold the mass flow rate in the pipes. A pseudocode for the virtual sensor calculation is seen in Alg. 1.

Algorithm 1: Pseudocode for virtual sensors in pipes of a radial pipe configuration.

```
for all end nodes in the system do
    Find the shortest route of pipes towards the substation using the Dijkstra
    algorithm [4];
    for all nodes along the shortest route do
        Place a virtual sensor in the upstream pipe the node connects to;
        Calculate the mass flow rate in the virtual sensor by summing up the
        total mass flow rate at the consumers in the downstream pipe network
        from the virtual sensor;
Filter out virtual sensor duplicates;
```

The pseudocode in Alg. 1 is for a radial pipe configuration, which is by far the most common configuration in DHSs. The approach is illustrated in Fig. 2, where crosses represent the virtual sensors in the pipes and dots are the nodes connecting the pipes and consumers. Applying this approach to all end nodes in a radial pipe network will ensure that all pipes in the system will have a virtual sensor.

In the sixth step, clusters of consumers are defined. These clusters will indicate which virtual sensors are of interest and can be used for FDD. Figure 3 shows how consumers along a small residential street connected to a main pipe can be clustered into one large cluster (blue area) and two smaller sub-clusters (green area). The virtual sensors of interest are shown in Fig 3 (black and red crosses). Larger clusters are divided into smaller sub-clusters as this allows for detecting and locating faults from a top-down approach. E.g. in Fig 3, if a fault occurs in the large cluster (black virtual sensor), then the sub-clusters (red virtual sensors) are investigated. Here a fault might only be detected in one of the sub-clusters, which narrows down the fault location.

Performing FDD on the individual consumers in the sub-cluster can indicate the final location of the fault. If only one consumer in the sub-cluster is operating in a faulty condition, that consumer will probably be the cause of the deviation in the virtual sensors. On the other hand, if multiple consumers operate under faulty conditions, the probability of similar faults at each consumer might be low and can therefore indicate faults in the upstream pipe network.

This paper proposes the use of a Specialised Agglomerative Hierarchical Clustering algorithm, which is a modification of the standard Agglomerative Hierarchical Clustering algorithm for which some of the earliest mentions are from Sibson [13] and Rohlf [11]. The specialized approach views the pipe network as a graph and uses the bottom-up hierarchical clustering method to cluster consumers in close proximity. The extension of the algorithm is proposed to avoid wrongly defined clusters, i.e., clusters of consumers that have no corresponding virtual sensor. This logic is defined as: If c represents a set of consumers, and $V(c)$ is a boolean representation of whether a virtual sensor exists that only aggregates information for the consumers in the set c, then a cluster produced with Agglomerative Hierarchical Clustering, which groups the consumers in set c, is accepted if and only if $V(c) = 1$. The extension of the clustering algorithm can be seen as a filter that eliminates clusters without a corresponding virtual sensor.

The Specialised Hierarchical Agglomerative Clustering algorithm is presented in Algorithm 2. C is a set of clusters at all hierarchical levels, R is a set of the indices of a cluster's children, and H is a set of distances between the children

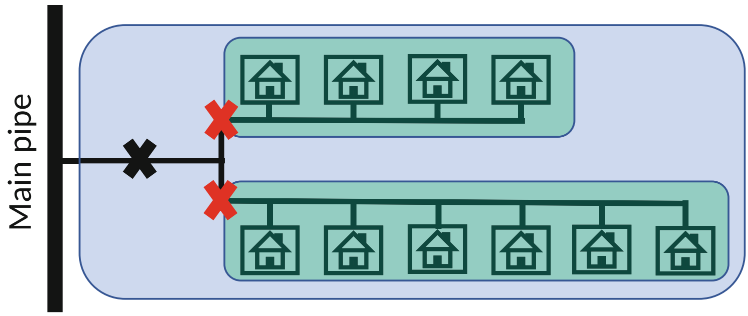

Fig. 3. Hierarchical clustering of consumers. The diagram shows one large cluster and two sub-clusters with their representing virtual sensors (black and red crosses). (Color figure online)

