Fault handling processes in district heating customer installations

Current and future solutions

Månsson, Sara

2018

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Månsson, S. (2018). *Fault handling processes in district heating customer installations: Current and future solutions*. [Licentiate Thesis, Department of Energy Sciences]. LTH, Lund University.

Total number of authors:
1

LUND UNIVERSITY

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

Fault handling processes in district heating customer installations

Fault handling processes in district heating customer installations
Current and future solutions

by Sara Månsson

Thesis for the degree of Licentiate in Engineering
Thesis advisors: Assoc. prof. Dr. Marcus Thern, Ass. prof. Dr. Kerstin Sernhed, Ass. prof. Dr. Per-Olof Johansson Kallioniemi
Faculty opponent: Dr. Jaime Arriagada

To be presented, with the permission of the Faculty of Engineering of Lund University, for public criticism in the M:E lecture hall at the Department of Energy Sciences on Thursday, the 17th of January 2019 at 13:15.

DOKUMENTDATABLAD enl SIS 61 41 21

Organization
LUND UNIVERSITY
Department of Energy Sciences
Box 188
SE–221 00 LUND
Sweden

Document name
LICENTIATE DISSERTATION

Date of disputation
2019-01-17

Author(s)
Sara Månsson

Title and subtitle
Fault handling processes in district heating customer installations

Abstract

When trying to integrate more fossil-free heat sources such as renewable energy and heat recycled from other processes into the current district heating systems, problems occur due to the high temperature levels in the systems. There are many reasons to why the temperature levels are high, but one frequently occurring issue is that the return temperatures of the customer installations are undesirably high due to faults in the substations or internal heating systems of the installations. This cause the return temperature of the entire district heating system to increase. Therefore, it is of great importance to be able to identify the customer installations with high return temperatures and eliminate the faults that are causing the high temperature levels.

The focus of this thesis has been to investigate what fault handling processes are currently being employed by district heating utilities to detect customer installations containing faults, and what could be done in addition to improve these. The first aim was to gather experiences from the existing district heating systems about how the utilities are working with their customers to decrease the system temperatures, and what faults are currently the most common ones. The second aim of the thesis was to investigate if the customer data that is available from the customers could be used to detect when and where a fault occurs. This was of interest since the large amount of customer installations in the district heating systems makes it hard to manually survey the performance of all installations. Therefore, automatic fault detection methods are needed.

The overall results show that the most important aspects of the faults handling process is to clearly involve the customer in the process and for the utilities to gain physical access to, and mandate to fix faults in, the customer installations. The results also show that the fault detection methods developed in the thesis are able to detect installations containing faults and that they may be a good addition to the already existing fault handling processes to simplify the work at the utilities.

Key words
District heating, Fault detection, Decreased system temperatures, Customer installations

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language
English

ISSN and key title
ISRN: <LUTMDN/TMHP-19/7104-SE>
ISSN: <0282-1990>

ISBN
978-91-7753-964-3 (print)
978-91-7753-965-0 (pdf)

Recipient's notes

Number of pages
126

Price

Security classification

Signature ______________________ Date 2018-12-21

Fault handling processes in district heating customer installations
Current and future solutions

by Sara Månsson

Lund University

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

Faculty of Engineering, Department of Energy Sciences

ISBN: 978-91-7753-964-3 (print)
ISBN: 978-91-7753-965-0 (pdf)
ISRN: <LUTMDN/TMHP-19/7104-SE>
ISSN: <0282-1990>

Printed in Sweden by E-husets tryckeri, Lund University, Lund 2019

Dedicated to my family
Lars – Marianne – Olof – Emma
Victor

List of publications

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

I **Fault handling in district heating customer installations: experiences from Swedish utilities**

S. Månsson, P-O. Johansson Kallioniemi, M. Thern, T. Van Oevelen, K. Sernhed
Submitted to Energy

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

S. Månsson, K. Davidsson, P. Lauenburg, M. Thern
Submitted to Energies

III **A machine learning approach to fault detection in district heating substations**

S. Månsson, P-O. Johansson Kallionimei, K. Sernhed, M. Thern
Energy Procedia, Volume 149, pp. 226-235

All papers are reproduced with permission of their respective publishers.

Acknowledgements

This work was financed by VITO in Belgium.

I would like to thank my supervisors for supporting me throughout the work and for always giving valuable comments and input to my work. I would also like to thank my PhD colleague Lisa for helping me understand that life as a PhD is not always about perfection - good enough is the new perfect.

I would also like to direct a massive *Thank you!* to my family and especially my boyfriend Victor. Thank you for always supporting me and having my back no matter what.

Contents

Chapter 1

Introduction

1.1 Background

District heating (DH) has been identified as an important part of the future energy systems and is currently one of the main heating alternatives in many countries. However, a large share of the international DH today is supplied by fossil fuels [1]. This is of course not desirable in the future energy systems where the main focus will be to utilize renewable energy sources and other innovative solution such as heat recycling. Some countries have already been very successful in supplying their DH systems with a large share of non-fossil fuels. For example, the main share of the heat supplied for DH in Sweden in 2017 was heat recycled from other processes (54.5 %), followed by heat from renewable sources (38.9 %) [2]. This clearly shows that there is already potential to phase out the fossil fuels being used in the DH systems.

The heat sources of the future energy systems poses a challenge to the existing DH systems. The temperatures of which the heat from renewable energy sources and heat recycled from other processes is delivered with is usually significantly lower than the temperatures of the heat delivered from fossil fuels. This means that the many of the DH systems are currently operated with temperatures that make it impossible to integrate other heat sources. However, systems that have been successful in integrating the new heat sources are also experiencing issues with higher than necessary system temperatures. This shows that there are other issues causing higher DH system temperatures.

There are many reasons to why the temperature levels of the DH systems are higher than they potentially could be, but one very important aspect is that many of the customer installations are delivering high return temperatures to the system. This is due to faults in the customer's internal heating system, or faults in the customer substation which transfers the heat from the DH system to the internal heating system. High return temperatures makes it impossible to decrease the supply temperatures, and so the installations containing faults need to be identified so that the overall system temperatures can be decreased.

Since there may be several thousands of customer installations in a DH system, it is important to be able to identify the installations containing faults in a methodical and efficient way. Therefore, it is of great interest to investigate how the DH industry is currently working to identify customer installations containing faults, as well as identifying what can be done to further improve this process. One way to do this is to use customer data to identify the faulty installations using different automated methods.

1.2 Objectives

The focus of this thesis is to investigate the current fault handling processes, as well as investigating how data analysis may be used to identify faulty installations automatically. Part of this work is to investigate already existing fault handling processes and work procedures that lead to successful identification and elimination of faulty installations. The aim is also to identify and implement different automated methods for detection of customer installations containing faults.

1.3 Limitations

There are different ways of connecting the internal heating system to the district heating system. In some countries, it is practice to have what is called a direct connection with no hydraulic separation between the internal heating system and the DH systems. Another solution is to create a hydraulic separation using one or more heat exchangers to transfer the heat to the internal heating system. This is called an indirect connection [3]. In this thesis, an indirect connection has been assumed. Furthermore, the studies in the thesis have been conducted in Sweden with a Swedish DH context in mind. Therefore, some of the theories and principles presented in the thesis are written from a Swedish perspective.

1.4 Outline of the thesis

The first chapter of the thesis presents some basic information about district heating which is included to provide the reader with some background information that is needed for the subsequent chapters.

Chapter 3 presents the DH customer installations, the components of the substation and internal heating system and how the system is controlled to deliver enough heat to the customer. The chapter also includes information about how metering is normally performed in the installation, and how and when the heat demand of the customer occurs.

The next chapter describes the importance of lowering the current system temperatures: the benefits associated with lower temperatures, what issues currently lead to higher system temperatures, and what may be done to avoid higher system temperatures in the future.

Chapter 5 presents a number of data analysis methods that may be used for fault detection of different processes. The chapter also includes an overview of fault detection methods that have been developed for the application of fault detection of DH customer installations, and is concluded with information about the methods that have been used in the papers included in this thesis.

Chapters 6 and 7 provide information about the methods that was used and the main results that were obtained in the papers included in this thesis. This is followed by a chapter containing some concluding discussion and suggestions for future studies within the field. The papers are appended at the end of the thesis.

Chapter 2

District heating

The following chapter aims to describe the basics of district heating systems and how the technology that is used today has been developed. A brief introduction to the future technology of district heating is also provided. The chapter also includes information about how the customers connected to the district heating systems are using heat, and when, where and why heat demands occur.

2.1 The fundamentals of district heating

District heating is a way to deliver heat from available heat sources to a location where heat is needed [3]. The heat source is utilized to heat water or steam, which is transported to the customer location in insulated, pressurized pipes. The distribution pipes form a pipe network that delivers heat to a number of different customers while allowing several heat sources to be connected to and deliver heat to the DH system. Depending on what heat sources are available in the proximity of the system, a multitude of different fuels and heat sources may be utilized, for example, waste heat from industrial processes, heat produced in combined heat and power (CHP) production, and heat from Waste-to-Energy (WtE) plants [1].

At the production site, the distribution water is heated to the supply temperature, T_s. When the hot water reaches the heat customer, the heat is transferred to the customer's internal heating system via a district heating substation which cools the distribution water to the return temperature, T_r. In this thesis, the entire customer system, including both the internal heating system and the DH substation will be called customer installation. The cooled water is then

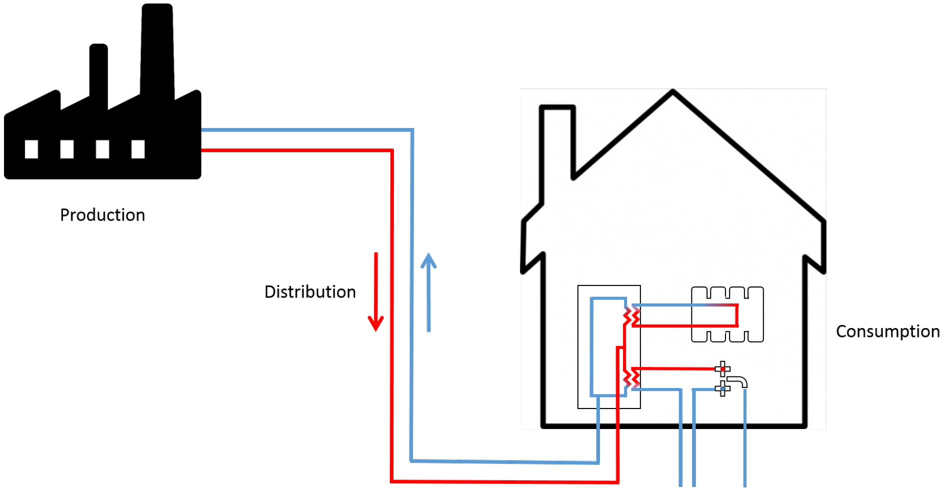


Figure 2.1: Schematic illustration of a district heating system

circulated back from the customer installation to the heat source where it is reheated before it is circulated into the distribution system again. A simplified illustration of a DH system can be seen in figure 2.1, including a simplified illustration of the substation and internal heating system of the customer installation. Roughly, the DH system can be said to be divided into production, distribution and consumption. The company, municipality or organization in charge of the DH system will in this thesis be referred to as a DH utility.

2.2 The development of district heating

The technology of district heating has developed continuously since it was introduced at the beginning of the 19th century. A common theme throughout the development has been the aim to decrease the overall system temperatures, while still being able to deliver heat to the customers without risking the security of supply. It has also been important to work towards material lean components and prefabrication, leading to reduced manpower requirements at construction sites [3].

The development towards lower temperature levels is strongly associated to different generations of how the DH systems were built, and the historical development of the systems is often said to have happened in three different generations [3]. The first generation of district heating utilized steam to distribute the heat to the customers. The heat was transported in steal pipes in concrete ducts and significant heat losses occurred in the distribution system due to the high temperature of the steam. In the second generation, the heat medium was still transported in pipes in concrete ducts, but instead of steam, pressurized hot water with supply temperatures mostly above 100 °C was used. This generation was also characterized by the use of tube-and-shell heat exchangers. In the third generation of district heating, which is the technology generation most commonly installed today, the heat medium is still pressurized hot water, but the supply temperatures are most often decreased below 100 °C. The characteristics of the third generation include pre-insulated pipes directly buried in the ground and compact plate heat exchangers [3].

In order for district heating to continue being a competitive, strong alternative on the heat markets, it is important that the technique continues to develop and that the temperature levels are decreased even further. The continued development has been defined as the 4th generation of DH (4GDH). In 4GDH, it is desirable to reach distribution temperatures of 50 °C/20 °C for the supply/return temperature [4]. The decreased system temperatures would allow for less heat losses in the DH systems, and for more low-temperature waste heat to be introduced into the systems. It would also be possible to introduce more prosumers, which are DH customers that also produce (comparatively small) district heating outputs [5].

2.3 Temperature levels in current district heating systems

The temperature levels of the district heating systems varies over the year since the customers' heat demands varies over the year. The supply temperature varies with the outdoor temperature and is decided by the heat provider. It varies from system to system depending on the pipe sizes in the system or if some customers have specific, high heat demands [3]. The return temperature is the aggregated result of the cooling performance of each individual customer located in the system [3]. Hence, the return temperature will never be lower than the temperature levels that the customers deliver to the system. Figure 2.2 displays an example of the supply (orange) and return (blue) temperature for a customer installation. As can be seen in the figure, the supply and return temperature levels are significantly higher than the desired temperature levels of the 4GDH systems. This can be seen as a general remark for the DH systems of today, since most systems operate with significantly higher temperatures than 50/20 °C. For example, the average temperatures of the Swedish DH systems

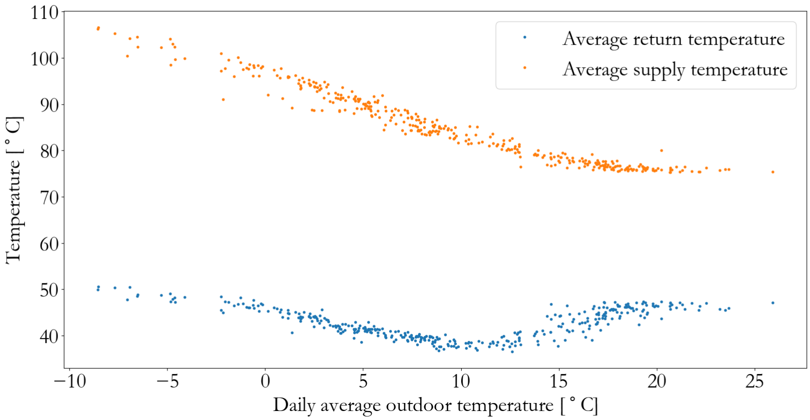


Figure 2.2: Supply and return temperature for a DH customer installation

during the last years have been 86 °C/48 °C for supply/return temperature [6].

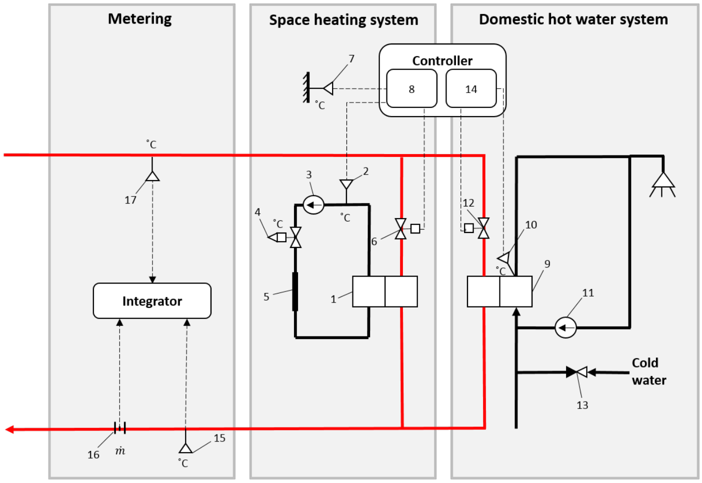
Chapter 3

District heating customer installations

The heat consumption in a DH system consists of two parts: space heating consumption and domestic hot water (DHW) preparation [3]. These two services are provided using two separate systems inside the building which both are supplied by district heating. The internal heating system and the DH system are hydraulically separated by a district heating substation. Depending on where the DH system is located, the heat demand from the customers' installations vary over the year due to varying outdoor temperatures and different social behaviours.

3.1 District heating substations

The district heating substation is the component that connects the customer's building to the district heating system. The substation transfers heat from the district heating system to the customer's internal heating system. The substation design and connection principle may vary depending on the heat demand and size of the building, as well as what the building is used for. The two most common connection principles are the parallel-connected and the 2-stage connected substation [7]. In the parallel-connected substation, there are two heat exchangers connected between the supply and return pipes, one for DHW and one for space heating. In the 2-stage connected substation, one heat exchanger is connected to the space heating system and two heat exchangers are connected to the domestic hot water system. In this solution, one heat exchanger preheats



Space heating system	
1	Heat exchanger
2	Supply temperature sensor
3	Circulation pump
4	Thermostatic radiator valve
5	Radiator
6	Control valve
7	Outdoor temperature sensor
8	Space heating control
Domestic hot water system	
9	Heat exchanger
10	Supply temperature sensor
11	Circulation pump
12	Control valve
13	Non-return valve
14	DHW control
Metering	
15	Return temperature sensor
16	Flow meter
17	Supply temperature sensor

Figure 3.1: Simplified illustration of a parallel-connected district heating substation

the incoming cold water using the return water from the space heating system before the final preparation of hot water takes place in the second heat exchanger [8]. Figure 3.1 displays a simplified illustration of a parallel-connected substation. The illustration also includes the components that are part of the customer's internal heating system. The substation components are usually described to include the heat exchangers, cirulation pumps, control valves, temperature sensors, controller, and heat meter which consists of supply and return temperature sensors and the flow meter.

In the figure, the substation and internal heating system components have been divided into three separate groups depending on where in the customer installation they serve their purpose: metering, space heating system, and domestic hot water system. These three parts, and the components of each part, will be described in detail in the following sections.

3.2 Space heating system

The space heating system consists of a number of components and the system design may vary in many different ways depending on what type of building the system is located in. The substation design in figure 3.1 displays the basic design of a space heating system using radiators to heat the room. Radiators are the most common type of room heaters and can be described as heat exchangers which exchange heat between the heated water and the air in the room which is to be heated [9]. The design temperatures of the radiator systems have traditionally been quite high with supply/return temperatures of 90/70 °C or 80/60 °C. Today, the systems are are designed for lower temperatures (e.g., 60/45 °C, 60/40 °C or 55/45 °C) [10]. To decide what space heating capacity to design the building system for, the design outdoor temperature is used. This is the lowest outdoor temperature to be considered when designing a particular building's heating system [3].

Some space heating systems also include ventilation. When this is the case, the incoming air will have the same temperature as the outdoor temperature. Therefore, the ventilation system requires an amount of heat to heat the incoming air to room temperature. This heating may be done using different solutions, e.g. by reusing heat from the air that is transported out of the building, or by heating the air using district heating [9].

In figure 3.1 the heat exchanger for the space heating system (1) is controlled by the control valve (6), which controls the amount of water that passes through the heat exchanger. The position of the valve depends on the outdoor temperature which is measured by an outdoor temperature sensor (7). The signal from the outdoor temperature sensor is connected to the controller (8).

When the outdoor temperature decreases, the supply temperature in the space heating system will have to increase to meet the higher heat demand of the consumer. The supply temperature will change according to a predetermined, but modifiable, control curve, which depends on the desired indoor temperature level [3, 11]. The increased temperature demand is met by changing the the opening position of the control valve so that the flow that passes through the heat exchanger at the DH side increases. To make sure that the correct supply temperature is achieved a temperature sensor (2) is mounted on the supply pipe of the space heating system, and the signal from the sensor is used in a control feedback loop which determines the position of the valve [12].

The water in the space heating system is circulated using a circulation pump (3). The flow in the radiators (5) is controlled by thermostatic valves (4), which measure the room temperature and opens or closes depending on how much heat is needed to meet the temperature demand in the room [9]. The heat demand in a room varies depending on several different factors, i.a. the desired indoor temperature, the amount of people and electrical appliances in the room, and the amount of heat contribution from the sun [3]. Therefore, different rooms require different amounts of heat.

When a new building is taken into use, it is important to properly balance the internal heating system. This balancing is done in order to have an even distribution of hot water in the radiators to satisfy the heat demand of every room in the building. In a hydronic system, this means that the valves are adjusted so that the flow in the radiator circuits located closest to the heat source is throttled, forcing the hot water to the more remote radiator circuits [13]. If the system is not in balance the supply temperature of the space heating system will have to increase. Otherwise, the system will not be able to deliver the required amount of heat to the peripheral rooms of the building [14]. However, the increased supply temperatures also lead to increased return temperatures from the internal heating system which is not desirable in the DH systems. This means that it is of great importance that the internal heating system is properly balanced when the building is connected to the DH system.

3.3 Domestic hot water system

The design of the system for preparation of domestic hot water may be slightly different depending on what country the building is located in, as well as how big the building in question is. The basic design includes a heat source where the water is heated, pipes where the heated water is transported, and the fixtures (e.g., valves and faucets) that control the flow of water in the system [9].

In figure 3.1, the flow of the DH water through the heat exchanger used for DHW preparation (9) is controlled by a control valve (12). The position of the valve is decided by the temperature of the water in the DHW system, which is measured by a temperature sensor (10) located close to the heat exchanger [15]. The design of the DHW control system is similar to the one used for space heating, but the temperature in the DHW system is not determined by the outdoor temperature. Instead, the control parameter is a setpoint value determined in the controller (14).

The setpoint value is determined according to certain temperature demands that have to be met for the domestic hot water. Different countries have different demands, but in Sweden the demands are that the temperature at each tap cannot be below 50 °C, due to the risk of bacterial growth of Legionella, or above 60 °C due to the risk of scalding [16]. Hence, the setpoint value is determined to comply with both of these demands.

Cold water is delivered to the building from the municipal water feed system and is then heated in the heat exchanger. Today, the most common solution when using district heating for DHW preparation is that the preparation occurs instantaneously in the heat exchanger. This means that there is an unlimited supply of hot water in the building as long as the DH system is working as it should [3]. The heated water is prevented from accidentally flowing backwards into the drinking water system by a non-return valve (13).

In the DHW system, it is important to have hot water available as soon as the heat demand occurs without delays, even when no DHW consumption occurs [3, 15]. Therefore, it is common practice to install a circulation pump (11) connected to the incoming cold water pipe when DHW preparation takes place in a larger building. This is done to make sure that hot water is available in the entire system at all times. When a circulation pump is installed in the DHW system, it may be possible to control the pump using the supply temperature to determine when to run the pump. By only running the pump when the supply temperature approaches the lower temperature limitation, it is possible to limit the energy losses of the system [9].

If no circulation pump is installed, the water will be standing still in the system, and the water temperature will decrease over time due to heat losses to the surroundings [9]. In systems where this is the case, the supply temperature sensor in the DHW system should be placed as close to the heat exchanger as possible. This is important to ensure a stable control of the DHW supply temperature level [3, 8, 7].

3.4 Metering in the substation

The integrator in figure 3.1 is part of the heat meter of the substation. The integrator uses the values measured by the two temperature sensors (15) and (13), and the flow meter (14) to calculate the amount of heat delivered to the internal heating system during a certain period of time. Temperature sensor (15) measures the supply temperature of the incoming district heating water, temperature sensor (13) measures the return temperature of the water being circulated back to the district heating system, and flow sensor (14) measures the mass flow rate of the recirculated water.

The frequency of which the meter estimates and updates the heat consumption using the integrator varies depending on the meter being used. Some meters do the estimations at regular time intervals, while others make the estimations depending on the flow rate [17]. The integrator is often able to calculate the amount of flow that passes through the system during the measuring interval, which supplies the DH utilities with the possibility to analyze how much DH water that passes through their customers' installations [3].

The supply and return temperatures are most commonly measured as instantaneous values, which means that they are not the average over a time period. This makes the temperature values less representative of the actual temperature levels than if they would have been an average for the temperatures measured during the metering interval [18].

Today, metering may occur on different levels in the customer installations. In apartment buildings, it is possible to perform metering in the individual apartments, or at an aggregated building level [3]. This provides different possibilities for the DH utilities in terms of billing and/or analysis of the meter reading. In buildings with metering on an apartment level, each customer is charged according to the actual heat being used in the apartment. However, this may cause inequity among the customers in the same building depending on where in the building the apartment is located. An apartment surrounded by other apartments will receive heat transmitted from the surrounding apartments, and may therefore use less heat [19]. However, individual metering would give greater insight into how the heating systems of each individual apartment is working compared to when the metering is performed at an aggregated building level.

In most modern industrial countries, heat metering is statutory [3], and according to the energy efficiency directive that became effective in 2012, DH customers should be charged according to the actual amount of heat that they use [20]. This means that the DH utilities are obliged to gather customer data in some way, and many modern meters contain features that allows for meter readings to be collected automatically to the DH utility. The collection may occur via optical cables or using wireless communication tools such as transmission over the mobile net or via the Internet [3]. The metering data is collected at the utilities and processed in different kinds of data management systems.

As the metering and transmission technologies improve, it becomes possible to increase the number of measuring instances that is collected by the utilities. Many utilities today collect meter values on an hourly basis, which allows for more detailed analysis of how the customers are using the heat and how well the customer installations are performing. The improved metering also allows for more different meter values to be collected at regular intervals, which is beneficial when performing analysis of the customer data.

3.5 Customer heat demands

The heat demands of individual customers depends on a number of different factors, including what the purpose of the building is, what space heating system is installed, and the size of the building. A general conclusion is that space heating is needed to maintain a comfortable indoor climate when the outdoor temperature decreases below the so-called balance temperature of the building. For temperatures above the balance temperature no space heating is needed since the desired indoor temperature is reached nonetheless. However, the heat demand of the building will most likely not be equal to zero, since there will still be a need for DHW preparation[3].

This means that for temperatures below the balance temperature DH will be used for both space heating and DHW preparation, while DH will only be used for hot water preparation for temperatures above the balance temperature. When considering this on the aggregated DH system level, the typical balance temperature of the buildings in the system determines when the heating season starts and ends. The temperature for which this occurs is normally called the threshold temperature [3].

