Data driven analysis of district cooling substations for performance diagnosis

Master's thesis in Sustainable Energy Systems

MOHAMMED BURHANUDDIN RABANI

DEPARTMENT OF ARCHITECTURE AND CIVIL ENGINEERING

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

MASTER'S THESIS 2022

Data driven analysis of district cooling substations for performance diagnosis

MOHAMMED BURHANUDDIN RABANI

Department of Architecture and Civil Engineering
Division of Building Services Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022

Data driven analysis of district cooling substations
for performance diagnosis
MOHAMMED BURHANUDDIN RABANI

Supervisor: Maria Jangsten, Department of Architecture and Civil Engineering
Examiner: Jan-Olof Dalenbäck, Department of Architecture and Civil Engineering

Master's Thesis 2022
Department of Architecture and Civil Engineering
Division of Building Services Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2022

Data driven analysis of district cooling substations
for performance diagnosis
MOHAMMED BURHANUDDIN RABANI
Department of Architecture and civil engineering
Chalmers University of Technology

Abstract

District cooling (DC) is a crucial technology for providing cooling for purposes such as space cooling for indoor environment, industrial processes, food storage industry. To improve the efficiency of the district cooling system (DCS) periodic maintenance of the components in the system is required. This requires detection of faults in the system and addressing these faults. Previously, fault detection has been done in district heating (DH) system but not as much on district cooling (DC). In this work, a large scale data driven approach is implemented to identify the performance of district cooling substations and detect faults.

The scope of this study includes all 180 substations from the city of Gothenburg, Sweden and the operational data from the primary side of the substation is analysed. Different signatures are analysed namely, energy signature, delta signature and return temperature signature to identify faulty behaviour in the substations. From previous research and domain knowledge, the most suitable methods for the data analysis are chosen and applied. Out of 180 substations, customer categories for 41 buildings are identified and their delta T signatures are studied for performance diagnosis.

The results show that the method implemented in this study can be used in the fault detection domain for DC substations on a large scale. A mix of domain knowledge and data driven tools are required to detect faults and analyse performance.

Keywords: District cooling, substations, performance diagnosis, fault detection, signatures

Acknowledgements

I would like to express my immense gratitude to my supervisor Maria Jangsten for her continuous support and invaluable feedback along this journey without whom this work would not have been possible. I would also like to thank the professors Anders Truschel, Torbjörn Lindholm and Jan-Olof Dalenbäck for their inputs and insights along the way.

I would also like to express my heartfelt thanks to my colleagues for motivating and always being cheerful throughout this journey.

Heartfelt thanks to my family for their support and prayers far from home. The acquaintances and friends I made in the department along the way made this thesis journey ever so joyous and look forward to. I am thankful for this opportunity given to me by my supervisor and making this study possible.
Thank you!

Mohammed Burhanuddin Rabani, Gothenburg, June 2022

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AHU	Air handling units
DC	District cooling
DCS	District cooling system
DH	District heating
BMS	Building management system
ETS	Energy transfer station
FDD	Fault detection and diagnosis
HVAC	Heating, ventilation and air conditioning

Nomenclature

Below is the nomenclature of indices, sets, parameters, and variables that have been used throughout this thesis.

Parameters

ρ	Density of water
C_p	Specific heat capacity of water

Variables

T_r	Return temperature from substation
T_s	Supply temperature to substation
Q	Cooling load of the substation
ΔT	Temperature difference on the primary side
$\dot{V}$	Volumertric flowrate

Contents

List of Figures

List of Tables

1
Introduction

This chapter includes a general introduction about the thesis topic, background and motivation for carrying out the thesis, the aim of the master thesis and finally the thesis outline.

In the past year alone, the electricity consumption in residential buildings increased by 40%, this was due to the pandemic. But as offices and commercial buildings remained unoccupied, energy consumption for maintenance of HVAC and cooling units remained unaffected [1]. With rise in global temperatures, the demand for cooling and generally space cooling is increasing. With this increase in cooling demand arises the problem of high electricity consumption. District cooling system (DCS) is an important component to reduce this electricty consumption when it comes to buildings, however faults occur from time to time in the system which require attention. A significant amount of research has been conducted in the research domain "fault detection and diagnosis" (FDD) in district heating (DH) substations, HVAC systems in buildings including chilled water systems based on data-driven methods like machine learning algorithms. FDD in district cooling (DC) substations is closely linked to both FDD in DH substations and HVAC systems, however, FDD research on DC substations specifically is still lacking. Fault detection in substations is fairly uncommon and maintenance is only carried out when customer satisfaction is not achieved, unlike air handling units (AHUs) and HVAC systems where regular maintenance is carried out [2].

1.1 Background and motivation

District cooling system (DCS) has been developed as an important technology to provide cooling for the indoor environment. It is mainly used for comfort cooling in residential, commercial and industrial buildings from a central source such as a river, chilled water storage and central chiller plants. The distribution network consists of pipelines which are generally underground. DC system utilises this energy distribution network to satisfy the customer needs [3].

A district cooling system can be divided into three components, a production plant, distribution network consisting of pipelines and building chilled water system which are connected to the distribution network by heat exchangers in the substations [4]. Faults occur in the substations and heat exchangers due to periodic operation and these must be addressed in order to improve the efficiency of the system.

In many of the previous studies on DH substations, a common approach to analyse the performance and detect faults is to study the energy signature, delta T signature and return temperature signatures [5],[6]. For DC substations the aforementioned approach has not been applied and it is unknown if performance diagnosis and identification of faulty operation can be done as efficiently as for DH substations.

1.2 Aim

The aim of this master's thesis is to evaluate commonly used fault detection methods for DH substations on the DC substations in Gothenburg district cooling system. A data driven approach is implemented to analyse the behaviour of substations to carry out performance diagnosis and fault detection.

1.3 Thesis outline

This thesis is divided into different chapters to help guide the reader into a clear understanding of the topic and the current problem faced by the present DC substations. In chapter 1, a brief introduction about the topic and idea behind the thesis is provided. The following chapter 2, includes a more technical background about the district cooling system and the related work done in the domain. A literature study is also carried out to find the most suitable method for the data. Chapter 3 includes the methodology adopted for the study and chapter 4 presents the results and a discussion on the results from this thesis. Chapter 5 includes the conclusions from the work carried out and future research is considered in chapter 6.