of a cluster. In this setting P is a set of the individual consumers, and $x_k \in X$ denote the consumers downstream of the virtual meter point k. N_c is the number of clusters, which reduce by 1 each time two clusters are agglomerated and it defines when the algorithm is done, i.e., when all points are agglomerated into 1 cluster. The distance between two points is the shortest distance along the pipes of the network, calculated with the Dijkstra algorithm [4]. The distance function denotes yields the maximum distance between all points of the two input clusters (complete linkage).

Algorithm 2: Specialised Hierarchical Agglomerative Clustering

Given: *A set of points to cluster* $P = \{p_1, ..., p_n\}$, *a set of potential clusters* $X = x_1, ..., x_k$, *and a distance function* $d(c_1, c_2)$ *Populate the set of clusters,* C *with the points in* P;
$C \leftarrow \{\}$;
H denotes the distance between the children of each cluster and R denotes the indices of the children of each cluster, both of which are empty for leaf clusters;
for $i \leftarrow 1$ **to** n **do**
    $C \xleftarrow{+} \{p_i\}$;
    $H \xleftarrow{+} \{\}$;
    $R \xleftarrow{+} \{\}$;
$Z \leftarrow C$;
$N_c = |C|$;
while $N_c > 1$ **do**
    $a, b = \underset{i,j}{\operatorname{argmin}}\, d(z_i, z_j)$ where $i \neq j$ **and** $\{z_i, z_j\} \subseteq X$;
    $C \xleftarrow{+} \{z_a, z_b\}$;
    $Z \xleftarrow{+} \{z_a, z_b\}$;
    $R \xleftarrow{+} \{a, b\}$;
    $H \xleftarrow{+} d(z_a, z_b)$;
    $Z \xleftarrow{-} \{z_a\}$ **and** $\{z_b\}$;
    $N_c -= 1$
Return C, R, and H

In step seven, a bound of normal operation is defined, and FDD is carried out. The bound of normal operation is defined as,

$$Target \cdot (1 \pm \beta) \tag{1}$$

The target in Eq. 1 is the simulated property value in the digital twin, and β is a ratio parameter of how much deviation is allowed due to small modelling errors. Faults are detected by investigating at each time step if the virtual sensor is outside the bound of normal operation simulated in the digital twin. Consecutive out of bound detections will be defined as one faulty event. Faults may arise as a consequence of different events or phenomena, abnormal boundary conditions, such as very low cooling efficiency, can increase modelling error and may be detected using the proposed method. Lastly, in step eight, the detected faults are located with a top-down approach using the hierarchical clusters, and corrective and maintenance actions can be initiated.

3 Case Study and Results

3.1 Description of Case Study

The digital twin-based FDD framework with virtual sensor employments presented will be implemented in a case study in a suburb of Odense, Denmark. The DHS operator, Fjernvarme Fyn, provided historical data and a Leanheat Network (LHN) model for the DHS, which was developed in collaboration with Danfoss, the developer of the software. The LHN model constitute the digital twin of the system.

LHN simulations are performed by hydraulic and thermal condition simulations and pressure and temperature optimisations, minimising pump- and heat production costs, based on defined material properties and boundary conditions at the substation and individual consumers. Simulation results from LHN provides properties at all nodes and pipes in the DHS [3].

Figure 4 shows the DHS of the suburb implemented in LHN, all 648 connected consumers are supplied via the substation. The DHS in the diagram has only radial connections because it is only operated in a radial manner, which is typically for DHSs. Nevertheless, most DHSs have meshed connections for redundancy, e.g. for when faults happen.