The heat demand related to DHW preparation is in many buildings rather constant during the year. However, daily fluctuations occur and in many cases it is also possible to see an hourly fluctuation. The fluctuations between days and during the day occur due to different social behaviours, e.g., that clear morning peaks occur since many people are getting ready to go to work at the same time [21].

A clearer seasonal dependence occur for the space heating demand. As mentioned, space heating is normally only needed during the heating season and the heat needed to heat the rooms in a building will vary with the outdoor temperature. This means that the heat power delivered to the building shows a clear temperature dependence, as can be seen in figure 3.2 which displays the heat consumed in a building located in Sweden as a function of the outdoor temperature. The values in the figure are the daily average heat power demand for the building during one year. This relationship is often called the heat power signature of the building [3].

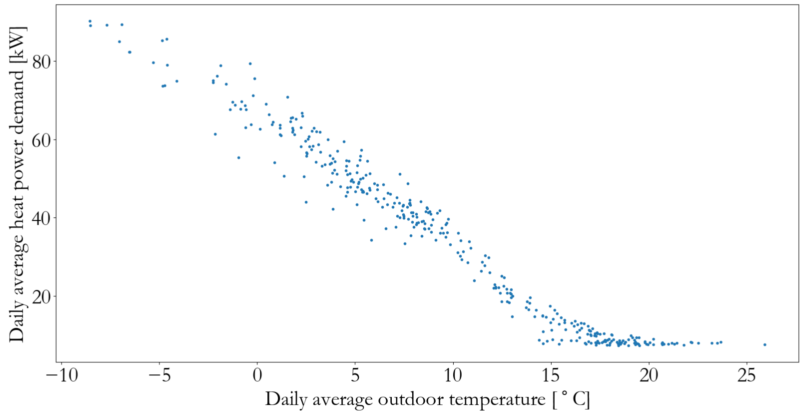


Figure 3.2: Heat power as a function of outdoor temperature

Further, the heat demand also varies over time due to the fluctuations in both the space heating and DHW demand. This dependence is illustrated in figure 3.3, which displays the hourly heat power demand for the same building as in the previous figure. As can be seen in the figure, the heat power demand is largest during the Swedish winter months, and decreases significantly during the summer months. In the figure, some values are equal to zero during days when there should be a heat demand in the building. These values originate from instances when the wireless connection between the heat meter and the utility was lost when trying to collect the values. When this occurs, the DH utility in question replace the missing values with zeros.

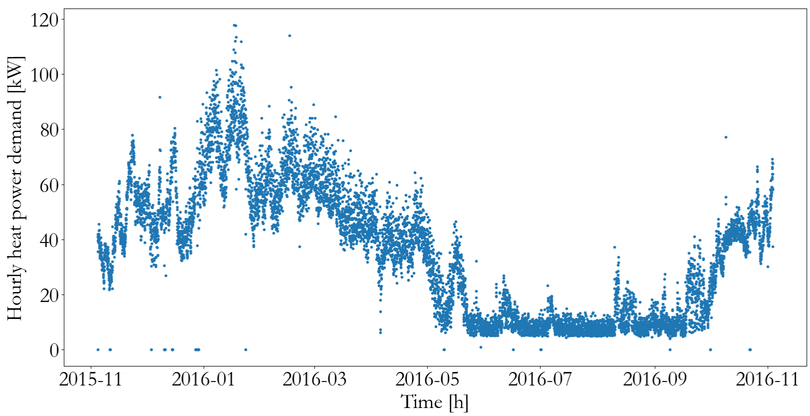


Figure 3.3: Heat power as a function of time

The heat demand of a building also varies depending on what type of customer is using the heat. Different types of customer requires heat at different occasions, which is clear when investigating the pattern of hourly heat power data during one week. Most dwellings have a heat pattern that is the same throughout the week with clear morning and evening peaks since this is when the customers are at home and have the largest heat demand. These buildings also have a rather constant space heating demand during the week [3]. The heat demand patterns of office buildings and schools often deviate from the patterns of dwellings, since many of these buildings have a ventilation system that is shut off during the night and/or weekends when no one is present in the buildings. This gives a clear difference in the heat demand pattern between weekdays and weekends, and also between day and night [3]. All of these variations arise due to the individual and collective social behaviors such as working hours which affect when people get ready for work in the morning, and when ventilation is needed in office buildings [21].

Irrespective of the type of building, the heat power delivered to the heat customer can be expressed using the following equation:

$$P_d = \dot{m} \cdot c_p \cdot (T_s - T_r) \tag{3.1}$$

where

P_d	= heat power	[J]
$\dot{m}$	= mass flow rate	[kg/s]
c_p	= specific heat capacity for water	[J/kg°C]
T_s	= supply temperature	[°C]
T_r	= return temperature	[°C]

Hence, the heat power depends on the mass flow of water in the pipes as well as the temperature difference between the supply pipe and the return pipe. This temperature difference is often called the cooling, or the delta T, (ΔT) of the substation. As can be seen in the equation, ΔT will decrease if the mass flow rate increases for the same heat power demand. Since the supply temperature will not change, a decrease of ΔT will lead to high return temperatures. It is also clear that if the cooling of the installation is insufficient, the mass flow rate through the substation will have to increase to meet the customer's heat demand. The cooling of a well performing substation should be largest in the winter and decrease with increasing outdoor temperatures [18]. Figure 3.4 displays the cooling pattern for a customer installation during one year.

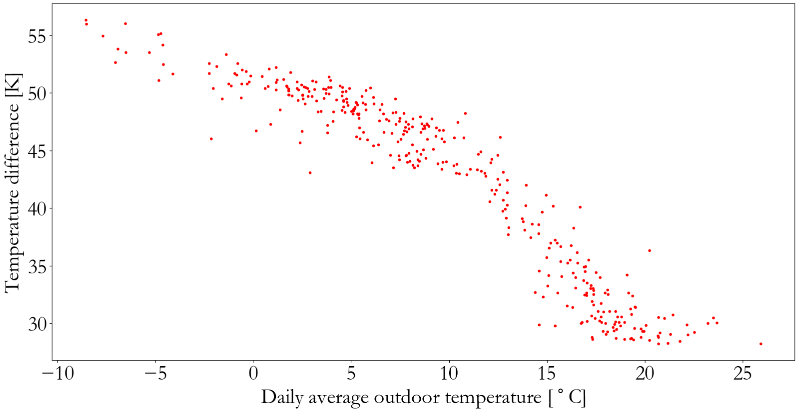


Figure 3.4: The cooling pattern for one DH customer installation

The relationship displayed in equation 3.1 is important to keep in mind when reading the next chapter of this thesis, which presents a number of issues that lead to higher flows and/or higher return temperatures in the customer installations and the importance of eliminating these issues to be able to reach lower system temperatures.

Chapter 4

Importance of lowering the system temperatures

Several studies have shown that the current temperature levels of the DH systems are higher than they theoretically could be [3]. As mentioned in section 2.3, the Swedish average temperature levels are 86 °C/48 °C. However, the theoretical temperatures that could be achieved from the existing substation technology is 70 °C/32 °C supply/return [18]. This clearly shows that it should be possible to decrease the system temperatures already in the existing DH systems, and not only in the future 4GDH systems. Despite this, many systems are operated with significantly higher system temperatures which depends on a number of reasons related to both the customer installations, and the DH system itself.

The following chapter presents the benefits of lower system temperatures in the existing and future DH systems, as well as the reasons to why the system temperatures currently are higher than they theoretically could be. The chapter is concluded with a review of measures that could be implemented in the current systems to decrease the system temperatures.

4.1 Benefits of lowering the system temperatures

There are a number of different reasons to why it is important to have lower system temperatures in the DH systems, and the decreased temperatures may have benefits in all parts of the system. Generally, the main focus is to decrease the return temperature levels primarily, since this enables a decrease of the supply temperature as well [3]. However, some benefits arise due to decreased return temperatures, some due to decreased supply temperatures, and others from a combination of the two [22].

Some of the main advantages related to decreased temperature levels appear in the production sites. The CHP plants benefit from lower supply temperatures since the power-to-heat ratio increases due to the increased power production in the plant that occurs when the temperature levels (and hence pressure) in the condenser is lower [22, 23]. Since electricity usually generates higher revenues than heating, the business case for CHP increases significantly. Systems where flue gas condensation is installed benefits from decreased return temperatures due to increased efficiency in the condensation process [24]. The coefficient of performance (COP) of large-scale heat pumps also increases when the distribution temperatures decreases [25].

There are also a number of heat sources available today that the DH systems cannot utilize due to the high system temperatures. An example of this is that many industrial processes generates large amounts of waste heat of temperature levels that are lower than the current DH system temperatures [4]. This means that in order to be able to use these heat sources in DH systems with high distribution temperatures, the temperature levels of the heat sources have to be increased [26]. This also applies for waste heat from, e.g., sewage treatment plants [22]. Other heat sources that delivers heat of insufficient temperature levels, and that would be possible to utilize if the DH system temperatures decrease, include solar thermal and geothermal heat sources [4].

Decreased system temperatures are also beneficial from a heat loss perspective. Due to the decreased temperature difference between the distribution pipes and the surroundings, the heat losses will also decrease [4]. Today, relative distribution heat losses in the DH systems range from 5-25 % depending on where the system is located since different countries have different building standards and hence varying standards of insulation. The relative distribution heat losses is the relationship between the annual heat losses and the annual heat input from the heat supply units [3].

When building new DH systems, the lower distribution temperatures are also beneficial due to the possibility to use other materials for the pipes. This will lead to lower installation costs compared to installing the pipes most usually used today [22]. The main reason to that the cheaper materials are not being used today is that they cannot tolerate the large temperature variations that occur in the distribution system during the year since the material will creep at elevated temperatures [3], a problem that would be avoided with lower system temperatures.

When obtaining lower return temperatures at a given supply temperature, it will be possible to reduce the volume flow rate in the system [27]. This is due to the fact that the volume flow rate is inversely proportional to the temperature difference between the supply and return pipe, i.e., if the temperature difference decreases the volume flow rate will have to increase to be able to deliver the same amount of heat power to the customers, and vice verse [3]. Hence, one benefit of having a large temperature difference between the supply and return temperature is that it will be possible to reduce the volume flow rate in the DH system. This decreases the need of pumping in the system, which means that there will be a smaller need of electricity in the system [22].

A number of studies have investigated the economical benefits of decreasing the DH system temperatures. Werner and Frederiksen have investigated the value of reducing the return temperature in 27 DH systems in Sweden [3]. The results show that the cost reduction varies significantly between the different systems, but that the typical value of reducing the return temperature in these systems is 0.15 EUR/(MWh,°C) [3]. Hence, a DH system with an annual heat sale of 1 TWh would save approximately 1.5 MEUR if the return temperature is decreased with 10 °C [27].

4.2 Issues resulting in high system temperatures

There are number of reasons to why the distribution temperature levels are higher than they theoretically could be. One is due to the tradition of using fossil fuels with high heat values, which are capable of delivering heat at high temperatures [28]. However, systems which do not utilize large amounts of fossil fuels in their production also experience higher than necessary temperature levels. This is due to a large number of issues in the DH distribution system which can be categorized as follows: (i) short-circuit flows between supply and return pipes in the system, (ii) low supply temperatures at peripheral substations because of high heat losses, (iii) faults in the customers' internal heating systems, and (iv) faults in the customers' substations [3].

The first issue, short-circuit flows between the supply and return pipes, occurs due to bypass circuits in the system. In the summer, when no space heating is needed, the water in the supply pipes will cool off if no tapping of hot water occurs. This leads to system temperatures which reduce the ability to prepare DHW instantaneously in the heat exchanger [29]. To avoid this problem, bypass valves between the supply and return pipes are commonly installed in the substations and/or other locations of the distribution system. When no space heating demand occurs the valves will open, allowing a small amount of supply water to flow into the return pipe and creating a circulation in the system. Although the bypasses are needed to be able to prepare DHW, they lead to increased return temperatures [30]. The second issue, low temperatures at the peripheral substations because of high heat losses, occurs when the heat demand and flow of the system is low and no bypasses are installed in the system. When the supply temperature decreases due to the small flow, the flow through the substation will have to increase to deliver the desired amount of heat. This leads to higher return temperatures [3]. Issues (iii) and (iv) relate to the customers' installations and will be treated in detail in section 4.3, since they are the main focus of the work conducted in this thesis.

4.3 Issues in customer installations resulting in higher system temperatures

The customer installations are an extremely important part of the DH systems, and governs both how much heat the systems have to be able to deliver and what temperature levels the heat should be provided with. As described in section 3.2, the internal heating systems have traditionally been designed for high supply and return temperatures and the DH temperatures have to match these temperatures to be able to deliver heat to the buildings. Further, equation 3.1 in section 3.5 clearly described the importance of having customer installations with a large ΔT to avoid an increase of the mass flow rate in the system. This relationship also applies for the opposite case: a large mass flow rate cause a smaller ΔT which results in higher return temperatures.

As described in section 4.1, decreased system temperature levels and mass flow rates entails a number of benefits for the DH system. Nonetheless, faults in the customer installations have been shown to cause both increased return temperature levels and increased mass flow rates. This indicates that as long as the customer installations contain these faults, it will be very hard to obtain the desired effects of lowered system temperatures.

An issue that complicates the elimination of the faults in the customer installations is that there is a multitude of different faults that may occur in many different parts of the customer installation. Some of the faults are further very hard to detect, especially the faults that do not impact the indoor climate and customer comfort significantly [31]. Faults that affect the customer comfort will be easier to detect, since the customer will become aware that something is wrong quicker if the space heating or preparation of DHW is not working as it should. The customer will, e.g., not notice if something is faulty in the heat meter, since the sensors connected to the meter are only used for heat metering purposes and do not affect the control of the substation [32]. However, the DH utility is affected by the fault since their metering data gets disordered.

The following sections introduce a number of faults that may occur in different parts of a customer installation. The faults have been divided into five different categories, depending on where in the installation a fault occurs. The first fault category, *Heat exchangers*, include faults related to the heat exchangers in the substation, and the pipe connections to the heat exchangers. The second category is *Controller and control system* which in this thesis includes faults in the controller itself, the temperature sensors measuring different temperatures, the wires from the sensors connected to the controller, and the software of the controller. The third and fourth categories are *Control valves* and *Actuators*. These categories are closely related to each other and to the control system, but have been treated as two separate categories since there is a number of different faults that may occur for both actuators and control valves. The last category describes faults that appear in the customers' internal heating systems.

4.3.1 Heat exchangers

One of the faults that occurs most frequently in literature is fouling of heat exchangers, where a deposit has formed on the heat transferring surfaces of the heat exchanger. The precipitation decreases the heat transferring ability of the heat exchanger and reduces the flow velocity through the heat exchanger [33]. This lead to poor heat transfer in the substation. However, moderate fouling may also increase the heat transferring ability of the heat exchanger, but only at low flows [34].

Leaking heat exchangers is another fault that occur with high frequency [32, 35, 36]. Leaking happens if, e.g., the seals between the heat exchanger and the pipes are not tight enough or if a packing shrinks due to exposure to excessive temperature variations [36].

There have also been occasions when the heat exchanger has been connected co-current instead of counter current [37, 38]. When it is connected co-current, the heat medium and the water in the internal heating system flows in parallel. When this is the case, the temperature of the supply water in the internal heating system will never exceed the temperature of the DH water that leaves the heat exchanger, i.e. the DH return temperature. In a counter current heat exchanger the flow in the internal heating system and the flow of DH water go in opposite directions through the heat exchanger which significantly improves the heat transferring ability in the heat exchanger. This enables the water in the internal heating system to reach temperatures very close to the temperature of the incoming DH water [39].

4.3.2 Controller and control system

The controller and control system as defined in this thesis may contain a number of different faults which are related to both the physical components and the control software. The controller itself may break down, leaving the system completely uncontrolled [36]. This may occur due to power outages or lightning strikes. There might also be faults related to how the controller is mounted when installing it, if the different wires are connected to the controller incorrectly. For example, the wire from the temperature sensor measuring the supply temperature in the space heating system may be confused with the wire from the temperature sensor measuring supply temperature in the DHW system. This interrupts the entire control system since the wrong information is delivered to the wrong location in the controller [37].

Problems related to the temperature sensors in the control chain occur quite often. One frequent fault is defective temperature sensors that occurs due to, e.g., bias changes or noise in the measurements [3, 32]. Other temperature sensor faults are temperature sensors placed on the wrong pipe delivering an incorrect signal to the controller, sensors being knocked down from their position, and sensors mounted loosely on the pipes where it measures the temperature which may cause large changes in the measurement signals [32, 37, 38].

As mentioned in section 3.3, the temperature sensor in the DHW system has to be placed close to the heat exchanger if no DHWC is installed. Without DHWC, the supply temperature in the DHW system will decrease over time, especially for the water located far away from the heat exchanger. If the temperature sensor is placed far from the heat exchanger the measured supply temperature will consequently be low. If the temperature falls below the set point value determined in the controller, the system will experience that there is a need for DHW preparation. However, since the water is standing still in the DHW system, it will take longer for the temperature front of the heated water to reach the temperature sensor. This creates a time lag in the system before the new supply temperature is measured correctly, causing the controller to overcompensate and increase the DHW supply temperature too much [40]. This will also cause the flow through the exchanger to be high, causing high return temperatures [41].

One control system fault related to the control of the DHW system, is that the set point values of the desired temperature levels are too high. If they are close to, or maybe even higher than the DH system temperatures, the control valve on the DH side of the heat exchanger will continuously stay in an open position since the controller will try to reach the desired temperature. However, this will not be possible and the consequence will be high mass flow rate and high return temperature from the installation [31, 42].

There is also control system faults related to the control of the space heating system. One common case is that the outdoor compensation curve in the controller is changed by the customer. The controller is easily accessible to the customer, and if the customer experiences that the space heating system is not delivering enough heat one first instinct may be to increase the desired indoor temperature level in the controller [43, 11]. However, this may not lead to any improvement since the outdoor compensation curve is not the fault in most cases - instead, other faults are most commonly causing the issue of poor heat delivery. The only consequence of the change of curve settings is that the return temperature increases [43].

4.3.3 Control valves

The control valves in the customer installation can either be mechanically self-operated or motorized [3]. The difference between the two is how the valve is controlled. The motorized valves are controlled by an external control signal sent to an actuator, and the self-operated valves are controlled by a self-control mechanism [44]. However, both kinds of valves contains the same components. A simplified illustration of a motorized valve can be seen in figure 4.1. As can be seen in the figure, the control valve consists of a body and bonnet which encase the inner, moving parts of the valve. The opening position of the valve is decided by the spindle and at the end of the spindle a disc is located, which is the part that restricts the flow through the valve. If the spindle moves into a closed position the disc gets in contact with the so-called seat of the valve, creating a leak-tight seal which do not allow any water to pass through the valve [45].

The faults related to valves are most commonly due to wear and tear in the moving parts of the valve. One common fault is that the control valve leaks in a closed position, due to that the disc and the seat do not close leak-tight

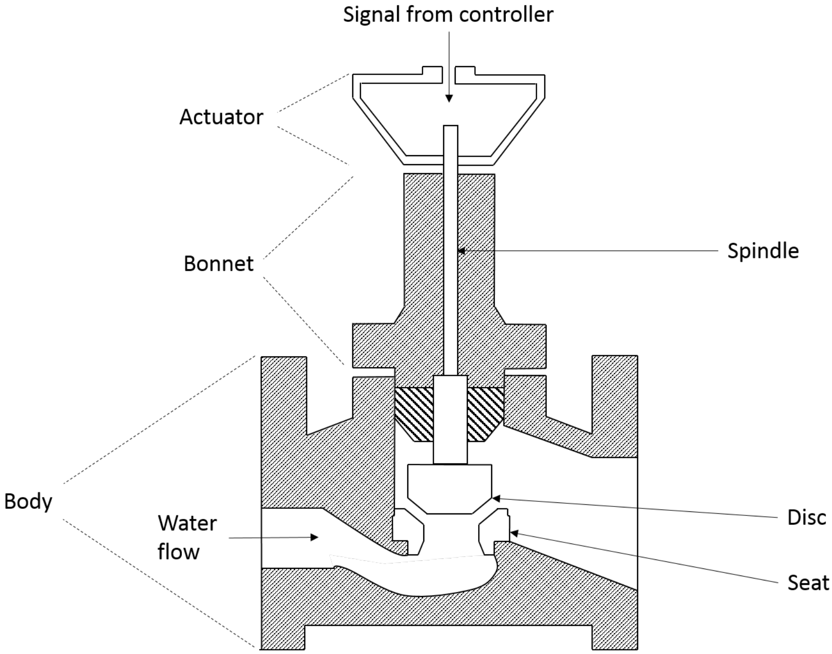


Figure 4.1: Simplified illustration of motorized control valve with actuator

after operating for some time [37, 38]. This might occur due to erosion of the valve disc, especially if there is a large differential pressure over the valve. A large differential pressure might lead to cavitation, i.e. the formation of vapour bubbles in the heating water when it passes through the valve. When these cavitation bubbles exit the valve they implode, creating water jets that may hurt the surface of the valve disc causing erosion [46].

Another common fault is that the valves are oversized [32, 31]. This means that the valve is dimensioned for a larger flow than necessary, which means that a small change of the valve's position will lead to a large change of the flow. This makes it hard to control the flow through the valve, leading to large flows and high return temperatures from the installation [37].

It is also quite common that the valves are seizing in an open or closed position, or that the valve is completely stuck [38]. These issues cause poor control of the flow in the system.

4.3.4 Actuators

Actuators are the components that determine the position of the control valves and are in turn controlled by a control signal from the controller, as can be seen in figure 4.1. The actuators transform the control signal into mechanical force and may be driven either electromagnetically or electromechanically [3].

There are some faults related to actuators that affect the installation's performance. The actuator may be broken, or due to wear and tear, it might happen that the connection between the actuator and the valve spindle is impaired [38, 37, 42]. This cause poor control of the flow in the installation.

There are also actuator faults related to the planning, design and execution of the customer installation. Actuators have different running time, depending on what valve they are controlling. The running time is the time it takes for the actuator to change the valve's position from completely closed to completely open [47]. In the DHW system, the running time should be short due to the rapid variations in heat demand and that the water should be available instantaneously which means that the valve must open and close rapidly to meet the demand. The actuator located in the space heating system should instead have a long running time since this part of the system requires a slow adjustment of the control valve to find the proper flow needed to meet the heat demand [37]. If the substation is assembled on site and not prefabricated it might happen that the actuator for the two parts of the system are mixed up, leading to a poor control of the entire internal heating system.

Another fault related to the planning and design of the installation is if the actuator is wrongly dimensioned for the pressure in the installation. If this happens, the actuator may not be able to open or close the valve due to the counteraction of the pressure [37].

4.3.5 Internal heating systems

The customer's internal heating system contains a number of different components where faults may occur. Some faults are related to the design of the installation, and these often occur in the DHW system. Some of these cannot be considered as pure faults, since they are a deliberate design choices. However, they lead to higher than necessary return temperatures.

One of these issues is the DHW circulation, or lack thereof. When there is DHW circulation or if the circulation pump is broken, the return temperature levels will increase if the substation is designed as in figure 3.1 since the circulation pipe is connected to the incoming cold water. This leads to a higher temperature of the incoming water on the customer side, causing poorer cooling performance in the heat DHW heat exchanger. As described in section 3.1, there may be a preheater installed in the DHW system. If this is the case, the circulation pipe should be connected after the preheater to increase the temperature difference over, and hence the cooling performance of, the preheater as much as possible [3]. Another issue that may increase the return temperature is if the DHW circulation cause unnecessarily large flows in the DHW system [38].

There are also many faults related to the space heating system. As mentioned in section 3.2, the design temperature levels of the buildings in a DH system may vary and may in some buildings be unnecessarily high. This cause the return temperatures to be higher than necessary due to an inherent fault that may be very hard to change [3]. It is also important that the space heating system is properly balanced so that sufficient heat reaches all room heaters in the building. However, many customer installations are not properly balanced today which cause poor cooling performance in the installations [31, 38, 42].

Another important aspect of the space heating system is that the thermostatic radiator valves are functioning as they should. However, broken radiator valves are frequently occurring in the installations, causing the flow through the radiator to be uncontrolled which lead to higher return temperatures [37, 38].

4.4 Measures for decreasing the system temperatures

As may be concluded from this chapter, there is a large need to identify and eliminate the issues that are currently leading to higher than necessary DH system temperatures. This may be done in a number of different ways, depending on what issue is in focus, in what part of the system the issue is located, and how much it will cost to eliminate the issue.

Many of the issues are related to how the current DH systems are being built. The bypass valves which are utilized during summer to maintain the supply temperature at a sufficient level are an example of this. It has been suggested that the future DH systems should not utilize this solution, but instead employ a three-pipe solution where two return pipes and one supply pipe are located in the same casing. The additional return pipe is utilized to recirculate the DH water to the production site when there is no heat load in the system [48]. This results in the same effect as when using bypass valves.

Another example is that many internal heating systems have been designed for higher temperatures than needed, as described in section 3.2. This makes it impossible to decrease the DH system temperatures below the requested temperature levels from these customer installations. Due to the development of more energy efficient buildings and internal heating systems such as floor heating and low-temperature radiators, the heat and temperature demands of installations connected to DH systems are expected to decrease during the coming years [49].

However, the larger share of the building stock will still consist of already existing buildings and the higher temperature demands that comes with them [4]. These buildings also contain many of the faults and issues discussed in section 4.3, which further increase the system temperatures and the return temperature in particular. This poses one of the largest challenges for the 4GDH technology and is therefore an extremely important aspect to consider.

Therefore, the focus of this thesis is on the faults and issues located in the customers' installations. First of all, it is important to investigate what the most common faults are and how they may be eliminated. This issue is discussed in detail in paper I. The second step is to be able to identify in what installations the faults are located. Papers II and III describe two different approaches to doing this utilizing customer data.

Chapter 5

Data analysis for fault detection

The large amount of customer data that has become available to the DH utilities since the energy efficiency directive became effective in 2012 provides the utilities with new possibilities to perform analyses of their DH systems. The customer meter data contain a lot of information about how the customer installations are performing, since the data include meter reading of heat consumption, mass flow rates, and temperature levels in the installation. Therefore it is possible to perform analysis of this data to identify the customer installations that are not working as they should.

When working with data analysis, a number of methods and approaches are available to choose from, including methods for detection of deviations in the behavior of a process. This also enables fault detection and fault diagnosis of the process; i.e., to detect when a fault occurs and to identify *what* fault is present. This is what many DH utilities are currently interested in achieving.

