Spot the difference!

On the way towards automated fault handling in district heating buildings

Månsson, Sara

2021

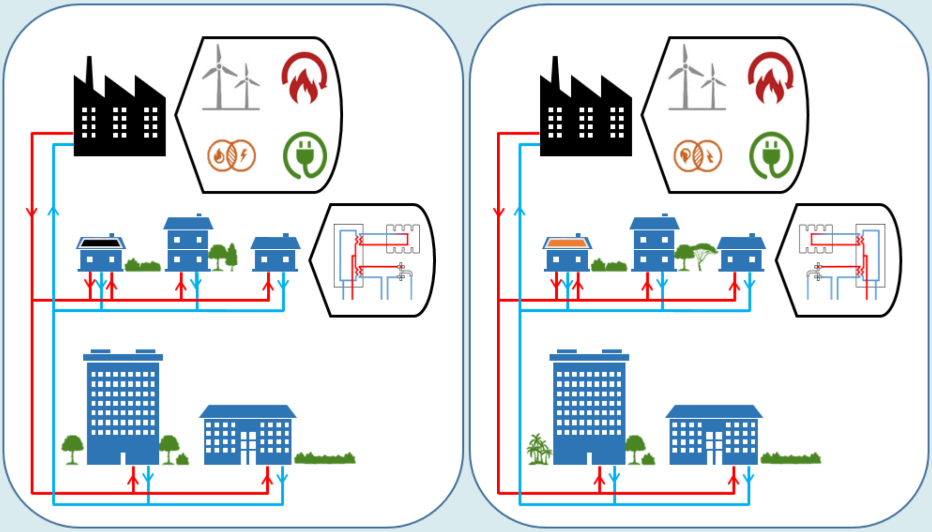
LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Spot the difference!

On the way towards automated fault handling in district heating buildings

SARA MÅNSSON
FACULTY OF ENGINEERING | DEPARTMENT OF ENERGY SCIENCES | LUND UNIVERSITY



Spot the difference!

Spot the difference!
On the way towards automated fault handling in district heating buildings

by Sara Månsson

Thesis for the degree of Doctor of Philosophy in Engineering
Thesis advisors: Associate professor Marcus Thern, Associate professor Kerstin Sernhed, Associate professor Per-Olof Johansson Kallioniemi
Faculty opponent: Professor Svend Svendsen

To be presented, with the permission of the Faculty of Engineering of Lund University, for public criticism in lecture hall KC:A at Kemicentrum on Friday, the 20th of August 2021 at 09:00.

DOKUMENTDATABLAD enl SIS 61 41 21

Organization
LUND UNIVERSITY

Department of Energy Sciences
Box 118
SE–221 00 LUND
Sweden

Author(s)
Sara Månsson

Document name
DOCTORAL DISSERTATION

Date of disputation
2021-08-20

Sponsoring organization
Vito
Boeretang 200
BE-2400 Mol, Belgium

Title and subtitle
Spot the difference!: On the way towards automated fault handling in district heating buildings

Abstract
This thesis focuses on how district heating utilities may work with fault handling in customer installations as a tool to reduce their return temperature levels. Reduced return temperatures would make it possible to improve the efficiency of the district heating systems and decrease their environmental impact. Faults in the customer installations cause high return temperatures that increase the return temperature of the entire district heating system. Thus, to reduce the system return temperatures, such faults must be detected, diagnosed, and corrected. The focus of the thesis has been to understand and identify the specific challenges that arise when integrating data analysis of district heating customer data as a natural part of the utilities' fault handling processes and find (parts of) the solutions to these challenges. The reason for this is that faults in the customer installation manifest themselves in district heating customer data. Thus, it is possible to analyse this data to detect faults.

The results show that a multitude of different faults may occur in a customer installation. However, some faults are more common than others, and these faults should be prioritized when developing fault detection methods using data analysis. There is also a need to develop a mutual way to name faults that occur in the installation. Today, the district heating utilities use different words when talking about the same faults, making it hard to compare results from different district heating systems. It also makes it hard to create labelled data sets, which are needed to develop and validate successful fault detection tools based on data analysis. However, two different data analysis methods for fault detection have been developed without access to labelled data. Both methods show that it is possible to utilize customer data analysis to detect faults in customer installations.

To realize the full potential of the fault detection tools, the utilities need to utilize them on a larger scale within their organization. Important organizational aspects include identifying a clear stakeholder in the fault handling process and calculating the economic value of eliminating faults. There is also a need to change the current fault handling processes to better align with automated fault detection tools. Therefore, an important result in this thesis is a suggestion for how a fault handling process based on data analysis may be designed.

District heating utilities may use the results of this thesis to improve their fault handling processes of customer installations, thereby reducing their return temperatures and improving the possibilities of district heating being part of the future energy systems.

Key words
Customer installations, Customer data, Data analysis, District heating, Experience from industry, Fault detection, Fault handling, Fault labelling, Return temperatures, Substation performance, Taxonomy

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title
0282-1990

ISBN
978-91-7895-865-8 (print)
978-91-7895-866-5 (pdf)

Recipient's notes

Number of pages
99

Price

Security classification

Signature

Date 2021-05-17

Spot the difference!

On the way towards automated fault handling in district heating buildings

by Sara Månsson

Lund University

A doctoral thesis at a university in Sweden takes either the form of a single, cohesive research study (monograph) or a summary of research papers (compilation thesis), which the doctoral student has written alone or together with one or several other author(s).

In the latter case the thesis consists of two parts. An introductory text puts the research work into context and summarizes the main points of the papers. Then, the research publications themselves are reproduced, together with a description of the individual contributions of the authors. The research papers may either have been already published or are manuscripts at various stages (in press, submitted, or in draft).

Cover illustration front: Spot the difference in a district heating system - find five differences in the two illustrations of a district heating system. Correct answers (from top-to-bottom and left-to-right in the illustration to the right): CHP icon upside down, roof in orange, different tree between the second and third house, substation mirror-inverted, different tree next to the tallest building.

Cover illustration back: A word cloud of my PhD thesis created at wordclouds.com.

Funding information: The thesis work was financially supported by VITO.

Faculty of Engineering, Department of Energy Sciences

ISBN: 978-91-7895-865-8 (print)
ISBN: 978-91-7895-866-5 (pdf)
ISRN: LUTMDN/TMHP-19/1164-SE
ISSN: 0282-1990

Printed in Sweden by Media-Tryck, Lund University, Lund 2021

Courage doesn't always roar.
Sometimes courage is the little voice at the end of the day that says
"I will try again tomorrow"
- Mary Anne Radmacher

Acknowledgements

My PhD project has been a journey down a long and winding (and sometimes very confusing and frustrating) road. Many have helped me stay sane and on track throughout the process, and I would like to thank all of you for your support, ”pepp”, and encouraging words. Some people have helped me more than others, and they deserve a more personal thank you - without you guys, I think I would have quit this project several years ago.

First and foremost, my supervisors. Thank you, Marcus Thern, my main supervisor throughout this project. Thank you for all discussions we’ve had about LaTeX, bird news, machine learning, pie charts, and life in general.

Thank you, Kerstin Sernhed, for being my co-supervisor and asking all the hard questions about my project. Thank you for the songs we’ve sung together, our talks, and for all the fun moments we’ve had together during conferences, coffee breaks, and everyday life at the office.

Per-Olof Johansson Kallioniemi, thank you for taking the role as my co-supervisor halfway through my project and for helping me whenever I’ve needed help, especially in the more technical aspects of my projects. And thank you for our discussions regarding beers, books, and films and our collaboration when switching to online teaching!

Patrick Lauenburg, the initiator behind this project and the person who was initially my main supervisor. Thank you for teaching me that I should keep a sane approach towards my work-life and for helping me build the foundation of this project.

I would also like to direct a big THANK YOU to all of my colleagues at the department for our interesting discussions in the lunchroom and for all of the fun moments we’ve shared together throughout the years. Thank you, Jens, Ted, and Alexandra, for all of our interesting discussions. I would also especially like to thank those of you who not only have been my colleagues but who have also become my close friends. To Lisa Brange, my first PhD partner in crime for your great company, for teaching me that good enough is the new awesome, for all of our great moments together (rose experiments, conference dinners, coffee breaks, etc.), and for being a great friend. To Nika Alemahdi, for being one of the most humble, generous, and inspiring people I've met, for always caring about me, and for all of our discussions regarding both PhD life and real-world life.

Thank you to the people at Vito in Belgium and especially to Robbe Salenbien. Thank you for bringing me to lunch, showing me life in Belgium, and introducing me to many great Belgian beers. Thank you to Dirk Vanhoudt for your inputs on my project and for helping me get in contact with people within the DH research area. Thank you also to Tijs and Jad for interesting discussions about both my project and other matters.

I would also like to thank some people that I have gotten to know during my project and that have contributed to my work in different ways. Thank you to the people involved in the Smart Energi project, who have given my research work a clear context during the last part of my project. I would also like to send a special thank you to Henrik Gadd for your help and support throughout my entire PhD project.

Of course, several people outside of my workplace(s) have been a great support during my project. Thank you, Caroline, Matilda, and Emma, for listening to me complaining about my project, coping with my frustration, encouraging me when I've needed it the most, and having (one too many) beers with me when I haven't coped. Thank you, Louise and Gabbi, for your PhD-to-PhD support throughout the years. It has been great being able to discuss PhD-related matters with someone who understands *exactly* what I'm going through. I would also like to thank my friends in the Dr. Nano group - Malin, Martin, Axel, Axel, Mårten, Joel, Anders, Emil, Calle, Marcus - for our Friday lunches!

Finally, to the most important people in my life - my family. Thank you, Mum, Dad, Olof and Emma. You mean the world to me, and I am so grateful for your support throughout this project. Thank you for all of our family dinners, listening to my complaints, and for your support. Thank you, Victor. I don't know if I would have finished this project without you. Thank you for being the most important person in my life, for always supporting me, for your unconditional love. You are the best.

Contents

Acknowledgements **v**

Popular summary **ix**

Populärvetenskaplig sammanfattning **xi**

List of publications **xiii**

1 Introduction **1**
- 1.1 Objectives 4
- 1.2 Limitations 4
- 1.3 Outline of thesis 5

2 District heating – basic concepts & fundamentals **7**
- 2.1 Heat supply in district heating 7
- 2.2 Current and future potential of district heating 9
- 2.3 Distribution of district heating 11
- 2.4 Heat loads in a district heating system 13
- 2.5 Benefits of decreased temperature levels 15

3 The customer installation **17**
- 3.1 Customer substations 17
  - 3.1.1 Customer substation ownership 18
  - 3.1.2 Customer substation connection principles 19
- 3.2 Space heating systems 19
- 3.3 Domestic hot water systems 22
- 3.4 District heating metering 22

4 Fault detection and diagnosis using data analysis **25**
- 4.1 Definition of fault 25
- 4.2 Fault Detection and Fault Diagnosis 26
  - 4.2.1 Model-based fault detection 26

4.2.2 Process history-based fault detection 27
4.2.3 Machine learning 27
4.3 Fault detection in district heating customer data 29

5 Methodology **33**
5.1 Data sets being used in the thesis 33
5.2 An initial approach – fault detection using data analysis 34
5.3 An attempt at validating the fault detection methods 36
5.4 Shifting focus – learning more about current fault handling processes . 37
5.5 A taxonomy for labelling deviations in customer data 40
5.6 Developing a workflow for fault handling 41

6 Results and analysis **43**
6.1 Faults in district heating customer installations 43
6.1.1 Most frequently occurring faults 45
6.1.2 Developing a taxonomy of fault labels 46
6.2 How district heating utilities work with fault handling today 48
6.3 How district heating utilities carry out fault detection today 50
6.4 Fault detection methods based on analysis of district heating customer data 52
6.5 Future fault handling processes at the district heating utilities 56

7 Concluding discussion **61**

References **65**

Scientific publications **75**

Popular summary

District heating is a common way of heating houses in Sweden today, which is both reliable and easy to use. However, the heating may not always work correctly, resulting in the house getting cold and not getting hot water in the tap when washing your hands. This type of problem is often due to something going wrong in the house's heating system. These faults not only affect those who live in the house but can also have a negative impact on the district heating system. Therefore, it is common for district heating companies to offer help to the customers who experience that something is wrong with the heating system in their buildings. There may also be faults in the houses that customers do not notice, and the district heating companies must find those faults as well. In this dissertation, I have investigated why district heating companies want to find and remove faults in the buildings, why faults occur in the buildings, and how district heating companies can work to find and correct more faults in the buildings.

So, why do district heating companies want to find more faults? I have investigated this by interviewing several district heating companies. They say that the biggest reason why they want to find faults is that the building's return temperature is affected. The return temperature is the water temperature that leaves the building after the heat from the district heating system has been delivered. Suppose the return temperature from the building is high. In that case, it means that the building has not been able to utilize the heat in the water in an efficient manner. Thus, the high return temperature has a negative impact on the entire district heating system, as the heat is not used efficiently in the system. If high return temperatures can be avoided, heat can also be produced more efficiently in several types of heat production plants. A more efficient heat production leads to reduced use of resources and more environmentally friendly heat production. Therefore, district heating companies need to find faults in buildings that lead to increased return temperatures.

The faults in the buildings can be due to many different things. I have investigated this in my dissertation by asking many Swedish district heating companies what the most common faults in their systems are. In a building connected to district heating, a district heating substation transfers the heat from the district heating system to the building's heating system. The substation consists of many different components such as valves, various meters, and a heat exchanger. These components may break or stop working correctly,

for example, if a temperature sensor has stopped working or if a mouse has gnawed off an electrical cord. Sometimes the error can also be due to the human factor. For example, suppose someone changes the settings in the substation's regulator, which determines how much heat is to be delivered to the building. In that case, the changed settings may cause an unnecessarily high energy consumption in the building.

To get more structured information about the faults in the buildings, I have developed a way to name the different faults. This way of naming faults makes it easier if the district heating companies would like to collaborate on how they handle faults and, for example, compare which errors are the most common ones in the industry. The naming is based on several different levels, where you must first describe where in the building the fault has appeared. Then you must explain which component it is that is faulty, what the fault is, if the fault has been corrected, and if so, what was done to correct the fault. By filling in this information every time you find a fault in a building, the district heating companies will know more about the faults that occur in their district heating system. They will also be able to help their customers even more by utilizing this knowledge.

The district heating substation also has a meter, which measures four different measurement values: the water temperature that enters the building, the return temperature from the building, how water passes through the building, and how much energy is used in the building. The measured values can be used to find faults in the buildings by analyzing the measurements with the help of computers. Faults may be detected in this way since a fault will create changes in the measured values, for example, if the fault causes the return temperature to increase. Because there are so many buildings in a district heating system, the data analysis must be fast. In this dissertation, I have examined two different data analysis methods, which show that this type of data analysis is possible. However, we need to know more about the customer data to create even better methods for fault detection. More information would make it possible to investigate whether the data analysis tools find faults in buildings that have had faults in them in reality.

Last but not least, it is also important that the district heating companies work with the faults systematically. Therefore, they may need to adjust the way they work today and, above all, look more actively for faults in the buildings. Some district heating companies do this today, and if they detect faults, they contact their customers to help them correct the faults. To further develop this work, I have developed a proposal on how a district heating company can work to find more faults in its district heating system. The workflow contains several sub-steps, where customer data analysis is one of the essential parts. To ensure that the workflow is possible to introduce in reality, I have conducted a workshop where several Swedish district heating companies participated and helped me evaluate the working method. In this way, we have developed a workflow that contains several smart solutions, which can help solve many of the problems related to fault detection and fault handling in the buildings.

Populärvetenskaplig sammanfattning

Fjärrvärme är ett vanligt sätt att värma upp hus i Sverige idag, som är både pålitligt och enkelt att använda sig av. Det kan dock hända att uppvärmningen inte fungerar som den ska, vilket kan resultera i att huset blir kallt, och att det inte kommer något varmvatten i kranen när man ska tvätta händerna. Ofta beror den här typen av problem på att något har blivit fel i det uppvärmningssystem som finns i huset. Dessa fel påverkar inte bara de som bor i huset, utan kan även påverka fjärrvärmebolaget på ett negativt sätt. Därför är det vanligt att fjärrvärmebolagen erbjuder hjälp till de kunder som upplever att något är fel i värmesystemet i deras byggnader. Det finns också fel i husen som kunderna inte märker av, och det är viktigt att fjärrvärmebolagen hittar de felen också. I den här avhandlingen har jag undersökt anledningar till varför fjärrvärmebolag vill hitta och ta bort fel i byggnaderna, varför fel uppstår i byggnaderna, hur man kan hitta fler fel i byggnaderna, och hur man skulle kunna arbeta för att åtgärda felen.

Varför vill då fjärrvärmebolagen hitta fler fel? Detta har jag undersökt genom att intervjua ett antal fjärrvärmebolag, som säger att den största anledningen till att de vill hitta fel är att byggnadens returtemperatur påverkas. Returtemperaturen är temperaturen på det vatten som lämnar byggnaden efter att värmen från fjärrvärmesystemet har levererats. Om returtemperaturen från byggnaden är hög, innebär det att byggnaden inte har kunnat ta tillvara på värmen i vattnet på ett effektivt sätt. Detta får en negativ påverkan på hela fjärrvärmesystemet, i och med att värmen inte används effektivt i systemet. Kan man undvika höga returtemperaturer kan man också producera värme mer effektivt i flera typer av värmeproduktionsanläggningar. Detta leder till minskad användning av resurser och en mer miljövänlig värmeproduktion. Därför är det viktigt för fjärrvärmebolagen att hitta fel i byggnader som leder till ökade returtemperaturer.

Felen i byggnaderna kan bero på många olika saker, och detta har jag undersökt i min avhandling genom att fråga ett stort antal svenska fjärrvärmebolag vilka de vanligaste felen i deras system är. I en byggnad kopplad till fjärrvärme finns det en fjärrvärmecentral som överför värmen från fjärrvärmesystemet till byggnadens uppvärmningssystem och består av många olika komponenter som ventiler, olika mätare och en värmeväxlare. Alla dessa komponenter kan gå sönder eller sluta fungera som de ska, till exempel om en temperaturmätare har slutat fungera eller om en mus har gnagt av en elsladd. Ibland kan felet också bero på den mänskliga faktorn, om någon till exempel ändrar inställningar i regulatorn i fjärrvärmecentralen som bestämmer hur mycket värme som ska levereras till byggnaden på ett sätt som leder till onödig energiförbrukning.

För att få mer strukturerad information om felen som finns i byggnaderna har jag tagit fram ett sätt att namnge de olika felen. Detta underlättar om fjärrvärmebolagen skulle vilja samarbeta kring hur de hanterar fel, och till exempel jämföra vilka fel som är de vanligaste inom branschen. Namngivningen är baserad på flera olika nivåer, där man först ska beskriva var i byggnaden har uppstått. Därefter ska man förklara vilken komponent det är som det finns ett fel på, vad felet är, om felet har tagits bort, och vad som i sådana fall gjordes för att ta bort felet. Genom att fylla i den här informationen varje gång man hittar ett fel i en byggnad, kommer det på sikt bli möjligt att skapa bättre dataanalysmetoder som kan hitta felen automatiskt. Fjärrvärmebolagen kommer också att veta mer om vilka fel som förekommer i deras fjärrvärmesystem, och de kommer att kunna hjälpa sina kunder ännu mer med hjälp av den kunskapen.

I fjärrvärmecentralen finns också en mätare, som mäter fyra olika mätvärden: temperaturen på vattnet som kommer in i byggnaden, returtemperaturen från byggnaden, hur stor volym vatten som passerar genom byggnaden, och hur mycket energi som används i byggnaden. Dessa mätvärden kan användas till att hitta fel i byggnaderna, genom att analysera mätvärdena med hjälp av datorer. Detta beror på att ett fel kommer att skapa förändringar i mätvärdena, till exempel om felet gör så att returtemperaturen ökar. I och med att det finns så många byggnader i ett fjärrvärmesystem är det viktigt att dataanalysen sker snabbt. I den här avhandlingen har jag undersökt två olika metoder för denna typ av dataanalys, som visar att denna typ av dataanalys är möjlig. Vi behöver dock veta mer om kunddatan för att kunna skapa ännu bättre metoder för feldetektering. Mer information skulle göra det möjligt att undersöka om dataanalysen hittar fel i byggnader som faktiskt har haft fel i sig i verkligheten. Ett slags facit helt enkelt.

Sist men inte minst är det också viktigt att fjärrvärmebolagen arbetar med felen på ett systematiskt sätt. Detta innebär att de kan behöva justera sättet de arbetar på idag, och framför allt leta mer aktivt efter fel i byggnaderna. Vissa fjärrvärmebolag gör detta idag, och då kontaktar de kunderna för att hjälpa dem hantera felen. För att bygga vidare på detta har jag tagit fram ett förslag på hur ett fjärrvärmebolag kan arbeta för att hitta fler fel i sitt fjärrvärmesystem. Arbetssättet innehåller flera delsteg, där analys av kunddata är en av de viktigaste delarna. För att se till så att arbetssättet är möjligt att införa i verkligheten har jag genomfört en workshop där ett antal svenska fjärrvärmebolag medverkade och hjälpte mig utvärdera arbetssättet. På så sätt har vi tagit fram ett arbetssätt som innehåller flera smarta lösningar, och som kan hjälpa till att lösa många av de problem som är relaterade till feldetektering och felhantering i byggnaderna.