The implementation of the framework on the case study results in a list of abnormalities, of which two are analysed to find the possible roots of the abnormalities. Due to not having a maintenance record, a more quantitative method of analysing the abnormalities could not be performed, where the abnormalities could have been compared with the maintenance record thereby confirming or denying them.

The data, from December 2022 to January 2023, contains consumer sensor measurements on a daily resolution at the 648 consumers and sensor measurements on an hourly resolution at the substation. Consumer data contains the energy consumption, volumetric flow rate and supplied and returned energy, which is the supplied and return temperature multiplied by the volumetric flow rate. The substation data contains supplied and returned energy, mass flow rate, pressure, and temperature.

The data will be formatted to have the correct units and a time series format for the model. The model will simulate hourly quasi-dynamic simulations representing the whole day, where the boundary conditions for the model will be return temperature and energy consumption at the consumers, supply temperature and return pressure at the substation, and pressure change at a critical node assumed to be 1.33 bar based on expert knowledge. The output from the model will be the mass flow rate in all the pipes of the network and consumers, which will be used for the later FDD and fault location investigation. This case

Fig. 4. Visualisation of DHS implemented in the LHN model. The substation located at the bottom supplies the 648 consumers through the pipe network, which represents the real configuration of the DHS. Red dots represent the nodes connecting the pipes and black houses represent the individual consumers. (Color figure online)

study will focus on two consumer clusters in the DHS, which will be divided into sub-clusters, to showcase the capabilities of the framework.

3.2 Consumer Clustering

To enable the monitoring of the DHS by a top-down approach, the consumers need to be clustered in several clusters and sub-clusters. This section will show how a cluster of seven consumers is defined (Cluster 1) using the proposed Specialised Agglomerative Hierarchical Clustering algorithm.

The bottom-up clustering method begins at the lowest level, i.e., each consumer is a cluster, which can be seen in the dendrogram in Fig. 5. Using the geographical data from the model, the method iteratively looks at the defined distance matrix of the edge (pipe) lengths and joins the two closest clusters into one cluster, until only one cluster is left. After the clustering, the dendrogram is used to define the number of sub-clusters by defining a distance threshold. For this clustering, a distance threshold of 60 m is chosen. This resulted in one overall cluster (blue) with two sub-clusters (orange and green), as seen in Fig. 5, but the distance threshold is ultimately a tuneable hyperparameter. Also in general, a series of distance thresholds must be selected to account for more upstream clusters.

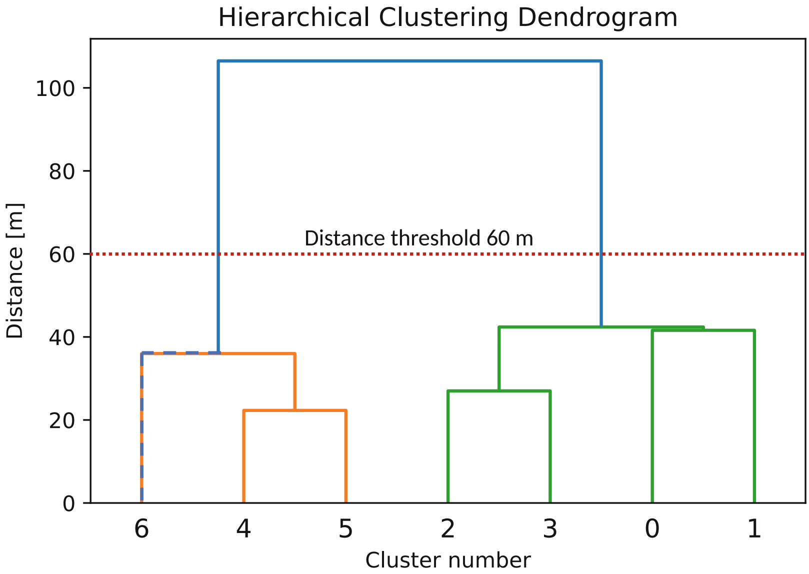

Fig. 5. Dendogram of Cluster 1; initially, all consumers are considered individual clusters. At each iteration, the two closest clusters are joined. The horizontal red dotted line shows the distance threshold is set to 60 m (Color figure online) and defines the two sub-clusters. Consumer 6 is connected with a dotted blue line, representing that the logic-based rule constraint rejects consumer 6 from the clusters containing consumers 4 and 5 as no virtual sensor exists that aggregates all three of them.