2
Theory

In this chapter a more comprehensive background is provided about district cooling system as a whole and for the district cooling network in the city of Gothenburg. Section 2.1 includes the background and technical details of the district cooling system in the city of Gothenburg. Section 2.2 describes the need for a state of the art approach to handle the data. Section 2.3 describes the literature review carried out in the study and the related work done in the domain.

2.1 Technical background

District cooling (DC) or district cooling system (DCS) is used to generate cooling centrally by utilizing chilled water or other cooling medium and utilise a distribution network to supply cooling energy to different buildings [7]. District cooling is used for different purposes such as space cooling in buildings, food supply storage and process cooling in industries [8]. Usually, the source of cooling is a large water body such as a river or lake. This type of cooling is known as free cooling.

When free cooling is not possible, cooling towers, absorption and compression chillers are used as a source for providing cooling. The district cooling system can be divided into the main system or production system, district cooling network consisting of pipelines and substations. The production plant and the distribution network are operated and controlled by the utility provider whereas the substation is controlled by the building owner.

District cooling is a suitable option when the cooling loads are high per surface area such as in densely populated cities having different use for cooling [7]. The cooling is necessary for space cooling, food supply chain, industrial processes and other uses. District cooling is mainly used to meet the space cooling demand in university campuses, hospitals, and office buildings, this is done by heat removal.

District cooling is used in commercial buildings to provide cooling because of its high investment cost but low operating costs [8]. It is more prevalent in regions where the outdoor temperature is high for most part of the year. Countries such as USA, the middle east, and Japan which experience hot summers require a significant amount of cooling for a comfortable indoor environment. At the same time, this consumes a lot of energy. USA and Japan have high demand for space cooling [8], Europe on the other hand has low space cooling demand but in the recent years with increasing global temperatures and consequent hot summers the demand for space cooling has increased substantially.

Space cooling is predominant during the day when commercial buildings are generally occupied and in operation, and when the outdoor temperature is high. In Europe, and Sweden in particular the share of space cooling in the service sector buildings was 14% and about half of these demands were satisfied with district cooling. District cooling systems are generally smaller compared to district heating in Sweden [9].

2.1.1 Low delta T syndrome

District cooling systems are designed to achieve a high delta T between the supply and return temperature. However, many systems fail to achieve this design objective. This phenomenon is termed as "Low delta T syndrome" in the industry. There are several reasons for a low delta T such as improper cooling in the cooling coils, faulty heat exchangers and this causes problems like high energy consumption in the substations. Studies have been done to identify and mitigate the causes of low delta T in buildings [10].

One such solution to improve the delta T of a DCS is to use heat pumps in buildings to transfer the excess heat to the return line. The heat pumps can be used directly in buildings which have a DCS or buildings which are only connected to the DCS return line [11].

2.1.2 District cooling system in Gothenburg

The district cooling system in Gothenburg is controlled and operated by the local utility company (Göteborg Energi AB). In the district cooling system of Gothenburg, both free cooling and chillers (both absorption and compression) are utilised to meet the cooling demands.In the city of Gothenburg, the Göta älv is the main source of cooling for most part of the year. During the summer months when the temperature of the river is high (≥ 5 °C), absorption and compression chillers are used for supplying cold water.
Figure 2.1 shows the map of the district cooling system in Gothenburg. The blue lines represent the distribution network and the capacities of the chilled water production plants are indicated by the numbers. The capacity of the entire network is 78.3 MW, previously the network separated by the river was not connected to the bigger network and the project of connecting the two networks via a underwater pipeline was recently accomplished. This was done to meet the increasing cooling demand of the city. The river is used for free cooling when the temperature of the water is ≤ 5 °C, this happens during the months of December-April. The cooling demand of different connected buildings depends mainly on the outdoor temperature but other factors such as location, building type, and period of operation play a role too.

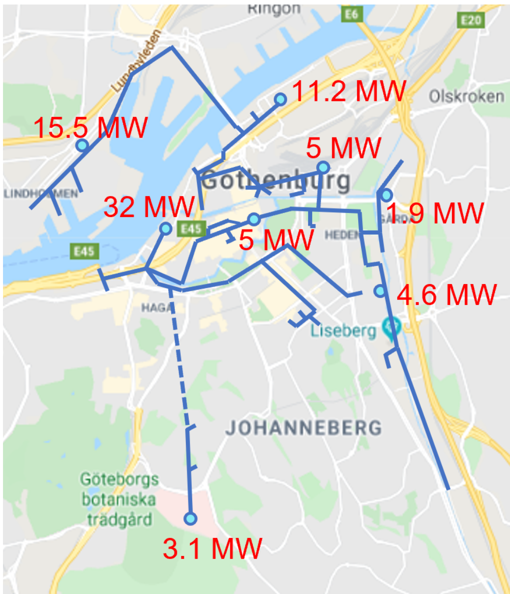


Figure 2.1: Map of district cooling network in the city of Gothenburg, 2021

Conventionally, production system (production plants) and distribution networks (pipelines) which are owned by the utility provider are monitored regularly, while substations owned by the building owner are considered to be working normally as long as customer comfort remains unaffected [2].

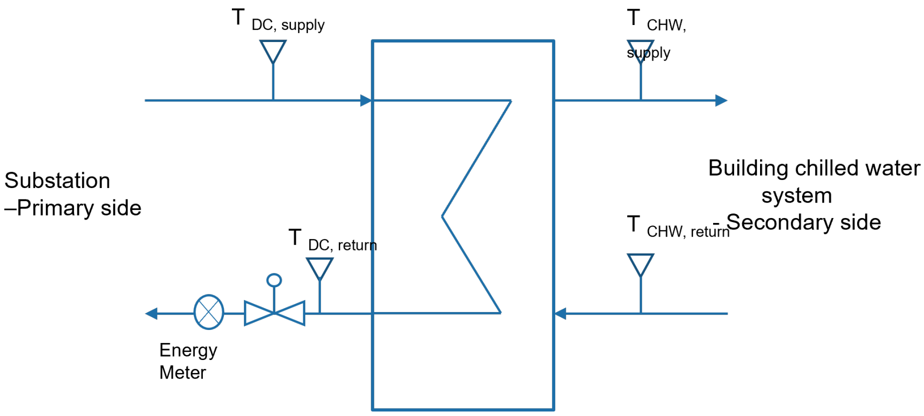


Figure 2.2: Schematic diagram of a substation

The above figure shows a schematic of a typical substation. The data analysed in this study is taken from the primary side of the substation as highlighted. The substation consists of control valves, measuring instruments such as flow meters, temperature sensors and energy meters. The locations of the different measurement devices and temperature sensors are also shown in the diagram.