List of publications

This thesis is based on the following publications, referred to by their Roman numerals:

I **A machine learning approach to fault detection in district heating substations**
Sara Månsson, Per-Olof Johansson Kallioniemi, Kerstin Sernhed, Marcus Thern.
Reprinted from Energy Procedia, vol. 149, pp. 226-235, 2018, with permission from Elsevier Science.

II **Automated statistical methods for fault detection in district heating customer installations**
Sara Månsson, Kristin Davidsson, Patrick Lauenburg, Marcus Thern.
Reprinted from Energies, vol. 12(1), nbr. 113, 2018, with permission from MDPI.

III **Faults in district heating customer installations and ways to approach them: Experiences from Swedish utilities**
Sara Månsson, Per-Olof Johansson Kallioniemi, Marcus Thern, Tijs van Oevelen, Kerstin Sernhed.
Reprinted from Energy, vol. 180, pp. 163-174, 2019, with permission from Elsevier Science.

IV **A taxonomy for labeling deviations in district heating customer data**
Sara Månsson, Ida Lundholm-Benzi, Marcus Thern, Robbe Salenbien, Kerstin Sernhed, Per-Olof Johansson Kallioniemi.
Reprinted from Smart Energy, vol. 2, 100020, 2021, with permission from Elsevier Science.

V **A fault handling process for faults in district heating customer installations**
Sara Månsson, Marcus Thern, Per-Olof Johansson Kallioniemi, Kerstin Sernhed.
Preprint submitted to Energies

Chapter 1

Introduction

District heating is a heating alternative that has the potential to satisfy both space heating and domestic hot water (DHW) demands in an environmentally friendly, resource-efficient, and cost-efficient way [1]. Therefore, district heating has been identified as playing a key role in the decarbonisation of the European heating sector [2]. The benefits of district heating are more prominent when the system is being supplied by renewable heat sources such as geothermal and solar thermal heat and excess heat from power production and industrial processes based on biomass or waste incineration [3]. The district heating systems further provide a possibility to increase the flexibility of the energy system by integrating the heating and electricity sectors if thermal energy storages and large-scale heat pumps are introduced into the energy systems [4]. Such integration creates important synergy effects between the different energy sectors that will be crucial in the future energy systems [5]. There is thus a need to implement district heating on a large scale in future energy systems.

Today, the district heating technology has had varying success in penetrating the national heat markets. The average utilization of district heating in the world and the European Union amounts to approximately 10 % [6]. However, the national figures vary significantly – in some countries, district heating corresponds to well below 10 % of the heat market. In contrast, some countries have a high penetration of district heating in the heat market with more than 50 % of the market shares [6]. In Sweden, district heating accounts for approximately 50 % of the heat market and is supplied by a significant share of renewable and recycled heat [7]. Thus, district heating is already a well-established and environmentally friendly heating method in Sweden. However, the recent development in the heat market has led to Swedish district heating being less competitive. Challenges include increased competition from other heating alternatives such as cheaper, more efficient heat pumps (particularly in small, private houses) and decreased heat demands from customers due to energy efficiency measures and warmer climates. District heating is also challenged by negative customer perceptions, which are related to the natural monopoly situation of district heating, as well as historically high prices [8]. These challenges also occur in other countries where district heating is an already well-established heating method. An additional challenge in these countries is the competition from other cheap heat sources, such as gas boilers [9]. In countries where district heating is not as common, extra challenges arise since the technology has to penetrate already existing heat markets and overcome barriers such as high capital investment costs, customer unawareness, and national legislations and policies [10]. It is thus crucial for district heating to face and overcome these challenges to stay competitive and acquire increased market shares. Otherwise, it will be difficult for district heating to live up to the role the technology has been identified to play in the future energy systems.

Although district heating is facing somewhat different challenges in different countries, the solutions to the problems could still be very similar. In the end, the main goal of dealing with these challenges is to achieve cost-competitive heating that is energy and resource-efficient, where heat is supplied with low greenhouse gas emissions [1, 11]. One determining factor in this work is to decrease the distribution temperatures in the district heating systems, and specifically the return temperature. By reducing the return temperature, it would be possible to reduce the supply temperature as well. However, it is essential to reduce the return temperature before reducing the supply temperature since it is crucial not to reduce the temperature difference between the supply and return temperature when decreasing the distribution temperatures [1]. Reduced distribution temperatures would result in lower heat losses from the system, the possibility to introduce low-temperature heat sources such as geothermal and excess heat from industries, and increase the power-to-heat ratio in combined heat and power (CHP) plants [3, 12]. These aspects improve both the cost competitiveness and environmental profile of the district heating system. Despite the importance of lower temperature levels, many current district heating utilities cannot decrease their system's temperature levels. One can then ask why this is the case.

Frederiksen and Werner [1] state that there are four main reasons why current return temperatures are higher than they theoretically could be: (i) Bypass flows between supply and return pipes in the district heating system, (ii) Low supply temperatures at peripheral substations because of high heat losses, (iii) Faults in customer heating systems, and (iv) Faults in customer substations. Bypass flows exist in most district heating systems and are created by connecting the supply and return media pipes using a bypass valve. The bypass makes it possible to maintain a sufficient supply temperature at the customer substations, irrespective of where in the system a substation is located by maintaining a flow in the system at all instances[13]. The bypass flows are especially important during periods of low heat load in the system. Low supply temperatures at the peripheral substations are often compensated by higher flows through the substation, causing higher return temperatures [1]. Issues (i) and (ii) are the responsibility of the district heating utilities and are already well known in the industry.

The focus of this PhD thesis is instead on the two fault categories related to the customer installations in the district heating systems: faults in customer heating systems and faults in customer substations. Potentially, each customer installation connected to a district heating system may contain faults in one or both of these categories. Previous studies have shown that as many as 74 % of the customer installations may contain faults [14]. These faults may thus be the foremost reason why the return temperature is higher than desired in a district heating system [1]. Therefore, faults belonging to one of these two categories must be corrected when working towards reduced distribution temperature levels.

Faults in the customer substations and internal heating systems occur in the interface between the district heating system, which is the responsibility of the district heating utility, and the installations connected to the system, which most often are the responsibility of the customer. Thus, there is a shared responsibility between the utility and the customers regarding these faults [1]. However, many of the faults do not affect customer comfort and will not be identified by the customers. In most cases, the faults mainly affect the efficiency of the district heating system, thus mainly affecting the district heating utility [15]. Therefore, utilities that want to decrease their distribution temperatures must work actively to detect, identify and correct these faults.

This thesis focuses on how district heating utilities may work with fault handling in customer installations as a tool to reduce their return temperature levels. The utilities can make use of customer data that is gathered for billing purposes to detect faults. Faults in customer installations will alter the behaviour of either the customer heating system or the customer substation, and this change of behaviour will manifest itself in customer data [14, 16]. The customer data consist of the parameters measured by the heat meter in the district heating substations: heat consumption, water volumetric flow, supply temperature, and return temperature [1]. By analysing this data and looking for deviating patterns, it is possible to detect faults in customer installations connected to district heating systems [15]. If making the analysis automated, it should also be possible to detect faults more rapidly than other types of manual fault detection. Fault detection in district heating customer installations using data analysis has been the focus of several recent research projects, for example, [14, 15, 17–19]. However, there are very few studies that focus on the ”bigger picture” of fault handling. That is, dealing with research questions such as: ’what is the state-of-the-art fault handling?’, ’how may the fault handling processes be improved?’, and ’how may fault detection based on data analysis fit into, and improve, the current fault handling processes?’.

The work conducted in this thesis approaches the problem of fault handling in district heating customer installations from a holistic point of view, and the results include both organizational and technical aspects of fault handling. The results include methods for fault detection based on data analysis, suggestions on how data analysis may be utilized as a natural part of the utilities’ fault handling processes, and proposals of frameworks that must be developed to improve fault handling processes. The point of view taken is that of the utilities – after all, they benefit the most from eliminating faults in the customer installations.

1.1 Objectives

The overarching objective of this thesis has been to contribute to improving the fault handling processes at current district heating utilities. The focus of the thesis has been to understand and identify the specific challenges that arise when integrating data analysis of district heating customer data as a natural part of the utilities' fault handling processes and to find (parts of) the solutions to these challenges. The results include specific fault detection methods, data quality, organizational aspects regarding how utilities are currently working with fault handling, and how these fault handling processes may be improved. Furthermore, the aim was to explore the possibilities to introduce this knowledge and these processes into future district heating systems. The specific objectives of the thesis were to:

1. Implement data analysis methods for fault detection on data from "real-world" customer installations connected to district heating, and identify challenges related to the implementation of such methods (Papers I, II and IV).
2. Investigate what faults occur in customer installations in district heating systems, and how data sets from these installations may be labelled in a standardized way to facilitate implementation and validation of fault detection methods (Papers III and IV).
3. Investigate how district heating utilities are currently working with fault handling to understand the fault handling processes that are in place today, and how these processes may be improved by implementing fault detection based on data analysis (Papers III, IV and V).

1.2 Limitations

The research work presented in this PhD thesis has been conducted from a Swedish perspective. Thus, the Swedish district heating context has been applied to the studies, regarding, e.g., organizational, technical and legislative aspects. Further, it has been assumed that high-resolution (i.e., hourly) district heating customer data is available when developing the fault detection methods. All customer data included in the studies have been collected from real-world district heating utilities.

1.3 Outline of thesis

Chapter 2 presents basic concepts and fundamentals regarding district heating. Special attention is being paid towards describing concepts that are important for the work conducted in this thesis. Chapter 3 focuses on the "main character" of the work conducted in this thesis - the customer installation and its components. Chapter 4 then focuses on concepts related to the analysis of data and fault detection specifically. In Chapter 5, the methodologies used in the research work and discussions regarding the choice of methods are presented. Chapter 6 summarizes the most important results of the thesis and an analysis of the results. The final chapter, Chapter 7, presents a concluding discussion of the research work and an outlook on future challenges and suggestions for continued research in this research area. The papers written and published during the course of the research work are appended at the end of the thesis.

Chapter 2

District heating – basic concepts & fundamentals

The main idea of district heating is to produce heat in one location to meet the heat demand in another location. The heat transfer is done by connecting the local heat sources to a pressurized pipe system that carries the heated water to the customer installations. A simplified illustration of a district heating system may be seen in Figure 2.1, where 1 represents production of heat, and 2 represents distribution of heat. 3 represents the customer installations connected to the district heating system, but to simplify the illustration only one installation has been drawn in the figure. 4 represents the substation located in the customer installation, and 5 represents the internal heating system of the installation, which includes both space heating and domestic hot water. The customer installation, substation and internal heating system will be described in further detail in Chapter 3. This chapter will focus on some basic concepts related to the district heating system. This includes what heat sources a district heating system may be supplied by, and the current and future potential of the district heating systems. The chapter further discusses how heat is distributed in the district heating system, and how the need for heat varies in the system. Furthermore, an explanation of the benefits of decreased temperature levels is included.

2.1 Heat supply in district heating

District heating makes it possible to utilize local heat resources that would otherwise go to waste to meet the heat demand of local heat customers. District heating systems are local or regional markets, as opposed to, e.g., the electricity market. Thus, the local conditions, such as geographical location and the available heat sources, have a significant influence on the system operation, its related economic situation, and related environmental aspects [1].

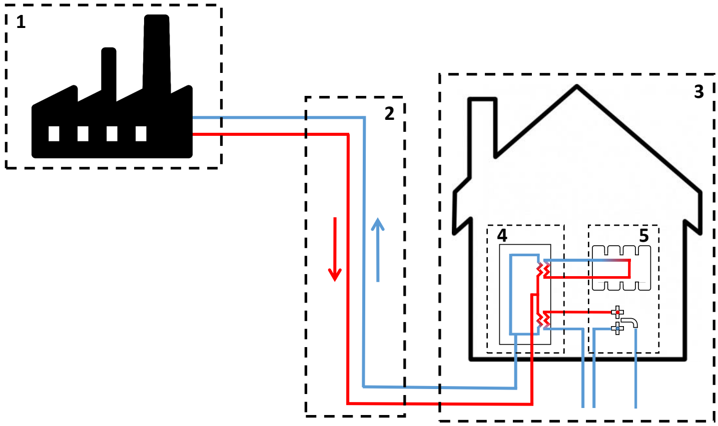


Figure 2.1: A simplified illustration of a district heating system, where 1 represents production, 2 represents distribution, 3 represents the customer installations, 4 represents the customer substation, and 5 represents the internal heating system.

Modern district heating systems provide considerable flexibility in what heat sources can be used to deliver heat. As opposed to the earlier district heating systems, which were supplied by central boilers employed to burn fossil fuels [1], there are typically a variety of different, decentralized production units in a district heating system today [1]. A district heating utility is responsible for monitoring the district heating system and ensuring that enough heat is supplied to meet the heat demand of the customers in the system. This way of operating the system makes it possible to operate the district heating system using the principle that the production units with the lowest operational costs should be used for baseload. The production units with the highest operational costs should instead be used for peak loads in the system [20]. This operation strategy makes it possible to achieve as low production costs as possible [1] and to include several different heat sources, depending on what local sources are available and what sources are beneficial to use from a market perspective. These aspects make it possible to utilize available heat resources in a very efficient way [1].

The heat source flexibility is visible in the international district heating supply statistics. A large share of district heating is still supplied by fossil fuels internationally (90 % globally and 70 % in the EU), primarily being utilized in CHP and boiler plants [6]. However, several European countries have shown that it is possible to utilize renewable and recycled heat sources in the district heating systems. In 2018, biofuels accounted for 62 % of the total energy input to the Swedish district heating systems [7]. During the same year, fossil fuels accounted for approximately 8 % of the energy input [21]. Other European countries, such as Norway, Iceland and Denmark, also have large shares of renewables in their district heating systems [22].

The share of renewable and recycled heat in the district heating systems varies between different countries due to, among other things, the availability of natural resources, market drivers, and national policies [23]. However, both current and future district heating systems are expected to be fuelled entirely by renewable and recycled heat in the future. These heat sources include geothermal and solar heat, heat from CHP and waste incineration, and waste heat from processes in industry and commercial buildings [3]. The decentralization of heat suppliers also makes it possible to incorporate so-called prosumers in the district heating systems. A prosumer may be defined as "a person who both consumes and produces a particular commodity" [24] – the commodity, in this case, being district heating. The most common ways that a prosumer in a district heating system produces heat are heat production in solar collectors and excess heat from facilities with great cooling demands. In current district heating systems with conventional district heating temperatures, it is common that the temperature of the heat that is delivered by prosumers has to be upgraded to match the temperatures in the district heating system [25].

2.2 Current and future potential of district heating

District heating has historically been introduced in the energy systems as a way to achieve fuel savings, replace individual boilers in buildings, increase efficiency in CHP plants, ensure the security of supply of heat, and replace oil with other fuels such as coal, biomass and waste [3, 23]. Today, and particularly in the future, district heating is identified as an integral part of a sustainable energy system. The European Union has identified district heating as an essential solution in future energy systems to help in decarbonizing Europe [5, 26]. The primary motivations for including district heating in such energy systems are its inherent fuel flexibility, its capability of utilizing residual heat from different processes, and its energy efficiency characteristics [3].

Despite the many benefits of district heating, the technology still accounts for a relatively small share of the total heat deliveries globally, and it has had varying success in penetrating the market. In countries with large market shares, such as Sweden, Denmark and Finland, district heating corresponds to approximately 50 % of the residential heat market [6]. In countries with small market shares, there are very few district heating systems due to a lack of awareness and competitiveness of the technology and missing incentives for introducing district heating in national legislations and policies [23]. Russia, China, and the European Union accounted for approximately 85 % of the district heating delivered in the world in 2014 [6], which shows that there is still a great potential to further increase the market shares of district heating in the future.

However, the current district heating systems have to develop to reach their full potential in the future sustainable energy system [27, 28]. Historically, district heating has developed through three generations that have entailed lower distribution temperatures, more material lean components, and the utilization of prefabricated components to reduce workforce requirements at construction sites [3]. All these factors have made district heating an increasingly competitive alternative on the heat market. To further strengthen this position today and in the future, there is a need to improve and further develop the district heating systems. The main point of improvement in the current district heating systems is to decrease the distribution temperatures.

Decreased distribution temperatures will entail several positive consequences, such as improved system efficiency, the possibility to introduce more renewable and recycled heat sources of low temperatures that it would otherwise not be possible to make use of, and reduced heat losses from the distribution systems [3, 23]. These benefits may also be obtained in new district heating systems if they are built and operated according to the specifications of the fourth generation of district heating (4GDH). The concept of 4GDH is defined and described in a paper written by Lund et al. [3]. The main idea behind 4GDH is to supply district heating at lower temperatures than the current district heating systems. The suggested supply and return temperature levels of 4GDH are 45–55 °C and 20 °C, respectively [3].

Thus, there are different national challenges related to the development of district heating, depending on how mature the technology is in the country. In countries where district heating corresponds to a large share of the heat market, the second and third generations of district heating (2GDH and 3GDH, respectively) are the most common technologies used [6]. These generations have many technological differences; although both generations distribute heat using hot water, the supply temperatures of the second generation are often above 100 °C, the water is distributed in pipes contained inside concrete ducts, and large shell-and-tube heat exchangers are utilized to deliver heat to the customers. The third generation typically has supply temperatures below 100 °C, the water is distributed in prefabricated, pre-insulated steel pipes that are directly buried in the ground, and compact substations using plate heat exchangers are installed in the buildings to deliver heat to the customers [3]. The main challenges in systems from these two generations regarding their role in the future energy systems are their high distribution temperatures, high distribution losses, inefficient equipment in buildings connected to the system, and the high capital investments required to remediate these issues [23]. In countries where district heating is not yet widespread, the technology should be implemented according to the specifications of 4GDH to enable the inclusion of renewable and recycled heat immediately. The main characteristics of the 4GDH are the low distribution temperatures that lead to the integration of renewable and recycled heat, lower distribution losses, and the use of pre-insulated, flexible pipes made from, for example, plastic materials [3, 28]. The challenges related to the development of entirely new district heating systems include high capital investments required to develop the infrastructure needed, lack of customer awareness of and confidence in the technology, and lack of legal frameworks and policies that promote the development of district heating [6, 23].

It is not only the technology that should develop if district heating should remain a competitive alternative on the heat market. The competition from other heating alternatives, such as heat pumps, is increasing as these technologies are becoming cheaper and more efficient. This competition, in combination with negative customer perceptions about the natural monopoly status of district heating, may cause customers to leave the district heating systems for other heating alternatives [8, 29]. One way to mitigate this problem is to improve the long-term relationships with the district heating customers while working to minimize the negative perceptions about district heating. One way of doing this is to work in a more service-oriented way towards the customers [8]. Many utilities are currently doing this by offering their customers maintenance agreements to help them make sure that their installations are performing well [30]. These maintenance agreements are typically designed to perform preventive maintenance, where the utilities perform maintenance in the buildings at preplanned, specific points in time. It is also common that the utilities perform corrective maintenance, which is defined as unplanned maintenance tasks performed to restore the functional capabilities of a failed or malfunctioning system [31]. However, the current maintenance schedules may often be expensive in terms of labour costs, costs of spare parts, and outage costs, and the costs will increase if maintenance has to occur frequently [32]. To reduce the costs related to maintenance, it would be beneficial to transition towards condition-based or predictive maintenance. Both strategies rely on data from the process to decide whether it is time to perform maintenance. Condition-based maintenance relies on specific conditions or thresholds. In contrast, predictive maintenance predicts the condition of a system in the future and decides whether maintenance will be needed or not based on that prediction [33, 34]. By introducing such maintenance schemes, it would be possible to work more actively with the district heating customers and their installations, thus creating a closer and improved customer relationship.