A number of different fault detection methods are being used by data scientists today, and the following chapter aims to provide a brief description of a number of them. Sections 5.1 - 5.4 aims to provide the reader with an introduction to some data analysis methods that are frequently used in the industry. The chapter then continues with an overview of a selection of fault detection and fault diagnosis studies that have been conducted and the methods that have been implemented. The chapter is concluded with a presentation of the specific methods that have been used to perform the work in papers II and III.

5.1 Piece-wise linear regression

Linear regression can be used to describe the relationship between two or more variables. The variable that is to be modelled is called the dependent variable and is often denoted Y. One single observation of the dependent variable is often denoted Y_i. The variable(s) that are used to model the dependent variable are called independent variables and are denoted X, or X_i for a single observation. These contain information about the behaviour of Y and are assumed to be known constants. A linear regression model also includes a number of unknown constants, which are called parameters and denoted by Greek letters. The parameters are estimated from the data and enters the model as simple coefficients on the independent variables [50]. In this thesis, the parameters are denoted as β_n, where $n = 0, 1, ...$

So called simple linear regression is one of the most common data analysis methods to investigate the relationship between a dependent variable and a single independent variable [51]. For simple linear regression, the true mean of the dependent variable, $\mathcal{E}(Y_i)$ changes at a constant rate as the independent variable increases or decreases. This relationship can be described as follows [50]:

$$\mathcal{E}(Y_i) = \beta_0 + \beta_1 X_i \tag{5.1}$$

In the equation, β_0 is the intercept with the y-axis and β_1 is the slope of the line [50]. However, the observations of the dependent variable Y_i will not all be equal to the true mean, but will deviate from $\mathcal{E}(Y_i)$ with an error ϵ_i. The random errors are pairwise independent, have zero mean, and equal variance σ^2 [50]. By introducing the error ϵ_i to the model described in equation 5.1, the mathematical relationship between the dependent and the independent variables for a simple linear regression model can be described by the following equation [52]:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i \tag{5.2}$$

The parameters β_0 and β_1 in equations 5.1 and 5.2 can be estimated in a number of different ways, the most common one being the method of least squares. This method minimizes the squared sum of the residuals, i.e., minimizes the sum of the squared distance between the observations and the sample mean [50].

When treating a data set with a large number of different observations Y_i, it may be hard to find one linear regression model which fits the entire data set well. If this is the case, it might be beneficial to divide the data set into smaller segments of data and model the relationship between the dependent and independent variable for each of the segments. This method is called piece-wise linear regression and the datapoints that divides the data into different segments are called breakpoints. If the breakpoint assumes the value H, the relationship between the dependent variable and the independent variable can be expressed using the following equation [53].

$$Y_i = \beta_0+\beta_1 X_i+\beta_2(X_i-H)\cdot I(X_i, H)+\epsilon_i,\ where\ I(X_i, H) = \begin{cases} 1 & X_i > H \\ 0 & X_i \leq H \end{cases} \tag{5.3}$$

5.2 Outlier detection

An outlier can be described as an observation that deviates significantly from the expected behaviour of the dependent variable [50]. Outliers are present in most data sets, and may occur for a number of different reasons: measurement or input error, data corruption or true outlier observation [54]. Outliers may cause problems when estimating the model parameters in equation 5.2, since their extreme values may dominate the parameter estimation completely [50]. Outliers can be handled in a number of different ways, depending on the nature of the outlier. Values that are clear deviations from the rest of the data set may be removed completely from the data set, while others that are not as distinct may be left in the data set for further analysis.

When leaving outliers in the data set, it is possible to investigate them using the mean and standard deviation of the data set. If the distance between a value in the data set and the mean of the data set is larger than a specified number of standard deviations, the value is considered an outlier [54]. If the specified number of standard deviations is denoted n, this can be described using equation 5.4.

$$|\mathcal{E}(Y_i) - Y_i| > n\sigma \tag{5.4}$$

The value of n should be decided from case to case, depending on the data set being analyzed. However, a common value of n is 3 since 99.7 % of the values should be within ±3 standard deviations from the mean if the variable is normally distributed [55]. This means that values that are further away from the mean than 3 standard deviations should be quite rare if the data set contains few outliers.

This kind of outlier detection can be done using different types of thresholds. Two different types of thresholds are commonly used: constant thresholds and linear thresholds. The constant thresholds consist of one value which is calculated from, e.g., the mean of the observations. The linear thresholds consist of linear regression models of the observations [55].

5.3 Machine learning

Machine learning (ML) can be described as a collection of algorithms that are not specifically programmed. Instead, the purpose is to allow the algorithms to learn and improve from experience to obtain a well performing algorithm. Therefore, the algorithms have a wide range of applications and are utilized within many different areas, e.g., email spam filtering, image recognition, fraud detection, and medical diagnosis [56].

When speaking of learning from experience in an ML context, the idea is that the algorithm is introduced to a number of different data samples to learn the behavior of the data in the samples [56]. The process of gaining the experience is called learning or training the model, and the samples used for this are called the training or learning set. There are a number of different ways to train the models, but the two most commonly used are supervised and unsupervised learning.

The ML algorithms trained using supervised learning are provided with labeled input and output data so that they can learn the relationships between the input and output variables. Both the input and the output data sets may consist of one or more variables, so called features. During the training, the algorithm tests a number of different combinations of parameters to find a model that is able to predict the labels of the output data when new, unlabeled input data is presented to the model [56]. If the output variable is continuous the learning problem is a regression problem, and if the output variable is discrete the learning problem is a classification problem [57].

The algorithms trained using unsupervised learning are instead learning from unlabelled data, where the algorithm seeks structure in the input training data without receiving any information about the expected output [58]. A common task for unsupervised algorithms is clustering, where the algorithm divides the data set into groups, or clusters, of similar data points [59]. Since the research method in Paper II utilizes supervised learning, unsupervised learning will not be described in further detail and the following sections are written assuming that supervised learning is used.

5.3.1 Creating a well performing ML algorithm

Before starting the training process, it is important to gather an amount of data that is representative of the behaviour that we are interested in modelling. As in linear regression, it is important to choose a number of input variables (features) that describe the output variable (target) well. It is also important to make sure that the data set does not contain data points that are not representative of the behaviour. This could for instance be outliers or missing values. Therefore, it is common practice to do some sort of data cleaning where abnormalities are replaced or removed from the data set before introducing the data to the model. It may also be necessary to do other types of data preprocessing, e.g., by scaling the features to similar ranges, creating polynomial features of the existing features, or aggregating them into a single feature that is more meaningful to the model [60].

The process of training aims to find a function $h : \mathcal{X} \mapsto \mathcal{Y}$ so that *h(x)* is able to predict the output variable y well [57]. When considering a continuous output variable, which is the case when looking at the customer data in district heating, this is the same as performing linear regression as described in section 5.1. By introducing more independent variables in the model, and denoting the parameters of the model using θ_i, equation 5.2 may be rewritten as:

$$h(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots \tag{5.5}$$

where x_i is the i:th independent variable or feature. If introducing $x_0 = 1$, equation 5.5 may be written as [57]:

$$h(x) = \sum_{i=0}^{n} \theta_i x_i = \theta^T x \tag{5.6}$$

where n is the number of input variables, and θ and x on the right-hand of the equation are vectors. In order to learn what parameters θ_i should be used in the model, the ML algorithm iterates over a number of different parameter combinations to test which one results in the smallest distance between the predicted value and the real value should. The test is done using a so called loss function, which will have its minimum value for the best parameter combination [57]. There are several different loss functions that are used in supervised learning, the most common one being the least-squares loss function [56].

When the model parameters have been obtained, it is important to test the model on a validation data set that the model has not seen before. This is done to test how well the model predicts unseen data, to decide which model to choose if more than one model is available, and to tune the so-called hyper-parameters of the model [61]. The hyper-parameters are not decided by the model; instead these parameters are decided by the user before training the model. An example of hyper-parameters is the learning rate, which decide how much the parameters are altered in each iteration step [62]. The hyper-parameters may have a great influence of the model performance, and so they need to be carefully chosen. After performing validation, it might be necessary to train the model again after adjusting the hyper-parameters. The last step is to test the model for completely new data, and to perform the prediction task that the model was created to do.

5.4 Data Mining

Data mining can be described as the task of extracting useful information in the form of interesting patterns and knowledge from large data sets [63]. The techniques have been applied within a large variety of research fields, including pattern recognition and prediction [64], as well as outlier analysis or anomaly mining [63]. The task of data mining may be performed using a number of different data analysis techniques. For example, different statistical models and machine learning methods have been applied to solve data mining problems [63].

The task of data mining often includes a number of steps, including the following [63]: data preprocessing, pattern discovery, pattern evaluation, and knowledge presentation. The process of identifying patterns is often iterative, and the main aim of the process is to identify patterns that are easy to understand and that provides meaning to a human interpreter.

5.5 Data analysis methods used for fault detection in DH customer installations

During the last years, a number of studies have been conducted to find ways to identify DH customer installations containing faults. Historically, one common analysis method has been to calculate the *overflow*, or overconsumption, of the individual installations [3, 18]. The overflow may be described as the additional amount of volume flow that is needed to meet the heat demand of a customer installation with a low ΔT, compared to if the installation had an ideal cooling ΔT_{ideal}. The ideal value of ΔT varies depending on the installation and the DH system that the installation is located in. The overflow is calculated according to equation 5.7.

$$overflow = V_{actual} - V_{ideal} = V_{actual} - \frac{E_{actual}}{\rho c_p \Delta T_{ideal}} \tag{5.7}$$

where

V_{actual}	=	Actual volume flow in the installation	[m^3]
V_{ideal}	=	Ideal volume flow in the installation	[m^3]
E_{actual}	=	Actual heat demand in the building	[J]
ρ	=	fluid density	[kg/m^3]
c_p	=	specific heat capacity for water	[J/kg°C]
ΔT_{ideal}	=	Ideal ΔT of the installation	[°C]

Since the overflow describes the additional amount of DH water that has to pass through the substation compared to the ideal case, an installation with a large overflow is more poorly performing than an installation with a smaller overflow. One advantage of the method is that the heat demand of the building is being used when calculating the overflow. This means that the overflows of each installation may be compared directly to each other and provides the DH utility with a priority list of the installations, since the ones that have the highest overflows have the largest impact on the DH system. The overflow method is most often not used to perform analysis on hourly values, but for monthly or maybe even yearly values, depending on how often the utilities are performing their analyses.

Recent studies have shown that there are plenty of possibilities to perform analysis on the hourly values as well. Gadd and Werner presented a novel method for fault detection in DH substations using outlier detection in the temperature difference signature, which may be described as daily average values of ΔT from one installation plotted as a function of the outdoor temperature [18]. The temperature difference signature consists of one average line which is estimated from the ΔT values of a number of substations with average cooling larger than 45 °C and two off-set lines, which are used as thresholds. If a substation has values located outside of the thresholds it may be poorly performing and should be further investigated. A similar approach was taken in a study conducted by Sandin et. al., who utilized hourly data to perform similar analysis for the heat power demand of 1 000 customer installations [55].

By evaluating the relationships between the different measurements from the meter device, it is possible to investigate how the customer installation is performing. This was done in a study conducted by Johansson and Wernstedt, who utilized n-dimensional analysis to perform status control of DH substations [65]. The method is called n-dimensional analysis since the relationships between n different variables may be investigated using the analysis method of the paper. In the study, the relationships that were to be investigated were already known, e.g. the relationship between the return temperature and the flow through the substation. It may also be possible to use data mining techniques to identify interesting relationships between parameters from the DH installations. Xue et. al. utilized clustering analysis to identify a number of different operating patterns [66]. These patterns were then analyzed using association analysis, which means that each of the pattern clusters were investigated to find correlations within the clusters. These correlations were then used to create a set of association rules that should be followed in customer installation containing no faults.

Pakanen et. al. have conducted a study where five different approaches to fault diagnosis of DH customer installations have been implemented and tested, including classification of installation performance indices and statistical tests [67]. In the study, fault detection is performed of the entire customer installation, a control valve, the heat exchanger, and a mud separating device. In his licentiate thesis, Yliniemi analyses the variance of the signal from a temperature sensor to detect when a fault occurs in the sensor [32]. An increase of the variance suggested a change of the measurement noise, which in turn indicated that a fault had occurred. Zimmerman et. al. also performed analysis of individual DH components in a study where they presented a method capable of detecting faults in pressure sensors and leakages in the DH systems using a Bayesian Network [68]. Bayesian Networks are used to describe cause and effect relationships between different parameters, and are capable of predicting the probability of a certain outcome based on a number of other events.

The machine learning applications for fault detection in DH customer installations is still quite few, but has gained a lot of interest lately. One example of a study utilizing ML methods was conducted by Abghari et. al., where a two-step classifier is utilized to identify manual changes of the temperature program of the internal heating system [69].An example from the DH field that is not sepcifically utilized for fault detection is heat load prediction using predictive ML modelling. This has been conducted in a study by Dalipi et. al. where the authors compared the performance of three different prediction models: Support Vector Regression (SVR), Partial Least Square (PLS), and random forest (RF) [70].

ML methods have however been used for fault detection in applications other than district heating. Among else, Araya et. al. have conducted a study where prediction-based and pattern-based classifiers were combined to perform fault detection of energy consumption data from buildings [71]. Another prediction-based fault detection method was used by Chen et. al, who created a model of the baseline behaviour of a building's energy consumption and then calculated prediction intervals for the baseline case that the future values should fall within to be consider normal [72].

This overview shows that a number of different data analysis methods may be applied to perform fault detection of DH data sets, and that the complexity of the methods varies. Many of the methods make use of DH domain knowledge which is beneficial since the results will be easier to understand and interpret for a person that has great knowledge of DH and DH systems, but little or no experience of data science methods. This is an aspect that has been considered as one of the main focuses when performing the work in this thesis.

5.6 Methods

In this thesis, a number of different data analysis methods have been used to perform fault detection in DH customer installations. Paper II focuses on outlier detection using limit checking with constant and linear thresholds, while the work in paper III was conducted using a machine learning approach. This section present the specific methods and evaluation tools that were used in the two studies.

5.6.1 Limit checking

Limit checking is a basic fault detection method that has traditionally been used in a number of different applications [73]. The theory behind limit checking is that a measured variable $Y(t)$ should stay within a certain tolerance zone when the process described by the variable is working as it should. The tolerance zone is limited by one maximum and one minimum threshold. The thresholds may be determined in different ways, e.g. by presetting constant maximum and minimum values, or by using prediction models to determine the thresholds [73]. In this thesis, limit checking was utilized to perform fault detection using both constant and prediction-based thresholds. The prediction model used was piece-wise linear regression.

5.6.2 Fault detection using prediction

Prediction algorithms have been used for a number of different tasks: weather forecasting, predicting house prices, and forecasting heat demands in DH systems. The algorithms may also be used for fault detection. To do this, the model is trained using a data set from a well performing process so that the model learns the behaviour and expected output of a well performing process. When a good model has been obtained, it is introduced to other input data sets, that might contain data from faulty processes. By comparing the model performance for the different data sets, it is possible to determine whether the data set contains faults or not [74].

Gradient Boosting Regression

There are a number of different regression algorithms that may be used for predicting a continuous target variable y. One of these is the Gradient Boosting Regression (GBR) algorithm, which is used for the research conducted in Paper II. In GBR, a number of weak predictors are combined to create an ensemble of predictors that is more accurate in its prediction than the individual predictors are separately [75]. The ensemble is created using an iterative process where the number of iterations are decided beforehand. In each iteration step a new predictor is added to the previously obtained ensemble. The new predictor is trained to minimize the prediction error of the previous ensemble, which means that the prediction ability of the ensemble increases with each iteration step [75].

5.6.3 TPOT - Tree-based Pipeline Optimization Tool

When working with applied machine learning, many different methods for data pre-processing and modelling may be utilized to perform the task at hand. It may be a daunting task to identify the correct combination, or *pipeline*, of pre-processing and modelling methods and so a number of tools have been developed that allows the user to automate parts of the work. One such tool is the Tree-based Pipeline Optimization Tool, which is implemented in Python and creates a pipeline using genetic programming [76].

5.6.4 Measures for model validation

There are a number of different ways to evaluate how well a model is able to describe the behaviour of a process. One basic measure is the coefficient of determination, or R^2 value. The R^2 value describes how much of the variance in the data that the model is able to describe [77]. The coefficient of determination is calculated according to equation 5.8, where SS_{res} is the sum of squares of the prediction residuals, and SS_{tot} is the total sum of squares which describes how much the observations vary around their overall mean. The R^2 value assumes values between 0 and 1, and a well performing model should have a coefficient of determination close to 1 since this indicates that the model is able to capture most of the variation in the data [77].

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{5.8}$$

Another commonly used measure is the Mean Absolute Error (MAE) which is calculated according to equation 5.9. The MAE describes the average magnitude of the errors in a set of predictions [78]. In equation 5.9, n is the number of test instances. The MAE value for a well performing model is expected to be as close to zero as possible.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |(actual\ value)_i - (predicted\ value)_i| \tag{5.9}$$

Chapter 6

Method

The work presented in this thesis include a number of different aspects which result in an interdisciplinary project which requires knowledge of many different research fields, including both DH theory and computer science skills. Therefore, the studies included in the thesis have been conducted using a collection of different techniques including qualitative interviews, quantitative research methodology, and different kinds of data analysis methods.

Paper I was conducted using both a qualitative and a quantitative approach to investigate the fault handling processes of the current DH industry. The study included a number of qualitative interviews, during which the utilities were asked to, i.a., describe why they wanted to detect faults in their customers' installations, how they detected the faults, how they encouraged their customers to correct the faults, and what the utilities gained from eliminating the faults. The study was conducted with a clear focus on work procedures that aimed to decrease the return temperature levels of the DH systems. The study also included a quantitative survey of what the most common faults were in different DH utilities in Sweden. The survey was based on the five fault categories described in section 4.3, and included questions about how common the different faults were, how the faults may be detected, and if it was possible to detect the faults in customer data. The results from the interviews and the survey were analyzed and are presented in detail in paper I.

Papers II and III aimed to investigate the possibility to use automated data analysis to identify customer installations containing faults. The data used in both studies originated from DH systems located in Sweden.

The study in paper II was conducted using automated limit checking. The aim was to detect customer installations that deviated significantly from the behaviour of a well performing installation. This was done by investigating three different variables that may be used as performance indicators: cooling in the installation, return temperature from the installation, and heat consumed in the installation. As described in section 5.6.1, limit checking may be used to check if a variable stays within a certain tolerance zone. In the paper, a reference case with well performing customer installations was chosen to create a tolerance zone describing the thresholds within which an installation should operate to be considered to be well performing. The fault detection method was implemented in the programming language R.

The study in paper III was conducted using a machine learning approach in the programming language Python, where TPOT was utilized to create a number of prediction pipelines. The pipeline that had the best R^2 and MAE values was chosen to be investigated further. The prediction algorithm in the pipeline was a GBR model which was trained to learn the behaviour of a well performing customer installation. In this study, the hourly mass flow rate was chosen as the output variable, i.e., the behaviour the model should learn. The input variables were chosen to be hourly values of the outdoor temperature, an average value of the outdoor temperature during the last 24 hours, hour of the day, hourly supply temperature, and hourly return temperature. The model was then introduced to data sets containing four different faults: (i) sudden decreases of the supply temperature to 60 °C, (ii) sudden decreases of the supply temperature by 10 %, (iii) drifting outdoor temperature meter, and (iv) drifting supply temperature meter. The performance of the predictions for the faulty and non-faulty data sets were compared to see if the model predicted differently for the different data sets. The comparison of the model performance for the different data sets was carried out using the sum and cumulative sum of the prediction residuals.

Finally, all three papers, as well as this thesis, are based on extensive literature reviews where a mixture of scientific papers, research reports, and other sources of knowledge have been reviewed.

Chapter 7

Main results

Paper i investigated fault handling in district heating customer installations from the DH industry's perspective: why were the DH utilities currently working with customer installations containing faults, and how were the successful utilities working to decrease the number of faults? The study was conducted in two parts. The first part consisted of a qualitative interview study with six Swedish DH utilities. The results from the interview study show that the two most important aspects of a successful fault handling process were to have a good and open customer relationship, and to gain physical access and mandate of the customer installations. Another important aspect was to create clear incentives for the customers to maintain their installations in a good condition. The second part of the study was conducted as a survey study which was sent out to 139 different DH utilities in Sweden, out of which 56 utilities replied. The results from the survey show that a number of different faults occurred in the DH customer installations, and that the distribution of the faults varied between the different systems. 33 % of the responding utilities stated that leakages was the most common fault in the customer installations, followed by faults in the customers' internal heating system which 31 % of the utilities answered was most common. However, these faults were also considered hard to identify when visiting a customer installation since the DH utilities in most cases only gained access to the substation and not the customer's internal heating system.

The aim of **paper ii** was to develop and test an automated fault detection method that was able to detect poorly performing customer installations using customer data. This method was implemented as an algorithm using the programming language R. The algorithm was then tested for a data set consisting of 3 000 customers in a DH system located in Sweden. The results showed that the algorithm developed performs well and is able to identify poorly performing customer installations rapidly. In total, 1 273 installations were found to be poorly performing, corresponding to 43 % of the installations. The installations that were found to be poorly performing all had a high overflow, which indicated that the identified installations had a large impact on the system efficiency.

Paper iii investigated how machine learning methods could be used to detect faults in district heating substations. The approach was to test if a substation prediction model trained on data from a well performing substation would predict differently when introduced to a data set from a substation known to contain faults. The data set used in the study consisted of hourly values during one year from a DH substation located in Sweden. The results showed that the different faults had different impacts on the model performance. The drifting supply temperature meter resulted in a very small deviation from the performance for the data containing no faults, while the sudden decreases of the supply temperature gave a clear indication that the model performance changed. The model performance also changed for the data set containing a drifting outdoor temperature meter fault. This implied that the model is capable of detecting faults in data sets, but that it might not be able to detect all different faults that may occur.

Chapter 8

Concluding discussion and future work

This thesis has presented a number of different approaches to working with faults in district heating customer installations. The approaches cover the current work procedures in the DH industry, but also includes two new methods for fault detection in the installations using customer data.

The main idea of the work conducted in the thesis has been to survey how the DH utilities that exist today may work to reduce the number of faults in customer installations. Therefore, it has been considered important to investigate how utilities with low system temperatures are working with their customers to obtain lower system temperatures from the existing customer installations. This provides an overview of work procedures that are actually working and that could be implemented by other utilities as well. An interesting aspect of this work that became clear when analyzing the material in paper I is that lower return temperatures may not be the only benefit when working with the customer installations. It will also be possible to improve customer relationships significantly with more satisfied customer as an effect.

Due to the focus on the existing DH systems, it has also been important to work primarily with solutions that are already implemented in the systems. An example of this is that the data used in papers II and III are customer data normally used for billing purposes. As described in section 3.4, this data normally only include four different variables: heat consumption, mass flow rate, and supply and return temperature. It could have been possible to include even more measurements in the analysis, but since it is not common practice to measure and collect more than the four basic variables in most DH system it was decided not to do so.

The large variety of faults presented in section 4.3 clearly shows that there are many different faults to eliminate in the customer installations. Today, a large share of the faults are identified when visiting the customer installation. However, the results from paper I showed that it may not be possible to detect some of the most common faults when visiting the customer installations, since many of the faults were located in the internal heating systems. Therefore, it will probably be necessary to be able to detect and identify these faults using different data analysis methods. To obtain this, measurements from the customer's internal heating systems will probably have to be included in the future fault detection analysis. Otherwise, the faults will be extremely hard to identify.

A main conclusion from papers II and III is that the possibilities for using customer data to detect poorly performing customer installations are large. The methods used in this thesis are not overly complicated and the focus has been to develop methods that are easy to understand and easy to relate to even though one is no data science expert. The results in paper III also shows that it might be necessary to combine a number of different fault detection methods to be able to capture all faults that may occur in a customer installation. Once again, data from the internal heating system may be beneficial to be able to detect a larger variety of faults.

The work in this thesis clearly shows that there is great potential to decrease the DH system temperatures. The utilities need to work actively to obtain this, and some of the methods and procedures might involve rather drastic changes in how the utilities are currently working. Nonetheless, when considering the greater picture, the number of benefits that comes with decreased system temperature may just be worth wile the effort.

In future studies, it would be interesting to continue working with the fault detection methods to see how well they scale and if the methods are applicable for many different DH systems. It would also be interesting to compare the results from the different fault detection methods to see if they detect the same installations. When having done this, the next step of the fault handling process would be to investigate how data analysis can be used to identify *what* is actually wrong in the customer installations. The first step should be to investigate if the customer data available today can provide any information about this. The problem is rather complex, and more measurements would probably have to be included in the analysis, but a starting point could be to investigate if it is possible to separate faults located in the substation and the internal heating system from each other. If so, the utilities would at least have an indication of whether the fault is accessible to them or not. One large issue related to this work is that many utilities have no or very sparse records of when and where a fault occurs. Other studies of interest would be to perform in-situ/lab experiments and/or simulations of customer installations to investigate what data is needed to detect different faults.

Chapter 9

References

[1] Sven Werner. International review of district heating and cooling. *Energy*, 137:617–631, 2017.

[2] Energiföretagen. Tillförd energi (Supplied energy). `https://www.energiforetagen.se/statistik/fjarrvarmestatistik/tillford-energi/`, 23-08-2017. Online, accessed: 2018-10-12. In Swedish.

[3] Svend Frederiksen and Sven Werner. *District heating and cooling*. Studentlitteratur AB, Lund, Sweden, 2013.

[4] Henrik Lund, Sven Werner, Robin Wiltshire, Svend Svendsen, Jan Eric Thorsen, Frede Hvelplund, and Brian Vad Mathiesen. 4th generation district heating (4GDH) - Integrating smart thermal grids into future sustainable energy systems. *Energy*, 68:1–11, 2014.

[5] Lisa Brange. *Technical and environmental analysis of prosumers in district heating networks*. Licentiate thesis, Department of Energy Sciences, Faculty of Engineering, Lund University, Lund, Sweden, 2015.

[6] Sven Werner. District heating and cooling in Sweden. *Energy*, 126:419–4299, 2017.