The graph representation of the seven consumers (Fig. 6), which is used to define the distance matrix, indicates that the orange sub-cluster is wrongly defined due to the locations of the virtual sensor, which is corrected with the extension of the algorithm. In the dendrogram (Fig. 5), the orange sub-cluster holds three consumers, but the virtual sensor only holds two of them. The logic-based rule extension of the Agglomerative Hierarchical Clustering will indicate that there is no virtual sensor which holds consumers 4, 5, and 6 and therefore disregards this cluster. Going one level down in the hierarchical clustering (Fig. 5), the logic-based rule will accept the sub-cluster containing consumers 4 and 5 and reject consumers 6, as there is a virtual sensor which holds consumers 4 and 5.

A second cluster (Cluster 2) located in another part of the DHS was also made. This cluster contains 13 consumers and was divided into three sub-clusters, two sub-clusters with four consumers and one with five (Sub-cluster 2.1, 2.2, and 2.3).

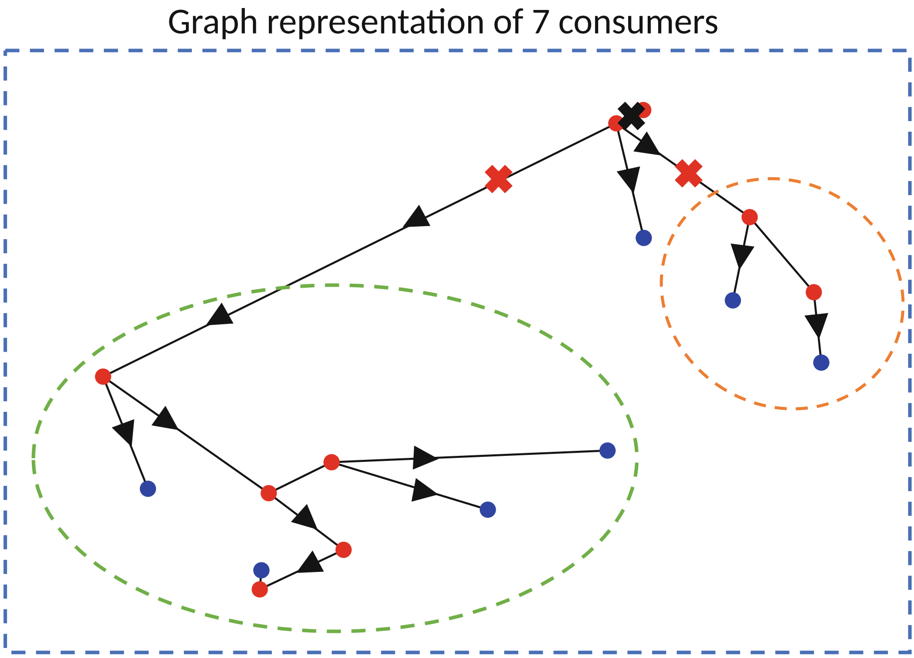

Fig. 6. Cluster 1 contains seven consumers (blue dots), nine nodes (red dots), and 11 pipes (black lines) represented as a graph. The virtual sensors of interest are defined as red and black crosses, where the belonging clusters are represented as dotted lines (blue, green, and orange). Black arrows show the direction of forward flow in the pipes. (Color figure online)

3.3 Fault Investigation and Location

This section will present the detected faults in the virtual sensors and locate the faults with a top-down approach using the hierarchical consumer clusters defined. The parameter β defines the size of the bound and must be calibrated to produce an appropriate number of faults, as there is a direct connection between the size of the bound and the number of faults detected. The bound of normal operation used in this paper is introduced in Eq. 1, with $\beta = 0.3$.