2.3 Distribution of district heating

Heat is transferred in the district heating systems through different combinations of mass flow and temperature difference [1]. Thus, the customers may affect the amount of heat delivered to the building in two ways: either by increasing the mass flow rate through the substation or by increasing the temperature difference over the substation. This relationship may be seen in Equation 2.1, which describes the heat transfer rate in a district heating customer installation [1].

$$\dot{Q} = \dot{m} \cdot (h_s - h_r) = \dot{m} \cdot c_p \,(T_s - T_r) = \dot{m} \cdot c_p \cdot \Delta T \tag{2.1}$$

where

$\dot{Q}$:	Heat transfer rate	[J/s]
$\dot{m}$:	Mass flow rate	[kg/s]
h_s :	Enthalpy of the incoming district heating water	[J/kg]
h_r :	Enthalpy of the outgoing district heating water	[J/kg]
c_p :	Specific heat capacity for water, av. temperature & const. pressure	[J/(kg · K)]
T_s :	Supply temperature	[K]
T_r :	Return temperature	[K]
ΔT :	Delta T, temperature difference between T_s and T_r	[K]

The role of the district heating supplier is to always provide the customers with the conditions that they need to be able to extract heat from the district heating system. This task is done by controlling the supply temperature and the differential pressure in the system [1]. The differential pressure is the pressure difference between the supply and return pipes and decides the direction of flow in the district heating system. Due to the pressure drop in the flow direction, the differential pressure will decrease further away from the heat supplier. Thus, the flow will travel in the direction away from the supplier. The pressure drop means that it may be hard to maintain sufficient pressure in the entire district heating system. This problem may be solved by installing decentralized circulation pumps in the system. Using such distributed circulation pumps has also been shown to reduce the overall power consumption of the circulation pumps [1, 35]. The pressure difference between the supply and return pipes is called the differential pressure. There has to be a differential pressure of at least 0.1 MPa at each heat delivery point to ensure that the equipment in the customer installation is working correctly [36]. The differential pressure should, however, not exceed 0.6 MPa, since this may cause noise problems [37].

The temperature difference between the supply temperature and the return temperature is decided by some factors in the district heating system. Both supply and return temperature depends on local conditions in the district heating system and will vary between different systems [1]. In general, systems with high distribution temperatures have narrow pipes or customers with high-temperature demands. Systems with low distribution temperatures are typically supplied by steam-based CHP plants or large heat pumps [1].

The supply temperature is determined by the heat suppliers when the water is heated at the supply units. The required temperature level depends on, among other things, the temperature requirements of the customers and their heat comfort demands. The customers' temperature requirements depend on design choices in the customers' internal heating systems and may vary depending on the age of the system and national standards. In general, older internal heating systems have higher design temperatures. In Sweden, many internal heating systems were designed according to older building standards, in which design temperatures of 90/70 °C or 80/60 °C supply/return were common. More modern internal heating systems are designed for lower internal distribution temperatures of 55-60/40-45 °C supply/return [1, 38]. The return temperature in the district heating system is mainly decided by the aggregated cooling performance of the individual customer installations. Thus, lower return temperatures from the customers will result in lower distribution return temperature [1].

2.4 Heat loads in a district heating system

Each customer installation connected to the district heating system will have a unique heat demand. The heat demand at each individual instance is determined by how large the need for space heating and domestic hot water is. The amount of heat power required to meet this heat demand is called the heat load of the building. The aggregated heat load of all customer installations in the district heating system, plus the distribution losses, is called the heat load of the system [1, 39].

There are three different ways of representing the heat load in a district heating substation (or system) graphically: heat load duration diagrams, consecutive heat load diagrams, and heat power signatures. The heat load duration diagrams display the daily or hourly heat loads sorted in descending order from left to right [1]. The consecutive heat load diagram

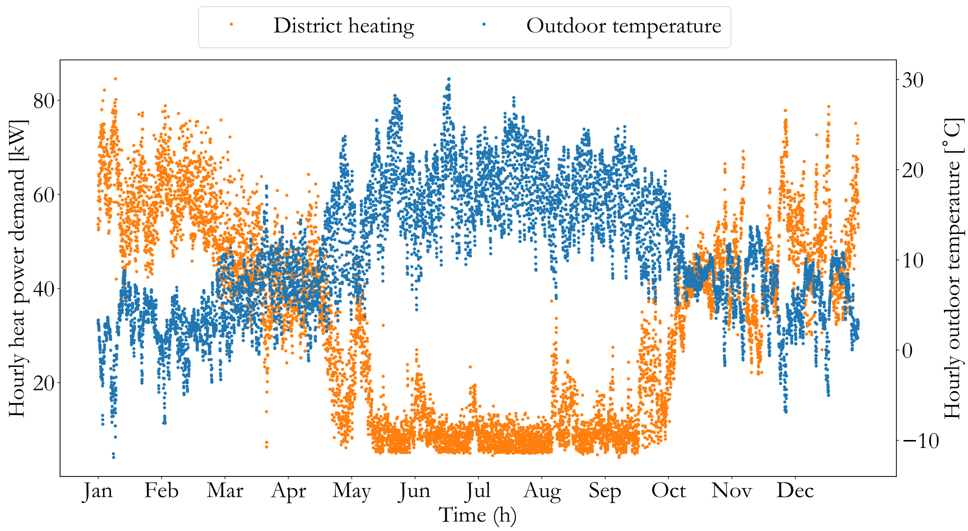


Figure 2.2: Illustration of a consecutive, hourly heat load diagram for a customer installation located in a district heating system in Sweden. The outdoor temperature is also illustrated in the figure.

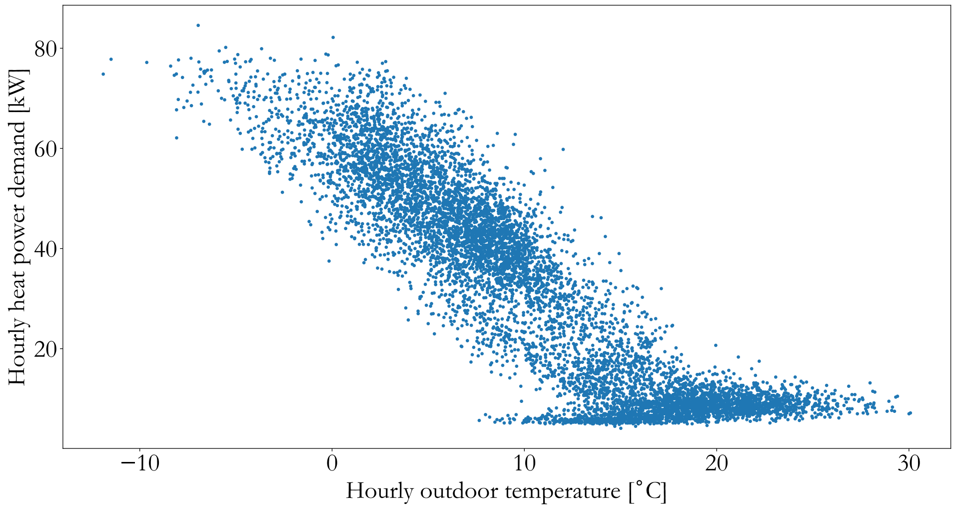


Figure 2.3: Illustration of an hourly heat power signature for a customer installation located in a district heating system in south Sweden.

represents the heat load as a function of time [1], and an example of this type of diagram may be seen in Figure 2.2. The Figure displays the hourly heat load as a function of time during one year for a customer installation in a district heating system in south Sweden (orange). The outdoor temperature is also displayed in the diagram (blue). The heat power signature is shown in Figure 2.3. Instead of representing the hourly heat load as a function of time, this diagram displays the hourly heat load as a function of the outdoor temperature. The data displayed in Figure 2.3 originates from the same customer installation as in Figure 2.2.

As may be seen in Figures 2.2 and 2.3, the heat load of a building varies during the year depending on the outdoor temperature; the space heating demand will be higher during cold winter days than during warm summer days when the main heat load will originate from the DHW use. Thus, the aggregated heat load in the district heating system displays seasonal variations and the outdoor temperature is the main driving force behind these variations [1]. The heat load also varies over days and weeks. Weekly variations usually depend on the fact that some types of customer installations have different heat load patterns during weekdays and weekend – for example, schools and office buildings. Daily heat load variations mainly depend on dynamic heat loads caused by the changes in the weather, and different social behaviours that generate heat demand peaks during mornings and evenings. Such behaviours include, for example, cooking and showering [1].

2.5 Benefits of decreased temperature levels

A reduction of the distribution temperatures is beneficial for many different aspects. The main benefit of decreased return temperatures is the fact that it will be possible to decrease the supply temperature while still being able to deliver the required amount of heat to the customers [1]. A reduced supply temperature would make it possible to introduce more low-temperature heat into the district heating systems. Such heat sources include renewable heat from solar thermal and geothermal resources, excess heat from industries, commercial buildings and sewage treatment plants, and heat from prosumers [3, 12, 20]. Thus, by reducing the supply temperature, the district heating systems would be able to use this low-temperature heat to a more extensive extent. The district heating utilities would then be able to deliver more environmentally friendly heat to their customers, and at the same time, make district heating more cost-competitive [3, 40]. Another aspect related to reduced supply temperatures is that the power-to-heat ratio in CHP plants increases at the same heat demand when the supply temperature is lower. An increased power-to-heat ratio means that more electricity is produced in the CHP, causing an increase of the revenues from the plant [41]. The coefficient of performance (COP) of large-scale heat pumps would also increase [42].

There are also benefits directly related to reduced return temperatures. A reduced return temperature is beneficial when flue gas condensation is installed in the district heating system. Decreased return temperatures generate higher energy efficiency in the condenser, yielding a higher energy output from the process [43]. By reducing the return temperature, it is also possible to decrease the flow rate in the system if the supply temperature is kept at a constant level. A decreased flow rate reduces the need for pumping in the system [1, 44].

Decreased distribution temperatures also yield several benefits that are important from a district heating system perspective. Lower temperature levels make it possible to use new pipe materials, such as plastic, since they will be less exposed to thermal stress when the temperatures are lower. These pipe types have lower related installations costs than conventional pipe types such as steel pipes. They may also be cheaper than the conventional pipe types [11, 45]. When distributing heat in a pipe system, heat losses occur due to the temperature difference between the water flowing in the pipes and the surroundings. This relationship may be seen in Equation 2.2, which describes the heat loss Q_{hl} from a single insulated pipe buried in the ground [46]. The equation shows that the heat loss from a pipe is proportional to the temperature difference between the water in the pipe and the temperature at the ground surface. It will thus be beneficial to reduce both the supply and the return temperature. However, the largest benefit in terms of heat losses will be obtained when decreasing the supply temperature since the largest heat loss will occur in the supply temperature pipe [47].

$$Q_{hl} = 2\pi k_g \left(T - T_0\right) \cdot h_1 \left(H/r_o, \beta\right) \tag{2.2}$$

where

Q_{hl} :	Heat loss	[W/m]
k_g :	Thermal conductivity of the ground	[W/(m·°C)]
T :	Average temperature in the pipe	[°C]
T_0 :	Temperature at the ground surface	[°C]
h_1 :	$\mathrm{h}_1 \left(H/r_o, \beta\right) = \ln\left(\frac{2H}{r_o}\right) + \beta$ = Heat loss factor	[-]
H :	Distance between center of pipe and ground surface	[m]
r_o :	Jacket pipe radius	[m]
β :	$\frac{k_g}{k_i} \ln\left(\frac{r_o}{r_i}\right)$	[-]
k_i :	Thermal conductivity of the pipe insulation	[W/(m·°C)]
r_i :	Carrier pipe radius	[m]

Except for decreasing the distribution temperatures, there are also other solutions to reducing the heat losses from the district heating systems. Such solutions include increasing the insulation of the pipes and using alternative pipe designs such as twin pipes instead of single pipes [47]. However, these solutions require an update of the infrastructure if implemented in an existing district heating system and may thus be more expensive than reducing the return temperatures.

Chapter 3

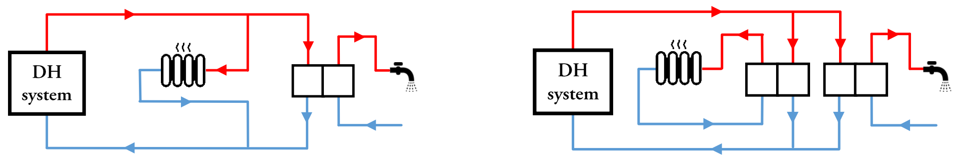
The customer installation

Each building connected to a district heating system contains two main components: the customer substation and the building's internal heating system(s). A district heating customer substation is the heat transferring interface between the district heating system and the customer's internal heating system. The substation consists of several different components, including heat exchangers, temperature sensors, controller, and control valves. The substation usually also includes a heat meter that measures the amount of heat delivered in the building.

The internal heating systems(s) may be described as the pipe system inside the house that delivers heat to all rooms and hot water to all taps in the building. There are two main types of internal heating systems: the space heating system and the domestic hot water (DHW) system. The space heating system delivers heat to each room in the building, while the DHW system delivers warm water to the taps. The internal heating systems include several different components, such as valves, temperature sensors, radiators, and taps. These components, including the substation and its components, are in this thesis referred to as the customer installation. The customer installations may vary significantly in size, and a multitude of different activities may occur in them. This variation means that each installation in a district heating system is somehow unique. These aspects also imply that all customer installations in a district heating system will have different heat demands [1]. In the following chapter, the customer installation and its components will be described in detail. This includes the customer substation, the space heating system, the DHW system, and the district heating meter.

3.1 Customer substations

The customer substation may be designed in many different ways, but the main difference in design depends on whether the substation includes heat exchanger(s) or not. These



(a) A simplified illustration of a substation with direct connection in the space heating system, and indirect connection in the domestic hot water system.

(b) A simplified illustration of a substation with indirect connection in both space heating and domestic hot water systems.

Figure 3.1: A simplified illustration of the two most common ways to create hydraulic separation in the district heating substation.

designs are referred to as direct or indirect connection. Figure 3.1 displays a simplified illustration of the most common solutions of direct and indirect connection principles. Figure 3.1a shows a customer substation with a directly connected space heating system and an indirectly connected DHW system. Figure 3.1b shows a customer substation with indirect connection in both space heating and DHW systems.

A direct connection does not include a heat exchanger, which means that there is no hydraulic separation between the district heating system and the internal heating system of the customer. Thus, the water circulating in the internal heating system is water from the district heating system. The exact opposite of this is the indirect connection, which includes a heat exchanger. The heat exchanger causes a hydraulic separation between the internal heating system and the district heating system [1]. Most DHW systems are supplied by indirect connections, while the connection design for the space heating systems to a large extent depends on national standards. While the indirect connection is the most common solution in, for example, Sweden and Finland, the direct connection is the most prevailing design in, for example, Denmark and Germany [1, 20].

3.1.1 Customer substation ownership

There are different business models regarding the ownership of the customer substation: the customer may be the owner, the district heating utility may be the owner, or there may be shared ownership. There are also other business models where the ownership of the substation depends on the type or size (with regards to heat power demand) of the building connected to the district heating system. In Sweden, the substation is most commonly owned by the customer [1]. This ownership structure means that it, on paper, is the customer's responsibility to make sure that their installation is performing as it should. However, the utility is most often the party that benefits the most from well-performing customer installations. According to Brange et al., the performance of the installation is often better when the utility is responsible for the maintenance of the installation [48]. However, not all customers sign up for the maintenance agreement offered by the utilities. Therefore, it is essential to create other incentives for the customers to work on the performance of their installations. One such incentive is to introduce a flow component in the price model, which means that the customer pays for the amount of water that passes through the substation. According to Equation 2.1, to transfer the same amount of heat to the building, the mass flow rate in the substation will increase if the return temperature increases. Thus, if a price component proportional to the mass flow rate is introduced in the price model, an incentive for the customer to reduce the return temperature from their installation is created [49].

3.1.2 Customer substation connection principles

The connection principle describes how the components in the customer installation are connected. Examples of connections principles are the parallel-connection, the two-stage connection, and different types of series connections. Today, the parallel-connected substation is the most common connection principle in Sweden and will be used as the standard example in this thesis. The interested reader is referred to [50] for an overview of these connection principles.

Figure 3.2 displays a detailed connection scheme for an indirectly parallel-connected district heating customer installation. The detailed corresponding list of numbered components may be found in Table 3.1. The components in Figure 3.2 are in this thesis divided into four different categories: primary system district heating, district heating metering, space heating system, and domestic hot water components. The components are divided into categories depending on where they serve a function in the customer installation. For example, all components contributing to the heating of the building are located in the space heating category.

3.2 Space heating systems

The purpose of the space heating system is to create thermal indoor comfort during the winter months. The system consists of four main components: heaters in the individual rooms, a system that distributes the heat in the building, a heat source (district heating in this case), and a control system [51]. In the case of district heating, the most frequently used heaters are hot-water radiators. However, in more modern space heating systems, it is increasingly common to install underfloor heating instead. Underfloor heating requires lower design temperatures than most radiators, making it an attractive and desirable option

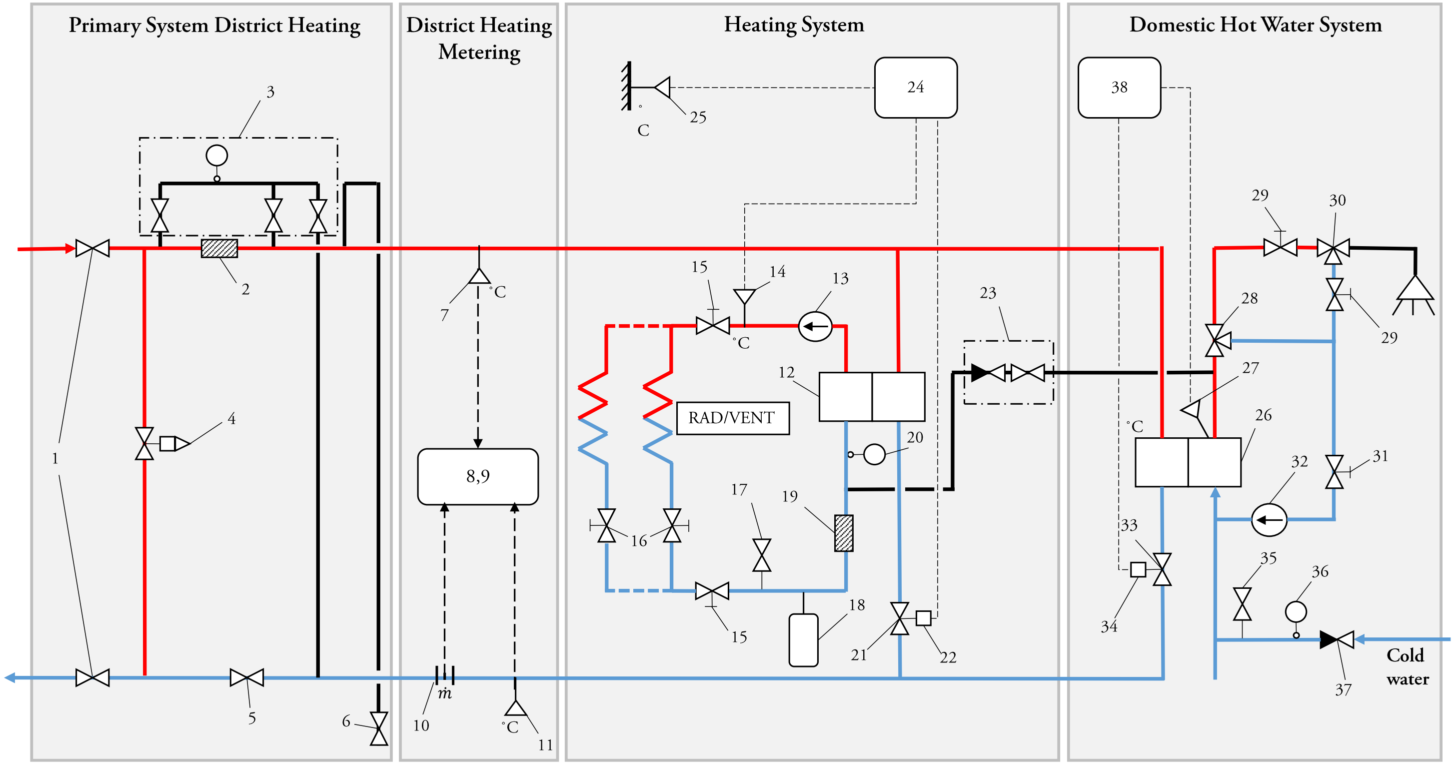


Figure 3.2: Detailed connection scheme of a parallel-connected district heating customer installation.

Table 3.1: Table of components in Figure 3.2.

Primary system district heating		**Space heating system**		**Domestic hot water system**	
1	Service connection isolating valves	12	Heat exchanger	26	Heat exchanger
2	Filter	13	Circulation pump	27	Temperature sensor, supply
3	Manometer	14	Temperature sensor, supply	28	Mixing valve
4	Bypass valve	15	Stop valve	29	Stop valve
5	Drain valve	16	Balancing valve	30	Tap water mixer with mixing valve
6	Vent valve	17	Safety valve	31	Balancing valve
District heating metering		18	Expansion vessel	32	Circulation pump
7	Temperature sensor, supply	19	Filter	33	Control Valve
8	Integrator	20	Manometer	34	Actuator
9	Communication unit	21	Control valve	35	Safety valve
10	Flow sensor	22	Actuator	36	Manometer
11	Temperature sensor, return	23	Filling valve with check valve	37	Check valve
		24	Controller/control	38	Controller/control
		25	Temperature sensor, outdoor		

for the future, low-temperature 4GDH systems. In some customer installations, usually non-residential buildings, ventilation systems are used for space heating [45].