[7] Svensk fjärrvärme. Fjärrvärmecentralen - kopplingsprinciper, rapport 2009:3 (The district heating substation - connection principles, report 2009:3). Report, 2009. In Swedish.

[8] Svensk fjärrvärme. Fjärrvärmecentralen - utförande och installation (The district heating substation - execution and installation). Technical regulations F:101, February 2014. In Swedish.

[9] Mats Dahlblom Catarina Warfvinge. *Projektering av VVS-installationer (Planning and design of HVAC installations).* Studentlitteratur AB, Lund, 2010. In Swedish.

[10] Patrick Lauenburg and Janusz Wollerstrand. Adaptive control of radiator systems for a lowest possible district heating return temperature. *Energy and Buildings*, 72:132–140, 2014.

[11] Natasa Djuric, Vojislav Novakovic, and Frode Frydenlund. Heating system performance estimation using optimization tool and bems data. *Energy and Buildings*, 40:1367–1376, 2008.

[12] Jonas Gustafsson, Jerker Delsing, and Jan van Deventer. Improved district heating substation efficiency with a new control strategy. *Applied Energy*, 87:1996–2004, 2010.

[13] Energimyndigheten. Energieffektivisering i stora värmesystem, publikation ET2013:20 (Energy efficiency in large heating systems, publication ET2013:20). Publication, 2013. In Swedish.

[14] Värmex konsult AB. Därför ska man injustera ett värmesystem (the reasons for balancing a heating system). `https://www.varmex.se/blogg/energieffektivisering/darfor-ska-man-injustera-ett-varmesystem/`, 2011-07-07. Online, accessed: 2018-10-31. In Swedish.

[15] Per-Olof Johansson. *Fjärrvärmeanslutna byggnaders värme- och varmvattensystem - samverkan, komfort och sårbarhet (District heating connected buildings heating and hot water systems - interaction, comfort and vulnerability).* Licentiate thesis, Department of Energy Sciences, Faculty of Engineering, Lund University, Lund, Sweden, 2007. In Swedish.

[16] Boverket. Boverkets byggregler (föreskrifter och allmänna råd), BFS 2011:6 (Boverket's building regulations - mandatory provisions and general recommendations, BFS 2011:6)), 2011, revised 2018. In Swedish.

[17] Yassin Jomni, Jan van Deventer, and Jerker Delsing. In *Proceedings of the 12th International Conference on Flow Measurement*, Guilin, China, 2004.

[18] Henrik Gadd and Sven Werner. Achieving low return temperatures from district heating substations. *Applied Energy*, 136:59–67, 2014.

[19] Helge Averfalk and Sven Werner. Framtida fjärrvärmeteknik. Research report 2017:419, Energiforsk, 2017. In Swedish.

[20] The European Commission. Directive 2012/27/EU of the European Parliament and of the Council - on energy efficiency, 2012. `https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:315:0001:0056:en:PDF`.

[21] Henrik Gadd and Sven Werner. Daily heat load variations in Swedish district heating systems. *Applied Energy*, 106:47–55, 2013.

[22] Patrick Lauenburg. *Improved supply of district heat to hydronic space heating systems*. Doctoral dissertation, Department of Energy Sciences, Faculty of Engineering, Lund University, 2009.

[23] Jeroen van Ruitenbeek and Johan van der Kamp. Efficient use of residual heat in combined heat & power systems. *European Energy Innovation*, Autumn 2013:10–12, 2013.

[24] Girts Vigants, Gundars Galindoms, Ivars Veidenbergs, Edgars Vigants, and Dagnija Blumberga. Efficiency diagram for district heating system with gas condensing unit. *Energy Procedia*, 72:119–126, 2015.

[25] Rasmus Lund, Dorte Skaarup Østergaard, Xiaochen Yang, and Brian Vad Mathiesen. Comparison of low-temperature district heating concepts in a long-term energy system perspective. *International Journal of Sustainable Energy Planning and Management*, 12:5–18, 2017.

[26] Fabian Bühler, Stefan Petrović, Kenneth Karlsson, and Brian Elmegaard. Industrial excess heat for district heating in denmark. *Applied Energy*, 205:991–1001, 2017.

[27] Robin Wiltshire. *Advanced district heating and cooling (DHC) systems*. Woodhead Publishing, Cambridge, UK, 2016.

[28] Sven Werner. 4dh, first phd student seminar - current and future temperature levels in district heating systems, March 2013.

[29] Dietrich Schmidt, Anna Kallert, Markus Blesl, Svend Svendsen, Hongwei Li, Natasa Nord, and Kari Sipilä. Low temperature district heating for future energy systems. *Energy Procedia*, 116:26–38, 2017.

[30] Marek Brand, Alessandro Dalla Rossa, and Svend Svendsen. Energy-efficient and cost-effective in-house substations bypass for improving thermal and dhw (domestic hot water) comfort in bathrooms in low-energy buildings supplied by low-temperature district heating. *Energy*, 67:256–267, 2014.

[31] Svenska fjärrvärmeföreningen. Avkylningen i ett fjärrvärmesystem, FVF 2000:03 (Cooling in a district heating system, FVF 2000:03). Project report, Svenska Fjärrvärmeföreningens Service AB, 2000. In Swedish.

[32] Kimmo Yliniemi. *Faults in district heating substations.* Licentiate thesis, Luleå University of Technology, Department of Computer Science and Eletrical Engineering, EISLAB, 2005.

[33] Bott T. Reg. Fouling. `http://www.thermopedia.com/content/779/`, 2011-02-02. Online, accessed: 2018-11-05.

[34] Janusz Wollerstrand and Svend Frederiksen. Försmutsningsförlopp i plattvärmeväxlare för fjärrvärmeabonnentcentraler (Fouling in plate heat exchangers for district heating substations). Report, Stiftelsen för värmeteknisk forskning, Stockholm, August 1995. In Swedish.

[35] Henrik Gadd and Sven Werner. Fault detection in district heating substations. *Applied Energy*, 157:51–59, 2015.

[36] Markus Alsbjer and Peer Wahlgren. Fjärrvärmecentraler 10 år - håller de måttet? (District heating substations 10 years - do they measure up?). Report 2011:9, Fjärrsyn, 2011. In Swedish.

[37] Helge Averfalk. Morgondagens effektiva fjärrvärme: en beskrivande litteraturstudie (The efficient district heating of tomorrow). Halmstad Högskola, 2014. In Swedish.

[38] Sven Werner. Feltyper i fjärrvärmecentraler strukturerade (Faults in district heating substations structured). Halmstad Högskola, 2013.

[39] Engineers edge. Parallel and counter flow designs. `https://www.engineersedge.com/heat_transfer/parallel_counter_flow_designs.htm`, n.d. Online, accessed: 2018-11-06.

[40] Karl Johan Åström. Reglerteknik – En elementär introduktion. Kapitel 8: Dödtidskompensering (Automatic Control Engineering – an elementary introcution. Chapter 8: Dead time compensation). Report TFRT-3173, Department of Automatic Control, Lund Institute of Technology (LTH), 1983. In Swedish.

[41] Svend Frederiksen and Janusz Wollerstrand. Thermostatic control of instantaneous water heaters in district heating substations. In *5th International Symposium on Automation of District Heating Systems*, Otaniemi, Espoo, Finland, 1995.

[42] Anders Fransson. Avkylningsarbete på göteborg energi ab 1995-2004 (optimizing of the district heating system in gothenburg 1995-2004 - practical application of r&d knowledge). Research report 2005:132, Svensk Fjärrvärme AB, 2005. In Swedish.

[43] Natasa Nord, Elise Kristine Løve Nielsen, Hanne Kauko, and Tymofii Tereshchenko. Challenges and potentials for low-temperature district heating implementation in norway. *Energy*, 2018:889–902, 2018.

[44] Samson AG. Introduction to self-operated regulators. Technical information L202en, Samson AG, 2000.

[45] Pumpportalen.se. Ventil (Valve). `http://www.pumpportalen.se/ventiler-ventil/`, 2015. Online, accessed 2018-11-12. In Swedish.

[46] Pumpportalen.se. Kägelventil (Globe valve). `http://www.pumpportalen.se/kagelventil/`, 2015. Online, accessed 2018-11-12. In Swedish.

[47] SweTherm AB. SweTherm Villaprefab - drift & skötsel RF/D standard (SweTherm houseprefab - operation & maintenance RF/D standard). Product sheet DRIFT:D5015, 2005. In Swedish.

[48] Helge Averfalk and Sven Werner. Novel low temperature heat distribution technology. *Energy*, 145:526–539, 2018.

[49] Marek Brand and Svend Svendsen. Renewable-based low-temperature district heating for existing buildings in various stages of refurbishment. *Energy*, 62:311–319, 2013.

[50] John O. Rawlings, Sastry G. Pantula, and David A. Dickey. *Applied Regression Analysis: A Research Tool*. Springer-Verlag, New York, 1998.

[51] Howard J. Seltman. *Experimental design and analysis*. Carnegie Mellon University, `http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf`, 2018. Online, accessed 2018-11-20.

[52] Iain Pardoe, Laura Simon, and Derek Young. 1.3 - The Simple Linear Regression Model. `https://onlinecourses.science.psu.edu/stat501/node/253`, 2018. Online, accessed: 2018-11-20.

[53] Kenneth O. Cogger. Piecewise linear modeling: Theory, guidelines, and applications. In *Management Science & Technology Symposium*. University of Kansas School of Business, 2006.

[54] Jason Brownlee. How to use statistics to identify outliers in data. https://machinelearningmastery.com/how-to-use-statistics-to-identify-outliers-in-data/, 2018. Online, accessed 2018-11-20.

[55] Fredrik Sandin, Fredrik Gustafsson, and Jerker Delsing. Fault detection with hourly district energy data. Project report 2013:27, Svensk Fjärrvärme AB, 2013.

[56] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. The MIT Press, Cambridge, USA, 2012.

[57] Andrew Ng. Cs229 lecture notes - supervised learning. http://cs229.stanford.edu/notes/cs229-notes1.pdf, 2018. Online, accessed 2018-11-21.

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

[59] Claude Sammet and George I. Webb. Clustering. In *Encyclopedia of Machine Learning and Data mining*. Springer, Boston, 2017.

[60] Jason Brownlee. How to prepare data for machine learning. https://machinelearningmastery.com/how-to-prepare-data-for-machine-learning/, 2013. Online, accessed 2018-11-21.

[61] Yufeng Guo. The 7 steps of machine learning. https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e, 2017. Online, accessed 2018-11-21.

[62] Hafidz Zulkifli. Understanding learning rates and how it improves performance in deep learning. https://towardsdatascience.com/understanding-learning-rates-and-how-it-improves-performance-in-deep-l 2018. Online, accessed 2018-11-21.

[63] Jiawei Han, Micheline Kamber, and Jian Pei. *Data mining - concepts and techniques*. Waltham, MA, USA, 2012.

[64] The Pennsylvania State University. Lesson 1(a).3 - the nature of the problem. https://onlinecourses.science.psu.edu/stat857/node/145, 2018. Online. Accessed 2018-12-19.

[65] Christian Johansson and Fredrik Wernstedt. N-dimensional fault detection and operational analysis with performance metrics. *DHC13 - The 13th International Symposium on District Heating and Cooling*, Copenhagen, Denmark, 2012.

[66] Puning Xue, Zhigang Zhou, Xiumu Fang, Xin Chen, Lin Liu, Yaowen Liu, and Jing Liu. Fault detection and operation optimization in district heating substations based on data mining techniques. *Applied Energy*, 205:926–940, 2017.

[67] Fault diagnosis methods for district heating substations. Research notes 1780, VTT Building technology, 1996.

[68] Nathan A Zimmerman, Erik Dahlquist, and Konstantinos G. Kyprianidis. Towards on-line fault detection and diagnostics in district heating systems. *Energy Procedia*, 105:1960–1966, 2017.

[69] Shahrooz Abgharia, Eva Garcia-Martina, Christian Johansson, Nicklas Lavesson, and Håkan Grahn. Trend analysis to automatically identify heat program changes. *Energy Procedia*, 116:407–415, 2017.

[70] Fisnik Dalipi, Sule Yildirim Yayilgan, and Alemayehu Gebremedhin. Data-driven machine-learning model in district heating system for heat load prediction: A comparison study. *Applied Computational Intelligence and Soft Computing*, 2016.

[71] Daniel B. Araya, Katarina Grolinger, Hany F. ElYamany, Miriam AM Capretz, and Girma T. Bitsuamlak. An ensemble learning framework for anomaly detection in building energy consumption. *Energy and Buildings*, 144:191–206, 2017.

[72] Bei Chen, Mathieu Sinn, Joern Ploennigs, and Anika Schumann. Statistical anomaly detection in mean and variation. *22nd International Conference on Pattern Recognition*, Stockholm, Sweden, 2014.

[73] Rolf Isermann. *Fault-Diagnosis Systems - An Introduction from Fault Detection to Fault Tolerance.* Springer, Darmstadt, Germany, 2006.

[74] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 2009.

[75] Alexey Natekin and Alois Knoll. Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, 7(21), 2013.

[76] Randal S. Olson, Ryan J. Urbanowicz, Peter C. Andrews, Nicole A. Lavender, La Creis Kidd, and Jason H. Moore. Automating biomedical data science through tree-based pipeline optimization. *European conference on the Applications of evolutionary computation*, 2016.

[77] The Pennsylvania State University. 1.5 - the coefficient of determination, r-squared. `https://newonlinecourses.science.psu.edu/stat501/node/255/`, 2018. Online. Accessed 2018-12-20.

[78] Claude Sammut and Geoffrey I. Webb. Mean absolut error. `https://doi.org/10.1007/978-0-387-30164-8_525`, 2011. Online. Accessed 2018-12-20.

Scientific publications

Co-authors are abbreviated as follows:
Per-Olof Johansson Kallioniemi (POJ), Marcus Thern (MT), Tijs Van Oevelen (TVO), Kerstin Sernhed (KS), Kristin Davidsson (KD), Patrick Lauenburg (PL).

Paper i: Fault handling in district heating customer installations: experiences from Swedish utilities

I did the conceptualization and developed the methodology for the interview and survey study together with KS. POJ also contributed to the methodology of the survey study. I wrote the main parts of the paper and performed the literature review of the paper. KS, POJ, TVO and MT contributed to the editing and revision 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. Me and KD also developed the software related to the study. The original drift and revision of the paper was conducted by me and MT. I also performed the literature review in the paper.

Paper iii: 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 POJ. I wrote the main parts of the paper, as well as conducted the literature study of the paper. MT, POJ and KS all contributed to the revision and editing of the paper.

Paper i

Fault handling in district heating customer installations: exeriences from Swedish utilities

Sara Månsson[1,2,*], Per-Olof Johansson Kallioniemi[1], Marcus Thern[1], Tijs Van Oevelen[1], and Kerstin Sernhed[2,3]

[1]Department of Energy Sciences, Faculty of Engineering, Lund University, P.O. Box 118, SE-221 00 LUND, Sweden
[2]VITO, Boeretang 200, BE-2400 MOL, Belgium
[*]Correspondence: sara.mansson@energy.lth.se; Tel.: +46-46 222 33 20
[3]EnergyVille, Thor Park 8310, BE-3600 Genk

Abstract

The district heating customer installations in current district heating systems contain a variety of different faults that cause the return temperatures of the systems to increase. This is a major problem, since the focus in the district heating industry is currently to decrease the overall system temperatures to be able to utilize more low-temperature heat. This study has been conducted in two parts, and the focus of the first part was to investigate how the district heating utilities that already have low return temperatures are working to keep their temperature levels down, and how they are involving their customers in this work. This was done by conducting a qualitative interview study, involving representatives from six Swedish district heating utilities. The results showed that the two most important elements in a successful work was to have physical access to and mandate of the customer installations, and to maintain a good and close customer relationship. The second part of the study involved a survey which was answered by representatives from 56 Swedish district heating utilities. The survey contained questions regarding what faults occurred in the customer installations most frequently, how these were identified, and how the utilities are working to eliminate them. The results showed that many faults occur in the customers' internal heating systems, or due to leakages somewhere in the installation. Overall, the results show that it is indeed possible to work close to and affect the customers to obtain lower return temperatures from the customer installations, and that the most common faults are faults that are rather easy to eliminate as long as the utilities gain physical access to the entire customer installations.

Keywords: District heating substations, Experience from industry, Fault diagnosis, Faults in substations, Poor substation performance

1 Introduction

Today, more than half of EU's energy is used for heating and cooling, and in 2016, approximately 75 % of the energy used for heating and cooling purposes was based on fossil fuels [1]. Hence, the heat supply sector needs to find solutions that reduce the fossil fuel dependence and utilize renewable heat sources instead. District heating (DH) has been identified as an important part of the future smart energy systems due to the ability of providing local, affordable and low-carbon heating [2]. DH will play an important role in increasing the energy efficiency of the energy systems due to the ability to make use of available heat sources that would otherwise go to waste, including large-scale heat pumps, large thermal storage, Combined Heat and Power (CHP) plants, solar thermal energy, geothermal energy and industrial waste heat [3, 4, 5].

In order to reach the full potential of district heating as described above, it is critical that the current technology develops further. This development has been described in a paper written by Lund et al., where the 4th generation of district heating (4GDH) has been defined [6]. In the paper, the authors have described that the future DH systems must be able to use recycled low-temperature heat and renewable energy to supply low-temperature district heating for space heating and domestic hot water preparation with low heat losses. This includes decreasing the temperature levels of the DH systems to levels around 50/20 °C supply and return temperature [6]. Today, the average temperature levels of most systems are much higher than this. For example, the Swedish DH systems have annual average temperature levels of 86.0 °C supply and 47.2 °C return [7].

The temperature levels of the DH systems depend on the climate where they are located, what heat sources are available in the vicinity, the age of the DH system, the age of the buildings in the system and the building standards applied when the building was constructed, as well as the temperature demands of the heat consumers in the systems [8, 9]. The supply temperature is determined by the heat supplier and varies with the outdoor temperature. The return temperature is determined by the aggregated result from all cooling processes in the individual customer installations (i.e., the internal heating system of the building and the DH substation) [8]. Hence, the return temperature will increase if the cooling processes in the installations are not working as they should.

The design of the customer installations varies, depending on what country and system the installation is located in. In some countries it is common practice to not have any hydraulic separation at all, so called direct connection, but the most common design is to have an indirect connection where a heat exchanger separates the water in the internal heating system from the water in the DH system [8].

The substation consists of a number of different components, and the design and connection principle of the substation also varies. The two most common connection principles are the parallel-connected and the 2-stage connected substation [10]. Figure 1 displays the general outline of a parallel-connected substation with hot water circulation in the internal heating system. The figure includes the most common components of the customer installation.

According to previous studies there are two common reasons to why the return temperatures from the customer's installation are higher than they theoretically could be: faults in the customer's internal heating system and faults in the customer's substation [8, 11, 12]. In fact, the existing DH substation technology enables temperature levels lower than this, but due to the poor cooling performance of the customer installations these temperature levels are seldom reached. This is a strong indication that the faults in the customer installations have to be detected and identified in order to decrease the current return temperature levels of the DH systems.

It will also be of great importance to be able to detect these faults at an early stage in the future 4GDH systems. If the faults are not detected rapidly the faults will prevail in the systems, making it impossible for the utilities to maintain the low temperature levels of 4GDH. Automatic fault detection methods could play an important role to mitigate these problems. Automated heat meter readings provide a lot of relevant data containing the signatures of faulty substations. This enables computer-automated analysis to evaluate the performance of individual substations and detect faults. To know what faults such a fault detection tool should be able to detect, it is important to know what the most frequently occurring faults are today and in what way they impact the return temperature levels.

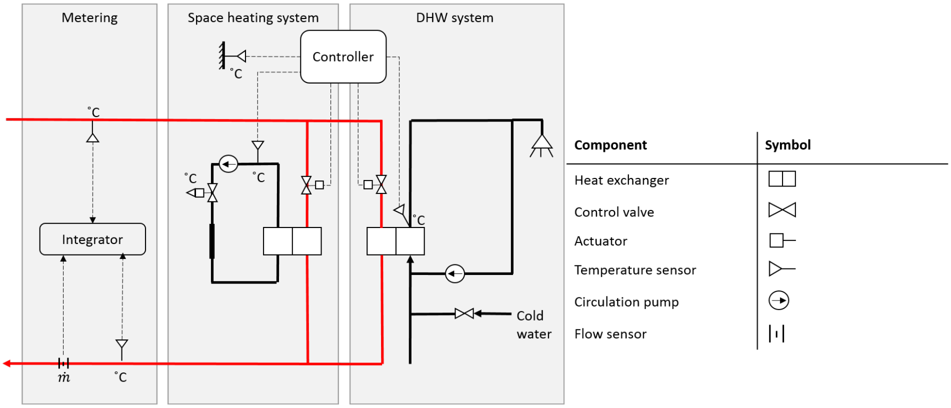


Figure 1: General outline of parallel-connected substation with hot water circulation

Two key issues to eliminating the faults is the utility's ability to communicate the information about the faults in the installations to the customers, and how to encourage the customers to correct the faults in their installations. Therefore, the purpose of this study was to investigate how utilities are currently working with their customers to decrease the return temperatures from the customer installations. In order to find good strategies for working with improving the cooling in the DH systems,interviews were carried out to identify the key successful measures that utilities are currently taking to reach lower return temperatures and how their customers are involved in the process. The aim was also to identify the most frequently occurring faults in the DH installations today, and how the utilities are working to eliminate this.

The study consists of a literature review, a survey which was sent to Swedish district heating utilities, and a qualitative interview study with representatives from Swedish DH utilities. The focus of the literature studies and the survey was to investigate what faults occur in the system, how often they occur, and to explore what strategies and methods the utilities use to eliminate the faults. The focus of the interview study was to investigate how the utilities are working with their customers and the customer installations today; what incentives they have to work towards decreased return temperatures, what incentives they provide their customers with so that they are interested in correcting faults in their installations, and how the utilities are working to eliminate faults in their customers' installations.

2 Methodology

2.1 Survey study

To investigate the faults that occur in the current DH system, a survey study was conducted. The first part of the study was to establish what faults to include in the survey. This was done by conducting a literature study, as well as an interview with a leading district heating expert. The literature review gave a comprehensive picture of the historical fault distribution, and what faults should be investigated further in the study. Information of the different faults was primarily collected among scientific articles and research reports conducted by various research institutes. The main purpose of the interview was to create a fundamental understanding of the faults, as well as identifying the faults that were most likely to occur in the modern district heating systems.

Based on the results from the literature review and expert interview, five different categories of faults were identified. The categories were based on what component or part of the customer installation that the fault occured in. The five different categories are: (i) heat exchangers, (ii) control system and controller, (iii) actuators, (iv) control valves, and (v) internal heating system of the customer.

These results were then used to create a survey which was sent to DH utilities in Sweden. In the survey, the utilities were asked to give information about how common certain faults were, how they went about to identify the different faults, and what the fault distribution looked like. The survey also included questions regarding information about the utilities themselves, their price models and service agreements, as well as how they utilized customer data. The price models and service agreements were of interest since they may be used as an incentive for the customers to work with faults in their installations more actively. The survey also provided plenty of space for the utilities to add additional faults that were not included in the survey to begin with, and the possibility to write down comments and explanations about the different faults. The survey was sent to 139 utilities that were all members of the Swedish non-profit industry and special interest organisation Swedenergy - Energiföretagen Sverige. 56 of the utilities answered the survey.

2.2 Interview study

The interviews were conducted during the fall of 2018, and representatives from six DH utilities were interviewed during the study.The participating utilities were all located in Sweden. The utilities were of varying sizes and delivered varying amounts of heat to their customers. The interviewees were people involved with, or responsible for, the operation and maintenance of DH production, distribution or end user.

The interview study was conducted using a qualitative, semi structured approach in order to obtain in depth knowledge of how the DH utilities were working with their customers in the context of decreasing the return temperature levels. Before the interview, a number of questions were prepared but the semi structured approach allowed for the interviewer to follow up the responses of the interviewee during the interview and ask new questions if needed. Before the interviews were conducted, a number of themes and questions were prepared in order to facilitate the interview process.

Four out of the six interviews were conducted face-to-face by the same interviewer. They were recorded and took approximately one hour to complete. The two remaining interviews were conducted as phone interviews by the same interviewer and recorded via the phone. The interviews were transcribed word for word, and the material obtained was then analyzed by dividing the answers thematically into four different categories: (i) incentives for the district heating companies, (ii) incentives for the district heating customers, (iii) access to and mandate of the maintenance of the district heating substations, and (iv) analysis methods for finding porrly performing customer installations.

3 Faults in customer installations

The literature review and the expert interview showed that a number of different faults may appear in the customer installations. However, several faults do not affect the comfort of the customer.

This means that a fault might be present in the substation without the customer experiencing any issues with the space heating or the DHW preparation [13].

In this study, five different categories of faults have been identified. The categories are based on what component or part of the customer installation the fault occurs in. The five different categories are: (i) heat exchangers, (ii) control system and controller, (iii) actuators, (iv) control valves, and (v) internal heating system of the customer. Below follows a review of the most common faults that occur for the different components or faults.

3.1 Heat exchangers

The category includes faults that are related to the heat exchangers of the substation. Depending on the design of the substation, there might be one, two, or more heat exchangers installed. All of these need to perform well for the substation as a components to perform well.

The literature shows that faults related to the heat exchangers themselves most commonly include fouling of heat exchangers [8, 13, 14, 15]. This can be described as unwanted deposits on heat transferring surfaces which cause a resistance to heat transfer and flow in the heat exchanger [16]. However, a study conducted by Wollerstrand and Frederiksen show that moderate fouling of the heat exchanger might lead to higher heat transfer at low flows [17].

Another fault related to the heat transferring capability of the heat exchangers is that some heat exchangers by mistake are installed so that the DH water and the water in the customer's internal heating system are flowing co-current instead of counter current [14, 15]. Other faults include leakages from the heat exchanger itself and uneven flows in heat exchangers connected in parallel [15, 18].

3.2 Control system and controller

The category of the control system and the controller includes the controller itself, the temperature sensors, the connections between the sensors and the controller, and the connections between the actuators and the controller. Hence, there is a number of components that may be faulty and so there is a number of different faults that may occur in the control system.

The controller itself may break down causing a completely uncontrolled system [14]. It might also be that the controller is installed incorrectly in the system. This occurs if the wrong input signal is connected to the wrong port on the controller. Since the controller is receiving an incorrect signal the control sequence of the installation will be disturbed [15].