Figure 7 illustrates the mass flow rate in the virtual sensors supplying Cluster 1 and Sub-cluster 1.1 and 1.2. Figure 7 shows that a large deviation in Cluster 1 occurs between the modelled mass flow rate and the calculated mass flow rate in the virtual sensor. This resulted in three fault detections (grey marked areas in Fig. 7). Going one level down in the hierarchy of the consumer clusters, it is evident that the fault in Cluster 1 is due to a fault occurring in Sub-cluster 1.1, as Fig. 7 shows faults detected in Sub-cluster 1.1 and not in Sub-cluster 1.2. FDD was done at the individual consumer level at last and showed that only one consumer was operating in a faulty condition. This may lead to the conclusion

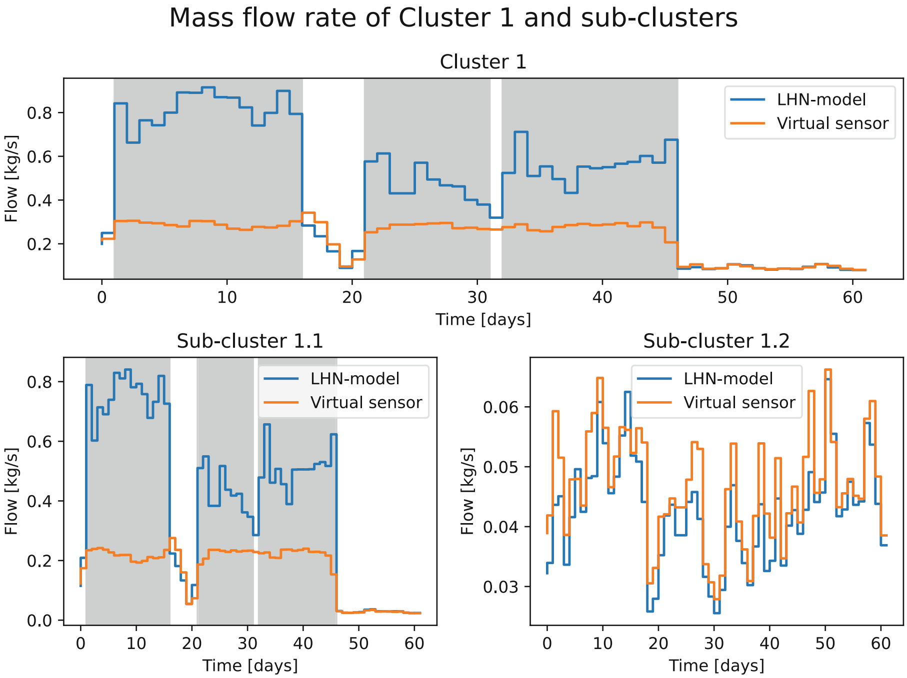

Fig. 7. Cluster 1; Shows the mass flow rate in the virtual sensors supply Cluster 1 and Sub-cluster 1.1 and 1.2. Grey marked areas represent faults detected. (Color figure online)

that it is a local fault at the consumer or its service pipe and is possibly not due to a fault in the upstream pipe network.

Figure 8 shows a consumer cluster of another part of the DHS. Cluster 2 contains three terrace houses with 13 individual consumers in total. Cluster 2 is located far away from the sub-station, and the consumers are the last connected to the pipe branch. Sub-cluster 2.1, 2.2, and 2.3 contains four, five, and four individual consumers, respectively. Figure 8 shows that two faults were detected in Cluster 2 (grey marked areas).