The amount of heat being delivered to the space heating system depends on the outdoor temperature. Lower outdoor temperature means that more heat is needed to maintain a comfortable indoor climate. The heat is adjusted by increasing or decreasing the supply temperature in the space heating system. This temperature is called the secondary supply temperature. This adjustment is made by controlling the valve that limits the flow through the heat exchanger. The relation between the outdoor temperature and the required secondary supply temperature is most often predefined by a linear relation [52]. In larger buildings, a circulation pump is normally installed in the space heating system to maintain circulation within the system [1].

3.3 Domestic hot water systems

The domestic hot water system prepares and delivers hot water to all delivery points (for example, showers and taps) in the building. The DHW demand occurs when someone opens a tap in the building. This generates a stochastic heat demand which is very hard to predict. However, the DHW system must always be ready to meet that heat demand. This is accommodated by keeping the set point value of the secondary DHW supply temperature at a constant level throughout the year. This ensures that the temperature level is always sufficient [45].

There are two main ways to prepare domestic hot water in an indirectly connected substation: using accumulation tanks, or instantaneous heat exchangers. Instantaneous heat exchangers are the prevailing method for DHW preparation in the Swedish DH systems of today [20]. Regardless of what preparation method is utilized, there is a distance in the building between the tap and the heat source. This distance can make it difficult to maintain the secondary supply temperature at a sufficient level in larger installations such as multi-family buildings. In larger buildings, such as multi-family buildings, this may be a problem, since it may entail long waiting times for the heated water to arrive at the tap. Therefore, it is common to install DHW circulation in buildings to reduce the waiting time and enable a fast and efficient withdrawal of DHW [53].

3.4 District heating metering

District heating metering takes place in the customer installation and is utilized for billing the district heating customers according to their actual consumption of heat [54]. In the European Union, billing should be done at least every month. Thus, the district heating utility has to collect meter readings from their customers at least once a month. In Sweden, it is more common that the utilities collect meter values daily, providing the utilities with a large amount of district heating customer data [14, 16]. There is also a requirement from the Swedish Energy Markets Inspectorate (Energimarknadsinspektionen) for the district heating utilities to, according to regulation EIFS 2014:2, provide their customers with information about their historical district heating consumption at least twice a year [55].

The heat meter is located on the district heating side of the substation. It consists of one flow sensor, two temperature sensors and one integrator that calculates the amount of energy delivered to the installation during a specific time. The flow sensor is typically placed in the return pipe, and there is one temperature sensor in the supply and return pipes, respectively [1].

Chapter 4

Fault detection and diagnosis using data analysis

Fault detection based on customer data analysis is one of the key components of the work conducted in this thesis. The interest in such solutions has increased as the information and communication technologies (ICT) have developed and improved in the district heating systems. By installing more connected sensors and smart meters and implementing different software tools, it is possible to obtain a smart district heating system where automated fault detection is a natural part of the work conducted by the utility. The following chapter will introduce the fundamentals of fault detection based on data analysis and some previous work that has been conducted within the district heating industry related to such tasks.

4.1 Definition of fault

When speaking of fault detection, it is essential to know what is considered to be a fault. According to Isermann, a *fault* may be defined as "*An unpermitted deviation of at least one characteristic property of the system*" [56]. A fault can occur in all types of systems, and due to many different reasons, for example, human-induced faults, mechanical faults, and technical issues [18]. A fault is an indication that the system is not performing as it should. If not corrected, a fault may cause both malfunctions (intermittent irregularities in fulfilment of a system's desired function) and failures (permanent interruptions of a system's ability to perform a required function under specified operating conditions) [56].

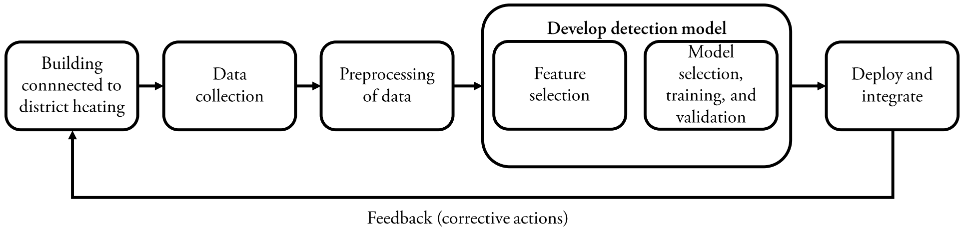


Figure 4.1: Implementation of a conventional FDD method, adapted for the case of buildings in district heating from [57], [61].

4.2 Fault Detection and Fault Diagnosis

Fault detection and diagnosis (FDD) may be described as a collection of methods developed to monitor a system's behaviour, to determine if a fault is present in the system, and to determine the characteristics and root cause of the detected fault(s) [57]. The idea behind FDD is that a fault should be detected and corrected at an early stage so that a system operator can perform counteractions such as maintenance or repair [58]. An illustration of an implementation of a general fault detection and diagnosis method may be seen in Figure 4.1. A number of sensors or metering points that collect data from the system of interest are needed to perform FDD. This data is then preprocessed, a procedure that prepares and transforms the collected data set into a suitable format for the data analysis step. Data processing may include several tasks, such as removal of duplicate values, feature extraction, and removal of outliers [59]. To develop a suitable fault detection model, one must identify what system parameters are essential to the system's behaviour and what system parameters will change if a fault is present in a system. These are the features that are of interest to analyse when performing FDD. The analysis of data, that is, the fault detection and diagnosis, may be carried out in many different ways. FDD methods may be broadly divided into two main categories: (i) Model-based and (ii) Process History-based [60].

4.2.1 Model-based fault detection

Model-based FDD methods are based on mathematical models based on a priori knowledge of the system of interest to model the system's output. The output of the model is compared to the actual measurements obtained from the system. The comparison is based on the difference between the two, that is, the residuals between the model output and the actual measurements. The analysis of the residuals indicates the system's performance; for example, very large or constantly fluctuating residuals may be an indication that the system is not performing as it should [62, 63]. These methods are thus based on the dependencies between different measurable signals, which will generate an altered output when a fault is present in a system [64].

The model-based fault detection and diagnosis methods may further be described as qualitative or quantitative [62]. Qualitative model-based methods are based on relationships derived from qualitative knowledge of the system at hand. These methods include, for example, expert knowledge of the system that gives information about whether the system is in a faulty state or not and limit checking methods that utilize threshold values that a measured variable should not exceed if being in a normal state. A disadvantage of the qualitative model-based methods is that they are not based on the system's underlying physics. This means that they generally are not capable of representing new system conditions. Quantitative model-based methods are based on mathematical models of the system's underlying physics, that is, the mathematical relationships between the inputs and outputs of the system. Thus, quantitative model-based methods consist of mathematical representations where the parameters in the model bear physical meanings. One disadvantage of this type of method is that the system complexity and lack of good data make it hard to develop an accurate model of the system [58, 60, 62].

4.2.2 Process history-based fault detection

The second broad category of fault detection methods, process history-based methods, includes methods that do not utilize a priori knowledge of the system to detect faults in the system. These methods utilize available historical data to derive models of the system, that is, mathematical relations between measured input and measured output signals. Process history-based methods include methods such as black-box and grey-box models. In black-box models, the model features and parameters have no physical meaning. In grey-box models, the parameter estimates can be traced to actual physical principles of the system being modelled [62]. One important class of methods that may be attributed to the process history-based methods are machine learning methods [65].

4.2.3 Machine learning

Machine learning (ML) methods may be applied to many different tasks, including email spam filtering, image recognition, and fraud detection on social media [66]. Several studies have also shown that ML may be successfully applied to fault detection and fault diagnosis problems [65, 67]. ML may be described as models that learn the behaviour of a system from data collected from the same system. The ML methods are thus capable of learning the behaviour of a system without being explicitly programmed [68]. This way of learning is especially advantageous when modelling real-world systems, which may be very complex and hard to model using analytical models based on the physics of the system [66].

Machine learning methods may be divided into two main categories, based on how the algorithm learns the system's behaviour: supervised learning and unsupervised learning. Both supervised and unsupervised methods have been shown to have applications in fault detection and diagnosis [65].

In supervised learning, labelled data sets are utilized to learn the behaviour of the system. Labelled data sets contain information about the desired target class. The target class is what the algorithm should learn to predict. Thus, the algorithm is trained in a supervised way by comparing its own output with the target outputs, learning the mapping between the inputs and outputs of the modelled system. The two main tasks carried out by supervised learning are classification and regression. Suppose the task is to predict a discrete value, such as if a fruit is an apple or a pear – then, we use classification methods. If the task is instead to predict a continuous variable, such as the price of the fruits at the grocery store, we use regression methods [69, 70].

In unsupervised learning, no labelled data sets are available. The lack of these data sets means that the algorithm has to discover hidden structures in the data on its own. Thus, all variables given are used as input to the algorithm. Unsupervised learning may be used for several different tasks, the most prominent one being clustering. Clustering is methods that use similarities in data to create groups, or clusters, of data points that are similar to each other [65, 69].

Selecting a data science assistant - TPOT

When performing data analysis using different ML applications, several different steps need to be carried out to obtain a well-performing algorithm, for example, feature engineering, model selection, and algorithm selection. This may be a very time consuming and challenging task before obtaining a well-performing ML application. These tasks are typically carried out today by allowing a data scientist to work with the data analysis until a satisfying result is obtained. However, this somehow contradicts the cornerstone of ML, which is to allow the computer to learn on its own without being specifically programmed. In recent years, researchers have therefore become interested in automating all steps of developing an ML application. This would also allow for less experienced data scientists to utilize the benefits of ML [71].

When speaking of ML applications, it is common to talk about ML pipelines. An ML pipeline may be described as the sequence of tasks carried out in ML, that is, collecting data, cleaning data, feature preprocessing, model selection, parameter optimization, and model validation. Developing a well-performing pipeline may be tedious and time-consuming work, which makes it an excellent candidate for automation. Thus, several research projects have focused on the development of such pipeline optimizers [71].

One such automated ML tool used in this thesis is the Tree-based Pipeline Optimization Tool, TPOT. The developers of TPOT describe the tool as a data science assistant that may be used to obtain a well-performing ML pipeline for a given problem by carrying out feature preprocessing, model selection, and hyperparameter optimization. TPOT may be used to solve supervised problems and is thus suitable for classification and regression problems. The tool is based on the scikit-learn library in Python, from which several different operators are collected to implement the pipeline. Since many different combinations of operators are possible, the task of TPOT is to try out and optimize different combinations to maximize the classification or regression accuracy. This optimization is done using genetic tree-based programming, which means that TPOT builds its pipeline by testing different combinations and sequences of operators, and the parameters of the operators, using tree-like structures to reach the best combination of operators, according to certain criteria [72]. An advantage of TPOT compared to other automated ML tools is that it provides a more flexible ML pipeline. This flexibility may, on some occasions, cause the model to overfit the data, which the user of TPOT has to be aware of [71].

4.3 Fault detection in district heating customer data

During recent years, the problem of fault detection customer installations has gained increasing interest. At the same time, more customer data have become available, which has made it possible to use more advanced fault detection and diagnosis algorithms. The possibility to use customer data analysis to detect faults has significantly improved since the European Energy Efficiency Directive became effective in 2012. The directive states that all energy customers should be billed according to their actual energy consumption [54], as opposed to earlier billing methods where the district heating utilities instead used estimations about their customers' consumptions to base the bills on [1]. This demand on energy metering includes consumption of district heating, and thus actual measurements of the district heating customers' consumption are now available. The Swedish district heating utilities are, according to regulation EIFS 2014:2, obliged to provide their customers with information about their historic consumption on a daily basis [55]. Furthermore, since 2015, all Swedish district heating customers are billed according to their actual use of district heating, as stipulated in the Swedish district heating law [73]. Although the demand for metering is usually of a lower resolution, most heat meters can measure the consumption with an hourly resolution [15]. Thus, there is a large amount of customer data available from Swedish district heating systems that may be utilized for fault detection purposes.

Earlier work on FDD in district heating customer installations primarily investigates how well the installation is performing by using the overflow method [14]. The overflow method assumes that there is an ideal value of the temperature difference over the substation, or cooling, that the substation should obtain. The overflow may then be described as the extra amount of district heating water that has to pass through the substation when the cooling of the substation is small, compared to the water volume that would be needed if the cooling had the ideal value during a specific time (day, week, month, year). The overflow may be calculated using Equation 4.1. The overflow method has mainly been utilized for analysis of monthly or annual values of the parameters collected from the customers' heat meters since it was implemented when meter values were still collected manually by the district heating utilities [1].

$$overflow = V_a - V_i = V_a - \frac{E_a}{\rho c_p \Delta T_i} \tag{4.1}$$

Where

V_a :	Actual water volume through installation during a specific time	[m^3]
V_i :	Ideal water volume through installation during a specific time	[m^3]
E_a :	Actual heat used in installation during a specific time	[J]
ρ :	Density of water	[kg/m^3]
c_p :	Specific heat capacity for water, av. temperature & const. pressure	[J/(kg $\cdot$ K)]
ΔT_i :	Ideal Δ T of the installation	[K]

In [14], Gadd and Werner utilize temperature difference signatures to detect faults. The temperature difference signature is a diagram that displays the daily average temperature differences as a function of outdoor temperature. The authors implemented the method manually on a smaller data set consisting of data from 140 different substations, out of which 14 substations were known to have faults during the investigated time. Ten of these substations were identified using the temperature difference signature, proving that the developed method is viable for detecting faults in district heating data. In a subsequent study, Gadd and Werner investigated what types of fault may be identified in customer substations. They also investigated how many of the analysed datasets may be considered to contain a fault or be showing symptoms of faults [15]. 74 % of the investigated data sets were identified as containing faults or showing symptoms of faults.

Other studies have shown that many different types of data analysis are applicable for fault detection in district heating customer data. Both model-based and process history-based methods have been utilized. Some developed methods focus on comparing the substation's performance to itself, whereas others compare the performance to a cluster of similar substations. Sandin et al. [16] utilize limit checking and outlier detection to detect faults in district heating customer data. Abghari et al. have developed a higher-order data mining approach to model a substation's behaviour on a weekly basis [74]. The method is based on sequential pattern mining and clustering of weekly heat demand patterns from individual substations, after which the behaviour of each cluster is modelled. The substations' behaviour profiles are then associated with different performance indicators. In [17], Calikus et al. utilize "the degree of abnormality" of heat power signatures to rank individual buildings. The heat power signatures are estimated using robust regression, and the ranking of the buildings is performed using three different approaches: outlier-based, dispersion-based and aggregated ranking. Calikus et al. further explore heat load patterns in [75], where the authors cluster heat load profiles to create 15 clusters of heat load patterns from which representative heat load patterns are extracted. The authors then identify heat load patterns that deviate significantly from the identified clusters' heat load patterns. In [19], Farouq et al. presents a reference-group based approach for detecting customer installations that display a deviating behaviour. The reference groups were based on a k-nearest neighbour approach, utilizing the return temperatures from the investigated installations to identify the reference groups for each installation in the data set. If an installation deviates significantly from its reference group, it is considered faulty. Xue et al. utilizes association analysis to identify correlations hidden in clusters of different operating patterns in two customer installations located in China [76]. These correlations are then represented as association rules, which may be used for fault detection. A large change in such a correlation indicates that a fault may be present in the installations.

All of the studies mentioned above focused on fault detection. Other studies have focused on individual faults and detecting faults in individual components or system changes. In [77], Abghari et al. propose a decision support system based on k-means clustering and support vector regression (SVR) to identify manual changes in the heating program used in the substation. In [78], Pałasz and Przysowa use an ensemble classifier to predict the failure of heat meters using data from real-world heat meters in a district heating system. Brés et al. produce synthetic data sets containing different faults using coupled building and system simulations [79]. These data sets are then utilized to develop a fault detection and diagnosis method based on binary decision trees. One significant result from the study was that it showed the value of having access to secondary side data when performing fault diagnosis. Skaarup Østergaard et al. use heat cost allocators in a multifamily building to calculate the return temperature from eight radiators in the building [80]. These results were then utilized as an indicator of faults in the individual radiators.

Chapter 5

Methodology

This thesis is about fault handling in district heating customer installations. The problem has been approached from a holistic perspective, using a multidisciplinary approach to investigate how the fault handling processes may be improved in the district heating industry. This perspective includes organizational aspects such as how utilities may work with fault handling and the development of data analysis tools that may be used to detect faults using district heating customer data. Due to the multidisciplinary nature of the project, several different methods have been used. The methods include programming to develop fault detection algorithms, interviews, literature studies, surveys, and workshops. By approaching the problem from different angles, it was possible to obtain a broader picture of challenges and solutions within the topic and reach the aims of the thesis work.

In this chapter, the methodologies used in the project will be presented. An overview of the methods used in each scientific paper that was written and published during the PhD thesis work will be presented along with relevant theoretical background related to the individual methodologies. A discussion regarding why the methodologies were chosen is also included in the chapter. The chapter is written in chronological order; that is, the methodology is presented in the order the studies were conducted. This structure allows the reader to follow the methodological choices that have been made throughout the project more easily.

5.1 Data sets being used in the thesis

When starting the project, the initial idea was to develop algorithms that could detect and diagnose faults in customer installation using real-world customer data. The reason for working with real-world data from active customer installations was that the aim was to develop methods that could be implemented rapidly after development. In theory, fault detection in district heating customer installations is ”only” an application area for data analysis. However, all such areas of applications come with their challenges and limitations. In the case of analysing district heating customer data, many of these challenges originate from the lack of information about the customer data. The challenges include not knowing when a specific fault has occurred, making it hard to validate and evaluate developed data analysis methods to make sure that they are identifying buildings that have contained a fault during the investigated time. The problem of not having labelled data often occurs in real-world applications of fault detection and may be solved in different ways. One approach to solving this problem is to generate artificial or synthetic data sets. To generate such data sets is a common strategy when developing fault detection algorithms for cases where labelled data is missing. Another common approach is to generate experimental data in a laboratory to obtain data with different faults and operating conditions. However, many of the fault detection methods developed for this type of data have been hard to transfer to the real-world application. The problems arise due to, among other things, the fact that the synthetic data sets do not capture the large variety that exists in real-world data. These data sets are thus often most useful when wanting to model a specific condition or setting [81]. Therefore, it was decided to use real-world data in this project to capture the complexity and variety in the district heating building population. Using these data sets would also make it possible to implement the developed methods in the district heating industry efficiently.

The data used in this thesis originate from two different district heating systems located in the south of Sweden. Both datasets contain hourly measurements of heat power consumption, mass flow, supply and return temperatures. The datasets also include timestamps of when the measurements occurred, as well as average hourly outdoor temperatures. Further, the datasets include unique numerical identifiers for each substation.

5.2 An initial approach – fault detection using data analysis

The first two papers in this thesis focused on developing algorithms for fault detection for real-world district heating customer data sets. As mentioned in the previous section, there are different ways to mitigate the fact that there is little information available regarding the data sets that are being investigated. One common approach to solving this problem is to model a well-performing entity's behaviour and then allow that model to perform predictions for subsequent data sets. Any faults are then detected as deviations from the normal behaviour in the model predictions [82].

This approach was explored in Paper 1 in this thesis, where a model for a well-performing customer installation was developed using a machine learning approach. The installation was selected from a data set consisting of approximately 1 000 customer installations after calculating the overflows for all installations and then visually inspecting the 10 installations with the best overflows. The most well-performing one, according to energy, mass flow rate, and temperature performance, without significant outliers and following an expected behaviour over the year, was then selected for the study. The method was based on modelling the mass flow rate of one district heating substation. Other modelling parameters were also considered, but after testing different combinations of input and output parameters, it was decided that the mass flow rate was the most suitable output parameter. In Equation 2.1, it is clear that either the mass flow rate or the temperature difference in the installation will change if there is heat demand in an installation. If the installation is not working as it should, it is likely that the temperature difference will be small, which means that the parameter that will vary is the mass flow rate. The mass flow rate thus gives a good indication of the performance of the customer installation. The fault detection method was developed in Python using TPOT, which resulted in a machine learning pipeline capable of carrying out data pre-processing and modelling. The model was trained and evaluated for a data set consisting of hourly measurements from the selected customer installation. Once a well-performing model of the specific installations had been obtained, the fault detection capability of the model was tested for the same installation. The fault detection capability was tested by introducing the model to data sets where artificial faults were induced. The model outputs for the faulty data sets were then compared to the outputs for the well-performing, original data set by calculating the residuals between the model outputs and the actual measurements from the customer installation. By comparing these residuals to each other, it was possible to investigate whether the model performance changed when introduced to a data set containing faults. Investigation of residuals is a common fault detection principle, for example, presented by Venkatasubramanian et al. in their comprehensive review of process fault detection and diagnosis methods [60], and Pakanen et al., who implement this type of method to perform fault detection on the control valve on the primary side of a district heating substation [83]. The comparison conducted in Paper I was made using both the hourly residuals and the cumulative sum of the residuals. The latter gave a possibility to investigate whether the residual pattern reflected a general pattern of deviation from the normal behaviour or only a random variation in the residuals.