With regards to temperature sensors, a number of different faults may occur. The temperature sensors may be broken, not sending any signal at all or a completely incorrect signal to the controller [14]. The temperature sensors might also be placed on the wrong pipe, leading to a disturbed control sequence since the sensor is measuring the incorrect temperature [8, 18]. They may also be assembled incorrectly, so that they are mounted loosely to the pipe where they are measuring the temperature [14, 15]. The measurements from the temperature sensors may also be distorted by noise or drift (bias change) [19].

When having no hot water circulation, the water in the DHW system will be stationary and will cool down over time. The temperature sensor in the DHW system will sense this and send a signal to the controller to increase the temperature again. Since there will be a time delay before the newly heated water reaches the sensor, the control signal to increase the water will continue for longer than necessary. This causes the DHW supply temperature to overshoot [20]. Therefore, the temperature sensor measuring the supply temperature in the DHW system should be placed as close to the heat exchanger as possible [10].

3.3 Actuators

The actuators control the position of the control valves in the customer installation. The actuators are connected to the valve via a valve stem. Due to wear and tear, the connection between the stem and the valve may become poor, or the actuator may break down. If this happens, the control valve will be stuck in the position it was when the actuator broke down. This might lead to too large or too small flow in the installation, causing no response to the heat demand in the building [21].

The actuators have a number of different parameters that need to be taken into consideration when dimensioning the actuator for the installation. Among else, it is important that the actuator is dimensioned for the correct pressure so that it has enough driving force to change the position of the valve [10].

The actuators also have different running times, which means that they open and close the valves with different rates. In the DHW system, the actuators should have a short running time since the valves need to be able to open and close rapidly when a demand for tap water occurs. On the other hand, in the space heating system the actuators need to have a long running time since the temperature in the space heating system should change slowly to obtain the desired room temperature. These actuators may be interchanged by mistake at installation, which cause poor control of the both systems [14].

3.4 Control valves

The faults related to the control valves in the customer installation may arise if the valve is overdimensioned. This fault is especially common for the DHW system in older buildings, due to a historical practice to overdimension the valves. This causes an undesirably large change in the flow when the valve position changes, and that the valve cannot control small flows [13, 18]. The valve may also be too small for the flow that is needed in the installation, which means that it will not be able to obtain larger flows than the valve is able to transmit [18].

Since the space heating system is not used during warmer months, the valve on the space heating system will be in the same position during this time. This may cause the valve to get stuck in a closed position [22]. It may also happen that the valves seize in an open position or that they are completely stuck [14].

Valves are also susceptible to wear and tear. Over time, the disc of the valve may erode due to cavitation if there is a large differential pressure over the valve [23], causing the valve to leak in a closed position [14].

3.5 Internal heating system of the customer

The category of faults related to the internal heating system of the customer include a large variety of faults, since the design of the internal heating system varies between the installations. In some DHW installations an additional heat exchanger, a pre-heater, is installed in series with the heat exchanger seen in Figure 1. This solution is used to pre-heat the incoming cold water to maximize the heat exchange in the installation [8]. If this is the case, the circulation pump should be installed between the pre-heater and the final heat exchanger. If it is installed before the pre-heater, the return temperature of the installation increases [14].

If the DHW circulation is missing or broken, the water in the DHW system will be standing still in the pipes. As described in section 3.2 this may cause a poor control of the DHW temperature, especially if the DHW temperature is placed at a distance from the heat exchanger. It is also important that the flow in the DHW circulation is not too large, since the cooling in the heat exchanger will decrease.

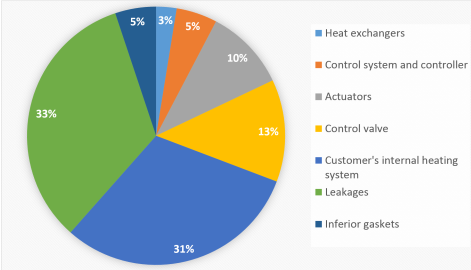


Figure 2: Distribution of the single most common fault in the utilities, sorted by category

The temperature demands in the customer installation is determined by the temperature levels for which the space heating system is dimensioned, and by the set point value that the DHW system requires to prepare hot water [10]. This means that if the temperature demands are high, maybe even higher than the temperature of the DH water, the flow on the DH side of the heat exchanger will always be large causing poor cooling and high return temperatures as a result [8].

When a building is connected to the DH system, the internal space heating system needs to be hydraulically balanced to make sure that the system is able to deliver the correct amount of heat to all heat delivery points in the system. This is done by adjusting the balancing valves in the system so that the opening position of the valves closest to the heat exchangers is smaller than the valves located further away in the building. This cause the flow to travel with the same flow rate as ti the ones closer to the substations, delivering heat to the more remote radiators. If this is not done correctly, the return temperature from the installation will increase [24].

The radiators in the space heating system are most commonly controlled by thermostatic radiator valves. If these are not working correctly, or are missing, the flow in the radiator system will be uncontrolled [15].

4 Results and analysis - Survey

The results from the survey study are based on the answers from the 56 utilities who answered the survey and that stated that they worked actively with faults in the customer installations. Three of the surveyed utilities answered that they do not work with the maintenance of the installations in their systems.

Figure 2 show the distribution of the faults that were reported to be the one most common fault in the utilities, sorted by category. As can be seen in the figure, there were two additional fault categories that emerged in the answers, that was not included in the survey from the beginning: leakages and inferior gaskets. Leakages include leaks in all parts of the customer installation and substation, and inferior gaskets include all gaskets that are used in the substation. The figure shows that the largest fault category according to the survey was leakages (33 %), closely followed by faults in the customers' internal heating systems (32 %).

As can be seen in Figure 2, the smallest overall fault category was the heat exchanger category. Figure 3 displays the fault distribution for this category. As can be seen in the figure, it is clear that leakages were experience to be the most common reason to faults in heat exchangers. The leakages include leaks that occur due to incorrect installation of the heat exchanger, and leaks in the pipings

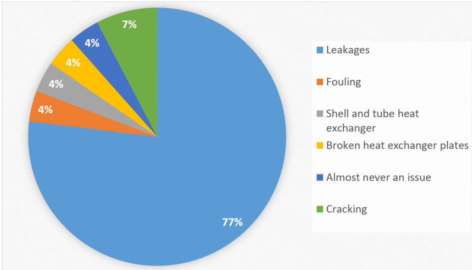


Figure 3: Distribution of the most frequently occurring faults related to the heat exchangers

connected to the heat exchangers. This indicates that the utilities may have answered "leakages" instead of "heat exchanger faults" when asked what category constitutes the largest fault category in their utility. A reason to this might be due to the way the question in the survey was formulated. When reading the word "heat exchanger", many people think of the entire DH substation, and not only the heat exchanger itself. It might be that the respondents have made this connection instead of thinking of the heat exchanger as one single component. The second most common faults was cracking of the heat exchanger plates (7 %), which the utilities have experienced when water hammers occurs in the system.

Faults in the control system and controller corresponded to approximately 5 % of the overall fault distribution. Although this was one of the categories that occur less frequently when regarding the overall distribution of faults, it was clear in the responses that the number of faults that may occur in this category is large and that many utilities do experience frequent occasions of faults in the control system and controller. Figure 4 displays the fault distribution of the faults related to the control system and the controller. As can be seen in the figure, 49 % of the utilities answered that the most common control system fault was that the controller was broken. The second most common fault was broken temperature sensors (16 %), and the third most common fault was temperature sensor giving the wrong signal (11 %). This indicates that a large share of the faults in the control system and controller was an effect of broken or poorly performing components. The utilities also experienced that incorrectly placed temperature sensors was a frequently occurring problem, corresponding to a total of 15 % of the faults (incorrectly placed outdoor temperature sensor, temperature sensors placed on the wrong pipe, and temperature sensors placed too far away from the pipe). All of these faults lead to poor control of the internal heating system of the customer, and they should all be recognized by the customer since they affect the customer comfort; either in terms of an unwanted change of the indoor climate if the temperature sensor is placed incorrectly, or in terms of unwanted DHW temperature levels. However, the utilities agreed that it was beneficial to be able to detect these faults at an early stage since they all lead to higher than necessary return temperatures. One interesting fault that emerged in the survey was that the customer itself may occur as a fault in the system. This problem occurred if the customer changed the set point values or control curve of the controller manually. The utilities experienced that this happened when the customers felt that their indoor temperature was too low, and changed the settings in order to try and increase the temperatures. In most cases, this did not give any effect since the low indoor temperature was due to faults located elsewhere in the installation and not due to how the system was controlled. The only effect of the change in the controller settings was that the flow through the substations increased, with increased return temperature levels as a consequence.

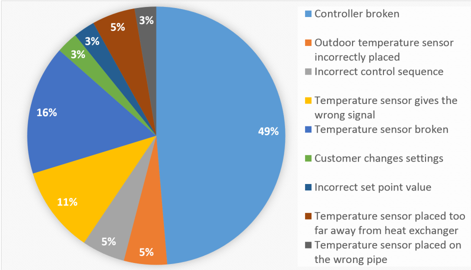


Figure 4: Distribution of the most frequently occurring faults related to the control system and controller

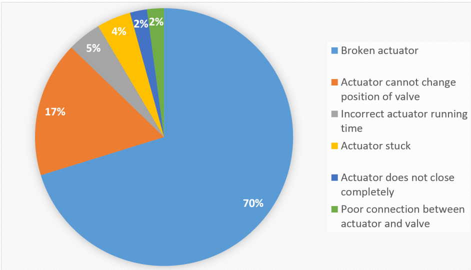


Figure 5: Distribution of the most frequently occurring faults related to the actuators

The fault distribution of the faults related to the actuators can be seen in figure 5. This category corresponded to approximately 10 % of the overall distribution of the fault categories. As can be seen in the figure, 70 % of the actuator faults were a result of broken actuators. Some of the utilities experienced that this fault may occur in the DHW system if the control valve it was connected to was too large. If so, the actuator had to operate constantly to attempt to control the flow, which caused the actuator to break down prematurely. The second most common fault for actuators was actuators that could not change the position of the valve, due to oversized valves or that the actuators were not dimensioned for the current pressure (17 %). The utilities experienced that this was a fault that the customer may not notice if it was located in the space heating system where the radiator thermostats control the final heat supply. Therefore, the utilities wanted to detect this type of fault at an early stage.

Figure 6 displays the fault distribution for the faults related to the control valves. In this category, overdimensioned control valves were the most common fault (35 %). However, some of the utilities never had experienced that this was a problem while some of the utilities experienced that this was their most common fault. The two second most common faults were control valves leaking in an open position (27 %) and control valves seizing in a closed position (27 %). The leaking control valves were most commonly due to wear and tear of the control valve itself, or in the connection between the control valve stem and the actuator. When this happened the valve could not close completely, causing a small seeping in the closed position. The seizing control valves most frequently

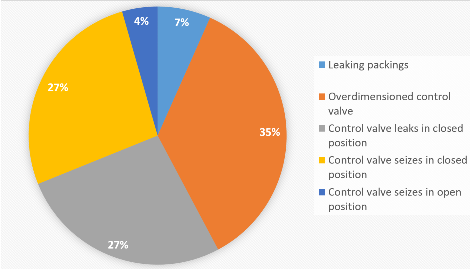


Figure 6: Distribution of the most frequently occurring faults related to the control valves

occurred in the space heating system at the beginning of the heating season, when the control valve was stuck in a closed position after the summer months when the space heating system was not used.

Figure 7 displays the fault distribution for the faults related to the customers' internal heating systems. The utilities experienced that the most common issue was poor balancing of the radiator system (32 %). Most utilities stated that the balancing of the radiator system was the customers' responsibility, but that they would be happy to advise the customers about the balancing. However, most of the utilities experienced that the main reason to poor balancing was that the customers were lacking knowledge of how to perform the system balancing and that they in some cases did not ask for professional help. The second most common fault was missing or broken radiator thermostatic valves (23 %). The utilities responded that many of the faulty thermostatic valves were located in older buildings where the valves had broken down due to wear and tear, causing poor space heating. Systems without DHW circulation was also mentioned as a frequently occurring issue (16 %), primarily in smaller houses. The consequences of this issue were increased return temperatures and poor DHW preparation. Another fault that was frequently occurring is that the set point values in the customers' systems were close to, or higher than, the DH supply temperatures (12 %). Some of the utilities mentioned that this issue may occur during the summer months due to low DH supply temperatures in the outer parts of the DH system, where it was harder to maintain higher supply temperatures. The utilities also stated that high customer radiator supply temperature set point values were an issue which led to a high overconsumption of flow in the installation, which indicates that it is necessary to find these installations as soon as possible.

50 of the 56 participating utilities utilized analysis of customer data in some way to detect deviating installation behaviour. There were a number of different ways to do this, and Figure 8 displays what the different utilities primarily measured to perform their analysis. The utilities had structured their analysis work in different ways. Some of the utilities performed analysis a few times per year, while other received instant alarms from their measurement system if a large deviation from the normal installation behaviour occurred. Others performed monthly controls of their customers' data to investigate whether the behaviour if the installation deviated significantly from its normal behaviour.

As can be seen in Figure 8, 19 of the utilities primarily performed analysis of the return temperature levels from their customers' installations to identify the poorly performing installations. The main method was to check the temperature levels, identify the installations with the highest return temperatures, and contact the customers in question. Some of the utilities administered the customer contact themselves, while others used subcontractors in the plumbing industry to con-

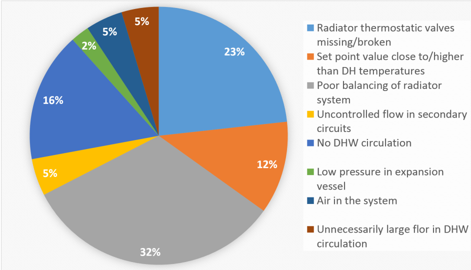


Figure 7: Distribution of the most frequently occurring faults related to the customer's internal heating system

tact the customer and investigate the installation further. Whether the utility performed the work themselves, or used a subcontractor, depended on the size of the DH system and if the number of employees at the utility allowed for doing this kind of maintenance work.

17 of the utilities stated that they performed analysis on a combination of measurement values from the customers' meters. This data was primarily used for billing purposes, but the utilities also utilized it for analysis. Nine of the utilities performed analysis of the cooling of the substation, while seven used the overconsumption, method to identify installations that was not performing well. The advantage of the overconsumption method s that it gives a possibility to rank the installations according to how large their impacts on the system are, since an installation with a larger heat demand will have a larger overconsumption if the installation is poorly performing in some way [11].

The utilities that investigated their customers' consumption (six utilities) often controlled how much heat their customers consumed during one month, compared to an expected heat consumption. This gave the utilities an indication that something may be wrong, but further investigation of the installation was needed since a change in the heat demand may occur due to natural reasons, e.g., new residents or an unusually cold month. However, the utilities also stated that in some cases it was not a change in the customer behaviour that had occurred; rather an incorrect heat demand estimation done by the utility.

As can be seen in figure 8, three of the utilities utilized so called QW values to investigate the performance of their customers' installations. The QW values are calculated by dividing the flow rate with the heat used in the building during, e.g., one month. This provided the utilities a figure that could be used to perform comparative analysis of installation data from installations of different sizes and heat demands.

5 Results and analysis - Interview study

The district heating utilities who participated in the interview study were all found to work with their customers' return temperatures in different ways. Some of the utilities put a lot of emphasis and work into reducing the costumers' return temperatures, while others considered their low return temperatures as an added bonus when working actively with other issues in their systems.

In general, the utilities agreed that it was most urgent to identify customers who had high return temperatures combined with a large heat demand. These customers were identified to have the largest impact of the system performance and were the ones that were most urgent to improve the

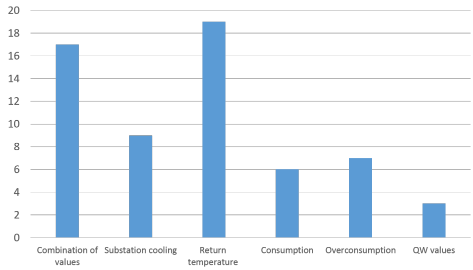


Figure 8: Distribution of different measurements used to identify poorly performing customer installations

performance of.

5.1 Reported incentives for the district heating companies

The interviews showed that there were different incentives for DH utilities to work with their customers' return temperatures. One common incentive was that the systems had flue gas condensation in their production facilities, which benefit from the low return temperatures since the efficiency of the heat production increases when the return temperatures from the DH system decrease.

There were also some incentives related to how the heat was used in the DH systems. In cases when a utility served as both heat distributor and heat transmitter, i.e. when a large amount of heat was transferred through the DH system each year to another system, it was important to decrease the customers' return temperatures mainly to prevent the need of increased pumping in the system.

5.2 Incentives for the district heating customers

From the interviews, it was clear that there are many different incentives that could be used to get the customers to address problems and faults in their installations. Some of the utilities utilized their relationships with their customers to create incentives, while other created incentives in terms of additional costs if the installations were not working correctly.

The overall most important factor to increase the customers' willingness to contribute to and work towards lower return temperatures was identified to be the relationship between the DH utility and their customers. The experience was that the customers were willing to pay the costs of correcting faults and problems in their installations, as long as the information about why and how was sufficient and easy to understand. This information was distributed to the customers in different ways, depending on what other measures the utility took to create incentives for their customers to work with their return temperatures. In the smaller DH systems, it was common practice to visit all customer installations once per year to examine the installations and substations on site. This provided the utilities with an excellent opportunity to talk to the customers and improve their customer relationships.

One common monetary incentive was to have a flow component in the basic price models for one or several types of customers. This meant an additional cost for the customers if a large amount of water passed through the substation during one month. The flow component did not constitute a large share of the final invoice, but the experience was that it was enough for the customers to become interested in different ways of reducing the flow through their substations. However, the experience was that the flow component itself was not enough to create an incentive for the customers; it was also of great importance to inform the customers of the reasons behind the flow fee and what they could do to decrease the amount of flow passing through their substation. Without this information, the flow component was seen as an additional cost only that the customers did not understand and therefore did not want to pay for.

In some companies the flow fee was used as a last and final action, if no other actions from the utilities encouraged the customer to improve the installation performance. If using the flow component in this way, the customers should first be informed about the high return temperatures and what this means to the customer and DH system, and be helped to find a solution to the problem. However, if the customer would not be susceptible to the information and proposed solution, a flow component might be introduced. The customer would then have to pay for the flow until the problem or fault in the installation has been solved. The utilities experienced that the need for this type of measures mostly occurred if the installation was not used by the owner of the installation but rented out to someone else. If so, the additional cost did not affect the owner him or herself since an increase of the rent for the tenant would cover the additional heating costs for the owner.

In general, the interviews showed that the customers were willing to work with the utilities to decrease their return temperatures in most cases, even though no clear incentive for each individual existed. However, there was a difference between the smaller and the bigger utilities where the smaller utilities experienced that it was easier to convince all of their customers to work with their installations. The smaller utilities also experienced that their customers were willing to work with their installations due to the collective responsibility to keep the DH prices down. The larger utilities experienced that it was relatively easy to target their largest customers, but stated that it was harder to work with and convince their smaller customers to improve their installation performance. This was both due to the large amount of man-hours would be needed to target all customer installations, as well as the fact that smaller installations have a smaller system impact when performing poorly. Hence, it is more cost efficient to target the larger customers first.

5.3 Access to and mandate of the maintenance of the district heating substations

One of the most important issues regarding the work with the customers' installations was to gain insight into how the installations performed, and get physical access to the customer substation. The general conclusion was that the utilities gained a lot by being allowed into the customers' properties to physically examine the substation. They also experienced that their customer relationships improved significantly since they not only gained physical access to the installation and substation - they also gained access to the customer. This meant that they were able to show and explain to the customer in person what happened in an installation if a fault occurred, and why it is beneficial to have an installation that performed as well as possible.

One common way of gaining access to the customers' installation was for the utilities to sign service agreements with the customers. This created a natural reason for technicians from the DH utilities to visit the customer's installation and make sure that the installation was performing well. The experience was that the outcome of the service visits was most beneficial when doing the visits at a regular basis, with at least one visit every other year. During the visits, the technicians investigated the performance of the installation by performing an inspection where the different parts of the substation were tested and made sure to work as they should. If something in the installation was broken, it was common practice to offer the customer some sort of solution. This could be that the utility replaced the broken component free of charge, or that they gave the customer an offer about the cost to replace the component. This gave the customer the possibility to decide if the utility should perform the work, or if someone from a plumbing company should get involved. The utilities stated that most customers chose to accept the offer from the DH utility rather than involving an external plumbing company.

Other ways of gaining access to the customers' installations at a regular basis was to offer the customers a free of charge survey of the installation when needed, and to allow the customers to call the utility at any time if something in their installations was not working correctly. This gave the customers a sense of security and trust towards the DH utility, which the utilities considered to be a big advantage in their customer relationships.

5.4 Analysis methods for finding poorly performing customer installations

During the interviews, it became clear that it was common practice to use different varieties of analysis methods to find the customers with undesirably high return temperatures. The methods varied in complexity, but common traits were that the methods were based on customer data from heat meters that were normally used for billing purpose. The different methods are summarized in the bullet list below and further explanation follows below.

- Monthly checks of customer billing data
- Quality index based on the installations performance
- Analysis of return temperature levels
- Analysis of flow
- Overconsumption method

The monthly checks of the customer billing data were used to identify whether a customer deviated significantly from its normal behaviour. In one utility, customer data was utilized to create a quality index which was based on the installation performance, as well as the location of the installation in the system. The quality index was then analyzed on a daily basis and if the index changed rapidly or the index pattern started to deviate from the normal behaviour, a service technician contacted the customer immediately to investigate the installation further.

The analysis methods were in most cases based on the temperature levels at the customer substation. The return temperatures or cooling performance of each individual customer installation were investigated, as well as the flow, to decide what customers to prioritize. The overconsumption method was also used by some utilities. The overconsumption can be described as the additional amount of DH water that has to pass through the substation in order to deliver enough heat when the installation contains a fault [8]. The overconsumption method takes into account the delta T of the substation, the flow through the substation, and the amount of heat being delivered to the customer. Hence, the customer with the largest overflow should be prioritized when using this method.

6 Discussion

The work conducted in this study clearly shows that high return temperatures and faults in the customer installations were viewed as important issues that needs to be addressed in order to enhance system efficiency in DH systems. It will also be of great importance to work with these issues in the 4GDH systems, to avoid unwanted increases of the system temperatures since lower system temperatures give less room for installations to deliver higher return temperatures than they are expected to. This means that faults in customer installations will have a larger system impact in the future systems. The interviews showed that the current DH utilities have a number of different work procedures for fault handling in customer installations that could be implemented in the future systems as well.

When investigating how the different utilities were working with fault handling today, it was clear that different utilities worked in different ways depending on size of the DH system, number of employees at the utility, and the amount of resources that could be allocated to this work. However, one of the most important aspects of the fault handling process was to create a close and good relationship with the customers. Although the utilities worked in different ways to obtain this, they all recommended to put an effort into providing their customers with clear incentives to why it is important to have a well performing installation and how to obtain this. In new and future DH systems, this could be done already from the start in order to create a good foundation to have well performing customer installations continuously during the operation of the DH system by signing service agreements with the customers and make sure to show the customers that the lower system temperatures are beneficial for the customers as well.

The utilities stressed the importance to gain physical access to the customers' installations. By achieving this, it was possible to gain insight into how the installations were performing and to further improve the customer relationship. If the utility also had a mandate to fix minor faults free of charge for the customer, such as replacing a broken temperature sensor, there were great possibilities to enhance the customer satisfaction while improving the overall system efficiency by eliminating faults in the installations.

The physical access to the customer installation was primarily achieved in two different ways: by signing service agreements with each individual customer, or to include a yearly inspection in the DH price. Some of the utilities included a small fee for the yearly inspections, while others made this free of charge. By including this in the price, the utilities had the possibility to visit all customers at least once a year. This gave them a large advantage in terms of access to and mandate of the customer installations, and gave large opportunities to identify and eliminate more faults in the customer installations. Therefore, it might be a good idea to include service agreements or inspections in the price model in the future DH systems.

From the answers from the interviews and the survey, it was clear that most utilities had different ways to find the poorly performing installations using customer data in different ways. The most common analysis method was to investigate the return temperature levels from the individual substations. This was due to the fact that most faults and problems in the installations caused higher flows and return temperatures.

Looking at the distribution of faults, the survey showed that leakages was experienced to be the most common category of faults, followed by faults in the customers' internal heating systems. These faults may be hard to detect during service visits, since the technicians may not be allowed by the customer to inspect all different parts of the installation. Therefore it would be beneficial to be able to detect these faults by investigating customer data. However, some of the faults may be hard to detect in the data that is available today, which in most cases is the data used for billing purposes. This indicates that it might be a good idea to implement more measurement points in the customer installations in the future systems. This is due to the fact that it might be very hard, maybe even impossible, to identify specific faults when only analyzing data from the DH system.

Regarding the results from the survey, it is important to keep in mind that the participating utilities were located in Sweden. This may have affected the results, since it is possible that some faults do not occur in the DH systems due to national conditions. One example of this is that fouling of heat exchangers seems to be a relatively unusual problem according to the utilities, while it is mentioned as one of the most common faults in the literature. One reason to that fouling is not as frequently occurring in the survey as expected may be that the overall water quality in the Swedish DH systems is high due to the measures that have been taken to improve the quality. Another reason may be that the study has been conducted in Sweden, where it is common practice to use compact plate heat exchangers which generally are less exposed to fouling.

The results from the survey may also have been affected by the response rate. 56 out of 139 utilities answered the survey, corresponding to a response rate of approximately 40 %. It might be that only utilities that experienced a frequent occurrence of faults answered the survey. However, when analyzing the survey responses it was clear that the participating utilities had different fault occurrence levels and that the responses also included utilities that did not experience a large amount of faults in their systems.

Regarding the results from the interview study, they should be seen as examples of how the Swedish DH utilities are currently working with fault handling in their customer installation. They may also be seen as good examples of how a utility may work with their customers and customer installations to keep the DH system temperatures down. It is possible that more interviews would have resulted in new responses. However, the last interviews did not produce any results that had not already occurred during the previous interviews. This indicated that the study was saturated, and that further interviews might not have contributed to a bigger picture of the issues investigated in this study.