2.2 A data driven approach

Smart meter readings in Sweden were introduced in 2015 after a change in the Swedish DH act (Fjärrvärmelag) [12]. This enabled the utility providers to bill their customers on the average monthly heat use. The metering devices on the primary side of the substation are used for data collection purposed by the utility provider. This in turn enables the use of data driven approaches to analyse and study the behaviour of the substations in DH networks [13], as well as DC networks. Data driven approaches have not been previously implemented on a large scale for DC systems.

2.3 Related work

In this section, the related work in the domain of fault detection is presented. The approach carried out in this thesis is inspired by similar studies for district heating. Månsson et.al [5] applied an automated method to detect faults in district heating substations based on a statistical approach. The study was tested on a dataset of 3000 substations from a district heating system in Sweden. Around 1273 installations were found to be poorly performing.

A reference based approach was previously adopted in [13], where different substations were grouped based on their operational behaviour. This reference based approach to detect faults has been studied on district heating substations but not on the district cooling substations. Since this also requires extensive prior domain knowledge, it is difficult to implement this method where sufficient information is not available.

Farouq et.al [13] analysed the behaviour of different district heating substations using a reference based approach. The approach was used to monitor the return temperature of 778 substations associated with multi-dwelling buildings. The reference group was used to detect outliers. The study found that 77 target substations were outliers, i.e., their return temperature were highly varied compared to those of in their reference groups.

Gadd and Werner [2] studied the faults by analyzing smart meter readings of 135 substations from two different district heating systems in Helsingborg and Ängelholm, Sweden. Around 74 % of the substations were found to have faults. The faults were categorized into three subgroups. The main conclusion of the study was that faults can be detected proactively using continuous commissioning of DH substations.

An extension of this work was carried out by Calikus et.al [14], where a data driven approach was implemented to analyse heat load profiles of different customer groups and detect anomalies based on deviations from the rest of the customer group.
Another study done by Gadd and Werner [12] analysed data from 140 substations using temperature difference signatures to detect faults and carry out quality assurance of the eliminated faults. Among the 140 substations studied, the fault frequency was found to be more than 6 %.

Ranking of substations based on their power signature was done by Calikus et.al

[6], this meant that substations with high dispersion were poorly performing and showed some sort of fault or issues in the substations. They presented a method where ranking of substations was done by measuring both dispersion and outliers in the power signature. This novel method proved to be better than the state-of-the-art outlier detection method which is based on z-score detection.

Gianniou et.al [15] performed a cluster based analysis on residential district heating data for households in Denmark. The main conclusions of the study were a majority of the customers can be represented by constant head loads. Seasonal change has not only an impact on the heat patterns but also consumption intensity and behaviours. The study also indicated that clustering-based forecasting for households is possible.

Danov et.al [16] implemented a method to determine the total heat loss coefficient, effective heat capacity and net solar gain of a building using linear regression approach. The method utilised data from nine public buildings in Spain and the method can be adopted for a larger set of buildings. In addition, Noussan et.al [17] carried out an analysis of a DH system to identify the main characteristics of heat load profiles and two main patterns were highlighted based on the hours of the day and outdoor temperature which mainly determined the heat demand of the building.

A common outcome from carrying out this literature study was that, there are of course limitations when choosing a novel method which has not been tested on the data before. However, analysing the behaviour of substations from different perspectives gives valuable results regarding the performance of buildings and detection of faults. In this thesis, the cooling energy, delta T and return temperature are studied as a function of outdoor temperature. Hereby referred to as, energy signatures, delta T signatures and return temperature signatures. The signatures were first referred in [8] where the cold loads were studied for hourly average values and daily average values.

3
Method

In this section, the methodology of the thesis is presented. A data driven approach is implemented to analyse the performance of substations. The different steps in the data analysis process are explained and presented in the following sections.

3.1 Data collection

In this study, data from the primary side of the substation (refer to figure 3.1) is considered for the analysis. Around 180 substations corresponding to all the substations in the city of Gothenburg are taken into consideration. In Sweden, the data from smart meters is collected automatically. This may result in values which may deviate from the real values due to connection problems [14]. The data consisted of hourly meter readings of volumetric flow, energy consumption, supply temperature and return temperature from the different buildings. The table below shows the data measured along with the units and measurement intervals.

Table 3.1: Data collected from the primary side of the substation

Data variables	Units	Measurement Interval
Cooling load	kWh/h	Hourly average
Volumetric flow rate	m^3/h	Hourly average
Supply temperature	°C	Instantaneous
Return temperature	°C	Instantaneous

The data was taken from the year 2021, from January 1st to December 31st for each hour, so it consists of 8760 meter readings for each substation. In addition, the installation IDs of the substations were also provided. These ID's are associated with the substation to help identify the substations. Since the different customer categories for all 180 substations are not available on record from the provider, it limits the extent of the analysis that can be carried out unlike the one done by Calikus et.al [14]. Therefore, a dispersion based approach is adopted which is explained in further sections.

3.1.1 Customer categories

Unlike DH buildings where the customer categories were previously available for the data as studied by Gadd et.al [2], the categories in district cooling buildings are difficult to obtain and include mainly commercial buildings. This data was available for around 41 buildings. The following building categories were analysed for their delta T signatures:

Table 3.2: Categories of different substations and the number of buildings in each category

Customer categories	No. of substations
Offices with restaurants	11
Offices	7
Schools and public administration	6
Healthcare and hospitals	16

It is important to note that the customer categories are not documented or available from the utility provider, this categorisation of different buildings was done by taking inspiration from other works done by Gadd et. al and Calikus et.al [2],[14]. After matching the substation ID's to the actual buildings and the building owners, the above categories were created.

3.2 Data handling

In this section, the approach carried on the data to analyze the behaviour of the different substation and the choice of algorithm is described. The energy signatures and delta T signatures are treated separately which is explained in the following sections. The initial analysis of the data set with the hourly and daily values was done on MATLAB and later python was used to carry out the pre-processing and analysis. A schematic of the data handling process is shown below.