Another common approach to fault detection is, as mentioned in Section 4.2.1, to introduce different limit checking methods where thresholds are utilized to analyse whether an investigated process is in a normal state or not. This approach was investigated in Paper II, where the aim was to investigate how simple statistical fault detection methods could be automated to detect faults in district heating customer installations. The limit checking approach has previously been implemented in the case of district heating in earlier research works. Gadd et al. utilize it to investigate the cooling performance of 135 substations [15], while Sandin et al. evaluate the heat power signatures of 996 customer installations to identify faulty behaviour [16]. When implementing limit checking, the thresholds may take both linear and constant values, depending on the investigated parameter. Constant thresholds are created by values that are somehow representative of the behaviour of the process being investigated, for example, the mean of the measured parameter, while linear thresholds may be created using a function that models the behaviour of the process of interest [58]. Both kinds of thresholds were utilized in Paper II. Three different parameters were investigated in the paper: delta T, return temperature, and heat power consumption. All three parameters were investigated as a function of the outdoor temperature. These three parameters are seen as good indicators of the performance of a customer installation. They are commonly investigated and analysed within the district heating industry to perform manual fault detection of district heating customer installations. The threshold values of the limit checking were calculated from a reference case of well-performing customer installations, which were selected based on their low yearly overflows. Linear thresholds were implemented for the delta T and the heat power consumption, and constant thresholds were selected for the return temperature. To create the linear thresholds, piecewise linear regression was utilized to predict the average behaviour of the reference case installations. Piecewise linear regression is utilized to fit multiple linear models to different ranges of the data that is being modelled. The different ranges are separated by breakpoints, which may be selected manually or automatically by an algorithm [16]. This modelling approach was selected due to the way that the district heating parameters vary with the outdoor temperature, as described in Section 3.2. The limit checking method was implemented as an algorithm in R, utilizing the standard packages. The algorithm was tested on a data set consisting of data from 3 000 customer installations in the same district heating system from April 2015 - March 2016. The fault detection algorithm included a data pre-processing step, and daily average values were used to perform the fault detection analysis.

5.3 An attempt at validating the fault detection methods

The work in Papers I and II was conducted using real-world data sets from real customer installations in real district heating systems. The fault detection methods did detect deviating behaviours in the data sets, which indicates that they have the capacity to detect faults in customer installations. After developing fault detection methods, a logical next step is to validate the methods. In this case, validation was carried out to investigate whether the fault detection methods were capable of detecting installations that had one or several faults in the substation or internal heating system. However, the validation of the methods was not as straightforward as anticipated. The reason for this was that none of the district heating utilities from which the data was obtained had a systematic way of recording information about faults that occurred in a customer installation at a specific point in time. This was also the case for other utilities that were contacted. The utilities did record if they had conducted some repair work or exchanged any components. However, this information was not related to the customer data and any deviations that might have occurred in it. Thus, a large amount of data was available, but it was not possible to know whether there were any faults present in these data sets. Neither was it possible to know which, if any, customer installations could be considered to be well-performing, at least not on a larger scale. Thus, no labelled data was available, making it hard to validate the fault detection methods on a larger scale. When reaching this conclusion, it was decided to shift the research focus towards improving the existing situation, so that labelled data from customers can be gathered by the utilities and the current fault handling process can be improved.

5.4 Shifting focus – learning more about current fault handling processes

The problem of obtaining more information about the customer data may be approached in different ways. One crucial aspect was already identified: to relate the customer data sets to other customer information already available, such as service and customer records. This raised the question of why this was not already the case at the utilities. Why did the utilities not know more about their customers and their related data? This indicated that there were several other aspects related to faults and customer data that would be of interest to investigate. When looking at the bigger picture, the work conducted in this thesis aims to contribute to the reduction of the return temperature levels in the current district heating systems and maintain the low return temperatures in both current and future systems by eliminating faults in customer installations. Thus, well-defined fault handling processes have to be in place at the utilities. It also means that it is essential to eliminate as many faults as possible and to eliminate new faults at an early stage. Fault detection and diagnosis based on analysis of customer data plays a significant role in this process. However, considering the current status of the knowledge of the faults in the current district heating systems, it was apparent that more work would be needed to implement data analysis methods in the district heating systems. A holistic approach to the fault handling process was needed that considered both fault detection and diagnosis methods, how to implement these in the fault handling work at the utilities, how to handle customer data and information related to faults in the customer installations, and how to involve both customers and utility employees in the fault handling processes.

The first step in this process was to investigate the current fault handling processes at current district heating utilities. That is, how are utilities currently working with their customers to reach lower return temperatures? The aim of Paper III was therefore to investigate, from a holistic point of view, what kind of measures, and combinations of measures, are implemented by utilities today to reach low return temperature levels. The aim was also to investigate how the utilities are actively working to eliminate faults in the customer installations. Another important aspect to investigate was what faults are most frequently identified when visiting the customer installations. By investigating this, it would be possible to know what kind of faults to prioritize in both fault detection and fault elimination schemes. Thus, the study conducted in Paper III focused on both organizational and technical aspects of the fault handling processes at the utilities that took part in the study.

The organizational aspects of the current fault handling processes were investigated in a series of interviews. Interviews may be conducted in many different ways. The interview format depends on the context in which it will be used and what kind of data the interviewer aims to obtain. One of the main differences in research methodology is whether the interviewer aims to obtain quantitative or qualitative data. Quantitative data may be described as measurable data and facts, which may be turned into statistics for analysis, and may be utilized to answer questions such as ”how many?”. Qualitative data instead describes the respondents' experiences of a matter and includes their interpretations of the matter. The gathering of this kind of data makes it possible to answer ”why” questions [84, 85].

Another main difference is how the interview is structured. There are three main types of interview structures: the structured interview, the semi-structured interview, and the unstructured interview. The structured interview provides the highest level of standardization in the interview format. The structured interview consists of a predetermined set of questions where each participant is asked the same set of questions in the same order. This methodology is typically utilized when intending to gather detailed and organized information on specific topics or issues. The structured interview typically generates quantitative data. An unstructured interview is the exact opposite of a structured interview: the interview is focused on a general topic and agenda, but no specific questions are prepared. Instead, the interview is conducted as a conversation between the interviewer and the interviewee(s). This type of interview is suitable when one is interested in developing new insights on a topic and gathering data on general themes instead of specific questions. The unstructured interview generates qualitative data. The semi-structured interview combines qualities from both the structured and the unstructured interview to explore a set of central topics while still allowing for exploration of general themes by asking follow-up questions. The interview is conducted using an interview guide containing a list of topics and questions that should be treated during the interview. Due to the nature of the semi-structured interview, it is possible to obtain both quantitative and qualitative data during these interviews [86].

In Paper III, a series of qualitative, semi-structured interviews were conducted. This approach was used since answers related to specific questions and themes were sought while still making it possible for both interviewer and interviewees to elaborate more on different topics if needed. Representatives from six Swedish district heating utilities participated. Five of the participating utilities were included in the study due to their low annual average return temperatures, indicating that they might actively work to keep their customers' return temperatures at low levels. By interviewing such utilities, it was possible to identify key aspects of already successful fault handling processes. The last participating utility had higher return temperatures than the rest. This utility was included to investigate whether there were any indications of differences in how utilities with low and high return temperature levels work with their customers and customer installations. Moreover, the size of the utilities varied, providing the possibility to investigate if there were any indications of possible differences in how large and small utilities respectively worked with fault handling. After conducting the six interviews, few new facts emerged and thus, the study was considered to be saturated.

The distribution of the most frequently occurring faults was investigated by conducting a web-survey study where Swedish district heating utilities were asked to report how often different faults appeared in their district heating systems. The questions included in the survey were based on a literature review of faults that may occur in a customer installation. A large variety of district heating literature was investigated and included technical reports, research papers, and district heating books. The survey was sent to 139 Swedish district heating utilities, out of which 56 utilities answered the survey. The surveyed utilities were all members of the Swedish district heating association, Swedenergy, in 2018, which represents approximately 98 % of the district heating supplied in Sweden [87].

A survey study was used since the goal was to obtain quantitative data on the fault distribution. Therefore, there was no need to collect more qualitative data. A survey study may be described as a type of structured interview and may be conducted using paper and pen, over the phone, face-to-face, or by using different forms of computer-aided tools. The latter has become increasingly common as the technology for web-based surveys has developed [88]. In a survey study, the respondents are consistently introduced to the same set of questions in the same order. The questions are often designed as close-ended questions, that is, questions that can only be answered by selecting from a number of predefined alternatives such as ”yes, no, maybe” or ranking scales with a various number of levels (for example, 1 to 5). This type of questions makes it possible to learn general patterns from a large group of people, for example, investigating general trends in how people behave, their experiences, and their opinions on different matters [89, 90].

In a survey study, there are some sources of errors that the survey creator has to be aware of. On an aggregated level, the sources of errors may be referred to as the total survey error. The total survey error affects the bias and variance of the survey statistics and consists of four main types of error: sampling error, coverage error, nonresponse error, and measurement error. Sampling error occurs because the survey respondents typically represent a part, or sample, of the population we are interested in surveying. Thus, the respondents' answers to the survey represent only a part of the populations' answers and not the answers of the entire population. Coverage error occurs when parts of, or individuals from, the population are missing from the sampling frame used to draw a representative sample. The sampling frame may be described as a list of individuals in the population that should be represented in the survey sample. Nonresponse error occurs when parts of the sample do not answer the survey, resulting in parts of the population not being represented in the survey statistics. Nonresponse error may take the form of item nonresponses or unit nonresponses. Item nonresponses occur where individual questions in the survey are not answered by the respondent, while unit nonresponses occur when respondents refuse the survey as a whole. Measurement error occurs when there are inaccuracies in the answers of the respondents, creating a difference between the true value and the values estimated from the survey responses [91].

The sampling and coverage errors of the survey conducted in Paper III may be considered as being small since the survey was sent to almost the entire population of district heating utilities in Sweden. The measurement error was also considered small since most questions in the survey were designed as close-ended questions. There was also room for the respondents to clarify their answers if needed. The nonresponse error may be considered more significant, since approximately 40 % of the utilities did not answer the survey. It might be that the utilities answering the survey were the ones that experienced many faults in their district heating systems. However, the likelihood of the nonresponding utilities not experiencing any faults at all in their systems is small. It was also clear from the survey responses that the utilities that answered the survey represented different frequencies of fault occurrences.

5.5 A taxonomy for labelling deviations in customer data

The work conducted in Paper III gave an overview of several aspects that were of importance to identify key aspects of successful fault handling processes. One such important tool is a standard method to describe or denominate faults. Therefore, Paper IV aimed to develop a deviation cause taxonomy that may be used to label deviating behaviours in the customer data. The purpose of the taxonomy was twofold: (i) to provide a structured and standardized way to label deviations in district heating customer data that are caused by faults in district heating systems, and (ii) to help to create labelled data sets that may be used to develop automated fault detection and fault diagnosis methods. The taxonomy was developed through a sequence of activities, where an initial draft was developed during an internal workshop conducted in the collaboration organization Smart Energi. Smart Energi is a Swedish initiative that creates collaboration opportunities for digital development in the energy industry [92]. The initial draft was then improved and extended by gathering more fault labels from literature reviews, service reports, business systems, and site visit reports from several district heating utilities. The proposed taxonomy was then evaluated by discussing its structure and contents with energy specialists and service technicians at district heating utilities involved in Smart Energi. The last step was to send the taxonomy for referral to 128 of the utilities that were members of Swedenergy in 2020. The referral was done by conducting a survey study regarding the contents and structure of the taxonomy. 46 individuals representing 44 of the member utilities answered the survey. In similarity with the survey study conducted in Paper III, the sample, coverage, and measurement errors may be considered small. Once again, the primary source of error was the nonresponse error, as described for the method in Paper III.

5.6 Developing a workflow for fault handling

After conducting the first four studies, one thing was clear: there was interest amongst the district heating utilities in implementing fault detection and diagnosis using data analysis. However, many utilities experienced that they did not have the appropriate frameworks and experience to do so. Therefore, Paper V aimed to develop a suggestion for a workflow that can be implemented when intending to utilize data analysis for fault detection and get more insights into the advantages (and drawbacks) that may arise during the practical implementation of this workflow. The first step of this process was to outline a preliminary suggestion for a fault handling workflow based on data analysis, using the experiences and information obtained in the previous studies. The suggested workflow was then evaluated and improved by conducting an online workshop. A number of district heating experts from eight Swedish district heating utilities participated. The experts were all somehow involved with fault handling at their respective utility. The group of experts included technical service personnel, people being organizationally responsible for fault handling at their utility, and people working with energy services that aim to help improve the fault handling at the utilities. By including several different utilities, it was possible to obtain a more holistic perspective of different organizational and technical challenges that may occur when implementing a new fault handling process at a utility. By having several different roles represented during the workshop, it was possible to investigate aspects and challenges related to specific parts of the organizations.

The workshop was carried out as a qualitative, semi-structured group interview. A group interview may be described as a systematic questioning of several individuals simultaneously in formal or informal settings [93]. The discussion is guided and monitored by a moderator that has prepared an interview guide with relevant topics. In this case, a qualitative, semi-structured approach was chosen due to the possibility to obtain answers regarding some specific themes while still having the possibility to ask further in-depth questions about the topics. It is also common practice that an observer is present during the group interview. The role of the observer is to observe the group interactions, and they may also help the moderator take notes during the interview [94]. An advantage of the group interview format is that the group setting provides a stimulating environment where the group may stimulate the interviewees to recall certain events or insights [93]. The collaborative aspect of the group interview is essential and is also seen as one of the main advantages of the data collection method [95]. By allowing the interviewees to interact with each other, it is possible to obtain information on collective views of a topic [94, 95]. Thus, the collective view of a topic is seen as more important than the aggregate view of the topic [95]. The group interview may also be challenging. Challenges include that some participants may be dominant in the discussions and that some participants may tend to conform to the responses of other participants, even though they do not actually agree with what has been said [93, 95]. Therefore, the moderator must be aware of these issues to ensure that all participants' opinions are voiced during the discussion [94].

During the workshop, the suggested workflow was discussed. The utilities' previous experiences of fault handling were also discussed to understand what challenges and opportunities may arise when implementing a fault handling process based on automated fault detection in an existing organization. These issues were also investigated by performing a short survey before the workshop. The survey included three questions listed below.

1. What methods are you currently using to detect faults in your customer installations?
2. What roles within your organization are currently involved in your fault handling processes? Which role has the main responsibility?
3. Are you using any digital tools to facilitate your fault handling process? If so, what tools?

Seven of the participating utilities answered the survey. The utility that did not answer the survey was allowed to answer the questions during the workshop. The results from the survey were then presented during the workshop and served as a basis for the discussion regarding the utilities' current fault handling processes.

Chapter 6

Results and analysis

This Chapter presents the most important results from research Papers I- V, as well as an analysis of the results. The results are presented in a ”bottom-up” approach, starting with the results related to the current situation of fault handling. This includes results regarding what faults occur in the customer installations and what faults were considered the most frequently occurring ones, and how utilities are currently working with fault handling. These results include both technical and organizational aspects. The results regarding the current fault handling processes are finalized with a presentation of how the utilities carry out fault detection today. The results related to future fault handling processes include an overview of the results from the thesis work where fault detection methods were developed and a presentation of the fault handling process developed in Paper V.

6.1 Faults in district heating customer installations

A district heating customer installation contains a multitude of different components, as described in detail in Chapter 3. All of these components may somehow malfunction or fail, which means there are many potential faults in a customer installation. When speaking of faults in customer installations, it is important to have a broader definition of the word fault. The most apparent faults are components that break or malfunction. However, due to the complexity of a real-world system such as a district heating customer installation, a fault should be defined according to the definition in Section 4.1. That is, a fault is an indication that a system is not working as it should. The definition of the word fault thus includes not only malfunctioning physical components but also, for example, changes in the system settings that cause the performance of the system to decrease. In the case of a district heating customer installation, a fault may therefore include anything that causes the performance of the installation to decrease. Such faults include a customer changing the settings in the controller, having no DHW circulation (which makes the DHW temperature hard to control), and leaking heat exchangers.

An overview of faults that may occur in a district heating customer installation is presented in Table 6.1. In the table, the faults are divided into five fault categories based on the faulty component: Actuators, Control system and controller (including sensors, wires, etc.), Control valves, Customer's internal heating system, and Heat exchangers.

Table 6.1: Faults that may occur in a district heating customer installation. Table compiled from results in Paper III [96].

Fault category	Fault description
Actuator	Actuator cannot change position of valve
	Actuator does not close completely
	Broken actuator
	Incorrect actuator stroke time
	Poor connection between actuator and valve
Control system and controller	Controller broken
	Customer changes settings
	Incorrect control sequence
	Incorrect set point value
	Outdoor temperature sensor incorrectly placed
	Temperature sensor broken
	Temperature sensor gives the wrong signal
	Temperature sensor placed on the wrong pipe
	Temperature sensor placed too far away from heat exchanger
Control valve	Control valve leaks in closed position
	Control valve seizes in closed position
	Control valve seizes in open position
	Leaking gaskets
	Oversized control valve
Customer's internal heating system	Air in the system
	Low pressure in expansion vessel
	No DHW circulation
	Poor balancing of radiator system
	Set point value close to/higher than district heating temperatures
	Thermostatic radiator valves missing/broken
	Uncontrolled flow in secondary circuits
	Unnecessarily large flow in DHW circulation
Heat exchanger	Broken heat exchanger plates
	Cracking
	Fouling
	Leakages
	Shell-and-tube heat exchanger

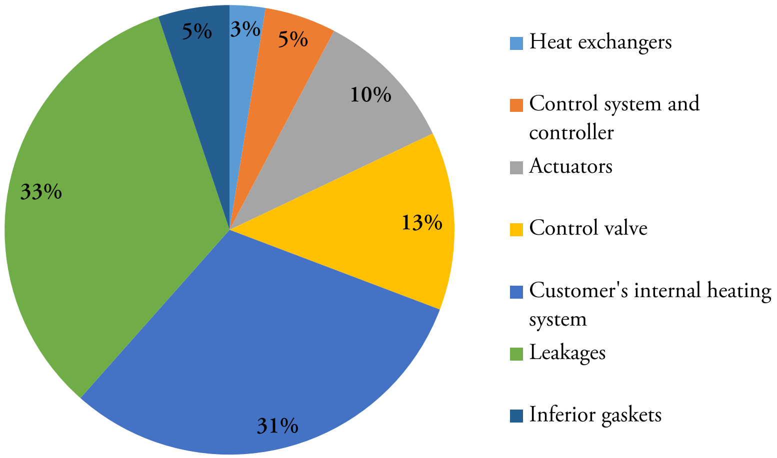


Figure 6.1: Distribution of the most frequently occurring fault categories [96].

6.1.1 Most frequently occurring faults

There is thus a large number of faults that need to be detected and corrected. It is possible that several different fault detection tools will be needed to carry out this task since each fault may generate very different deviations in the customer data. Thus, it is possible that several analysis tools would have to be developed to detect the full range of faults. It would therefore be valuable to have a possibility to prioritize between the different faults, starting with the ones that occur most frequently. This group of faults is likely to have a large impact on the district heating system due to its abundance. They should thus be prioritized in the fault handling processes and the development of the fault detection methods.

Figure 6.1 shows how the utilities answered when asked which fault category, out of the five categories presented in Table 6.1, their most frequently occurring fault belonged to. As may be seen in the figure, two additional categories have been added to the categories in Table 6.1: Leakages and Inferior gaskets. These two categories emerged when analysing the responses from the survey study conducted in Paper III. Leakages include all leakages that may occur in all parts of a customer installation. Inferior gaskets include all malfunctioning or failing gaskets that appear in a customer substation and internal heating system. As seen in the figure, the most frequently occurring faults belonged to the leakages category, closely followed by faults in the customer's internal heating system. This fault distribution indicates what faults should be prioritized when developing fault detection algorithms and working with fault handling in customer installations.

6.1.2 Developing a taxonomy of fault labels

Another interesting aspect that emerged from the surveys and interviews is that there were some discrepancies in how the different utilities named and classified faults. Some utilities associated particular faults with fault categories other than those to which the faults had been assigned in the survey study. It was also common that the utilities reported the symptom of a fault rather than the fault itself. An example of this is that many utilities reported that leakages were their most common fault. However, a leakage is a symptom that occurs due to a fault, for example, broken pipes or inferior gaskets. Thus, there were several indications that a mutual, clearly structured nomenclature, or taxonomy, for faults was needed. A taxonomy would facilitate the discussion on different faults and make it possible to compare results and statistics within the district heating industry and research. From both research and industrial perspectives, a bonus is that if all utilities were to use the same, structured way to name faults, it would be possible for the utilities to gather valuable information about their customer installations and the faults that occur in them. This information would, in turn, facilitate the creation of labelled data sets by labelling specific, deviating data patterns with a fault label. A fault label may be described as one or a few words that describe the details about the cause of a deviation in data, that is, the fault that caused the deviation. The taxonomy for faults in district heating customer installations should consist of a number of different fault labels, clearly describing the faults.