7 Conclusions

This study has investigated how the DH utilities are currently working with fault handling in customer installations in order to reach lower return temperature levels. The study has also investigated what the most frequently reported faults are and how the utilities are working to identify these faults. The main conclusions of the study are described in the bullet list below.

- To obtain low return temperature levels from the customer installations, it is of significant importance to gain physical access to the installations. This gives the opportunity to not only fix minor faults, but also to get to talk to the customer face to face.
- Utilities with low return temperature levels are working close to their customers to obtain a good customer relationship. A common way to obtain this is to have service agreements with the customers or to include free of charge inspections in the DH agreement.
- It is important to create clear incentives to why the customer should work with faults in the installation. The incentives should be easy to understand and combined with substantial information.
- It may be beneficial for DH utilities to help the customer to identify and correct minor faults in the installation. This encourages the customers to contact the utilities the next time there is an issue in the installation which improves the fault detection rate.
- Most of the utilities utilized customer billing data in some way to identify poorly performing substations.
- The most common faults in the customer installations lead to higher flows and/or higher return temperatures.

- The most frequently reported fault categories in the study are leakages and faults in the customers' internal heating system.

References

[1] The European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - An EU strategy on heating and cooling. (COM(2016) 51 final), 2016. `https://ec.europa.eu/energy/sites/ener/files/documents/1_EN_ACT_part1_v14.pdf`.

[2] United Nations Environment Programme (UNEP). District energy in cities - Unlocking the potential of energy efficiency and renewable energy. Technical paper, United Nations Environment Programme, France, 2015.

[3] Henrik Lund, Anders N. Andersen, Poul Alberg Østergaard, Brian Vad Mathiesen, and David Connolly. Fromelectricity smart grids to smart energy systems - a market operation based approach and understanding. *Energy*, 42:96–102, 2012.

[4] Henrik Lund, Frede Hvelplund, Brian Vad Mathiesen, Poul A. Østergaard, Per Christensen, David Connolly, and et al. Coherent energy and environmental system analysis (CEESA). Research report, Department of development and planning, Aalborg University, Denmark, 2011. ISSN/ISBN: 978-87-91404-15-3.

[5] Brian Vad Mathiesen, Henrik Lund, David Connolly, Henrik Wenzel, Poul A. Østergaard, Bernd Möller, Steffen Nielsen, Iva Ridjan, Peter Karnøe, Karl Sperling, and Frede Hvelplund. Smart energy systems for coherent 100% renewable energy and transport solutions. *Applied Energy*, 145:139–145, 2018.

[6] H. Lund, S. Werner, R. Wiltshire, S. Svendsen, J. E. Thorsen, F. Hvelplund, and B. Vad Mathiesen. 4th generation district heating (4gdh): Integrating smart thermal grids into future sustainable energy systems. *Energy*, 68:1–11, 2014.

[7] S. Werner. District heating and cooling in sweden. *Energy*, 126:419–429, 2017.

[8] S. Frederiksen and S. Werner. *District Heating and Cooling.* Studentlitteratur, Lund, Sweden, 2013.

[9] K-M. Steen, U. Sagebrand, and H. Walletun. Att använda fjärrvärme i industriprocesser (To use district heating in industrial processes). Research report 2015:55, Svensk fjärrvärme AB, Stockholm, 2015. In Swedish.

[10] Svensk fjärrvärme. Färrvärmecentralen - utförande och installation (The district heating substation - implementation and installation). Technical regulations F:101, Svensk fjärrvärme AB, 2014. In Swedish.

[11] H. Gadd and S. Werner. Achieving low return temperatures from district heating substations. *Applied energy*, 136:59–67, 2014.

[12] Henrik Gadd and Sven Werner. Fault detection in district heating substations. *Applied Energy*, 157:51–59, 2015.

[13] Svenska Fjärrvärmeföreningen. Avkylningen i ett fjärrvärmesystem, (The cooling in a district heating system). Project report FVF 2000:03, Svenska Fjärrvärmeföreningens Service AB, Stockholm, 2000. In Swedish.

[14] Sven Werner. Feltyper i fjärrvärmecentraler strukturerade (Types of faults in district heating substations structured), 2013. Halmstad Högskola. In Swedish.

[15] Helge Averfalk. Morgondagens effektiva fjärrvärme: en beskrivande litteraturstudie (The efficient district heating of tomorrow), 2014. Halmstad Högskola. In Swedish.

[16] T. Reg Bott. Fouling. `10.1615/AtoZ.f.fouling`, 2011. Online. Accessed: 2018-10-25.

[17] Janusz Wollerstrand and Svend Frederiksen. Försmutsningsförlopp i plattvärmeväxlare fär fjärrvärmeabonnentcentraler (Fouling in plate heat exchangers for district heating substations). Report, Stiftelsen för värmeteknisk forskning, Stockholm, 1995. In Swedish.

[18] Kimmo Yliniemi. *Fault detection in district heating substations.* PhD thesis, Luleå University if Technology - Department of Computer Science and Electrical Engineering, 2005.

[19] Kimmo Yliniemi, Jan van Deventer, and Jerker Delsing. Sensor fault detection in a district heating substation. Proceedings of the 10th IMEKO TC10 International Conference on Technical Diagnostics, 2006.

[20] Karl Johan Åström. Reglerteknik – En elementär introduktion. Kapitel 8: Dödtidskompensering (Automatic Control Engineering – an elementary introcution. Chapter 8: Dead time compensation). Report TFRT-3173, Department of Automatic Control, Lund Institute of Technology (LTH), 1983. In Swedish.

[21] Avkylningsarbete på Göteborg energi AB 1995-2004 (Optimizing of the district heating system in Gothenburg 1995-2004 - practical application of R&D knowledge). Anders fransson. Report FOU 2005:132, Svensk Fjärrvärme AB, 2005.

[22] Markus Alsbjer and Peter Wahlgren. Fjärrvärmecentraler 10 år - håller de måttet? (District heating substations 10 years - Do they measure up?). Report 2011:9, Svensk Fjärrvärme AB, 2011. In Swedish.

[23] Pumpportalen.se. Kägelventil (Globe valve). `http://www.pumpportalen.se/kagelventil/`, 2015. Online, accessed 2018-11-30. In Swedish.

[24] Energimyndigheten. Energieffektivisering i stora värmesystem (Energy efficiency in large heating systems). Publication ET2013:20. In Swedish.

Paper ii

Automated Statistical Methods for Fault Detection in District Heating Customer Installations

Sara Månsson[1,2,*], Kristin Davidsson[1], Patrick Lauenburg[1], and Marcus Thern[1]

[1]Department of Energy Sciences, Faculty of Engineering, Lund University, P.O. Box 118, SE-221 00 LUND, Sweden
[2]VITO, Boeretang 200, BE-2400 MOL, Belgium
[*]Correspondence: sara.mansson@energy.lth.se; Tel.: +46-46 222 33 20

Abstract

In order to develop more sustainable district heating systems, the district heating sector is currently trying to increase the energy efficiency of these systems. One way of doing so is to identify customer installations in the systems that have poor cooling performance. This study aimed to develop an algorithm that was able to detect the poorly performing installations automatically using meter readings from the installations. The algorithm was developed using statistical methods and was tested on a data set consisting of data from 3 000 installations located in a district heating system in Sweden. As many as 1 273 installations were identified by the algorithm as having poor cooling performance. This clearly shows that it is of major interest to the district heating companies to identify the installations with poor cooling performance rapidly and automatically, in order to rectify them as soon as possible.

Keywords: Automatic fault detection, district heating, substation performance

1 Introduction

Sub-optimal performance of faulty components in the district heating (DH) systems of today forces the DH system temperatures to be above those theoretically needed to deliver enough heat to the DH customers [1]. The return temperature of the DH system These high temperatures cause heat losses in the system and prevents the utilization of low-grade heat such as industrial excess heat [2]. Finding the faulty components is key to increase the energy efficiency of the DH systems.

Previous studies concludes that there are a number of issues in the customers' installations that cause increased return temperature levels in the DH system [1, 3, 4, 5, 6, 7, 8]. The customer installation includes the internal heating system of the customer's building and the substation, which transfers heat from the DH system to the internal heating system.

The internal heating system is most commonly divided into two separate systems: the space heating system and the system for domestic hot water (DHW) preparation. Both of these systems include a number of different components, including control valves, temperature sensors, actuators, pumps, and piping. The space heating system also includes a heat transferring unit in the rooms that are to be heated, e.g. radiators or floor heating, and the DHW system includes faucets where hot water tapping occurs [9].

The substation design varies depending on different national requirements and/or traditions. In some countries, it is common practice to have a so called direct connection where the DH water is directly used in the internal heating system. In other countries, the indirect connection is the most common solution, where the substation separates the DH system from the customer's internal heating system. If an indirect connection is used, the heat is transferred to the internal heating system via one or several heat exchangers [1]. This study focuses on installations with indirect connections, which means that the substations include heat exchanger(s), temperature and flow sensors, control valves, pumps, controller and control system, and a heat meter which delivers data to the DH utility that operates the DH system [10].

The issues, or faults, in the customer installations may occur in a number of different components and include faults and problems such as fouling of heat exchangers, broken temperature sensors, control valves stuck in an open position, temperature sensors placed on the wrong pipe, DHW circulation connected before the pre-heater, high return temperatures from the customer's internal heating system, and high set point values in the customer system [1, 5, 6, 7, 11, 12, 13]. All of these faults may not be seen as an actual fault where something is broken, but they still lead to high return temperatures.

As the DH utilities are now aiming to decrease their overall system temperature, it becomes increasingly important that the vast majority of the faulty installations are detected and corrected. Traditionally, the poorly performing installations have been detected using manual analysis methods [1]. Due to the vast amount of customer installations in the DH systems, this can be a time-consuming task. Hence, the main focus has been to detect the installations that has the largest impact on the system energy efficiency. This has primarily been done by investigating the excess flow, or overflow, i.e. the difference between the anticipated ideal flow and the actual flow in the installations [1, 14, 15]. A large overflow indicates that the customer installation is poorly performing, which means that the installations with the largest overflow have been prioritized when conducting the system analysis.

In order to increase the detection rate, the DH utilities can utilize the increasing amount of customer data that has become available during the last years due to different national legislations and international directives, e.g., the Energy Efficiency Directive adopted by the EU [16]. This data is collected using the meter device of the substation, and includes supply and return temperature levels, volume flow rate, and the heat amount being used by the customer [1]. The data is usually used for billing purposes, but contains a lot of information about how the customer installations are performing in terms of energy efficiency. This indicates that it is also possible to detect customer installations that are not performing as expected using this data. By investigating the measurements originating from the installations, it is also possible to detect the faults shortly after they have occurred.

Using this customer data, Gadd and Werner used a statistical fault detection method that uses the temperature difference signature of the substation to detect temperature difference faults [3]. In the study, 140 substations were investigated, out of which 14 were found to be poorly performing. In a subsequent study [4], Gadd and Werner investigated data from 135 different substations manually. In this study, the results showed that 74 % of the substations contained faults, including unsuitable load patterns, low annual average temperature differences, and poor control of the substations. A similar approach was taken by Sandin et. al. who used statistical methods, e.g., limit checking and outlier detection, in order to identify poorly performing installations [8].

A different fault detection approach was presented by Johansson and Wernstedt, using visualization methods to show the operational functionality of individual installations [17]. In order to aid the subjective visual interpretation, a number of performance metrics describing the relationship between different installation variables were calculated. Xue et. al. proposed a fault detection method based on data mining techniques [18]. The authors investigated data from two installations using cluster analysis and association analysis. The combination of the two methods generated a set of association rules that should be fulfilled in a well performing installation.

Yliniemi et. al. presented a method of detecting faults in temperature sensors in a DH substation [19]. The method was developed in order to detect increasing noise levels in the sensor readings. Zimmerman et. al. presented a method capable of detecting faults in pressure sensors and leakages in the DH systems using a Bayesian Network [20]. Pakanen et. al. uses a series of different methods in order to detect faults in the entire customer installation, as well as in three different components of the substation: a control valve, a heat exchanger, and a mud separating device [21].

The previous studies show that there has been some success in obtaining fault detection methods that can detect the poorly performing customer installations. However, many of the presented methods require an amount of manual handling and/or interpretation. Some of the methods also presumes a certain amount of understanding for more advance computer science and data handling methods. The focus of this study has been to try and eliminate as many manual stages as possible in the fault detection process. The focus has also been to utilize knowledge that is already commonly used in the DH industry to create a fault detection tool that is easy to understand and interpret.

This study presents an automated statistical method for fault detection, which has been developed to detect faults automatically in larger data sets. The method utilizes linear regression and outlier identification, testing the performance of the installation in terms of temperature difference over the installation, the return temperature from the installation, and the heat extracted in the installation.

Further, the method is tested in a case study using real data from a DH system located in Sweden. The data set consists of hourly measurements during one year from 3 000 customer installations. The results from the case study shows that 43 % of the customer installations in the data set are poorly performing.

2 Performance of the customer installation

When investigating the performance of the customer installation, a number of different measurements are of interest. One of these is the temperature difference between the primary supply and return temperature in the installation. This temperature difference is often called the cooling, or deltaT (ΔT), of the installation. The cooling depends on both the supply temperature of the DH system, and the return temperature from the customer's internal heating system. A DH system with higher supply temperature naturally has a larger delta T if the installations are able to cool the DH water well. However, high ΔT's in a system with high supply temperatures do not necessarily mean that all customer installations are performing well in terms of cooling. When having high supply temperature levels, it is also important to investigate the return temperature levels from the installations to see how well they are actually performing. An installation with high return temperatures from the customer's internal heating system will not be able achieve a good cooling performance, since the return temperature level from the installation will never fall below the return temperature level in the internal heating system [1, 3, 10].

Figure 1 displays the cooling pattern as a function of the outdoor temperature for a well performing installation. The grey circles in the figure represent the average cooling in the installation during one day. The red line represents the average cooling in the same installation as a function of the outdoor temperature. The data that the figure is based on originates from a customer located in a city in the south of Sweden where the climate is relatively mild. This explains why figure 1 does not display any values for outdoor temperatures below -5 °C. The average supply and return

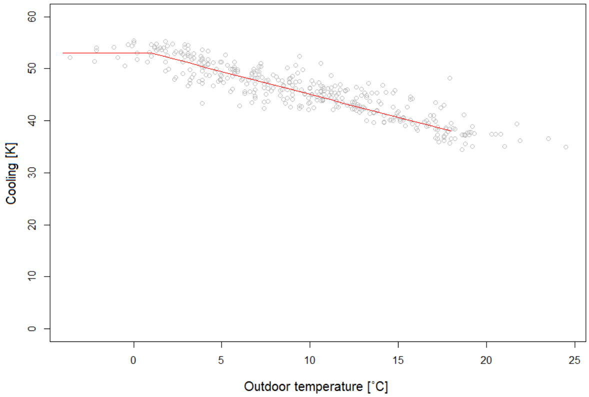


Figure 1: Daily averages of the cooling values for a well performing DH installation

temperatures for the DH system that the installations is located in are 90 °C/49 °C. Hence, the temperature levels of this DH system are quite high and the DH utility is currently working in different ways to obtain lower system temperatures, including performing analysis of the cooling of the customer installations.

As can be seen in the figure, the daily cooling values display a somewhat scattered structure, and the value of ΔT is not necessarily the same for days with the same outdoor temperature. This reflects the fact that the heat demand in the same installation varies from day to day, primarily depending on the varying need for DHW preparation [22].

As can be seen in the figure, the cooling of this customer installation fluctuates around 50 °C values for outdoor temperatures below 0 °C and then decreases linearly for higher outdoor temperatures. This is visible in the figure where the inclination of the cooling curve (red line in figure) increases for outdoor temperatures above 0 °C. The scattering of the daily cooling values also increases. For temperatures above approximately 15 °C, the need for DH decreases since there is no need for space heating. DH is merely used for DHW preparation at these temperatures, or for industrial processes that require heat during the entire year. This means that the amount of heat that the customer uses is less temperature dependent. Days with an outdoor temperature above 15 °C are in this study called non-heating days. The temperatures where the cooling curve inclination starts to decrease, and the space heating ceases, are called breakpoints. The breakpoints may have other values than in this study, depending on which DH system is being investigated. There may also be more than one breakpoint, depending on the nature of the cooling pattern.

Another measure that may be considered when investigating the customer performance is the heat that is transferred to the building, i.e. the heat that is needed to meet the heat demand of the customer. This heat demand can be calculated according to [1]:

$$P_d = \dot{V}\rho c_p(T_s - T_r) = \dot{V}\rho c_p \Delta T \,[\mathrm{W}] \quad (1)$$

where

$\dot{V}$ = volume flow [m^3/s]
ρ = density [kg/m^3]
c_p = specific heat capacity for water [J/(kg·°C)]
T_s = supply temperature [°C]
T_r = return temperature [°C]

The heat being transferred is proportional to the volume of water that passes through the substation and the ΔT of the installation. This implies that an increase of one of these parameters would enable a decrease of the other. As it is highly desirable to increase the ΔT, the focus should be on improving the value of ΔT for as many individual customer installations as possible. Furthermore, it is highly desirable to lower the volume flow, since a lower flow rate results in a lower pressure drop in the distribution pipes [1]. The reduced pressure drop allows for smaller pipe dimensions to be used, which in turn may reduce the building costs of the DH system. A decrease of the pressure drop will also mean that less pump work is needed in the systems [1, 23].

The decreased volume flow may also be beneficial for the DH customers. Many DH utilities have decided to include a flow price component in their price models for some or all of their customers [24]. This means that the customers pay for the amount of water mass flow that passes through the customer substation each month, and it is often introduced as an incentive for the customers to improve the performance of their installation [25].

As mentioned in section 1, one of the historically most common methods to identify poorly performing customer installations has been to calculate the overflow, or overconsumption, of DH water for each installation. According to equation (1), an installation with a smaller, and hence poorer, value of ΔT will need a larger volume of heat medium to pass through it in order to extract the same amount of energy as an installation with a larger value of ΔT. The extra amount of heat medium is the overflow, which can be calculated using equation 2.

$$overflow = V_{actual} - V_{ideal} = V_{actual} - \frac{E_{actual}}{\rho \cdot c_p \cdot \Delta T_{ideal}} \left[m^3\right] \tag{2}$$

where

V_{actual} = actual annual volume [m^3]
V_{ideal} = ideal annual volume [m^3]
E_{actual} = actual annual energy [J]
ρ = fluid density [kg/m^3]
c_p = specific heat capacity [J/(kg · K)]
ΔT_{ideal} = ideal temperature difference [°C]

ΔT_{ideal} is the ideal temperature difference between the supply and return temperature when the installation is performing ideally. This value may vary depending on what DH system is being investigated, since it depends on the temperature difference between the return and supply temperatures. In this study ΔT_{ideal} was chosen to be 45 °C, in accordance with [3].

3 Description of fault detection method

3.1 Data Set

The data used in this study was gathered from the business system of a DH utility in Sweden. The time interval was April 2015 - March 2016 and it was important to use data from one year since the heat loads of the customer installations are different throughout the year. The data set included data from the 3 000 installations that had the largest energy consumption in the system. The installations with the largest energy consumption should be the ones with the largest system impact, and these should be prioritized when performing maintenance work on the DH system in order to improve the system performance. The data set consisted of installations with a wide variety of purposes. The internal heating systems were designed differently for the different buildings depending on purpose of the installation. Hence, no general information was given regarding what specific equipment was used in the space heating and DHW system, e.g. the type of room heaters and if the installations had DHW circulation.

The investigated data set contained values for energy consumption, volume passing through the installation, return temperature and supply temperature for each of the 3 000 installations. In addition, the installation ID for the installation where the installation was located, the postal code of the installation, and the outdoor temperatures during the investigated time interval was collected from the business system of the DH utility.

The DH system from which the data was gathered was said to contain weak and non-weak areas. The weak areas were areas known to have a low differential pressure, which means that in these areas it might be harder to deliver heat to the customer installations. This is due to the fact that there has to be a certain amount of differential pressure over the installation for the control valves of the installation to work as they should. If the differential pressure is too low, the installation will not receive the amount of heat that is needed at the installation's maximum heat load since the control valves will not operate as they should [10]. The weak areas were well known by the company, and their locations were based on measurements, calculations and simulations of different operation modes of the system. A poorly performing installation had a larger system impact on the DH system if it was located in an area with low differential pressure, than an installation in an area with high differential pressure. Therefore, it was important to be able to distinguish between the installations located in the different areas, and in this study, the postal code of the area where the installation is located was used to determine if it was located in an area with high differential pressure, or in an area with low differential pressure.

3.2 Data preprocessing

The data set being investigated in this study originally consisted of hourly values for energy, volume, return temperature and supply temperature for 3 000 customer installations in a DH system located in Sweden. Since the hourly values were found to contain some errors which affected the analysis result, the data was first investigated in order to find the errors and handle them according to the nature of the error. After this, the hourly meter readings were converted into daily values as follows: the volume and energy were calculated as the total daily heat use, the supply temperature was calculated as the daily mean value of the measured values, and the return temperature was calculated from equation 1. The reason to why the return temperature was calculated in this way was that the measured values of the supply temperature from this DH system were often more reliable than the measured values of the return temperature. This meant that the return temperature values would be more reliable if calculated from 1. These daily values were then analyzed in order to find the poorly performing installations. A schematic illustration of the data handling process, from the hourly values to the results from the fault detection algorithm can be seen in figure 2, and table 1 displays the data included in the data set. The additional data consists of the information about the customers and investigated time interval that were described in section 3.1.

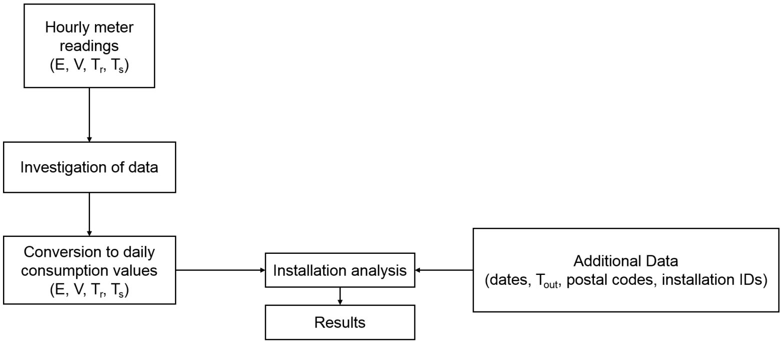


Figure 2: Schematic view of the data handling process

Table 1: Data included in the data set.

Quantity	**Unit**	**Time Aspect**
Energy	kWh	Daily
Volume	m^3	Daily
Cooling	°C	Daily
Return temp.	°C	Daily
Dates	-	Daily
Outdoor temp.	°C	Daily
Installation ID	-	Fixed
Postal Code	-	Fixed
Weak area	-	Fixed

3.3 Choice of Programming Language

The automatic fault detection algorithm developed in this study was developed in the programming language R. R is mainly used for statistical computing and graphics, and include functions for linear and non-linear modelling, time-series analysis and clustering [26]. Furthermore, R is an open source program and it is therefore easily accessible to the public.

3.4 Three Signatures

To determine if an installation was poorly performing, three different criteria were used: the cooling performance, the return temperature level, and the energy consumed in the building. The three criteria were used to create three signatures; one cooling signature, one return temperature signature, and one energy signature. The signatures consisted of a reference case, and threshold values which were used for outlier detection.

Since this study primarily focused on identification of installations with poor cooling performance, the cooling and return temperature signatures were considered to be of greater importance than the energy signature. This meant that the installation did not necessarily have outliers according to the energy signature to be considered as a poorly performing installation. The energy signature was mainly developed in order to provide the user of the analysis tool with extra information regarding the installations which were identified as poorly performing.

3.5 Reference Cases

The reference cases of the three signatures were created using data from customers in the investigated data set. This made sure that the reference cases were representative of the system in question. The reference cases contained data from the installations with lowest overflows in the data set.

The reference cases for the cooling and energy signatures were created using piece-wise linear regression with one breakpoint for the non-heating days. Piece-wise linear regression is typically used to model the relationship between two or more variables in large data set, where it may be hard to find a linear regression model that explains the relationship well for all data. When this is the case, the data set may be divided into smaller segments of data and fit a linear regression model to each of these segments. The segments are divided at the breakpoints, and the combination of the individual regression models is the final piece-wise linear regression. The mathematical relationship for a piece-wise linear regression model with one breakpoint H may be described according to equation 3 [27].

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i - H) \cdot I(X_i, H) + \epsilon_i, \quad where\ I(X_i, H) = \begin{cases} 1 & X_i > H \\ 0 & X_i \leq H \end{cases} \tag{3}$$

In the equation, β_n, n=0,1,2 is the parameters of the regression model, Y_i is the dependent variable being modeled, X_i is the independent variable which is used to model the dependent variable, and ϵ_i is the model error. In the cooling signature the cooling of the substation was modeled as a function of the outdoor temperature, and in the energy signature the energy consumption was modeled as a function of the outdoor temperature.

The breakpoint for the piece-wise linear regression was determined by visually inspecting the data set. As could be seen in figure 1, there was a clear change of the inclination of the cooling patterns for lower outdoor temperature levels. This pattern was also visible in the energy consumption patterns. However, when comparing the energy consumption patterns it was important to keep in mind that different customers consume different amounts of heat. Therefore, it was important to scale the heat consumption so that they were on the same scale. This was done by dividing all heat consumption values for one customer with the maximum heat consumption for that individual customer. After this was done, the breakpoints were determined to assume the values of the outdoor temperature where the inclination of the two patterns changed. This meant that one linear regression modelled the behaviour from the lowest outdoor temperature of the data set to the breakpoint temperature, and the other linear regression modelled the behaviour from the breakpoint temperature to the temperature at which the non-heating days started.

The values for temperatures corresponding to non-heating days were treated separately from the rest of the values. For the cooling signature, the summer days were not considered, since the cooling values for these days only gave information about the size of ΔT for DHW preparation in the installations. This may vary greatly from day to day depending on the need for hot water, and hence the analysis of ΔT for summer days is not of interest in this study. For the energy signature, the median of the values for summer days was used to create a constant threshold for outlier detection.

The pattern of the return temperature was completely different from the pattern for the cooling and energy signatures, and linear regression was not a good choice of method for the return temperature reference case. Instead two constant thresholds were calculated, one for heating days and one for non-heating days. Both of the thresholds were calculated as the mean of the return temperature for the reference case installations.