Figure 3.1: Schematic of data handling process

3.3 Data pre-processing

Data collected from smart meters usually contains erroneous values which makes the analysis step difficult. Therefore, before proceeding to the analysis step the data must be cleaned or pre-processed. Different ways of cleaning the data are available. Some common methods used are employed in this study to find the right fit for our data. After obtaining the meter readings from the different substations, the three signatures were plotted namely, Energy signature, Delta T signature and Return temperature signature as a function of outdoor temperature. Since the return temperature and supply temperature are measured instantaneously, the values are more prone to erroneous readings. Therefore, before plotting the delta T signatures for the substations the delta T was calculated from the flow rate and cooling energy using the equation.

$$\dot{Q} = \dot{V} * \rho * C_p(T_r - T_s) = \dot{V} * \rho * C_p(\Delta T) \tag{3.1}$$

where,
V = volumteric flow (m^3/s)
ρ = density (kg/m^3)
C_p = specific heat capacity of water (J/kg.°C)
T_s = supply temperature (°C)
T_r = return temperature (°C)

As previously mentioned, the data set consisted of hourly average values for energy consumption, flow rate, and instantaneous values of supply and return temperature for 180 substations in the DC system of Gothenburg, Sweden. The hourly values were converted into daily average values to study the behaviour of the substations on a daily basis. Therefore, the hourly meter readings were converted into daily average values as follows: the energy consumption was calculated as the total daily cooling load, delta T was calculated from equation 3.1.

After obtaining the daily average values, the three signatures were analysed again for the 180 different substations. The daily average signatures of energy load and delta T were chosen for further analysis since the daily values describe the behaviour of the substation on a daily basis. Chapter 4 includes a comparison between the hourly signatures and daily signatures. For further analysis, the energy signature and delta T signature are taken into consideration since they give a more clear representation of the building's cooling demand and performance of the substation respectively. The return temperature signature, is though an important parameter to study when it comes to fault and performance diagnosis is not further analysed in this study. The hourly values of return temperature were used to obtain the signatures of the 180 substations. The findings from the analysis of hourly return temperature are presented in chapter 4.

3.3.1 Pre-processing of energy signatures

Energy signature (ES) methods have been used for characterizing heat load behaviour of buildings in many different studies [16],[17]. Outliers may be described as values which deviate from the expected behaviour of the parameter. A common method for detecting outliers in a data set is by finding the z-score of the data. This was tested on the data set to find outliers. The z-score is calculated as follows and the choice of the outlier threshold is described after:

$$z = \frac{y_i - \mu}{\sigma} \tag{3.2}$$

where,
y_i = cooling load,
μ = mean,
σ = standard deviation of the data

Given the outdoor temperature x_i, the cooling load y_i, and the predicted cooling load $\hat{y}_i$, the residual is calculated as shown:

$$r_i = \hat{y}_i - y_i \tag{3.3}$$

The standard threshold for data points to be considered as outliers is 3σ from the mean or a z-score of 3 [14]. The outliers are determined based on the following condition,

$$f(x_i, y_i) = \begin{cases} outlier, & if\ r_i - \mu \geq 3\sigma \\ inlier, & otherwise \end{cases} \tag{3.4}$$

This approach of detecting outliers is appropriate for a data set which is normally distributed. For a dataset which falls outside the normality distribution a different approach must be adopted. A dispersion based approach to find the degree of dispersion in the energy signature is therefore adopted in our method which gives an indication of the behaviour of the substation which will be explained in the next section.

3.3.2 Pre-processing of delta T signatures

The delta T signatures have to be treated differently than the energy signatures because of the nature of the signatures. The calculation of delta T is explained in the previous section 3.3. After calculating the delta T signature, the data still has to be pre processed for missing values and mathematical errors. The missing values were omitted from the data, this made the analysis simpler and robust since missing delta T indicates when the cooling power is zero and there is no flow in the substation. Values which are outliers can then be easily identified by setting thresholds. The choice of threshold requires some domain knowledge. The most common choice of thresholds from literature study was found to be three standard deviations (3σ) away from the mean value [6]. Any data points outside this threshold are considered as outliers.
After establishing thresholds, the delta T signatures can be studied for faults such as low delta T.

3.4 Data analysis

After data pre-processing, the dataset consisting of energy load, delta T and return temperatures are used to obtain the signatures as function of outdoor temperature. Each category of signatures requires a different approach and this is described in the following subsections.

3.4.1 Energy signatures

3.4.1.1 Evaluation based on clustering

K-means clustering algorithm was chosen to evaluate the energy signatures. Initially, this algorithm was tested on one substation. K-means clustering classifies the data according to the user selected clusters. To select the optimal number of clusters, different methods can be employed [18]. One such method, known as "elbow method", was used to determine the optimal number of clusters for our substation. In this method, an arbitrary value of K (number of clusters) is chosen and the sum of squared errors is calculated. This is done for different values of K, the higher the number of clusters the lower is the sum of squared errors.

3.4.1.2 Evaluation based on linear regression

The energy signatures are studied to understand the cooling demand of a building with respect to the outdoor temperature. The cooling demand of the building is mainly from comfort cooling or space cooling of the indoor environment, and this increases with the outdoor temperature. Naturally the cooling required is higher in summer than winter. The temperature at which the cooling demand increases with outdoor temperature was identified to be 10 °C. This also depends on the type of building and its time of operation. After the threshold was identified, the data was taken for outdoor temperature $\geq$ 10 °C. The remaining data is not taken into consideration since most substations are either not operating below this outdoor temperature or have low flows in the system causing the meter to not pick up the readings.

The next step is to normalise the data to carry out a comparison. Since different substations have different scale of operation throughout the year due to building type, customer needs and time of operation, it is imperative to scale the signatures to a uniform standard to enable a comparison of their behaviour. This was done by using the following function in python.

$$\mathit{MinMaxscaler()} \tag{3.5}$$

Linear regression model, also known as Ordinary Least squares (OLS) method was chosen to study the behaviour of the energy signatures of the substations. This is a typical approach used to estimate energy signatures. A robust regression approach was carried out in [6], however, this method requires a prior knowledge of inliers and outliers in the data which is not always the case. Therefore, a linear regression model was adopted.

Linear regression is used to fit a model to minimize the residual sum of squares between the observed data points and the data points estimated by linear approximation. After the model has been tested on our data, the next step is to find the degree of dispersion in the signatures. This is done by finding the R^2 score of each signature. R^2 is called the coefficient of determination and can be described as a way to evaluate the performance of the linear regression model. The mathematical expression of R^2 score is shown below:

$$R^2 = 1 - \frac{SSE}{SST} \tag{3.6}$$

where,
SSE = the sum of squares of residual errors and,
SST = the total sum of errors.

An R^2 score of a perfect model is 1.0. This means that the predicted values are the same as the actual values. After measuring the dispersion of the different signatures, the substations are ranked based on decreasing order of R^2 values.