Investigating the three sub-clusters of Cluster 2, in Fig. 8, shows faults were detected only in Sub-clusters 2.2 and 2.3. Notice that some of the faults in Cluster 2 and Sub-cluster 2.2 and 2.3 happen in the same time span, indicating a correlation of the faults. An investigation of the individual consumers in the two faulty sub-clusters showed that faults only were detected at one out of five consumers in Sub-cluster 2.2 and at three out of four consumers in Sub-cluster 2.3.

A visualisation of Cluster 2 and its sub-clusters can be seen in Fig. 9. As faults were both detected in Sub-cluster 2.2 and 2.3, the location of the fault could be upstream in the pipe network from Sub-cluster 2.2. A precise location

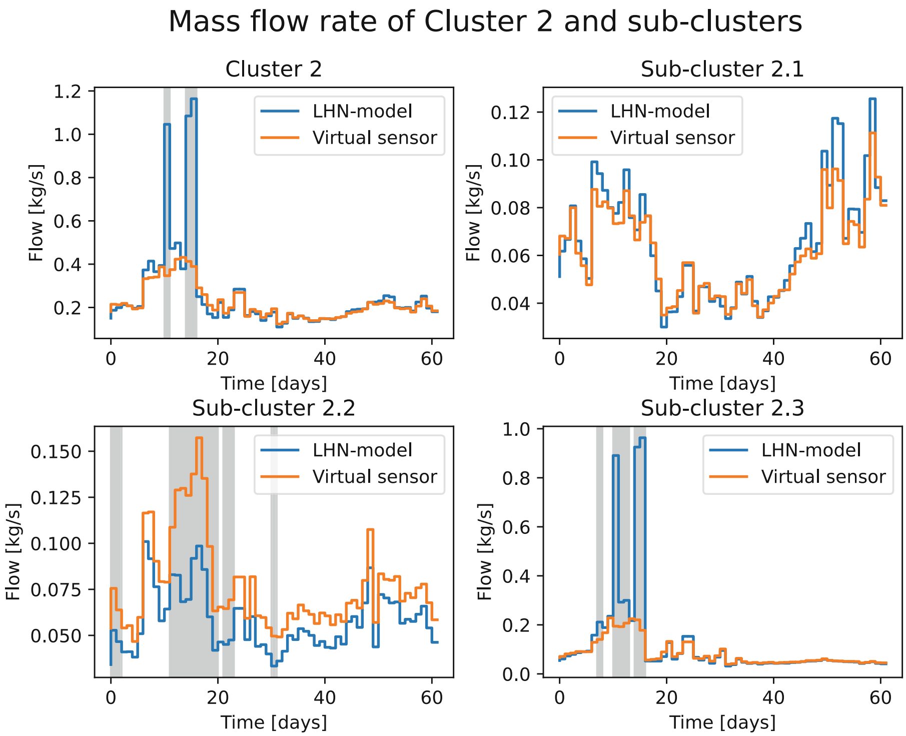

Fig. 8. Cluster 2; Shows the mass flow rate in the virtual sensors supply Cluster 2 and Sub-cluster 2.1, 2.2, and 2.3. Grey marked areas represent faults detected. (Color figure online)

of the fault or multiple faults can be difficult to determine and will have to be found by a thorough investigation.

One possibility could be a fault had occurred in the pipe between Sub-cluster 2.1 and 2.2 as faults were detected in both sub-clusters downstream from Sub-cluster 2.1. This fault could be due to pipe leakage or damage to the pipe insulation, among other things. A fault in this pipe could be the reason for the faults detected further downstream in the pipe network and would have a cascading effect of introducing multiple faults.

Another scenario could be that a fault had occurred in the pipe between Sub-cluster 2.2 and 2.3 and that the single fault detected in Sub-cluster 2.2 is due to a local fault at the faulty consumer. It is unlikely that the faults at three out of four consumers in Sub-cluster 2.3 are due to local faults, which increases the possibility of a fault had occurred further upstream in the pipe network, possibly the main pipe going into the terrace house.