3.6 Identification of Deviating Values

The performance of the installations not included in the reference cases was investigated to identify deviating behaviours and values, so called outliers. Outliers may be described as values deviating significantly from the expected behaviour of a certain parameter [28]. The outliers were identified using the mean and the standard deviations of the reference case values. If the distance between the mean and a certain value was larger than three standard deviations for a certain outdoor temperature the value was considered to be an outlier, in accordance with [8, 29]. This created thresholds located ±3 standard deviations from the mean. For the cooling and energy signatures, piece-wise linear regression was used to model the mean of the reference case and so the thresholds were also linear. For the return temperature signature the mean was modeled using a constant value, resulting in constant thresholds. The three resulting signatures can be seen in figures 3-5. The reference data set for the return temperature included some measurements that were unreasonable considering the nature of the data. As may be seen in figure 4, the return temperature levels sometimes fell below 0 °C, which is not plausible in DH systems. These values have appeared when the return temperature was calculated, as described in section 3.2. Since the values were calculated in this way, these values were not considered further in the analysis.

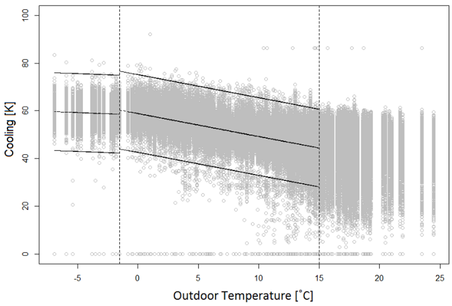


Figure 3: Cooling signature (black lines) and the cooling values for the substations in the reference case (grey circles). Values located outside the black lines may be considered to be outliers.

4 Description of fault detection algorithm

The algorithm developed in this study consisted of a main function and several different functions which performed different parts of the analysis. The structure of the program can be seen in figure 3, and the different functions will be described in detail in the following sections.

4.1 Main algorithm

The main algorithm, *mainAnalysis*, loaded and structured the data which was used to perform the analysis. There were also a number of constant parameters that were needed to perform the analysis, e.g., the ideal cooling values. These parameters were determined in *mainAnalysis*, and can be find in the following bullet list.

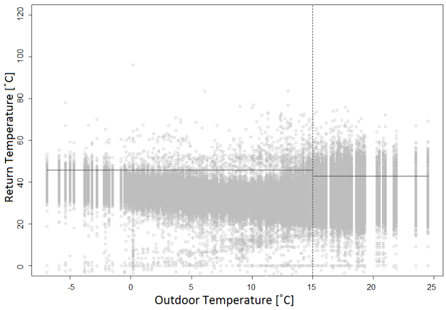


Figure 4: Return temperature signature (black lines) and the return temperature values for the substations in the reference case (grey circles). Values located outside the black lines may be considered to be outliers.

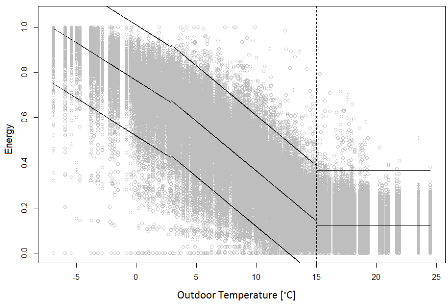


Figure 5: Energy signature (black lines and the energy values for the substations in the reference case (grey circles)). Values located outside the black lines may be considered to be outliers.

- The outdoor temperature after which the days were considered to be non-heating days. In this study, the non-heating days occurred for outdoor temperatures higher than 15 °C.
- The ideal cooling value. In this study, ΔT_{ideal} was chosen to be 45 °C.
- The specific heat capacity of the heat medium of the system. In this study, the heat medium was water and so c_p= 4185.5 J/(kg·°C) was used.
- The number of breakpoints that should be used for the linear regression, in this study one breakpoint due to the nature of the heating and cooling patterns (figure 1).
- The temperatures which should be used as breakpoint values for the linear regression. In this study two different breakpoints were used, one for the energy signature and one for the cooling

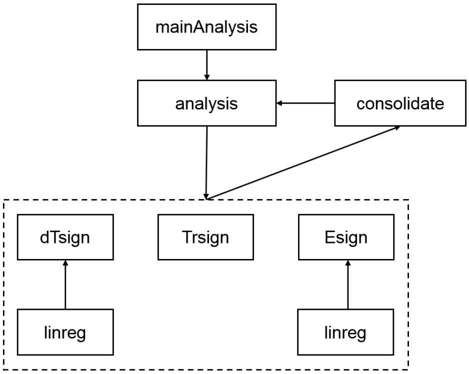


Figure 6: Structure of the algorithm used for identification of poorly performing customer installations

signature. The breakpoint for the energy signature was located at 3 °C, and the breakpoint for the cooling signature was located at -1 °C.

- The threshold values for the return temperature signature, for heating days 45.7 °C and 42.4 °C for non-heating days.
- The share of installations that should be used in the reference case. 25 % of the substations were used for the reference case in this study.
- The number of outliers that were acceptable before the installation was considered to be poorly performing in each signature, in this study 0 outliers was used.
- The number of standard deviations which were used to create the reference case. In this study, 3 standard deviations were used.

Some of these parameters were system specific, and will have different values depending on which DH system is being investigated. In *mainAnalysis*, it is also possible to determine if all three signatures should be investigated when running the analysis program. This made it possible to perform analysis of one, two, or three of the signatures, depending on what results the user was interested in.

4.2 Analysis function

The first subfunction, *analysis*, was to developed to perform a number of different tasks. The first was to identify which days that were heating days and which were non-heating days, and divide the entire data set in two subsets according to if the data belonged to a heating or a non-heating day. The next task of *analysis* was to determine what installations to include in the reference case. This was done by ranking the installations according to their overflow, and using the share of installations that was determined in *mainAnalysis* and had the lowest overflows. The next task of *analysis* was to determine the number of outliers in each of the three signatures. This was done using three separate functions, *dTsign*, *Trsign* and *Esign*. The function *consolidate* compiled the results and a created a list of the results. This was returned to *analysis*, which in turn returned the results to *mainAnalysis* from which the results were presented to the user.

4.3 Linear Regression

The function *linreg* was used to create piece-wise linear regression models of the reference case values, and calculating the standard deviation values of the reference cases in the energy and cooling signatures. The breakpoints were placed at the predefined breakpoint value between heating and non-heating days. The parameters of the linear regression models were approximated using the least squares method. The non-heating days were treated differently in the different signatures, and this will be described in further detail in sections 4.4 and 4.6.

4.4 Cooling Signature

The function *dTsign* was implemented to create the cooling signature. A linear model of the reference case values, as well as two linear thresholds, were created using *linreg*. The linear model of the reference case and the linear thresholds constituted the cooling signature of the analysis program. The values of the non-heating days were not included in the cooling signature, and were not considered when investigating the number of outliers in the signature. To find the outliers, *dTsign* compared the cooling values of each individual installation to the cooling signature and used the linear thresholds to identify the installations which had outliers according to the cooling criterion. The results were compiled in a matrix containing information about the installations and the number of outliers they had according to the signature.

4.5 Return Temperature Signature

The return temperature signature was implemented in a function called *Trsign*. In the function two constant thresholds were used to perform the outlier detection, one for the heating days and one for the non-heating days. The thresholds were calculated using the mean value of the return temperatures for the reference case installations. To identify the outliers, the return temperature values of each individual installation was then compared to the constant threshold values and the results were compiled in a matrix.

4.6 Energy Signature

The function *Esign* was implemented to identify installations that had outliers in the energy signature. The function was similar to the function *dTsign*, described in section 4.4, but some additional features were added. As described in section 3.5, the different installations consumed different amounts of heat depending on the size of the building and what the building was used for. To enable a comparison of the different installations in the data set, the heat consumption values were first scaled so that they were of the same order of magnitude.

After scaling the values, a piece-wise linear regression was created for the energy reference case for the heating days using *linreg*, and the standard deviation for the reference case substations was calculated. For the non-heating days, the median of the heat consumption values was used to create a constant reference case, and linear thresholds were calculated for the heating days using the standard deviation of the reference case. *Esign* then identified outliers for both heating and non-heating days and the results were compiled in a matrix.

4.7 Compilation of Results

The functions *dTsign, Trsign* and *Esign* returned the results matrix to the function *consolidate*, which compiled the results from the three signatures to identify the installations that were poorly performing. When running the algorithm with all three signatures, the cooling and return temperature signatures were primarily considered when compiling the results (as described in section 3.4. This resulted in that the compiled result list possibly contained installations that had no outliers at all according to the energy signature. If only two signature functions have delivered result lists, the *consolidate* function created the final result list from the available results. If the algorithm only investigated the installations using one of the signature, *consolidate* used the results matrix from that signature function. When more than one signature was used, *consolidate* compared the installation IDs in the result lists from the different signature functions. If the installation appeared in more than one list, it was included in the final result list with poorly performing installations.

Due to the nature of the DH system that the data originated from, two result lists were created. One contained installations located in the weak areas of the system, and one contained installations located in the non-weak areas of the system. The installations in the list were ranked according to the overflow of the installations. The list contained the installation IDs, the number of outliers for each of the three signatures, if the installation was located in a weak or non-weak area, and the overflow of the installations. The compiled list was then returned to the *analysis* function and presented to the user in the *mainAnalysis* function.

5 Results

5.1 Output from Analysis Algorithm

The customer installation analysis algorithm rapidly identified the poorly performing substations, even though it investigated a large amount of data (3 000 installations with 365 values for each variable plus additional data). The total number of installations being identified as poorly performing was 1 273, corresponding to approximately 43 %.

Tables 2 and 3 display the five installations that were identified by the algorithm to be the most poorly performing installations in the weak and non-weak areas. In order not to violate the customer privacy, all installations IDs and their corresponding postal codes have been anonymized. As can be seen in table 3, installation number nine did not have any outliers in the energy signature but was still considered to be poorly performing due to the large amount of outliers in the other two signatures. 109 customer installations were identified in the weak areas, and 1 164 installations were identified in the non-weak areas.

Table 2: The five most poorly performing customer installations in the weak areas.

Installation number	**1**	**2**	**3**	**4**	**5**
Installation ID	XXXXXX19	XXXXXX61	XXXXXX77	XXXXXX26	XXXXXX91
Outliers dT	243	210	148	112	232
Outliers Tr	312	354	255	136	330
Outliers energy	4	4	71	78	9
Weak area	Yes	Yes	Yes	Yes	Yes
Overflow [$\mathbf{m^3}$]	21 346	17 948	17 089	12 652	12 641

Table 3: The five most poorly performing customer installations in the non-weak areas.

Installation number	**6**	**7**	**8**	**9**	**10**
Installation ID	XXXXXX54	XXXXXX93	XXXXXX45	XXXXXX41	XXXXXX21
Outliers dT	277	277	169	201	277
Outliers Tr	366	366	316	363	366
Outliers energy	131	123	18	0	38
Weak area	No	No	No	No	No
Overflow [m^3]	125 581	102 885	64 824	57 121	55 927

5.2 Well and Poorly Performing installations

Figures 7-9 gives a visual representation of one well performing and one poorly performing customer installation. The red circles in the figures are values that originate from the poorly performing installation and the blue circles are values that originate from the well performing installation. The reference case values are represented as grey circles, and the signature for each case is also visualized in the figures (black lines). As can be seen in the figure, the values of the well performing installation do not appear as outliers, while a large share of the values for the poorly performing installation appears outside of the thresholds determined by the functions.

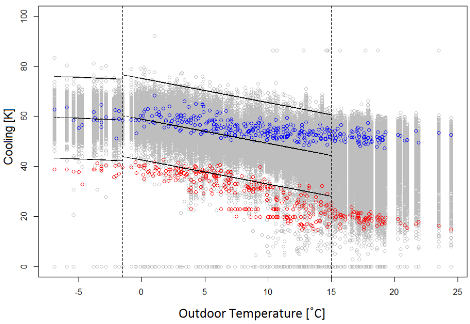


Figure 7: Example of one well performing (blue circles) and one poorly performing (red circles) substation according to the cooling signature

6 Analysis and Discussion

The results from the customer installations analysis algorithm states that 43 % of the total amount of the investigated installations are performing poorly. This indicates that it is possible to make large improvements in the overall DH system performance, if the poorly performing installations are fixed. In the future DH systems this will be necessary to maintain the low temperatures that enable the use of other and more efficient heat sources. Therefore, some sort of automatic installation analysis method will most likely be a necessity in all future DH systems.

Even though the entire system was not investigated, the number of investigated installations is

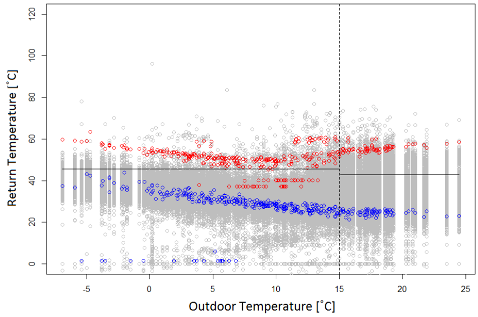


Figure 8: Example of one well performing (blue circles) and one poorly performing (red circles) substation according to the return temperature signature

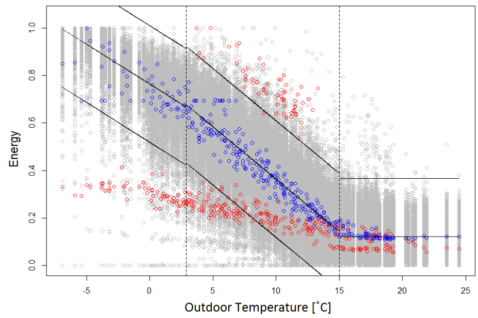


Figure 9: Example of one well performing (blue circles) and one poorly performing (red circles) substation according to the energy signature

larger than those of previous studies. The use of the analysis tool could easily be extended to an entire DH system by using larger data sets. The results from this study clearly shows that it is possible to identify poorly performing DH customer installations in large data sets, without having to perform the time consuming manual analysis that a lot of DH companies are relying on today.

Regarding the results in tables 2 and 3, it is clear that the installations presented all have a large overflow. Two of the installations in table 3 have overflow values over 100 000 m^3, which means that a very large additional amount of DH water is passing through the installation each year. These installations are located in the non-weak areas of the system and so have a smaller impact on the system performance than they would in a weak area. However, the large overflows are a clear indication that there is room for significant improvements of the system efficiency if action is taken to fix the installations in question. From table 2, it is clear that there are a number of installations located in weak areas that have a very large overflow per year. Due to the increased amount of flow that passes through the installation to meet the heat demand of the customer, the pressure drop over the installation will increase. This means that the areas where it is already hard to maintain the differential pressure levels will be even more affected by installations that are not working as they should, giving further incentives that the installations in the weak areas should be fixed as soon as possible.

The analysis method used in this study is based on customer data and patterns that are well familiar to the DH utilities. This is a big advantage since the method is easier to understand compared to other, more advanced computer science methods. When developing a new analysis method, it is important to keep the end user in mind and by utilizing what is already well familiar within the industry, it will be easier to also implement the developed method and the understanding for it would be more intuitive. For example, figures similar to 7-9 give a nice overview of the performance of the installations using patterns that are well familiar to people in the DH industry, and could be used to present the results to the end user of the algorithm.

The share of poorly performing installations that was identified in this study deviates from the shares that have been identified in previous studies. For example, Gadd and Werner show that 74 % were poorly performing [4]. One of the reasons to why this is the case, might be that the definition of what is considered to be a well performing installation is different in the different studies. This means that the reference case will have different properties, depending on the definition of a well performing installation. This will have a large impact on the result, since the reference case values decide the threshold lines. If the reference case contained less installations, it might be that the standard deviation of the reference case values is smaller, hence creating a narrower reference case signature. A narrower signature results in more installations being identified as poorly performing.

Another reason to the deviation might be that a larger data set includes more installations of different purposes which uses heat differently during the day. The different building types have heating patterns that might differ largely from each other due to which type of heating control system that is used in the buildings. Some types of buildings have a very wide heat pattern, e.g. office buildings, since their daily average heat demand is significantly lower during days when nobody is present in the premises. This means that an office building can have a wide range of different values for one single outdoor temperature, and this causes a wider heat pattern. If the reference case includes installations with these wide heat patterns, they may contribute to the reference case being wider than it would be if only installations with a narrow heat pattern were to be included. Since the data set in this study contains a larger number of installations compared to the previous studies, it is likely that the reference case in this study contains a larger variation of buildings. This indicates that the thresholds in the reference cases of this study will be more allowing than for a smaller data set, and therefore less values in the data set will be identified as outliers. However, the number of installations that was identified as poorly performing in this study is still very high. This clearly indicates how important it is to be able to identify the poorly performing installations in a quick and efficient way in order to be able to improve the overall performance of the DH systems.

In this study, the ideal value for ΔT was chosen to be 45 °C, in accordance with the results from other studies. However, this is a parameter which is very dependent on the system that is being investigated since the temperatures of the system might not allow that a value of 45 °C is obtained. If the temperature difference between the supply and return temperature is, e.g., 30 °C then it will not be possible to obtain the ideal ΔT used in this study. It might also be that a DH system has very high supply temperatures, which means that it is "easier" to obtain high ΔT's. This indicates that in some systems, the ideal ΔT in this study might be small in a system with high supply temperature. Therefore, it is of great interest to find another way to determine what value should be used as the ideal ΔT if this analysis program is to be used in other DH systems.

When considering the fault detection method used in this study, there are some variables and approaches that could be done differently. One example is the value of the ideal ΔT as described above, and another is how the reference case installations were chosen. In this study, the reference case was decided based on the overflow of the installations. Hence, the installations with the lowest overflow were included in the reference case. There might be other ways to determine the reference case installations, e.g. by choosing the installations with the lowest return temperatures. It might also be interesting to rank the poorly performing installations differently. In this study, they were ranked according to their overflows. Other possibilities would be to rank them according to the number of outliers in each signature, or to rank them according to the total number of outliers in all signatures. It would also be possible to choose the breakpoint of the piece-wise linear regression differently. In this study, one breakpoint was used and the value of this was determined by visual inspection of the customer data. The choice of breakpoints could have been further improved by using statistical methods.

7 Conclusions

The analysis tool developed in this study is a first version of an automatic DH analysis tool. The algorithm developed was able to rapidly identify poorly performing installations in a large data set. The main conclusions from the paper may be find in the bullet list below.

- The algorithm could identify poorly performing installations in a data set containing data from 3 000 customer installations.
- The algorithm identified 1 273 installations to be poorly performing. 109 installations were located in weak areas, and 1 164 installations were located in non-weak areas.
- The overflow of the identified installations were very large and indicates that the poorly performing installations have a large impact on the system efficiency.
- The method used in this paper has the end user in focus and utilizes data and data patterns that are well familiar to people in the DH industry.

Although the analysis algorithm performs well for the task at hand, there is some room for improvement:

- The algorithm would benefit from being able to separate different building types from each other. This would provide the user with more detailed information about the performance of the district heating installation. This could be done by using hourly meter readings instead of daily values.
- The parameters used in this study should be tested further. For example, the breakpoints of the linear regressions are decided manually, and the values of these could probably be improved by using, for example, statistical methods.

References

[1] Svend Frederiksen and Sven Werner. *District Heating and Cooling.* Studentlitteratur, Lund, 2014.

[2] Henrik Lund, Sven Werner, Robin Wiltshire, Svend Svendsen, Jan Eric Thorsen, Frede Hvelplund, and Brian Van Mathiesen. 4th generation district heating (4gdh) - integrating smart thermal grids into future sustainable energy systems. *Energy*, 68:1–11, 2014.

[3] Henrik Gadd and Sven Werner. Achieving low return temperature from district heating substations. *Applied Energy*, 136:59–67, 2014.

[4] Henrik Gadd and Sven Werner. Fault detection in district heating substations. *Applied energy*, 157:51–59, 2015.

[5] Sven Werner. Feltyper i fjärrvärmecentraler strukturerade (Types of faults in district heating substations structured), 2013. Halmstad Högskola. In Swedish.

[6] Helge Averfalk. Morgondagens effektiva fjärrvärme: en beskrivande litteraturstudie (The efficient district heating of tomorrow), 2014. Halmstad Högskola. In Swedish.

[7] Natasa Nord, Elise Kristine Løve Nielsen, Hanne Kauko, and Tymofii Tereschenko. Challenges and potentials for low-temperature district heating implementation in norway. *Energy*, 2018:889–902.

[8] Fredrik Sandin, Jonas Gustafsson, and Jerker Delsing. Fault detection with hourly district energy data. Research report 2013:27, Svensk Fjärrvärme AB, 2013.

[9] Catarina Warfvinge and Mats Dahlblom. *Projektering av VVS-installationer (Planning and design of HVAC installations)*. Studentlitteratur AB, Lund, Sweden, 2010. In Swedish.

[10] Energiföretagen Sverige. District heating substations - design and installation. Technical regulations F:101, Energiföretagen Sverige AB, 2016.

[11] Anders Trüschel. Värdet av injustering (the value of balancing). Research report 2005:134, Svensk Fjärrvärme AB, 2005. In Swedish.

[12] Kimmo Yliniemi. *Fault detection in district heating substations.* Licentiate thesis, Luleå University of Technology, 2005.

[13] Markus Alsbjer and Peter Wahlgren. Fjärrvärmecentraler 10 år - håller de måttet? (district heating substations 10 years - do they measure up?). Research report 2011:9, Svensk Fjärrvärme AB, 2011. In Swedish.

[14] Henrik Gadd. *To analyse measurements is to know!* PhD dissertation, Lund University, Faculty of Engingeering (LTH), 2014.

[15] Oddgeir Gudmundsson and Jan Eric Thorsen. Benefits of using dp controllers in district heating substations. `http://grijanje.danfoss.com/PCMPDF/VFIQA102_benefits-using-dp-controllers_180614_lores.pdf`, 2018.

[16] Council of European Union. Eu directive 2012/27/eu, 2012. `http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012L0027&from=EN`.

[17] Christian Johansson and Fredrik Wernstedt. N-dimensional fault detection and operational analysis with performance metrics. In *DHC13, the 13th International Symposium on District Heating and Cooling*, Copenhagen, Denmark, 2012.

[18] Puning Xue, Zhigang Zhou, Xiumu Fang, Xin Chen, Lin Liu, Yaowen Liu, and Jing Liu. Fault detection and operation optimization in district heating substations based on data mining techniques. *Applied energy*, 205:926–940, 2017.

[19] Kimmo Yliniemi, Jan van Deventer, and Jerker Delsing. Sensor fault detection in district heating substations. In *10th IMEKO TC10 International Conference on Technical Diagnostics*, Budapest, Hungary, 2010.

[20] Nathan Zimmerman, Erik Dahlqvist, and Konstantinos Kyprianidisa. Towards on-line fault detection and diagnostics in district heating systems. *Energy Procedia*, 105:1960–1966, 2017.

[21] Jouko Pakanen, Juhani Hyvärinen, Juha Kuismin, and Markku Ahonen. Fault diagnosis methods for district heating substations. Technical report, VTT Building technology, Oulo, Finland, 1996.

[22] Henrik Gadd and Sven Werner. Heat load patterns in district heating substations. *Applied Energy*, 108:176–183, 2013.

[23] Pengfei Jie, Xiangfei Kong, Xian Rong, and Shangqun Xie. Selecting the optimum pressure drop per unit length of district heating piping network based on operating strategies. *Applied Energy*, 177:341–353, 2016.

[24] Stefan Petersson and Cilla Dahlberg Larsson. Samband mellan flödespremie och returtemperatur (relationship between flow price component and return temperature). Technical Report 2013:25, Svensk Fjärrvärme, 2010. In Swedish.

[25] Henrik Gåverud, Kerstin Sernhed, and Annamaraia Sandgren. Kundernas uppfattning om förändrade prismodeller (the customers' perception of changing price models). Report 2016:301, Energiforsk AB, 2016. In Swedish.

[26] The R Foundation. R: What is r? `https://www.r-project.org/about.html`, n.d. Online. Accessed 2018-12-06.

[27] Kenneth O. Cogger. Piecewise linear modeling: Theory, guidelines, and applications. In *In Management Science & Technology Symposium*. University of Kansas School of Business, 2006.

[28] John O. Rawlings, Sastry G. Pantula, and David A. Dickey. *Applied Regression Analysis: A research tool*. Springer-Verlag, New York, 1998.

[29] Jason Brownlee. How to use statistics to identify outliers in data. `https://machinelearningmastery.com/how-to-use-statistics-to-identify-outliers-in-data/`, 2018. Online, accessed 2018-11-20.

Paper iii

Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 149 (2018) 226–235

Energy Procedia

www.elsevier.com/locate/procedia

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

A machine learning approach to fault detection in district heating substations

Sara Månsson[a,b,*], Per-Olof Johansson Kallioniemi[a], Kerstin Sernhed[a], Marcus Thern[a]

[a]*Department of Energy Sciences, Faculty of Engineering, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden*
[b]*VITO, Boeretang 200, BE-2400 MOL, Belgium*

Abstract

The aim of this study is to develop a model capable of predicting the behavior of a district heating substation, including being able to distinguish datasets from well performing substations from datasets containing faults. The model developed in the study is based on machine learning algorithms and the model is trained on data from a Swedish district heating substation. A number of different models and input/output parameters are tested in the study. The results show that the model is capable of modelling the substation behavior, and that the fault detection capability of the model is high.

Keywords: District heating substations, fault detection, machine learning

1. Introduction

As the district heating (DH) systems are developing towards the 4th generation, it becomes increasingly important that the components of the systems are performing as well as possible [1]. In the existing DH systems, there are components that are poorly performing, and they need to be identified and addressed. One component that has been

* Corresponding author. Tel.: +46-46- 222 33 20.
E-mail address: sara.mansson@energy.lth.se

10.1016/j.egypro.2018.08.187

found to contain many faults is the DH substation located at the customers' installations [2]. Faults in substations could for example include valves that are stuck, fouled heat exchanger areas in the heat exchanger, malfunctioning temperature transmitters, but also wrong settings in the control system [3].