3.4.2 Delta T signatures

The delta T signatures are analysed to identify abnormal or faulty behaviour in the substations. This can be identified by a low delta T, which has previously been studied by Jangsten et. al [19]. Delta T in DC system is generally calculated from the return temperature and supply temperature according to the following equation,

$$\Delta T_{DC} = T_{return} - T_{supply} \tag{3.7}$$

However, as already stated due to the return and supply temperature being measured instantaneously the readings are prone to errors and jumps. Typical delta T for district cooling systems is about 9-12 °C. This is most economical to the system and has low operating costs [7]. 10 °C is the design delta T of Gothenburg DC system.

Under each customer category, a reference group of substations was chosen which was identified as well performing, and it is with respect to this reference group that the performance of other substations within the customer category is evaluated. The thresholds were set in place for delta T, calculated based on the mean and standard deviation of the delta T values. The data points which fall outside these thresholds are considered as outliers and signatures which show deviation from the reference case are identified as poorly performing. Using this approach, poorly performing substations can be identified and studied.

Based on the different customer categories as highlighted in table 3.2, the substations are grouped accordingly and their performance based on their delta T is studied. The subgroups are dealt with individually. This is mainly done because dealing with such a large number of substations becomes cumbersome and tricky.

Each subgroup has its own average delta T and based on the data from each substation within the subgroup, the mean and standard deviation with respect to the reference group are used to set the thresholds to determine the performance of the substation. This is done to prevent the influence of skewed readings while establishing thresholds.

4

Results and Discussions

In this chapter, the results of the study are presented along with a discussion about the findings. This chapter is divided into three sections, each presenting the results from one of the three signatures. section 4.1 includes the findings from the return temperature signature, section 4.2 includes the results from the analysis of energy signatures and section 4.3 includes the results from the analysis of delta T signatures.

4.1 Return temperature signatures

The return temperature signatures were studied for hourly values of the return temperature. This was done initially to observe the nature of the signatures. The return temperature from the primary side of the energy transfer station as mentioned previously was used to plot the signatures. It is however not possible to show the signatures for all 180 substations, therefore a few selected substations which showed different trends are presented below. The variation of the hourly return temperatures versus the outdoor temperature of different substations can be observed below.

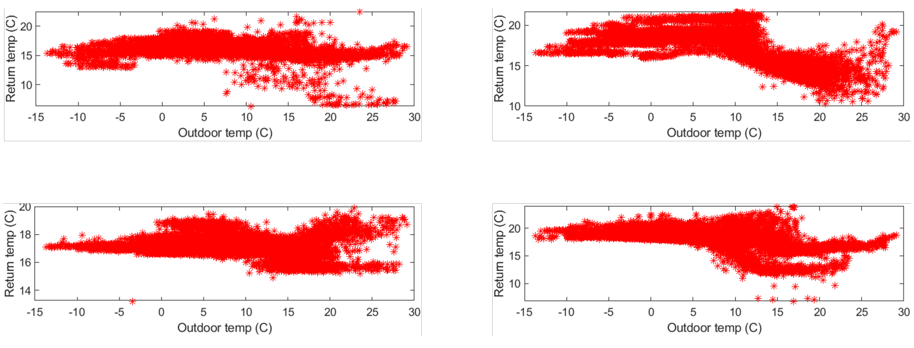
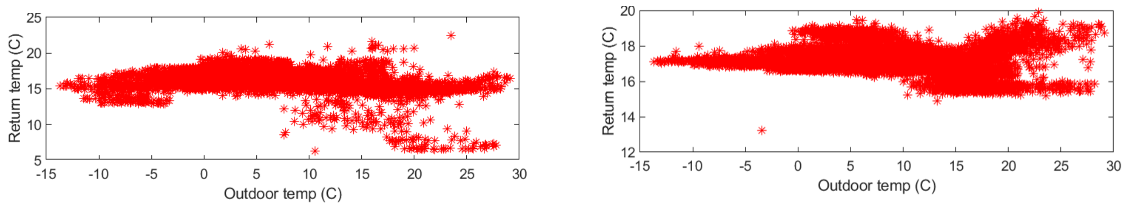


Figure 4.1: Return temperature of different substations in hourly resolution (*note the y-scale is different for each figure*)



(a) *Return temperature signature of a substation with low return T*

(b) *Return temperature signature of substation with high return T*

Figure 4.2: Return temperature signatures of substations which show variations with outdoor temperature

In the above two substations, we can observe how the return temperature varies hourly with respect to the outdoor temperature. In the first substation, there are periods when the return temperature is very low ($\leq$ 10 °C) for high outdoor temperatures this may be either due to jumps in the meter readings or when the substation is turned off during non operational hours (particularly at night). A low return temperature indicates that there is not sufficient heat transfer taking place in the walls of the heat exchanger, this may be due to faults such as fouling. Another reason is that since the data points in the figure include all hours, including when the flow is zero, the data is misinterpreted. It is only interesting to analyse return temperature signatures when the flow is greater than zero (when there is a cooling demand). Therefore, to proceed with further analysis of the return temperature signatures a method of removing data when the flow is zero would be needed, which was decided to be outside the scope of this thesis.

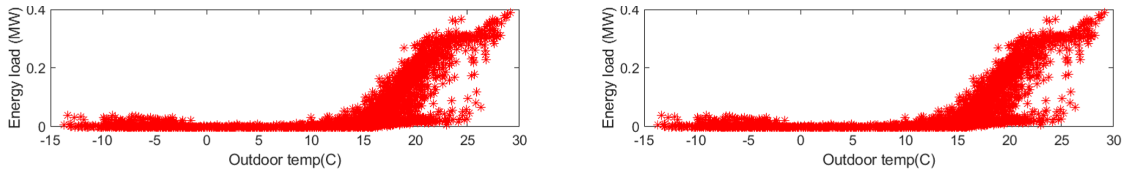
4.2 Energy signatures

4.2.1 Data pre-processing

The results from data pre-processing are used to study outlier behaviour and deviating values in the energy meter readings in the substations. Based on the method adapted, the deviating values were observed and identified. As previously stated, values which were identified as outliers were in fact periods when the substation was under periods of normal operation. This was confirmed based on prior domain knowledge and close inspection.

4.2.2 Evaluation

Here the findings of the energy signatures are presented both for the hourly resolution and the daily resolution. After analysing the energy signatures of different substations, it was found that most substations have an increasing cooling demand with an increase in outdoor temperature. Since it is not possible to show all the substations analysed in this section, typical examples are chosen and presented below. The hourly and daily energy signature are first presented to show the general trend of the data points.