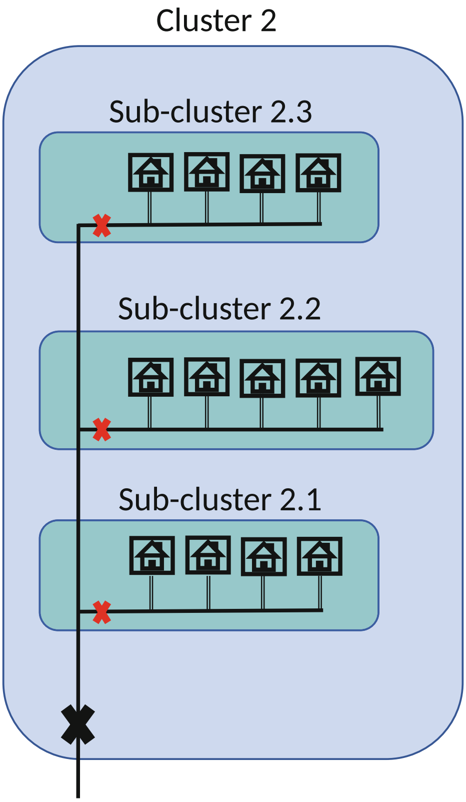

Fig. 9. Visualisation of the pipe network configuration of Cluster 2 and Sub-Cluster 2.1, 2.2, and 2.3. The blue area represents Cluster 2, and the black cross is its virtual sensor. Green areas represent the sub-clusters, and red crosses their virtual sensor. (Color figure online)

3.4 Conclusion and Future Work

Faults occurring in the complex domain of DHSs must be detected, and corrective actions must be made to ensure the cost-effectiveness of the system, avoid disturbance of loads, and lower the overall energy losses and CO_2 emissions of the system. To target these objectives and locate faults, this paper proposes a digital twin-based FDD framework with virtual sensor employments to detect, locate faults in DHSs

The framework defines virtual sensors in the DHS's pipes measuring the mass flow rate, where real sensors are absent, by summing up the total mass flow rate at consumer sensors downstream in the pipe network from the virtual sensors. Faults were detected by the framework by investigating residuals between the calculated and digital twin-simulated mass flow rate in pipes using a bound of normal operation defined as $Target \cdot (1 \pm \beta)$, where the ratio parameter (β) can be calibrated.

To define which virtual sensors are of interest and should be monitored, this paper proposes a Specialised Agglomerative Hierarchical Clustering algorithm that validates discovered clusters with information about the virtual sensors. The framework shows a great ability to detect faults in places where real sensors are absent, and it uses a top-down approach to narrow down the location of the fault. Faults found in the case study of this paper could not be confirmed and diagnosed since Fjernvarme Fyn did not have a maintenance record of abnormalities at the locations of detected faults by the framework. For future work, the framework should be implemented on a large-scale DHS, with a focus testing how well it generalises and performs quantitatively, using metrics like precision and recall.

Acknowledgements. This work is supported by the "Proactive and Predictive Maintenance of District Heating Systems" and "IEA DHC TS4", funded by the Danish Energy Agency under the Energy Technology Development and Demonstration Program, ID number 64020-2102 and 134-22011, respectively. We also thank Peer Andersen, Lasse Elmelund Pedersen, and the rest of their team at Fjernvarme Fyn A/S for their assistance with the data and the model. Also, thanks to Johan Peter Alsing from Danfoss A/S for assisting us with Leanheat Network.