Therefore, the results of this thesis include a taxonomy of different fault labels that may be used to label deviations in district heating customer data. The overall structure of the taxonomy is presented in Figure 6.2. The taxonomy includes information about what fault

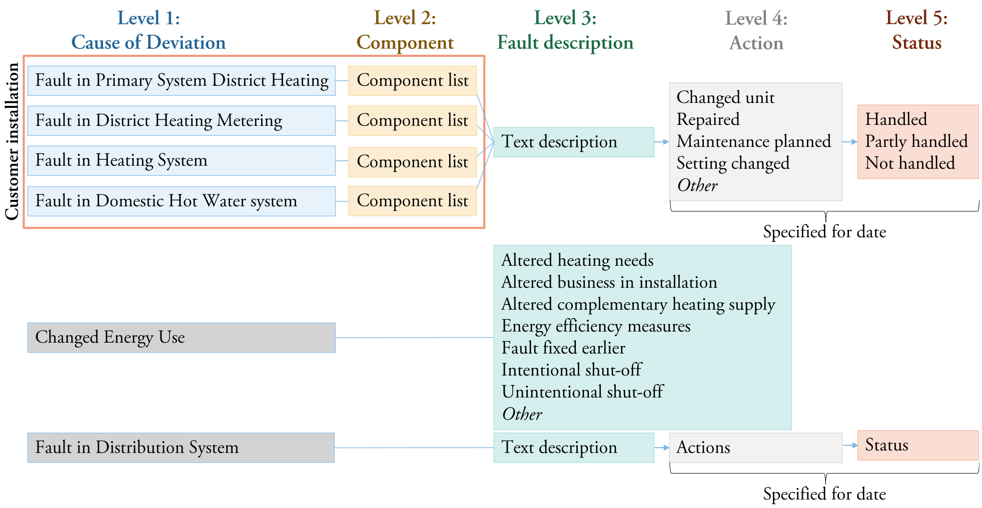


Figure 6.2: Overall structure of the deviation cause taxonomy [97].

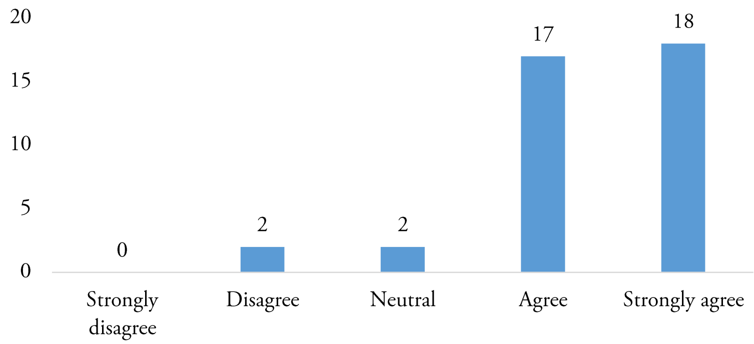


Figure 6.3: The taxonomy covers the most common issues that generate deviations in district heating customer data. 39 of the respondents answered this question [97].

has occurred in an installation and if something has been done to correct it, what has been done to correct it, and when the fault was corrected. By utilizing the taxonomy regularly, a large amount of labelled data from many different district heating systems would become available. These data sets would make it possible to, first and foremost, validate already developed fault detection methods and improve their accuracy significantly. It would also be possible to develop more general fault detection (and fault diagnosis) methods that may be implemented in many different district heating systems. The full description of the taxonomy and its contents may be found in Paper IV.

The taxonomy was well-received among the utilities that responded to the taxonomy referral survey. They stated that the taxonomy covered the most common reasons why a deviation in customer data occurs, as shown in Figure 6.3. The figure shows how the responding utilities answered when asked whether they think the taxonomy contains the essential fault labels. Most utilities agreed or strongly agreed that the taxonomy covered the most frequently occurring faults. Two utilities disagreed that this was the case. Their reason for this was that many faults might occur in an installation, but not all may cause a deviation in customer data.

The responding utilities also showed great interest in using the taxonomy to register information about the faults in their customer installations, which may be seen in Figure 6.4. The figure shows how the responding utilities answered when asked whether they would be interested in using the taxonomy in their fault handling processes. The results show that most of the responding utilities would be interested in using the taxonomy to register faults. Ten utilities answered that their interest in using the taxonomy was neutral. Some of these utilities stated that they could not give an answer about their interest since they did not get any information regarding how and if the taxonomy could be implemented in a user interface. This lack of context may also be why three utilities stated that they were not interested in using the taxonomy and shows the need to put the taxonomy into a clear context when presenting it.

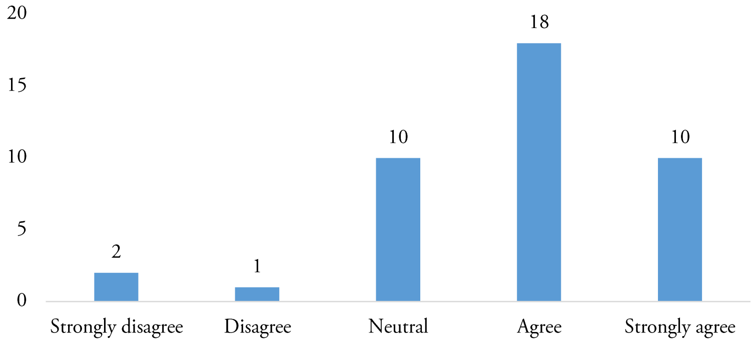


Figure 6.4: I/we would be interested in using the taxonomy to register faults. 41 of the respondents answered this question [97].

6.2 How district heating utilities work with fault handling today

The results showed that faults in customer installations are a common problem and that many different faults may occur in the buildings. When asked about faults in their customer installations, the utilities stated that this is a common problem that they would like to work against. Their main incentives for working with this problem were related to production and distribution conditions in their district heating systems that benefited from decreased return temperatures. The incentive was also related to an economic value in a reduction of the return temperatures. Previous research has shown that the cost reduction gradient for reducing the return temperature may amount to up to 0.5 € per MWh and °C, but that it varies between different district heating systems [1]. If a utility is interested in knowing the economic value of reduced return temperature in its own system, it must calculate the cost reduction gradient for its specific district heating system. However, few utilities participating in the studies conducted in this thesis had done this. This meant that they had little knowledge of the actual economic value of reduced return temperature levels in their systems.

The results related to the organizational aspects of fault handling further showed that few utilities had a clear stakeholder in their fault handling processes. The lack of stakeholder meant that no one actually demanded that the fault handling process was to be carried out. The utilities experienced this to be a problem since no one followed up on the fault handling process and made sure that the work was being carried out. The lack of stakeholder also made it hard to allocate resources to the fault handling process, both in terms of person-hours and economic resources. Many utilities, therefore, mainly conducted the fault handling process if time was available and had no clear organizational structure for fault handling. A problem related to the latter is that there may be a lack of clarity regarding what part of the organization should carry out the actual fault correction. It is common practice in a district heating utility to divide the responsibility for heat meters and the rest of the substation between two different divisions. Thus, if there is a fault in the heat meter, the fault should be treated by the division responsible for heat meters. This example shows the need to allocate the correct fault instances to the correct utility division, which means that the division should be allocated resources to perform the assigned work. However, the utilities expressed that they currently had no structured way of performing these kinds of allocations.

There was thus room for improvement in the current fault handling processes, particularly regarding the organizational aspects. These challenges may also contribute to the current large number of faults in the district heating systems. However, the results in the thesis work also showed that some utilities had succeeded in reaching lower return temperatures by working actively with fault handling in their customer installations. These utilities had actually investigated the value of reduced return temperatures and were therefore aware of the negative impact increased return temperatures would have on their system's performance and economics. A significant contributor to their success in reducing the system temperatures was that they involved their customers in the fault handling process. The customer involvement was done by creating different types of incentives for their customers to eliminate faults in their installations. One way to create a monetary incentive was to introduce a component in the price model directly related to the mass flow. This led to an increased district heating tariff if the return temperature from the installation was high, since this will cause volume flow through the substation to be high as well (according to Equation 2.1). Thus, the customer would have to eliminate any faults causing a high return temperature to receive a lower district heating tariff. Another contributing factor, generating an incentive for the customers to work with faults in their installations, was the relationships with their customers. Most of the utilities with low return temperatures had a smaller number of customers, making it possible to maintain a more personal relationship. The personal relationship was created by performing regular service visits and providing the customers with clear information about why and how they should eliminate faults in their buildings. Other solutions were to offer the customers a free of charge survey of their installation when needed and encourage the customers to call the utility whenever the heating of the installation was not working as it should. Another important aspect of the successful fault handling processes was to gain physical access to the substations and internal heating systems to look for faults in the building. This aspect may seem trivial, but several of the utilities stated that they currently experienced issues with gaining physical access to the installation – either due to the customer not wanting them in their installations or due to lack of maintenance agreements. It was not very common that the utilities had physical access to the customer's internal heating system, since they were mainly responsible for the performance and maintenance of the district heating customer substation. However, utilities that were currently successful in fault handling stated that they also, at least in some cases, were allowed to monitor the internal heating systems as well.

Many of the solutions mentioned above match very well with how smaller utilities were working with fault handling. The larger utilities stated that some of these solutions were hard to implement due to the large number of customer installations in their district heating systems. The fault handling process thus required a large number of person-hours, especially for the detection of faults. These utilities would benefit from having a clear organization for fault handling, with resources and person-hours allocated for this type of work.

6.3 How district heating utilities carry out fault detection today

The results related to how utilities are currently working with fault detection show a need for, and interest in, fault detection based on customer data analysis in the district heating industry. This need was identified during the interviews, surveys and workshop conducted in this thesis. However, most utilities currently identified most faults when conducting service visits to their customer installations. These service visits were part of the customers' maintenance agreements. This way to detect faults meant that most faults were identified after they had already occurred and had already impacted the system negatively. An important aspect of this approach to detect faults is that fault detection was not the district heating utilities' primary objective of the service visits. The main objective of these visits was instead to maintain a good relationship with their customers and maintain the customers' confidence in district heating. Detecting faults in the customer installations was thus primarily seen as a bonus to the other benefits of having maintenance agreements.

There were, however, several utilities conducting fault detection using other methods. Some of the utilities were already, to some extent, analysing customer data to identify customer installations with undesirably high return temperatures. However, their analysis methods were primarily based on analysis of individual customer installations, and most utilities performed the analysis manually. A common method for doing this was to extract top-ten-lists of installations with poor performance, for example, by analysing the return temperatures or overflows of the installations. These customers were then contacted to help them improve the performance of their installation.

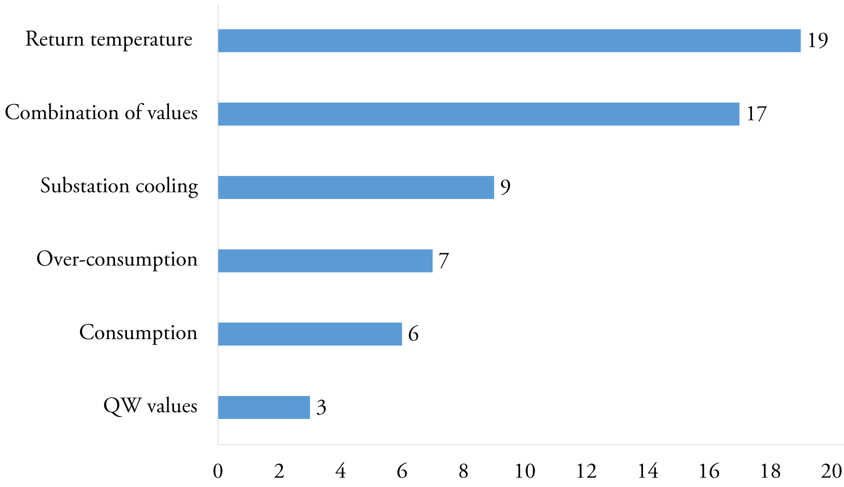


Figure 6.5: Different measurements/values Swedish utilities used to identify poorly performing customer installations [96].

Other common methods for fault detection were monthly checks of customer data, analysis of quality index based on the installation performance, analysis of return temperature levels, analysis of flow, and the over-consumption or overflow method. The parameters being investigated are summarized in Figure 6.5. *Combination of values* refers to a combination of the different parameters measured in the heat meter that were analysed together to investigate the performance of the installation. The QW values are calculated by dividing the flow rate with the heat used in the building during a specific time, for example, during a week or a month. This calculation results in a number that makes it possible to compare customer installations of different sizes and heat demands with each other.

The fault detection methods listed above were based on identifying large deviations in one or some of the parameters measured by the heat meter in the customer substation. This means that the current fault detection methods were primarily based on comparing new data to historical data. The current fault detection methods were thus primarily suitable for corrective maintenance planning methods, and to some extent, also condition-based maintenance [33, 34]. However, to reach the full potential of condition-based maintenance, the methods would have to become more automated and analyse multiple customer installations at the same time. The current methods would not be suitable for predictive maintenance, where a model capable of predicting the future states of the customer installations would be needed.

The methods currently being used for fault detection made it possible for the utilities to detect the customer installations that deviated most significantly from the expected behaviour. However, the utilities experienced that they were currently not capable of detecting faults that had a continuously negative effect on the system but that were not clearly performing poorly when a human being was analysing the customer data. Such faults include, for example, faults occurring in the space heating system that are not clearly visible in the customer data when a DHW load is present in the system as well, and long-term drifting values in measurement equipment. These faults may only be possible to identify during night-time when there is no DHW load, and if the night-time data is not being investigated, it will be challenging to identify these faults. Therefore, an automated fault detection method would be beneficial for the detection of such faults. The utilities also asked for methods capable of detecting imminent faults; that is, detecting when the performance of a customer installation starts to deviate from the normal behaviour. Such deviations may be an indication of an imminent fault and should thus be investigated immediately. The latter type of fault detection aligns with predictive maintenance schemes and would be an interesting aspect of future fault detection methods. The utilities thus expressed a need for, and a clear interest in, more rapid and automated fault detection methods in order to be able to speed up their fault detection processes.

6.4 Fault detection methods based on analysis of district heating customer data

The studies conducted in Papers I and II were focused on automated fault detection in district heating substations and internal heating systems. Two different fault detection approaches were investigated, and the results from both papers indicated that there is great potential in detecting such faults using automated data analysis methods. The studies resulted in two fault detection methods implemented on real-world district heating customer data, and both methods were capable of detecting deviations in the data.

The results from Paper I include a description of the ML pipeline that was developed in the study using TPOT. The pipeline may be seen in Figure 6.6. As may be seen in the figure, the ML pipeline consisted of two main steps: data pre-processing and modelling.

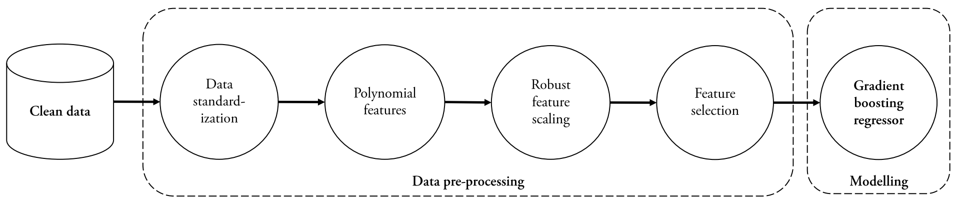
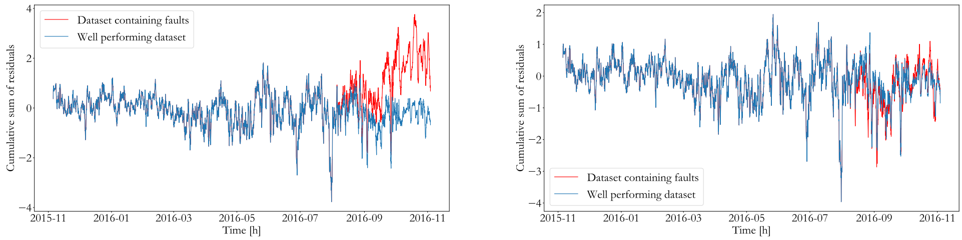


Figure 6.6: Illustration of the TPOT pipeline developed in Paper I [98].



(a) Cumulative sum of residuals for data set with drifting outdoor temperature meter [98].

(b) Cumulative sum of residuals for data set with drifting supply temperature meter [98].

Figure 6.7: Figure of the cumulative sum of residuals for data sets containing two different faults: a drifting outdoor temperature meter and a drifting supply temperature meter. Blue lines represents the residuals between the mass flow rates from the data set that did not contain faults and the model outputs for this data set. Red lines represents the residuals between the mass flow rates from the data set where faults had been induced and the model outputs for this data set.

The data pre-processing step comprised four different steps: data standardization, generation of polynomial features, robust feature scaling, and feature selection. The modelling step consisted of the regressor that predicted the mass flow rate in the installation, which in this case was the Gradient Boosting Regressor (GBR). The full description of the pipeline and its components may be found in Paper I.

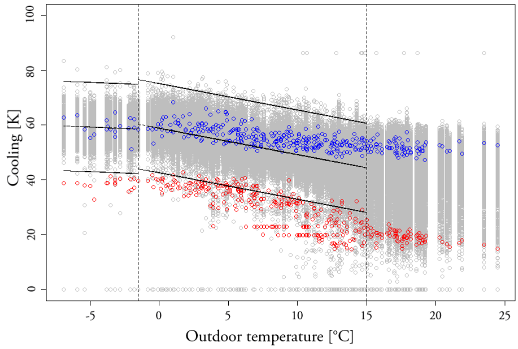
Furthermore, the results from Paper I showed that the developed method could detect faults that were artificially induced in data from a well-performing customer substation. Although not being implemented on a data set known to contain actual faults, the method still shows great promise for fault detection. It also shows promise for fault diagnosis since the method displays different behaviours for different faults. This may be seen in Figure 6.7 where the cumulative sums of residuals for two of the investigated faults (drifting outdoor temperature meter and drifting supply temperature meter) are displayed. In the figure, the blue line represents the residuals between the mass flow rates from the data set that did not contain faults and the model outputs for this data set. The red line represents the residuals between the mass flow rates from the data set where faults had been induced and the model outputs for this data set.

As may be seen in the figure, the residuals are different for the two different faults. Both faults were induced in the data sets after August 2016. Shortly after this, the cumulative sum of residuals for the data set with a drifting outdoor temperature meter starts deviating from the cumulative sum of residuals for the original data set. This deviation may be seen in Figure 6.7a. In October 2016, the deviation was quite significant, which indicates that the model predictions for the data set containing a fault have changed compared to the predictions for the original data set. When comparing these conclusions to the cumulative sum of residuals for the data set with a drifting supply temperature meter in Figure 6.7b, it is clear that the difference in the model's behaviour is not as significant between the original and the faulty data set. Thus, there is an indication that the developed model predicts the mass flow rate differently for different faults. This behaviour is precisely what is expected from a fault diagnosis algorithm - the capability to distinguish between different faults.

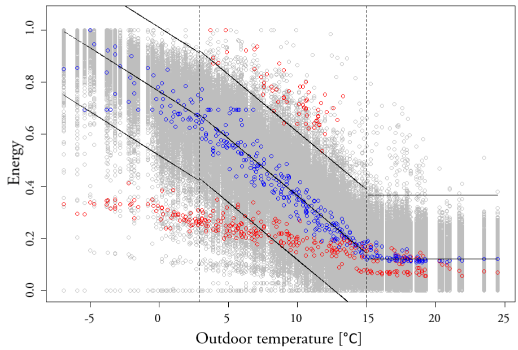
The algorithm developed in Paper II was implemented on a data set from 3 000 real-world customer installations. The results show that a large number of customer installations (1273 or 43 %) were identified as poorly performing by the algorithm. Figure 6.8 displays a visualization of the limit checking methods that were implemented for the three investigated parameters: delta T, return temperature, and energy consumption. As seen in the figure, linear thresholds were used for delta T or cooling (Figure 6.8a) and energy consumption (Figure 6.8b, values are normalized to make it possible to compare installations of different sizes with each other), while a constant threshold was used for the return temperature (Figure 6.8c). In Figure 6.8, the grey circles represent the reference case values, and the black lines represent the thresholds. The blue circles represent the values of a well performing substation and the red circles represent the values of a poorly performing substation.

As may be seen in the figure, the values of the well performing customer installation are located inside the threshold values as is expected if the installation is working as it should. In contrast, a large share of the values of the poorly performing installation is located outside the threshold values. This indicates that the installation is not performing as it should, since a well performing installation should have its main share of the values inside the thresholds, as explained in Section 5.2.