(a) *Energy signatures for hourly cooling load for a substation*

(b) *Energy signature for daily average cooling load for the same substation*

Figure 4.3: Comparison between hourly and daily energy signatures

In the above figure 4.3, the energy signature for one substation is shown for both the hourly cooling load and daily average values. We can observe that for lower outdoor temperatures, the cooling load is almost absent or very low. This was the case for most substations in the study.

4.2.2.1 Evaluation based on clustering

K-means clustering was initially chosen to study the performance of the substations from the energy signatures. However, the choice of selecting an optimal number of clusters for such kind of signature data is difficult. Clustering works well when working with time series data where the cooling load is distributed over a time period and as a function of time such as days,weeks, or even months. Figure 4.4 shows k-means algorithm applied to a particular substation. We can see that choosing the optimal number of clusters is not possible when cooling load is studied as a function of outdoor temperature. The purple stars represent centroids from which the distance of other data points is calculated and clusters are classified based on this distance also known as sum of squared errors.

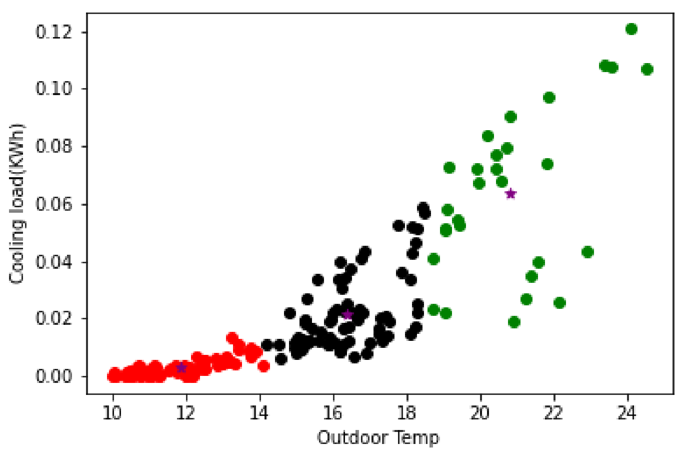


Figure 4.4: Clustering of energy signature for a particular substation

4.2.2.2 Evaluation based on linear regression

When the outdoor temperature reaches a certain limit, the cooling load increases almost linearly. For this reason, the cooling load for outdoor temperature ≥ 10 °C is considered and linear regression is carried out. Figure 4.5 shows the cooling load of different substations with respect to the outdoor temperature, the scales have been normalised to enable comparison since different substations have different cooling demands throughout the year.

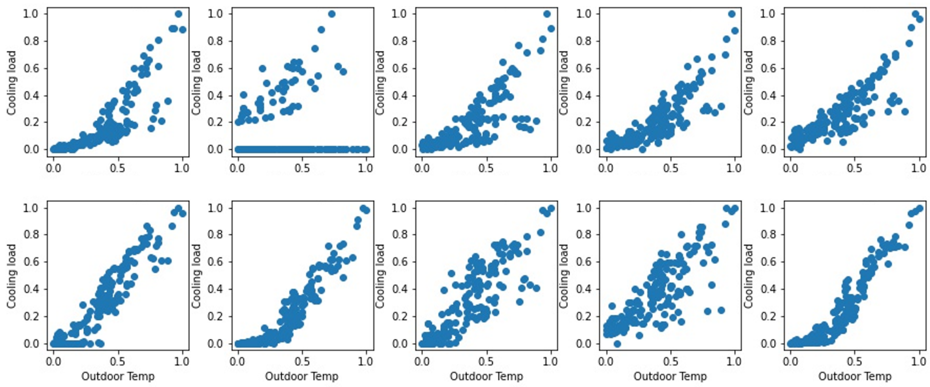


Figure 4.5: Energy signature for a set of substations, the scales have been normalised to enable comparison(for outdoor temperature greater than 10°C, 1.0 on the x-axis represents an outdoor temperature of 30°C)

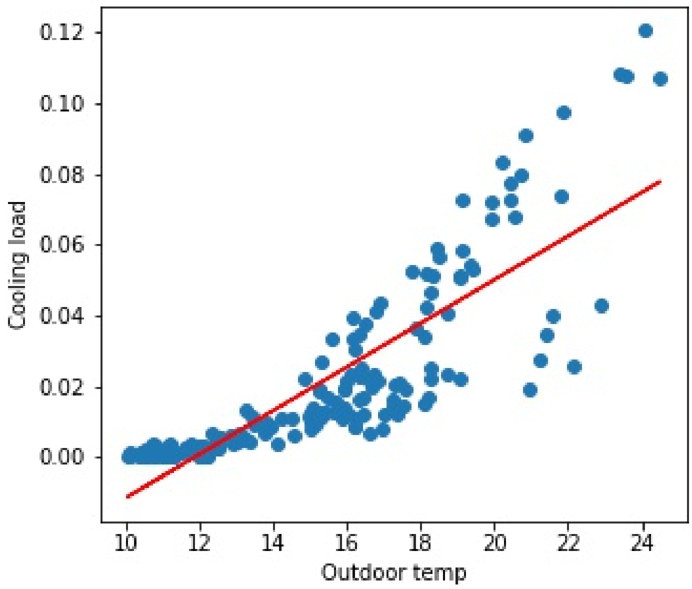


Figure 4.6: Energy signature of one substation after carrying out linear regression.(non-normalised data)

In figure 4.6, the cooling load of a substation is shown with respect to the outdoor temperature. Linear regression model fits the line of best fit for the given data. The R^2 score of this model gives the dispersion of the data points from the fitted line. From the R^2 value, the performance of the substation can be estimated.
The regression results are similarly found for all the substations and studied. Substations which had an R^2 score between 0.7 and 0.9 indicate normal operation or typical operation. This can be either due to comfort or space cooling in buildings such as offices, schools, hotels etc. Substations with lower R^2 value indicate atypical operation and possible faults. By atypical operation, it means that the substation has two modes of operation; a base load operation and comfort cooling. The typical value of R^2 for such type of substations is between 0.4-0.6. An example of such a signature is shown below.