The faults in the substation cause poor cooling performance, which means that the heat from the DH system is not transferred in an optimal way in the customer substation. This leads to a need to increase either the flow or the supply temperature to be able to transfer the heat needed. This causes the energy efficiency of the entire DH system to decrease, especially if many of the substations are poorly performing [3]. Previous studies have shown that large shares of the substations in different DH systems are not performing well [4], [5], indicating that this is an issue that should be treated to as soon as possible in order to increase DH efficiency.

Historically, the most common way of detecting poorly performing substations has been to perform a manual analysis of customer data. This task can be extremely time consuming, due to the vast amount of data that is investigated when performing the analysis. Because of this, the substations with the largest heat demand in the DH system are normally prioritized when performing the analysis, leading to a large share of poorly performing substations being undetected. Therefore, well performing, automatic substation fault detection methods are needed.

The first step towards obtaining a well performing automatic fault detection method based on algorithms is to model the behavior of a substation in a good way. This can be done by identifying relationships between the measurements from the DH substation in order to establish patterns that occur when a substation is not performing optimally. Machine learning (ML) techniques have become one of the most used collection of techniques when modelling the relationship between different parameters. ML techniques are algorithms that are not specifically programmed; instead the purpose is to let the algorithms learn and improve from experience in order to obtain a well performing algorithm. Supervised learning techniques are a collection of ML methods that uses labeled data to train models that are used for a multitude of different problems, the most common ones being classification and regression modelling [6]. The labeled data consists of an input data set and an output data set, which both can contain one or more variables. During the training phase, both input and output data (labeled data) is presented to the model in order for it to learn the relationship between input and output data. When the training is finished, the model should be able to model the relationship well enough to be able to predict the labels of the output data when new, unlabeled input data is presented to the model [7]. Since the data from the district heating customers consist of numerical values, regression modelling should be used to predict the labels of the output values.

In this study, an ML regression model based on Gradient Boosting Regression (GBR) is used to predict the mass flow of one individual, well performing substation. The aim of the study is to develop a model that predicts the mass flow on an hourly basis, using as few input parameters as possible. When a good model has been obtained, a data set containing faults is introduced to the model in order to see if the model will make a different prediction for the faulty data set than for the data set that the model is created for

2. Background and related research

The heat demand in a building mainly originates from two different needs: the use of space heating and domestic hot water preparation [3]. These needs have to be met at all times, which means that the DH utilities need to make predictions about the heat demand in their systems in order to meet the customer demand. The prediction can be done for the entire system at once, as well as for the individual buildings in the system. Several studies have been conducted in the area, but the most recent studies focus mainly on statistical and ML prediction models. The statistical methods include Seasonal Auto-Regressive Integrated Moving Average (SARIMA) models, and Multiple Linear Regression (MLR) models, [8], [9], [10]. The neural network models are typically based on Boosting Regression (BR), Random Forest (RF), Support Vector Machines (SVM), Multi-Layer Perceptrons (MLPs), and k-Nearest Neighbor (k-NN) [11].

Dalipi et. al. described how different ML algorithms could be used to predict the heat load of a DH system using data from a waste incineration plant [12]. The authors compared three different ML algorithms (Support Vector Regression (SVR), Partial Least Squares (PLS), and RF) to each other in order to define a heat load forecaster that performs well in terms of low error and high accuracy. Fumo presented a review of the basics of building energy estimation, in which he mentioned two main approaches: forward (classical) approach and data-driven (inverse) approach [13]. The forward approach utilizes the already known mathematical equations describing the physical relationships of a system and known inputs to predict an output parameter. In contrast, the data-driven approach consists of methods where measured variables are used as input for ML algorithms that uses the data to learn the behavior of the system.

Prediction methods have also been used in different contexts in order to detect anomalies in datasets. Araya et. al. presented a method that combined prediction-based and pattern-based classifiers to determine whether a consumption pattern was anomalous or not [14]. Chen et. al. described a statistical predictive method for detecting anomalies in building energy consumption [15]. The authors created a model of the baseline behavior of a building's energy consumption and calculated prediction intervals (PIs) for the mean and variance of the predicted data. The PIs described an estimate of an interval that future values should fall within. Future real values were considered as anomalies when their mean and/or variance fell outside the Pis. A similar approach was used in a study conducted by Baldacci et. al., where a prediction method using linear regression was used to detect anomalies in natural gas consumption [16].

The previous studies show that the most common techniques for prediction modelling in DH contexts are techniques that utilize patterns in existing DH data in order to identify important relationships between different variables. However, prediction methods have been used to detect anomalies in other energy-related research areas. The novelty of the work conducted in this study is that it combines the concepts of prediction modelling for forecasting in DH with the concepts of fault detection used in other fields.

3. Method

3.1. Data set

The data set used in this study consisted of hourly values during one year (November-November) from the primary side of one DH substation in a DH system in Sweden. The set originally contained hourly meter readings for one year for each measurement variable, but since some of these measurements were incomplete due to different reasons, the final number of instances was 8726. The variables, or features, that were included in the data set can be found in Table 1. $T_{out,24}$ refers to the average value of the outdoor temperature during the previous 24 hours.

Table 1. Features included in the data set.

Feature	Notation	Unit
Outdoor temperature	T_{out}	˚C
Outdoor temperature, 24 hour average value	$T_{out,24}$	˚C
Hour of the day	t	h
Supply temperature	T_s	˚C
Return temperature	T_r	˚C
Mass flow per hour	$\dot{m}$	m^3/h

The data set was divided into one test set and one training set. The ratio between the number of meter readings in the training and test sets was 80/20.

3.2. Programming language

In this study, Python and the Scikit-learn package was used to conduct the modelling and fault detection. Python is an interpreted, object-oriented, high-level programming language, often used for data analysis and scientific computing. Python contains a wide range of different packages that can be used for these tasks, the Scikit-learn package being one of them. Scikit-learn is one of the most used packages for machine learning and provides a range of ML algorithms, including algorithms for regression, clustering, dimensionality reduction and preprocessing of features [17].

3.3. TPOT

Developing a machine learning algorithm can be a time consuming task, due to the fact that all data sets are unique and contain different challenges. There is almost always need for some sort of manual preprocessing of the data, to make sure that the data quality is sufficient. The variables (or features) may have to be modified in some way, e.g., scaling or introduction of polynomial features, and a well-performing predictor has to be chosen from the numerous methods that are available amongst the ML techniques [18].

In order to bypass some of the issues that traditionally arise when developing an ML model, a number of tools have been developed that allows the user to automate parts of the data analysis. One of these tools is the Tree-based Pipeline Optimization Tool (TPOT), which is an automated machine learning tool. TPOT is implemented in Python and creates combinations, or pipelines, of data transformations and machine learning models using genetic programming [18]. The tool optimizes the performance of the entire pipeline, using pre-existing algorithms in novel ways. Before using TPOT, the user must manually prepare the data for modelling by making sure that the data contains no missing or mislabeled values. TPOT then optimizes feature selection, feature preprocessing, feature construction, model selection, and parameter optimization. The last step of the ML process, the model validation, is carried out by the user.

3.4. Model Validation

The model validation can be carried out in a number of different ways. In this study, two different measures of difference between two variables were used: the coefficient of determination and the mean absolute error.

The coefficient of determination, or R2 value, can be calculated using Equation 1. In the equation, SS_{res} is the sums of squares of the residuals, and SS_{tot} is the total sum of squares which is proportional to the variance of the data. The R^2 value is a number between zero and one, and a well performing model is expected to have a value close to one.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{1}$$

The mean absolute error, MAE, is the averaged sum of the absolute value of the residuals between actual values and predicted values can be calculated using equation 2. The MAE value for a well performing model is expected to be as close to zero as possible.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|(actual\ value)_i - (predicted\ value)_i\right| \tag{2}$$

3.5. Determining what TPOT pipeline to use

When running TPOT, it is important to make sure that the tool provides a well-functioning pipeline, with a model that captures the behavior of the substation well. In order to make sure that such a pipeline is obtained, TPOT was used on 16 different training sets and 16 different test sets to produce 16 different pipelines. The next step of the selection process was to introduce the obtained pipelines to the same training set, in order to be able to compare their performance for the same data. After fitting the pipelines to the labeled training data, they were introduced to the same test set and the R^2 and MAE values were then calculated. The resulting values are displayed in Table 2. The pipeline that had the highest R^2 value and the lowest MAE value for the test set was picked as the one to be further investigated in this study. The best performing pipeline was pipeline number 13 (highlighted in grey in Table 2), which had an R^2 value of 0.986 an and MAE of 0.100.

Table 2. R^2 and MAE values for the 16 different pipelines, used for the same test set. The best performing pipeline is highlighted in grey.

TPOT number	R^2	MAE
1	0.963	0.139
2	0.972	0.132
3	0.959	0.153
4	0.974	0.139
5	0.980	0.102
6	0.971	0.128
7	0.959	0.150
8	0.968	0.127
9	0.964	0.135
10	0.963	0.151
11	0.972	0.141
12	0.960	0.145
13	0.970	0.133
14	0.986	0.100
15	0.959	0.148
16	0.962	0.143

3.6. Final pipeline

The best performing pipeline, that was chosen to be further investigated in the study, consisted of two major components: pre-processing of data, and modelling, as illustrated in Fig. 2.

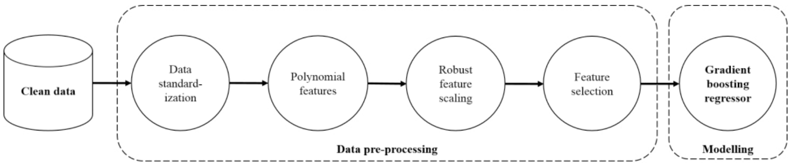


Fig. 1. Illustration of the TPOT pipeline

The first component, data pre-processing, consisted of four different steps. The first step consisted of a method for data standardization, which means that the features were standardized by removing their individual means and scaling to unit variance. This was done in order to make the distribution of the individual features look more or less like normally distributed data, since many machine learning algorithms (including the one used in this study) expects normally distributed data. If features of other distributions are introduced to the algorithms, there could be a risk of overestimating some features and excluding others that still would be useful in the next step.

In the second data pre-processing step, more features were generated. The generated features were polynomial and interaction terms of the original features. If the original features were labelled x_1, x_2, and so on, the additional features were generated as, e.g., the polynomial term x_1^2, and the interaction term x_1x_2. This kind of terms are normally introduced if no good model can be found when only using the original features. In this study, polynomial terms up to the second order of the original terms and first order of the interaction terms were included.

The next step of the data pre-processing scaled the features robustly, which means that all features, both the original and the generated ones, were standardized using a method that was robust to outliers. This was done in order to obtain a dataset where the ranges of the features were kept within the same limits. Similar feature ranges are important when modelling features that are normally measured in different units, e.g., when comparing the supply temperature to the flow rate per hour. Due to the different scales of the features, it might seem as if one feature is not varying as much as the other. This implies that the model should not put equal emphasis to the features' variation when fitting the model, causing a poorer fit of the data.

The last step consisted of a method for feature selection, where the features that were finally used to model the substation behavior were selected using the family-wise error rate. This meant that some of the features, either original or generated features, was not used to model the behavior of the substation. This is a common procedure in when using ML as a prediction tool, since it is important to keep the model as simple as possible. The features that were not contributing with new information to the model was therefore excluded.

The second main component of the pipeline was the regressor that predicted the behavior of the substation. The regressor was the Gradient Boosting Regressor (GBR). Gradient boosting is an iterative process, where regression models are added one by one to a so-called ensemble of methods. The model that is added in each iteration step is trained to minimize the regression error that the rest of the ensemble produces when modelling the data [19]. This means that in each iteration step, the added model improves the prediction ability of the ensemble. The final ensemble is obtained when a specified number of iterations has been performed. The number of iterations vary depending on how complex the data structure is, but in this study, 100 iterations was needed to obtain a model that did not improve significantly in terms of prediction ability when adding new models to the ensemble.

3.7. Test of variables

Since there are a number of plausible parameter combinations that could be used to model the substation behavior, a parameter test was conducted. In the test, the final pipeline was introduced to a number of different combinations of input and output parameters in order to see how well it could learn the substation behavior. The combinations were evaluated using the R^2 and MAE values.

The most promising combinations of the variable test are found in Table 3. The input/output combination that gave the highest R^2 value and the lowest MAE value was combination number 5 (highlighted in grey).

Table 3. Parameter combinations that were tested with the model. Combination number 5, highlighted in grey, obtained the best R^2 and MAE value.

Combination	Input parameters	Output parameters	R^2 value	MAE value
1	T_{out}, T_r, T_s, t	$\dot{m}$	0.9703	0.1337
2	$T_{out,24}$, T_r, T_s, t	$\dot{m}$	0.9555	0.2027
3	T_{out}, T_s, $\dot{m}$, t	T_r	0.8839	1.1091
4	$T_{out,24}$, T_s, $\dot{m}$, t	T_r	0.8903	1.0348
5	$T_{out,24}$, T_{out}, T_s, t	$\dot{m}$	0.9740	0.1301
6	$T_{out,24}$, T_{out}, T_s, t	T_r	0.8841	1.0555
7	$T_{out,24}$, T_{out}, T_s, t	T_r, $\dot{m}$	0.9296	0.5857
		T_r	0.8868	1.0412
		$\dot{m}$	0.9723	0.1301

3.8. Inducing faults

In order to test the fault detection potential of the model, the model was introduced to a dataset consisting of the parameters in the best performing parameter combination presented in section 3.7. To obtain a faulty dataset, two faults that are commonly known to occur in DH customer values were induced in the dataset from the well performing substation.

The first fault was related to communication problems between the meter and the DH utility's database, and occurs when the connection between the meter and the utility's database is lost. When this happens, many utilities choose to replace the missing meter reading with a fixed value. In this case, a value of 60 ˚C was inserted randomly in the data set for the supply temperature. Since this was an extremely low value comparing to the original dataset, another fault was also induced where the original value of the supply temperature was decreased by 10 %.

The second fault that was induced was a drifting meter fault, which can be described as a "…gradual change in output over a period of time which is unrelated to any change in input" [20]. The drifting meter fault was induced for two different meters: the outdoor temperature meter and the supply temperature meter. The fault was induced as a gradually increasing addition to the original meter readings. For both variables, the maximum increase of the values was 1.08 ˚C.

In order to evaluate how the model performs when it was introduced to a dataset containing faults, the prediction performance of the model for the faulty dataset was compared to the prediction performance for the original dataset. The hourly residual between the real values and the predicted values for the faulty dataset was computed, as well as the hourly residual between the real values and the predicted values for the original dataset without faults. The cumulative sums of these residuals were also calculated using a rolling window that included the residuals from the last 24 hours.

4. Results

4.1. Model performance, well performing data set

Fig. 2 (a) displays the actual values obtained from the substation (blue dots), and the predicted values obtained from the model (red dots) as a function of time. As can be seen in the figure, the model is not able to capture all of the rapid changes that occur in a DH substation. However, the R^2 value of 0.986 is deemed to be good enough to continue with the model. Fig. 2 (b) displays the predicted values as a function of the real values. The relationship between the variables displays a more or less linear behavior, which is expected when a model that describes the main share of the variation in the original data has been obtained.

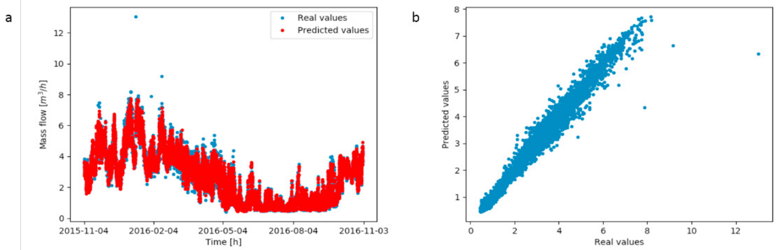


Fig. 2. (a) Real and predicted values for the entire data set. (b) Predicted values as a function of the real values.

4.2. Model performance, data sets containing faults

Fig. 3(a) displays two residuals: the residual between the real values and the predicted values for a well performing substation (blue line), and the residuals between the real values and the data where a communication fault has been induced (red line). Fig. 3Fig. 4 (b) displays the cumulative sum over a 24 h interval of the same residuals. As can be seen from the residuals and the cumulative sum of the residuals, the residuals of the dataset containing faults deviate significantly from the residuals of the well performing dataset. This can especially be seen in Fig. 3 (b). The deviations were further investigated and were found to correspond to days where the 60 ˚C fault had been induced.

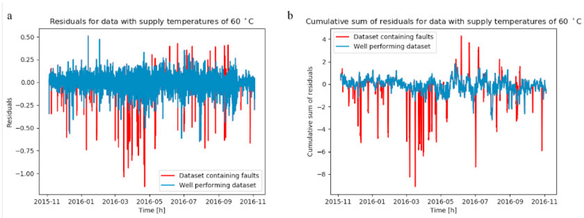


Fig. 3. (a) Residuals for data with supply temperatures of 60 ˚C. (b) Cumulative sum of residuals for data with supply temperature of 60 ˚C.

Fig. 4 (a) and (b) displays the residual and the cumulative sum of the residuals for the data where supply temperature instances have been decreased by 10 % (red lines). The figures also show the residual for the well performing dataset (blue lines). The difference between the residuals of the well performing dataset and the faulty dataset is not large, but the cumulative sum displays a larger difference.

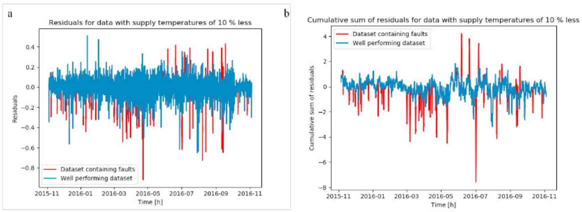


Fig. 4. (a) Residuals for data with supply temperatures that have been decreased by 10 %. (b) Cumulative sum of residuals for data with supply temperature that have been decreased by 10 %.

Fig. 5. displays the residual and cumulative sum of a dataset containing measurements from a drifting outdoor temperature meter (red lines), as well as the residual and cumulative sum for the well performing dataset (blue lines). For the residual, the deviation from the well performing dataset is not clearly visible, but the cumulative sum in Fig. 5. (b) displays a clear deviation approximately one month after the fault was induced in the data.

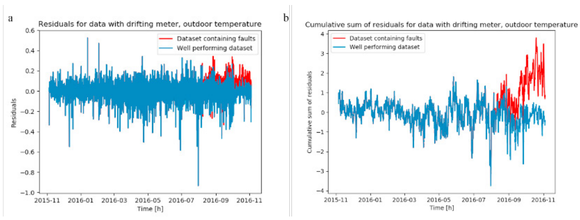


Fig. 5. (a) Residuals for data with drifting meter, outdoor temperature. (b) Cumulative sum of residuals for data with drifting meter, outdoor temperature.

Fig. 6. (a) displays the residuals for data induced with a drifting supply temperature meter (red line), and Fig. 6. (b) displays the cumulative sum of the residual (red line). In the same figures, the residual and cumulative sum of the residual for the well performing substation are displayed as well (blue lines). Here, there is no clear deviation from the residuals and cumulative sum of the well performing substation.

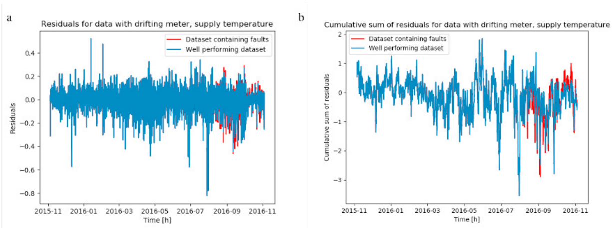


Fig. 6. (a) Residuals for data with drifting meter, supply temperature. (b) Cumulative sum of residuals for data with drifting meter, supply temperature.

5. Analysis and Discussion

The results of this study show that it is possible to model the behavior of a DH substation using machine learning algorithms, specifically using the Gradient Boosting Regressor. As can be seen the variable test, the model predictions were most accurate when having the flow rate as output parameter, and outdoor temperature, time of day, and supply temperature as input parameters. Fig. 2. (a) shows that the model manages to capture the main part of the substation's behavior, except for the most extreme values in the original dataset. This is confirmed by Fig. 2. (b), which displays the predicted values as a function of the real values. The values display a more or less linear relationship, which is expected (within a reasonable error limit) when a model is able to capture the largest part of the behavior of the output variable. The values that are deviating significantly from the linear behavior in Fig. 2. (b) are coupled to the most extreme values of the real dataset, which the model was not able to capture.

When testing the fault detection capability of the model, it is clear that different parameters have different impact on the model's performance. The fault representing communication error between the substation and the DH utility's database is clearly visible when comparing the residuals of the faulty dataset and the well performing dataset (Fig. 4 (a) and (b)). However, when comparing this to the faulty dataset containing an induced drifting supply temperature meter in Fig. 6., it is clear that the deviation from the original dataset is not as large. When comparing these conclusions to the results in Fig. 4. (a) and (b), it is clear that the residuals from the dataset containing supply temperature measurements that have been decreased by 10 % show a larger difference from the well performing substation than in the case of the drifting supply temperature meter. The deviation is not as large as for the data where 60 ˚C was induced, but this was expected since the 60 ˚C fault was extreme in this particular DH system and dataset.

Based on the figures presented in the results, the largest difficulty of obtaining a well performing fault detection method that works automatically will be to find an evaluation algorithm that can distinguish a sudden change in the residual and/or cumulative sum. When looking at Fig. 4. (b), it is clear that the deviations from the cumulative sum of the well performing dataset in most cases do not deviate significantly from the most extreme residuals from the well performing dataset. However, the most extreme deviations in the residual of the well performing dataset occur during summer, when the behavior of the substation will be different due to the fact that no space heating demand occur in the customer installation. When looking at the first part of the cumulative sum in Fig. 4. (b), from November until May, it is clear that the deviation from the residual of the well performing dataset is large, and it might be necessary to detect the deviations during heating season separately. One way of doing this could be to monitor the change of the residual and/or cumulative sum that the model produces continuously and flag a fault if a rapid change occurs.

The model is more sensitive to changes in measurements from the outdoor temperature meter. As can be seen in Fig. 5. (b), the cumulative sum of the residual of the faulty dataset deviates significantly from the residuals of the well performing dataset approximately one month after the fault is induced. The fault induced had a maximum value of 1.08 ˚C, and even though this can be considered as a small (and reasonable) change in the measurements of the outdoor temperature, the model responded clearly to the change. This is very promising for continued model development and fault detection studies.

6. Conclusion

In this study, a model for the flow rate in a DH substation has been created, using the outdoor temperature, supply temperature and time of day as input parameters. The model is based on a number of data pre-processing steps, as well as a regression method called Gradient Boosting Regressor. When introduced to datasets containing different faults, the model predictions change, which shows great promise for continued fault detection studies using this model.

Acknowledgements

This work is part of the TEMPO project which is funded by the European Union's Horizon 2020 Programme under Grant Agreement no. 768936.

References

[1] H. Lund, S. Werner, R. Wiltshire, S. Svendsen, J. E. Thorsen, F. Hvelplund och B. Vad Mathiesen, ”4th Generation District Heating (4GDH) - Integrating smart thermal grids into future sustainable energy systems,” *Energy,* vol. 68, pp. 1-11, 2014.

[2] H. Gadd, *To analyse measurements is to know!,* Lund: Lund University, 2014.

[3] S. Frederiksen och S. Werner, District heating and Cooling, Lund: Studentlitteratur, 2013.

[4] H. Gadd och S. Werner, ”Achieving Low Return Temperature from District Heating Substations,” *Applied Energy,* vol. 136, pp. 59-67, 2014.

[5] H. Gadd och S. Werner, ”Fault Detection in District Heating Substations,” *Applied Energy,* vol. 157, pp. 51-59, 2015.

[6] C. Sammut och G. I. Webb, ”Supervised learning,” i *Encyclopedia of Machine Learning and Data Mining*, Boston, Springer, 2017.

[7] M. Mohri, A. Rostamizadeh och A. Talwalkar, Foundations of machine learning, Cambridge: MIT Press, 2012.

[8] S. Grosswindhager, A. Voigt och M. Kozek, ”Online short-term forecast of system heat load in district heating networks,” i *Proceedings of the 31st International Symposium on Forecasting*, Prague, Czech Republic, 2011.

[9] T. Fang och R. Lahdelma, ”Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system,” *Applied energy,* vol. 179, pp. 544-552, 2016.

[10] T. Catalina, V. Iordache och B. Caracaleanu, ”Multiple regression model for fast prediction of heating energy demand,” *Energy and buildings,* vol. 57, pp. 302-312, 2013.

[11] H.-x. Zhao och F. Magoulés, ”A review on the prediction of building energy consumption,” *Renewable and sustainable energy reviews,* vol. 16, pp. 3586-3592, 2012.

[12] F. Dalipi, S. Yildirim Yayilgan och A. Gebremedhin, ”Data-Driven Machine-Learning Model in District Heating System for Heat Load Prediction: A Comparison Study,” *Applied Computational Intelligence and Soft Computing,* vol. 2016, 2016.

[13] N. Fumo, ”A review on the basics of building energy estimation,” *Renewable and sustainable energy reviews,* vol. 31, pp. 53-60, 2014.

[14] A. D. B, K. Grolinger, H. F. ElYamani och M. A. Capretz, ”An ensemble learning framework for anomaly detection in building energy consumption,” *Energy and Buildings,* vol. 144, pp. 191-206, 2017.

[15] B. Chen, M. Sinn, J. Ploennigs och A. Schumann, ”Statistical Anomaly Detection in Mean and Variation,” i *22nd International Conference on Pattern Recognition*, Stockholm, Sweden, 2014.

[16] L. Baldacci, M. Golfarelli, D. Lombardi och F. Sami, ”Natural gas consumption forecasting for anomaly detection,” *Expert systems with applications,* vol. 62, pp. 190-201, 2016.

[17] W. McKinney, Python for data analysis - Data wrangling with Pandas, NumPy, and IPython, O'Reilly, 2018.

[18] R. S. Olson, R. J. Urbanowicz, P. C. Andrews, N. A. Lavender, L. C. Kidd och J. H. Moore, ”Automating Biomedical Data Science Through Tree-Based Pipeline Optimization,” i *European conference on the Applications of evolutionary computation*, 2016.

[19] A. Natekin och A. Knoll, ”Gradient boosting machines, a tutorial,” *Frontiers in neurorobotics,* vol. 7, nr 21, 2013.

[20] W. Bolton, Newnes instumentation & measurement pocket book, Oxford: Newnes, 2000.