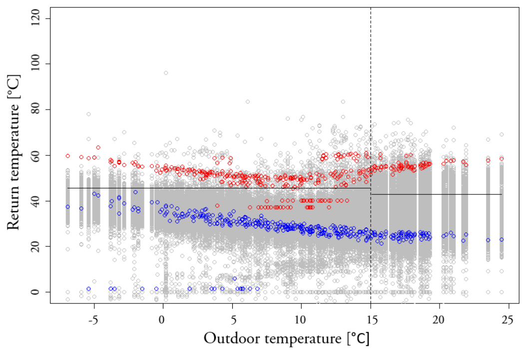
When moving forward to start validating the results from the fault detection studies, some issues occurred due to the lack of labelled data, especially in Paper II. All the identified installations had a large overflow, indicating that they were not performing as well as possible, compared to an ideal case (in this case, the ideal case was a cooling in the substation of 45 °C). This indicated that the developed algorithm detected installations that did not have an optimal performance. However, to truly validate the method, it would have to be implemented on data sets known to contain specific faults at specific occasions – that is, being implemented on labelled data. Implementation on this type of data would make it possible to analyse whether the algorithm detected installations that actually contained or had contained a fault. It would also be possible to improve the algorithm's performance to reduce the number of false alarms. However, no such data was available from the district heating utility. It was possible to validate smaller batches of the results by hand by visualizing the data and discussing data and data patterns with district heating experts and service technicians. However, this kind of validation is very time-consuming, which means it is not applicable for validating larger data sets. Labelled data would thus contribute to improving both the validation process and the fault detection methods themselves. This is also clear when comparing the fault detection capabilities in the district heating domain to other related and similar domains such as fault detection in the heating, ventilation, and air



(a) The limit checking method using linear thresholds for the cooling of the substation as a function of the outdoor temperature [99].



(b) The limit checking method using linear thresholds for the energy consumption (normalized) as a function of the outdoor temperature [99].



(c) The limit checking method using constant thresholds for the return temperature as a function of the outdoor temperature [99].

Figure 6.8: Visual representation of the limit checking methods implemented in Paper II. Grey circles represent the reference case values, black lines represent the limit checking thresholds, blue circles represent the values of a well performing substation, and red circles represent the values of a poorly performing substation [99].

conditioning (HVAC) domain. Several research papers related to fault detection in water chillers, for example, [100–104], successfully utilize the same data set from the ASHRAE RP-1043 project to validate their developed fault detection methods. ASHRAE RP-1043 was a research project which conducted several experimental studies which resulted in labelled data sets that contained both faulty and fault-free data [100].

The lack of labelled data does not only affect the access to data that is known to contain faults – it also affects the access to data that is known to originate from a customer installation without faults. There was thus difficulty in finding well-performing installations for the reference case in both Paper I and Paper II. In these studies, this problem was solved by analysing the data set for customer installations with low, even negative, overflows and then visualizing the data from these installations. In this way, it was possible to obtain customer data that could be used to develop the fault detection methods in both Paper I and Paper II. However, this aspect is something that needs to be further considered when implementing the methods in a real-world district heating system. The choice of reference case installations will become very important when doing so.

6.5 Future fault handling processes at the district heating utilities

The results showed that all utilities, irrespective of size, agreed that they want to work more systematically with fault handling. Some key factors in the work towards obtaining more systematic fault handling processes have been identified – to calculate the economic value of eliminating faults in the system, identify a clear stakeholder in the fault handling process, and allocate resources to the fault handling process. An aspect related to this is that the results from the taxonomy referral survey study showed that the utilities are interested in adapting and improving their fault handling processes on a larger scale to detect more faults in their customer installations. Thus, there is a need for tools and frameworks in the district heating industry that may help the utilities improve their fault handling processes.

By implementing automated fault detection tools, the utilities saw an opportunity to detect and correct more faults. However, this was not the only opportunity that they identified. They also saw a possibility to develop new energy services and business models, making it possible to improve both the fault detection capability in the system and the customer relationships at the same time. One such solution was that they saw an opportunity to develop a new type of maintenance agreement, where data analysis is utilized to detect faults in the customer installations. The utility would then reach out to the customers immediately and offer them help to correct the fault in their installation. This would simplify the fault handling process for the customers – all they would have to do is to say yes to a service visit to get the fault corrected. It would also create an actual incentive for the customers to correct the faults in their installations. By working in this manner, it would be possible to find and correct more faults compared to today, and the customer relationships would be strengthened.

There was also a need to develop structured, holistic approaches to fault handling where utilities can eliminate a large number of faults. These approaches include an entire cycle of fault handling – from automated data analysis methods, to fault correction, to reporting faults, to improving data analysis methods. A suggestion for this type of fault handling process is presented in Paper V. The suggested fault handling process may be seen in Figure 6.9. The process is structured as a cycle to make the most of the information obtained during the process. The fault handling process starts at the top of Figure 6.9 with the collection of customer data and customer information (step 1). The following steps are then carried out: analysis of customer data (step 2), decides on continued fault handling (step 3), creates work order (step 4), performs site visit (step 5), creates a report from site visit (step 6), and reports back to different systems (step 7). Therefore, it is important to carry out each step of the process and not exclude any of them. It is also important to decide when a fault has been treated and corrected to close a case of fault handling. Some deviations in customer data may require several iterations of the fault handling process before being considered "solved". This iteration may have to happen if a deviation in data has more than one root cause; that is, more than one fault is present in the installation. In this workflow, it is suggested that this decision should be carried out in step 2 of the process after an entire fault handling cycle has been completed. Two essential parallel tracks are not part of the fault handling process itself: other indications of faults and documents decision. All steps of the fault handling process, including the parallel tracks, will be described in further detail below.

In step 1 of the process, customer data and customer information are collected. This data will serve as input to the data analysis tool.

In step 2 of the process, data analysis is carried out to detect deviating behaviours in customer data. This requires an automated data analysis tool capable of performing fault detection and a data analyst surveying the analysis. The output from the data analysis tool should be an indication that there may be a fault in a customer installation. There may also be other indications of fault in a fault handling process, for example, if a customer contacts the utility about a problem with heat comfort in the customer installation. Therefore, other indications of faults are included in the workflow as well. These indications may serve as input for the data analysis tool if there is a need to analyse data before deciding to visit the customer installation. They may also serve as direct input for step 3 of the fault handling process, for example, if the fault is urgent and has to be treated as soon as possible.

In step 3, a service admin (short for service administrator) decides whether a fault indication should be further treated in the fault handling process or not. To make this decision, the

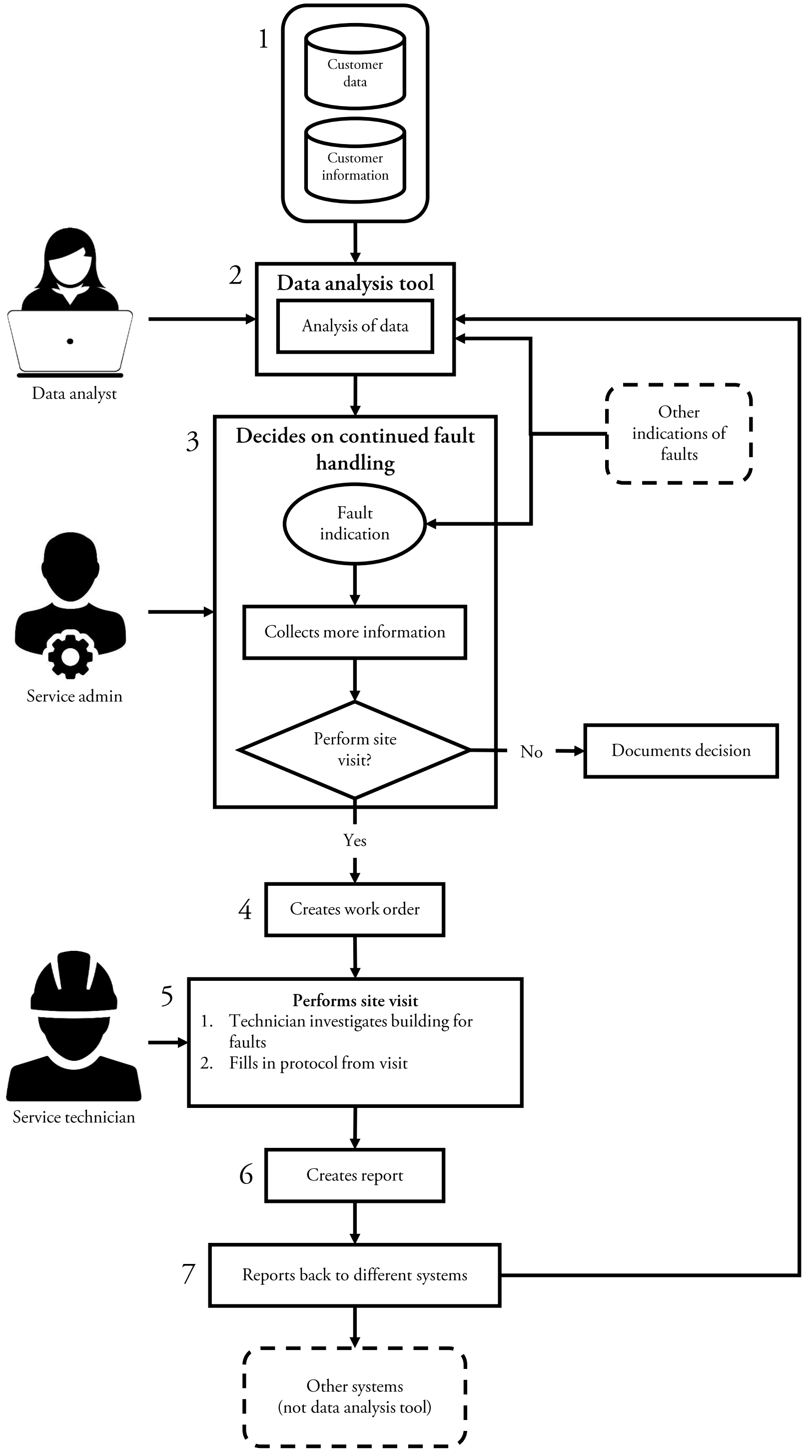


Figure 6.9: Suggested workflow for fault handling of district heating customer installations using analysis of customer data [105].

service admin may have to collect more information about the customer installation by contacting the customer, visualizing different parameters of customer data, or reading previous service protocols. The service admin also has to decide if the fault should be prioritized or not, depending on the system impact and severity of the fault. If it is decided that the fault should *not* be treated further in the process, this decision must be documented. The reason for this is that if an indication of a fault has occurred multiple times for the same installation without the fault being treated further, it may be valuable to visit and investigate the installation for faults.

In step 4, the service admin creates a work order for the fault indication that is to be further handled in the process. The work order should include information about the customer installation, such as what components are installed in the installation and if faults existed in the installation on previous occasions. There could also be instructions regarding preparatory work that has to be carried out before visiting the installation. For example, if customer data has to be investigated further before performing the visit. The work order should also specify, if needed, what part of the organization should carry out the work order.

In step 5, a site visit is performed by a service technician that inspects the installation. The service technician also fills out a service protocol during the visit, preferably using a service application, where information obtained during the visit is recorded. The recorded information should be in a standardized format. In this thesis, it is suggested that the taxonomy developed in Paper IV should be utilized to record information about identified faults. However, it is also important to record if no fault has been identified in the installation, since this provides valuable input for the data analysis tool.

In step 6, a report from the service visit is created. The information included in the report should be based on the information in the service protocol.

In step 7, the site visit report is reported back to the district heating utility. Its contents may be used for several different purposes, for example, billing the customer for the work that has been carried out and updating the record of components installed in the installation. Thus, there is a need to report the information to several different systems, such as the utility's business and customer systems. These systems also include the data analysis tool. By reporting back to this tool, it is possible to create labelled data sets by relating the information in the service protocol to deviations in the data. This work procedure will generate data sets that may be utilized to improve the performance and accuracy of the fault detection method.

Chapter 7

Concluding discussion

The work conducted in this thesis has treated the subject of fault detection in district heating customer installations. The objective of the thesis has been to develop solutions that help improve the fault handling capabilities of the district heating industry by suggesting both organizational and technical solutions to the problem. More specifically, the solutions have been centralized around customer data analysis and the challenges that arise when implementing such methods. Therefore, the faults in focus have been the ones that cause a change in district heating customer data. The problem has been approached by developing methods for fault detection based on data analysis and identifying key frameworks needed for the successful implementation of such tools. It was, therefore, important to investigate current fault handling processes to identify the largest potentials for improvement. Several such areas were identified and included a taxonomy for labelling deviations in district heating customer data, organizational aspects such as clear stakeholders in the fault handling process, and the importance of creating incentives for the customers to eliminate faults in their installations.

The work in this thesis has been conducted with both current and future district heating systems in mind. In current systems, it is essential to detect and eliminate faults causing high return temperatures so that they may transition towards decreased distribution temperatures, thereby allowing a more environmentally friendly and resource-efficient heat supply. In future systems, which would be designed to operate at low distribution temperatures, there will be a need to rapidly detect, diagnose, and correct new faults that occur in the customer installations. If the faults go undetected, there is a risk that unwanted increases in the distribution temperature levels occur. Increased temperature levels would have a negative impact on the entire system since there is a risk that it would not be possible to efficiently utilize the low-temperature heat sources that future district heating systems will be supplied by. Thus, there are lower margins for temperature changes in future systems, making rapid fault detection even more crucial in the future. Thus, successful fault handling processes serve a purpose in both types of systems, but their implementation will be somewhat different. In current systems, they will have to be implemented step-wise to allow the organization to adjust to and realize the new fault handling processes. In future systems, improved fault handling schemes should be implemented from the start of the operation of the district heating system to reduce the possibility of significant fault occurrences. The results of this thesis may thereby contribute to increased energy efficiency and decreased environmental impact of both current and future district heating systems.

Fault detection in district heating customer installations is essential due to the negative impact that these faults have on the system's efficiency. As shown in Paper III, more than 30 different faults belonging to five different fault categories may occur in an installation. The faults presented in this thesis may not even be the complete list of faults; other faults may occur if, for example, the substation is designed to fill a particular purpose in a building that requires specific components to be installed in the installation. The results related to the most frequently occurring faults may also vary between different district heating systems. The utilities who participated in the study were all Swedish, and the results may thus be affected by the national characteristics of Swedish district heating. Other countries may have a different distribution of faults and may also have other fault categories represented in their statistics.

The results also show that the utilities' primary way of detecting faults in their customer installations, the service visits, is not the optimal way to perform fault detection. Even though faults are detected, the utilities mainly see this as a bonus to the other benefits obtained when performing service visits (improved customer relationships and customers' trust in the services provided by the district heating utility). There is thus a need to develop methods dedicated to the rapid and automated detection of faults. This topic was explored in the first two papers of this thesis, which showed that it is possible to detect deviations in district heating customer data using data analysis. Both methods use data collected from customers' heat meters, which is the data most utilities currently have access to. The methods are thus analysing variables the utilities already have access to and are familiar with. This is advantageous since new fault detection methods should be easy for the utilities to assimilate and implement in their fault handling processes. There will otherwise be a risk that the developed methods are not being used actively in the industry.

However, the results in this thesis have also shown that there are challenges in developing methods for fault detection in customer installations. The main challenge is the lack of labelled data, making it hard to validate the developed fault detection methods. The deviation cause taxonomy developed in Paper IV may be utilized to help solve this problem by labelling deviations in customer data with the fault labels presented in the taxonomy. This requires information on what fault label should be utilized for each deviation. The information may be gathered by allowing the technical service personnel to record information about identified faults in a service application, where the taxonomy is implemented in a user interface. The deviation may then be labelled with the correct fault label, creating data sets that have been labelled by experts within the district heating industry. Such data sets are valuable inputs for both the development and validation of fault detection methods. Even though few labelled data sets are available today, minor adjustments to the current fault handling process would make it possible to obtain such data sets. A suggestion is to implement (at least parts of) the fault handling process suggested in this thesis – more specifically, to implement the steps related to creating service reports with standardized information and reporting structured information back to the utility where the deviations in data are labelled with the fault label in question. This involves the implementation of a user interface based on the taxonomy developed in Paper IV.

The results regarding how the utilities are currently working with faults in customer installations show that some organizational aspects of fault handling could be improved. An important aspect is that few utilities have a clear stakeholder in the process, which results in difficulties in allocating resources for fault handling and a lack of follow-up of the fault handling work within the organization. A way to clarify who should take the lead on the fault handling process would be to investigate the value of reduced return temperatures in the district heating system. Identifying such a value would clarify what part within the organization benefits the most from it and should thus be the stakeholder in successful fault handling processes.

Furthermore, an outcome of this PhD thesis is that there is a need to approach fault handling in customer installations from a holistic point of view within a utility. It is not possible to allow parts of an organization to work individually with these issues ”when they have time to spare”. The work has to be conducted in a structured way regularly, and someone has to be responsible for the work being carried out. By implementing a fault handling process such as the one described in Paper V, it would be possible to work actively with fault handling and methodically eliminate faults. The automated fault detection methods currently being developed, both in this thesis and in other research, are primarily suitable for condition-based or corrective maintenance schemes. Therefore, the fault handling process suggested in this thesis is developed with these types of fault detection methods in mind. However, the fault handling process would also be suitable to use in a predictive maintenance scheme. The key to developing this type of maintenance scheme is that the automated fault detection methods improve and develop towards predictive fault detection instead. The rest of the fault handling process could be more or less the same in a predictive maintenance scheme. The developed fault handling process further assumes that the utility has access to the customer installation so that a fault may be corrected. However, the results show that this is not always the case. On most occasions, the utility has access to the customer substation but not to the internal heating system of the customer. The lack of access is a problem since the results in Paper III show that one of the fault categories, where many of the faults occur, is the customer's internal heating system. Future maintenance agreements could thus be designed to also include maintenance of the internal heating system. Such agreements would be beneficial for both the utility and the customer, who would be able to receive help to fix a fault in their installation, no matter where in the installation the fault is located.

An important aspect of the results presented in this thesis is that they are not aiming to give definitive answers to how fault handling is best conducted or what faults are the most common ones. District heating operates like a local heat market, and each district heating system has its own unique conditions and characteristics in terms of size, heat supply, and distribution methods. In addition, each customer installation in a district heating system is unique, making the handling of faults a very complex problem. It is thus impossible, and would not make sense, to provide solutions suitable for all fault instances, for all systems, at all times. The results, however, aim to provide suggestions and recommendations of how fault handling processes may be improved to help solve the problems that the utilities are currently facing. They also give an overview of different aspects that a utility would have to consider when starting to work in a more organized manner with fault handling. However, the utility would have to identify its specific conditions, such as what business systems and other systems are involved in current fault-related work, and what parts of the organization should be involved in and in charge of the fault handling process, to be able to succeed with their fault handling.

The main limitation of the work in this PhD thesis is that it has been conducted from a Swedish perspective. This context may have affected the generality of the results in terms of the national characteristics and maturity of the Swedish district heating technology. This is mainly the case when discussing the results related to fault distribution, since local conditions may affect these results. An interesting extension of the work conducted in this thesis would be to investigate the fault distributions of other countries. The results related to fault detection methods, organizational aspects of fault handling, and the deviation cause taxonomy are of a more general nature. For example, the fault handling process is designed to fit the organization of many different utilities. The taxonomy may easily be adapted if a connection principle other than the parallel-connection is used. These results are therefore more easily transferable to the context of other countries as well.

The work conducted in this thesis has been performed from the utility's point of view, since they have the most to benefit from eliminating faults in customer installations. However, this work also affects the customer and requires close collaboration between the utility and the customer. It would thus be interesting to investigate the fault handling problem from the customer perspective in future studies, especially questions related to maintenance agreements and incentives for correcting a fault. Other future studies may include using the deviation cause taxonomy to collect labelled data and develop fault detection methods using this data as input. It would also be interesting to implement the developed fault handling process and investigate the challenges and opportunities that arise when implementing the process in practice. Related to this is to develop a method for utilities to investigate the value of structured fault handling processes. This method should include the economic value of reduced return temperatures and other values such as improved customer relationships.

References

[1] Svend Frederiksen and Sven Werner. *District Heating and Cooling*. Lund, Sweden: Studentlitteratur AB, 2013.

[2] Brian Vad Mathiesen et al. *Towards a decarbonised heating and cooling sector in Europe - Unlocking the potential of energy efficiency and district energy*. Aalborg University, 2019.

[3] Henrik Lund et al. "4th Generation District Heating (4GDH) - Integrating smart thermal grids into future sustainable energy systems". In: *Energy* 68 (2014), pp. 1–11.

[4] Henrik Lund et al. "Smart energy and smart energy systems". In: *Energy* 137 (2017), pp. 556–565.

[5] United Nations Environment Programme. *District energy in cities - Unlocking the potential of energy efficiency and renewable energy*. Paris, France: United Nations Environment Programme, 2015.

[6] Sven Werner. "International review of district heating and cooling". In: *Energy* 137 (2017), pp. 617–631.