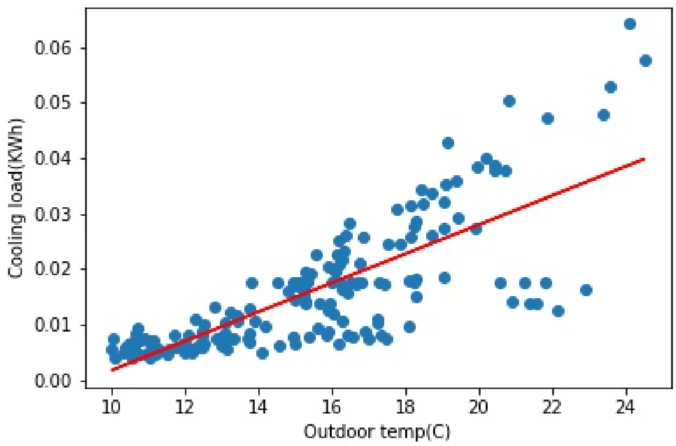


Figure 4.7: Example of substation with two modes of operation

Substations with R^2 score lower than 0.4 depict faults such as leakages. This can be seen in the figure below, where a large number of data points are concentrated around 0 KWh cooling load. Figure 4.8 shows one such substation where faults are observed.

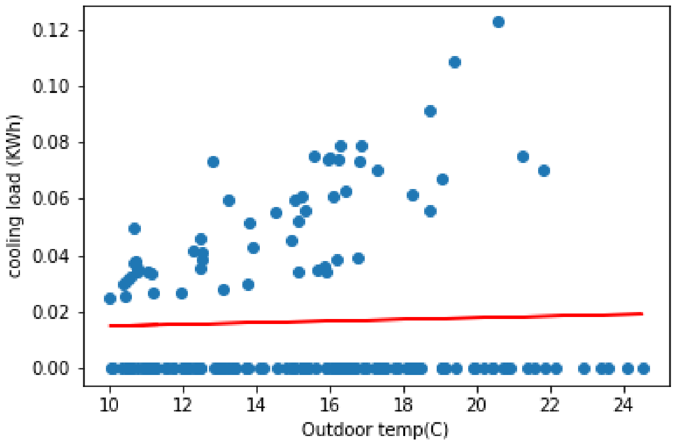


Figure 4.8: Example of a substation with faults

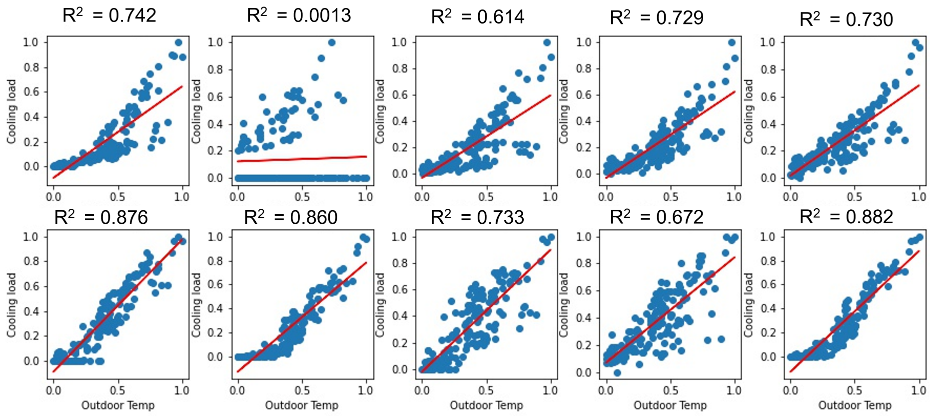


Figure 4.9: Example of Linear regression carried out on a set of substations, the substation ID's are not shown for anonymity

Figure 4.9 shows the trend of different substations with respect to outdoor temperature. The substations follow a linear trend with respect to outdoor temperature when it comes to energy signatures. After running the linear regression model, the dispersion of the different data points with respect to the fitted line are calculated, this is shown in figure 4.9 where the regression line is fit for different set of substations and the regression score is stored as the R^2 value. We can also notice that one particular substation which has a large number of data points which lie close to zero cooling load, this indicates some kind of fault in the substation such as leakages. This is also confirmed by a low R^2 value for that particular substation.
It is difficult to group substations based on their regression values and mode of operation since different types of buildings fall into overlapping categories. Substations with R^2 lower than 0.4 show faulty behaviour such as leakages or fluctuating flows which can be observed in few substations. This problem is fairly common and often under reported. After carrying out the regression analysis on all 180 substations, the results were summarised in a histogram.

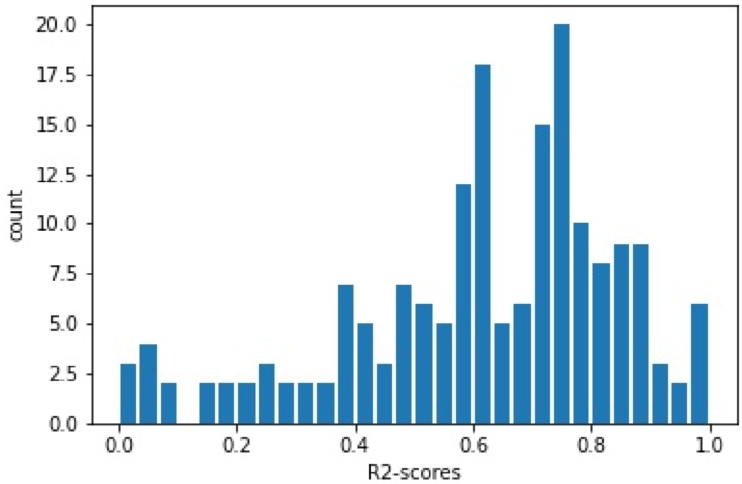


Figure 4.10: Distribution of R^2 score for different substations

From the figure 4.10 we see that most substations have an R^2 score between 0.6-0-8. Out of 180 substations studied, around 53 substations show a high R2 value indicating normal operation. In our data, substations which are not operated throughout the year have an R^2 value of 1.0. Substations which have a low R2 value which is lower than 0.4 indicate some sort of faults and poor performance. These substations can be identified from their ID's and the faults can be addressed.

4.3 Delta T signatures

This section presents the results from the analysis of the delta T signatures. The delta T signatures like the energy signature were analysed for the daily average values for the purpose of uniformity.

4.3.1 Data pre-processing

Since delta T for the substations was calculated from the flow rate and cooling energy as described in equation 3.1, the instantaneous supply and return temperature readings have no influence on the delta T signatures. This also eliminates the possibility of erroneous readings to some extent. Mathematical errors from the calculation step are omitted from the data. There are data points which are zero or close to zero for the calculated delta T's this is because the cooling power is zero but there is small amount of flow picked up by the meter resulting in this problem.