References

1. Bahlawan, H., et al.: Detection and identification of faults in a district heating network. Energy Convers. Manage. **266**, 115837 (2022). https://doi.org/10.1016/j.enconman.2022.115837
2. Buffa, S., Fouladfar, M.H., Franchini, G., Lozano Gabarre, I., Andrés Chicote, M.: Advanced control and fault detection strategies for district heating and cooling systems-a review. Appl. Sci. **11**(1) (2021). https://doi.org/10.3390/app11010455
3. Danfoss: Concept guide: Leanheat network concept and modeling elements(2021). Accessed 9 Sept 2023
4. Dijkstra, E.W.: A note on two problems in connexion with graphs. Numer. Math. **1**(1), 269–271 (1959)
5. EU: Commission recommendation (eu) 2019/786 of 8 May 2019 on building renovation (2019). https://bit.ly/30nxBs5
6. Katipamula, S., Brambley, M.R.: Review article: methods for fault detection, diagnostics, and prognostics for building systems-a review, part ii. HVAC&R Res. **11**(2), 169–187 (2005). https://doi.org/10.1080/10789669.2005.10391133
7. Mattera, C.G., Quevedo, J., Escobet, T., Shaker, H.R., Jradi, M.: A method for fault detection and diagnostics in ventilation units using virtual sensors. Sensors (2018). https://www.mdpi.com/1424-8220/18/11/3931
8. Månsson, S., Kallioniemi, P.O.J., Sernhed, K., Thern, M.: A machine learning approach to fault detection in district heating substations. Energy Procedia **149**, 226–235 (2018). https://doi.org/10.1016/j.egypro.2018.08.187, 16th International Symposium on District Heating and Cooling, DHC2018, 9-12 September 2018, Hamburg, Germany
9. Pakanen, J., Hyvärinen, J., Kuismin, J., Ahonen, M.: Fault diagnosis methods for district heating substations. VTT Tiedotteita - Valtion Teknillinen Tutkimuskeskus (1996)

10. Pedersen, A.S.H., Ustrup, S.E., Mortensen, L.K., Shaker, H.R.: Data validation for digitally enabled operation maintenance of district heating systems. In: 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), pp. 1–7 (2022). https://doi.org/10.1109/ICECCME55909.2022.9988721
11. Rohlf, F.J.: Single-link clustering algorithms (1987)
12. 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 AB (2013)
13. Sibson, R.: SLINK: an optimally efficient algorithm for the single-link cluster method. Comput. J. **16**(1), 30–34 (1973). https://doi.org/10.1093/comjnl/16.1.30
14. Søndergaard, H.A.N., Shaker, H.R., Jørgensen, B.N.: Automated and real-time anomaly indexing for district heating maintenance decision support system (preprint). SSRN Electron. J. (2023). https://doi.org/10.2139/ssrn.4344182
15. Sun, W., Cheng, D., Peng, W.: Anomaly detection analysis for district heating apartments. J. Appl. Sci. Eng. **21**, 33–44 (2018). https://doi.org/10.6180/jase.201803_21(1).0005
16. Xue, P., et al.: Fault detection and operation optimization in district heating substations based on data mining techniques. Appl. Energy **205**, 926–940 (2017). https://doi.org/10.1016/j.apenergy.2017.08.035
17. Yoon, S., Choi, Y., Koo, J., Hong, Y., Kim, R., Kim, J.: Virtual sensors for estimating district heating energy consumption under sensor absences in a residential building. Energies (2020). https://www.mdpi.com/1996-1073/13/22/6013
18. Yu, W., Patros, P., Young, B., Klinac, E., Walmsley, T.G.: Energy digital twin technology for industrial energy management: classification, challenges and future. Renew. Sustain. Energy Rev. **161**, 112407 (2022). https://doi.org/10.1016/j.rser.2022.112407, https://www.sciencedirect.com/science/article/pii/S136403212200315X
19. Zimmerman, N., Dahlquist, E., Kyprianidis, K.: Towards on-line fault detection and diagnostics in district heating systems. Energy Procedia **105**, 1960–1966 (2017). https://doi.org/10.1016/j.egypro.2017.03.567, 8th International Conference on Applied Energy, ICAE2016, 8-11 October 2016, Beijing, China