[7] Energimyndigheten. *Energiläget 2020*. ET 2020:1. Eskilstuna, Sweden: Energimyndigheten, 2020. In Swedish.

[8] Kristina Lygnerud. "Challenges for business change in district heating". In: *Energy, Sustainability and Society* 8.20 (2018).

[9] The European Commission. *An EU strategy on heating and cooling*. Communication Volume ID: COM(2016)51. Brussels, Belgium: The European Commission, 2016.

[10] International Energy Agency. *Heating*. 2020. URL: `https://www.iea.org/reports/heating` (visited on 02/23/2021).

[11] Dietrich Schmidt et al. "Low Temperature District Heating for Future Energy Systems". In: *Energy Procedia* 116 (2017), pp. 26–38.

[12] Fabian Bühler et al. "Industrial excess heat for district heating in Denmark". In: *Applied Energy* 205 (2017), pp. 991–1001.

[13] Annelies Vandermeulen et al. “A theoretical benchmark for bypass controllers in a residential district heating network”. In: *Energy* 151 (2018), pp. 45–53.

[14] Henrik Gadd and Sven Werner. “Achieving low return temperatures from district heating substations”. In: *Applied Energy* 136 (2014), pp. 59–67.

[15] Henrik Gadd and Sven Werner. “Fault detection in district heating substations”. In: *Applied Energy* 157 (2015), pp. 51–59.

[16] Fredrik Sandin, Jonas Gustafsson, and Jerker Delsing. *Fault detection with hourly district energy data*. 2013:27. Stockholm, Sweden: Svensk Fjärrvärme AB, 2013.

[17] Ece Calikus et al. “Ranking Abnormal Substations by Power Signature Dispersion”. In: *Energy Procedia* 149 (2018), pp. 345–353.

[18] Shahrooz Abghari. “Data mining approaches for outlier detection analysis”. Doctoral thesis. Karlskrona, Sweden: Department of Computer Science, Blekinge Institute of Technology, 2020.

[19] Shiraz Farouq et al. “Large-scale monitoring of operationally diverse district heating substations: A reference-group based approach”. In: *Engineering Applications of Artificial Intelligence* 90.103492 (2020).

[20] Patrick Lauenburg. “Improved supply of district heat to hydronic space heating systems”. Doctoral thesis. Lund, Sweden: Division of Efficient Energy Systems, Department of Energy Sciences, Faculty of Engineering, Lund University, 2009.

[21] Energimyndigheten. *Energiläget 2020 i siffror*. 2020. URL: `http://data.worldbank.org/indicator/SP.DYN.LE00.FE.IN` (visited on 02/26/2021). Data retrieved from World Development Indicators. In Swedish.

[22] Euroheat & Power. *2015 Country by country statistics*. 2016. URL: `https://www.euroheat.org/wp-content/uploads/2016/03/2015-Country-by-country-Statistics-Overview.pdf` (visited on 02/26/2021).

[23] IRENA, OECD/IEA and REN21. *Renewable Energy Policies in a Time of Transition Heating: and Cooling*. IRENA, OECD/IEA and REN21, 2020.

[24] Collins English Dictionary. *Definition of prosumer*. 2021. URL: `https://www.collinsdictionary.com/dictionary/english/prosumer` (visited on 11/19/2020).

[25] Lisa Brange. “District heating development - Prosumers and bottlenecks”. Doctoral Thesis. Lund, Sweden: Division of Heat Transfer, Department of Energy Sciences, Faculty of Engineering, Lund University, 2019.

[26] Konstantinos Kavvadias, Juan Pablo Jiménez-Navarro, and Georg Thomassen. *Decarbonising the EU heating sector -Integration of the power and heating sector*. EUR 29772 EN. Luxembourg: European Union, 2019.

[27] Henrik Lund et al. “Future district heating systems and technologies: On the role of smart energy systems and 4th generation district heating”. In: *Energy* 165 (2018), pp. 614–619.

[28] Helge Averfalk and Sven Werner. “Essential improvements in future district heating systems and technologies”. In: *Energy Procedia* 116 (2017), pp. 217–225.

[29] Urban Persson and Sven Werner. “Heat distribution and the future competitiveness of district heating”. In: *Applied Energy* 88 (2011), pp. 568–576.

[30] Euroheat and Power. *District Heating in Buildings*. Brussels, Belgium: Euroheat and Power, 2011.

[31] Anthony M. Smith and Glenn R. Hinchcliffe. “2 - Preventive Maintenance—Definition and Structure”. In: *RCM - Gateway to world class maintenance*. Butterworth-Heinemann, 2004. Chap. 2, pp. 19–37.

[32] Hongjiu Liu and Jiaxuan Wu. “Research on Preventive Maintenance Strategy of Elevator Equipment”. In: *Open Journal of Social Sciences* 6 (2018), pp. 165–174.

[33] Marcus Bengtsson. “On condition-based maintenance and its implementation in industrial settings”. Doctoral Thesis. Västerås, Sweden: Department of Innovation, Design and Product Development, Mälardalen University, 2007.

[34] A.S. Nazmul Huda and Soib Taib. “Application of infrared thermography for predictive/preventive maintenance of thermal defect in electrical equipment”. In: *Applied Thermal Engineering* 61 (2013), pp. 220–227.

[35] Hai Wang et al. “Optimization modeling for smart operation of multi-source district heating with distributed variable-speed pumps”. In: *Energy* 138 (2017), pp. 1247–1262.

[36] Euroheat & Power. *Guidelines for District Heating Substations*. Brussels, Belgium: Euroheat & Power, 2008.

[37] Svensk Fjärrvärme AB. *Fjärrvärmecentralen - utförande och installation*. F:101. Stockholm, Sweden: Svensk Fjärrvärme AB, 2014. Tekniska bestämmelser. In Swedish.

[38] Maria Jangsten et al. “Survey of radiator temperatures in buildings supplied by district heating”. In: *Energy* 137 (2017), pp. 292–301.

[39] Henrik Gadd and Sven Werner. “Daily heat load variations in Swedish district heating systems”. In: *Applied Energy* 106 (2013), pp. 47–55.

[40] Søren Djørup et al. “District heating tariffs, economic optimisation and local strategies during radical technological change”. In: *Energies* 13.5 (2020).

[41] Kari Sipilä. “Cogeneration, biomass, waste to energy and industrial waste heat for district heating”. In: *Advanced District Heating and Cooling (DHC) Systems*. Ed. by Robin Wiltshire. Woodhead Publishing, 2016. Chap. 3, pp. 45–73.

[42] Joana Neves and Brian Vad Mathiesen. *Heat Roadmap Europe: Potentials for Large-Scale Heat Pumps in District Heating*. Aalborg, Denmark, 2018.

[43] Girts Vigants et al. “Efficiency diagram for district heating system with gas condensing unit”. In: *Energy Procedia* 72 (2015), pp. 119–126.

[44] Markus Köfinger, Daniele Basciotti, and Ralf-Roman Schmidt. “Reduction of return temperatures in urban district heating systems by the implementation of energy-cascades”. In: *Energy Procedia* 116 (2017), pp. 438–451.

[45] Patrick Lauenburg. “Temperature optimization in district heating systems”. In: *Advanced District Heating and Cooling (DHC) Systems*. Ed. by Robin Wiltshire. Woodhead Publishing, 2016. Chap. 11, pp. 223–240.

[46] Petter Wallentén. “Steady-state heat loss from insulated pipes”. Licentiate thesis. Lund, Sweden: Department of Building Physics, Lund Institute of Technology, Lund University, 1991.

[47] Alessandro Dalla Rosa, Hongwei Li, and Svend Svendsen. “Method for optimal design of pipes for low-energy district heating, with focus on heat losses”. In: *Energy* 36 (2011), pp. 2407–2418.

[48] Lisa Brange, Marcus Thern, and Kerstin Sernhed. “Risks and opportunities for bottleneck measures in Swedish district heating networks”. In: *Energy Procedia* 149 (2018), pp. 380–389.

[49] Christer Wirén. *Enkla, tydliga och informativa prismodeller för fjärrvärme till småhus (Simple, clear and informative price models for district heating to single-family houses)*. Report Värmegles 2005:15. Stockholm, Sweden: Svensk fjärrvärme AB, 2005. In Swedish.

[50] Svensk Fjärrvärme AB. *Fjärrvärmecentralen - Kopplingsprinciper*. 2009:3. Stockholm, Sweden: Svensk Fjärrvärme AB, 2009. In Swedish.

[51] Catarina Warfvinge and Mats Dahlblom. *Projektering av VVS-installationer*. Lund, Sweden: Studentlitteratur AB, 2016. In Swedish.

[52] Jonas Gustafsson and Fredrik Sandin. “District heating monitoring and control systems”. In: *Advanced District Heating and Cooling (DHC) Systems*. Ed. by Robin Wiltshire. Woodhead Publishing, 2016. Chap. 12, pp. 241–258.

[53] Water Efficiency Labelling and Standards Regulator. *Consideration of hot water circulators for inclusion in the WELS scheme: Background research project*. Canberra, Australia: Department of the Environment, Water, Heritage and the Arts, 2008.

[54] The European Commission. *Annex to the Commission recommendation on the implementation of the new metering and billing provisions of the Energy Efficiency Directive 2012/27/EU no C(2019) 6631 final*. 2019.

[55] Energimarknadsinspektionen (The Swedish Energy Markets Inspectorate). *Energimarknadsinspektionens föreskrifter och allmänna råd om mätning, rapportering och debitering av levererad värmeenergi (fjärrvärme)*. 2014. URL: `https://lagen.nu/eifs/2014:2` (visited on 04/27/2021). In Swedish.

[56] Rolf Isermann. "Terminology in fault detection and diagnosis". In: *Fault Diagnosis Applications*. Berlin, Heidelberg, Germany: Springer, 2011. Chap. 13, pp. 321–323.

[57] You-Jin Park, Shu-Kai S. Fan, and Chia-yu Hsu. "A Review on Fault Detection and Process Diagnostics in Industrial Processes". In: *Processes* 8.1123 (2020).

[58] Rolf Isermann. *Fault-Diagnosis Applications*. Berlin, Heidelberg, Germany: Springer Berlin Heidelberg, 2006.

[59] Suad A. Alasadi and Wesam S. Bhaya. "Review of data preprocessing techniques in data mining". In: *Journal of Engineering and Applied Sciences* 12 (2017), pp. 4102–4107.

[60] Venkat Venkatasubramanian et al. "A review of process fault detection and diagnosis Part I: Quantitative model-based methods". In: *Computers and Chemical Engineering* 27 (2003), pp. 293–311.

[61] Mathworks. *Designing Algorithms for Condition Monitoring and Predictive Maintenance*. 2021. URL: `https://www.mathworks.com/help/predmaint/gs/designing-algorithms-for-condition-monitoring-and-predictive-maintenance.html#mw_5d264a05-dce5-4ade-bb8e-82f3f34d2af0` (visited on 02/26/2021).

[62] Srinivas Katipamula and Michael Brambley. "Methods for Fault Detection, Diagnostics and Prognostics for Building Systems - A Review Part II". In: *HVAC&R research* 11 (2005), pp. 169–187.

[63] Ron J. Patton, Jie Chen, and Sofus B. Nielsen. "Model-based methods for fault diagnosis: Some guidelines". In: *Transactions of the Institute of Measurement and Control* 17 (1995), pp. 73–83.

[64] Rolf Isermann. "Model-based fault-detection and diagnosis – status and applications". In: *Annual Reviews in Control* 29 (2005), pp. 71–85.

[65] Ndeye Lo, Jean-Marie Flaus, and Olivier Adrot. "Review of Machine Learning Approaches In Fault Diagnosis applied to IoT System". In: *International Conference on Control, Automation and Diagnosis ICCAD'19*. Grenoble, France, 2019, pp. 1–6.

[66] Jafar Alzubi, Anand Nayyar, and Akshi Kumar. "Machine Learning from Theory to Algorithms: An Overview". In: *Journal of Physics: Conference Series* 1142.012012 (2018).

[67] Yaguo Lei et al. “Applications of machine learning to machine fault diagnosis: A review and roadmap”. In: *Mechanical Systems and Signal Processing* 138.106587 (2020).

[68] A. L. Samuel. “Some studies in machine learning using the game of checkers”. In: *IBM Journal of Research and Development* 3.3 (1959), pp. 210–229.

[69] Ethem Alpaydin. *Introduction to Machine Learning*. Cambridge, USA: MIT Press, 2010.

[70] Andrew Ng. *CS229 Lecture Notes - Supervised Learning*. URL: `http://cs229.stanford.edu/notes2020spring/cs229-notes1.pdf` (visited on 03/04/2021).

[71] Jonathan Waring, Charlotta Lindvall, and Renato Umeton. “Automated machine learning: Review of the state-of-the-art and opportunities for healthcare”. In: *Artificial Intelligence in Medicine* 104.101822 (2020).

[72] Randal S. Olson et al. “Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science”. In: *Proceedings of the Genetic and Evolutionary Computation Conference 2016*. GECCO ’16. Denver, Colorado, USA, 2016, pp. 485–492.

[73] Infrastrukturdepartementet RSED E. *Fjärrvärmelag (2008:263)*. `https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/fjarrvarmelag-2008263_sfs-2008-263`. 2008.

[74] Shahrooz Abghari et al. “District heating substation behaviour modelling for annotating the performance”. In: *Machine learning and knowledge discovery in databases*. Vol. 1168. Springer International Publishing, 2020, pp. 3–11.

[75] Ece Calikus et al. “A data-driven approach for discovering heat load patterns in district heating”. In: *Applied Energy* 252.113409 (2019).

[76] Puning Xue et al. “Fault detection and operation optimization in district heating substations based on data mining techniques”. In: *Applied Energy* 205 (2017), pp. 926–940.

[77] Shahrooz Abghari et al. “Trend analysis to automatically identify heat program changes”. In: *Energy Procedia* 116 (2017), pp. 407–415.

[78] Przemysław Pałasz and Radosław Przysowa. “Using Different ML Algorithms and Hyperparameter Optimization to Predict Heat Meters’ Failures”. In: *Applied Sciences* 9.3719 (2019).

[79] Aurélien Brés et al. “Coupled Building and System Simulations for Detection and Diagnosis of High District Heating Return Temperatures”. In: *Conference: Building Simulation 2019: 16th Conference of IBPSA*. Rome, Italy, 2019.

[80] Dorte Skaarup Østergaard et al. "Test and evaluation of a method to identify heating system malfunctions by using information from electronic heat cost allocators". In: *Energy and Buildings* 184 (2019), pp. 152–162.

[81] Gerrit Bode et al. "Real-world application of machine-learning-based fault detection trained with experimental data". In: *Energy* 198.117323 (2020).

[82] Liangwei Zhang. "Big Data Analytics for Fault Detection and its Application in Maintenance". Doctoral thesis. Luleå, Sweden: Operation and Maintenance Engineering, Luleå University of Technology, 2016.

[83] Jouko Pakanen et al. *Fault diagnosis methods for district heating substations*. Oulu, Finland, 1996.

[84] Melissa J. Goertzen. *Applying Quantitative Methods to E-book Collections*. ISSN 0024-2586. Chicago, USA, 2017. Library Technology Reports, Vol. 53, No. 4.

[85] Jaber F. Gubrium and James A. Holstein. *Handbook of Interview Research*. Thousand Oaks, USA: Sage publications, Inc., 2001.

[86] Chauncey Wilson. *Interview Techniques for UX Practitioners*. Boston, USA: Morgan Kaufmann, 2014.

[87] Energiföretagen Sverige. *Tillförd energi*. 2021. URL: `%5Curl%7Bhttps://www.energiforetagen.se/statistik/fjarrvarmestatistik/tillford-energi/%7D` (visited on 11/04/2020). In Swedish.

[88] Kathryn White. "Web-Based Questionnaire". In: *Encyclopedia of Quality of Life and Well-Being Research*. Ed. by Alex C. Michalos. Dordrecht, Netherlands: Springer, 2014.

[89] Amy Blackstone. *Principles of Sociological Inquiry – Qualitative and Quantitative Methods*. Saylor Foundation, 2012.

[90] Dana Lynn Driscoll. "Introduction to Primary Research: Observations, Surveys, and Interviews". In: *Writing Spaces: Readings on Writing*. Vol. 2. Anderson, USA: Parlor Press, 2011.

[91] René Bautista. "An Overlooked Approach in Survey Research: Total Survey Error". In: *Handbook of Survey Methodology for the Social Sciences*. New York, USA: Springer, New York, 2012.

[92] Smart Energi. *Smart Energi*. 2018. URL: `https://smartenergi.org/en/om-smart-energi/` (visited on 04/16/2021).

[93] Michael S. Lewis-Beck, Alan Bryman, and Tim Futing Liao. "Group interview". In: *The SAGE encyclopedia of social science research methods*. Vol. 1-0. Thousand Oaks, USA: Sage Publications, Inc., 2004.

[94] P. Gill et al. “Methods of data collection in qualitative research: interviews and focus groups”. In: *British Dental Journal* 204 (2008), pp. 291–295.

[95] Rana Muhammad Dilshad and Muhammad Ijaz Latif. “Focus Group Interview as a Tool for Qualitative Research: An Analysis”. In: *Pakistan Journal of Social Sciences (PJSS)* 33.1 (2013), pp. 191–198.

[96] Sara Månsson et al. “Faults in district heating customer installations and ways to approach them: Experiences from Swedish utilities”. In: *Energy* 180 (2019), pp. 163–174.

[97] Sara Månsson et al. “A taxonomy for labeling deviations in district heating customer data”. In: *Smart Energy* 2.100020 (2021).

[98] Sara Månsson et al. “A machine learning approach to fault detection in district heating substations”. In: *Energy Procedia* 149 (2018), pp. 226–235.

[99] Sara Månsson et al. “Automated statistical methods for fault detection in district heating customer installations”. In: *Energies* 12.113 (2019).

[100] C. Y. Zheng and W. Chen. “A novel modelling method for fault detection and diagnosis in large water chillers without water flow measurements”. In: *International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2015)*. Phuket, Thailand, 2015, pp. 518–522.

[101] Zhuozheng Wang et al. “A Novel Fault Diagnosis Approach for Chillers Based on 1-D Convolutional Neural Network and Gated Recurrent Unit”. In: *Sensors* 20.2458 (2020).

[102] Yalan Wang et al. “A practical chiller fault diagnosis method based on discrete Bayesian network”. In: *International Journal of Refrigeration* 102 (2019), pp. 159–167.

[103] Jiangyan Liu et al. “Data-driven and association rule mining-based fault diagnosis and action mechanism analysis for building chillers”. In: *Energy & Buildings* 216.109957 (2020).

[104] Zhanwei Wang et al. “Fault detection and diagnosis of chillers using Bayesian network merged distance rejection and multi-source non-sensor information”. In: *Applied Energy* 188 (2017), pp. 200–214.

[105] Sara Månsson et al. “A fault handling process for faults in district heating customer installations”. In: *Preprint submitted to Energies* (2021).

Scientific publications

My contributions to the publications

Co-authors are abbreviated as follows:
Marcus Thern (MT), Kerstin Sernhed (KS), Per-Olof Johansson Kallioniemi (POJK), Kristin Davidsson (KD), Patrick Lauenburg (PL), Tijs van Oevelen (TVO), Robbe Salenbien (RS), Ida Lundholm-Benzi (ILB).

Paper I: A machine learning approach to fault detection in district heating substations

I did the conceptualization and developed the methodology of the paper in collaboration with MT and POJK. I wrote the main parts of the paper, as well as conducted the literature study of the paper. MT, POJK and KS all contributed to the revision and editing of the paper.

Paper II: Automated statistical methods for fault detection in district heating customer installations

I did the conceptualization and developed the methodology of the paper in collaboration with KD and PL. KD and I also developed the software related to the study. The original draft and revision of the paper was conducted by me and MT. I also performed the literature review in the paper.

Paper III: Faults in district heating customer installations and ways to approach them: Experiences from Swedish utilities

I did the conceptualization and developed the methodology for the interview and survey study together with KS. POJK also contributed to the methodology of the survey study. I performed all gathering of data in the interview and survey studies, as well as the analysis of this data. I wrote the main parts of the paper and performed the literature review of the paper. KS, POJK, TVO and MT contributed to the editing and revision of the paper.

Paper IV: A taxonomy for labeling deviations in district heating customer data

I did the conceptualization and developed the taxonomy together with ILB and POJK. I developed the methodology for the survey study together with POJK, MT and KS. I wrote the main parts of the paper and performed the literature review of the paper. KS, POJK, RS and MT contributed to the editing and revision of the paper.

Paper V: A fault handling process for faults in district heating customer installations

I developed and performed the methodology for the workshop together with KS. I also developed the fault handling process, and wrote the paper. POJK and MT contributed to the editing and revision of the paper.

Printed by Media-Tryck, Lund 2021 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

Department of Energy Sciences
Faculty of Engineering
Lund University
ISBN 978-91-7895-865-8
ISSN 0282-1990
ISRN: LUTMDN/TMHP-19/1164-SE