4.3.2 Evaluation

The performance analysis of the substations based on the delta T signatures is explained in the following subsection. Before the results of the different customer categories are presented, the hourly and daily delta T signatures for two different substations are shown below. We observe the variation between the hourly and daily delta T values, the second substation which has a delta T of 15 °C does not necessarily imply that there are faults in the substation. It may be possible that this substation has been designed for a delta T of 15 °C rather than for 10 °C as can be seen from the first substation.

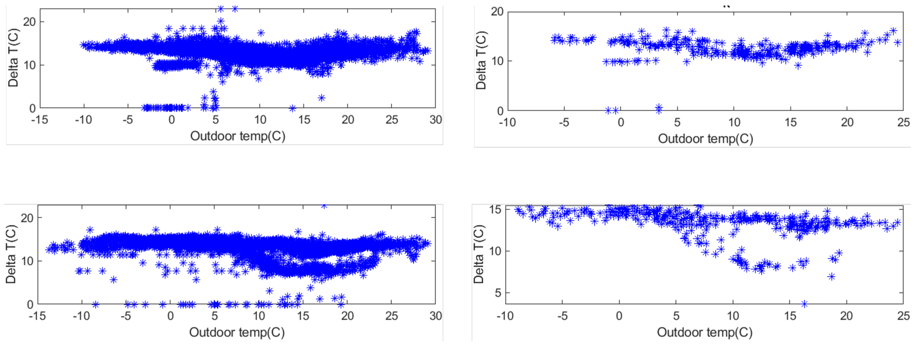


Figure 4.11: Delta T signatures of different substations *(Hourly and daily values on the left and right respectively, note that y-scale is different for each figure)*

For further analysis, the substations are grouped in different customer categories and a reference group is selected and the performance of the substations under each customer category is studied. The figure below gives a representation of delta T signatures for different substations with respect to the design delta T of the DC system.

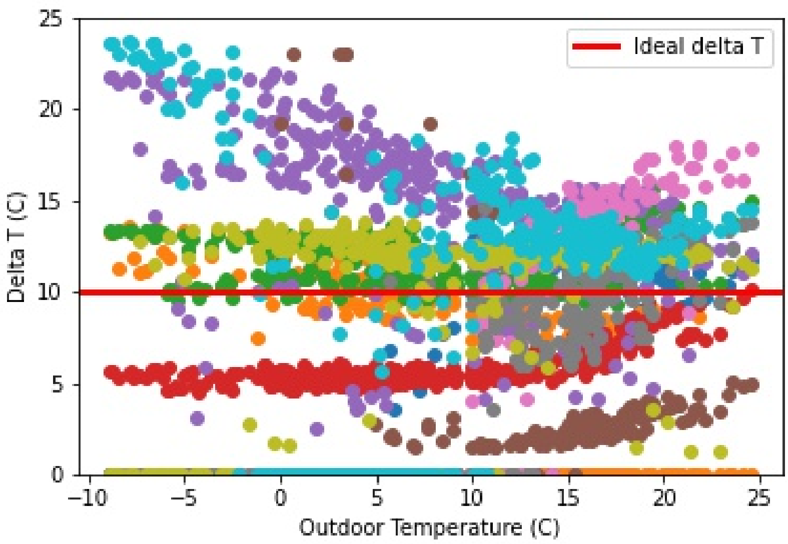


Figure 4.12: Delta T of different substations with respect to the ideal delta T

Figure 4.12 shows the delta T of different substations with reference to the ideal or design delta T, this ideal delta T was chosen based on domain knowledge and from the literature [7]. We observe some substation have a delta T which is much lower than the ideal delta T or the design delta T for which the substation is designed for. This low delta T causes a higher flow rate from the substation reducing its

efficiency. Identifying these substations with low delta T is crucial since a low delta T indicates faults in the substation such as improper cooling in the cooling fans and coils of the building, faults in heat exchanger. A low delta T is compensated by an overflow in the system which in turn leads to increased water flow rate.
To solve this issue of low delta T in substations, regular inspections of heat exchangers is required and maintenance of the valves is necessary periodically. Continuous monitoring of supply and return temperatures must be done to prevent low delta T syndrome and achieve the design delta T [7]. As already pointed out, customer comfort is the main criteria when it comes to dealing with faults in substations. We can see from this study that through a data driven approach fault detection can be done automatically and robustly.
The figures below shows the different customer categories with their delta T signatures, each subgroup has its average delta T and the performance of each substation within the subgroup is studied.

4.3.2.1 Category 1: offices with restaurants

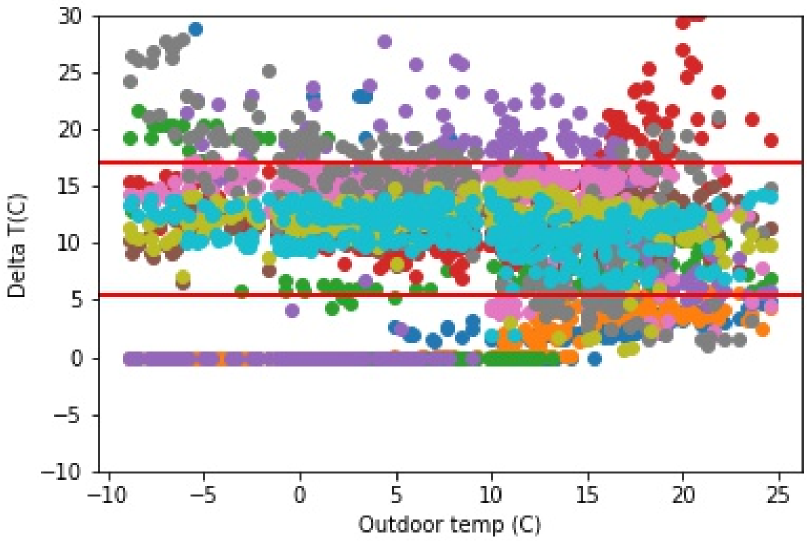


Figure 4.13: Delta T of substations in category 1: offices with restaurants. Each color represents an individual substation.

In the above figure, the upper and lower thresholds were calculated from the mean and standard deviation of the data and found to be 16.9 °C and 5.5 °C for the substations. We notice that few substations perform poorly for higher outdoor temperature, this is observed from the high number of data points outside the thresholds. Values which lie outside the thresholds can be considered as outliers. The data points which have zero delta T can be disregarded since those indicate when the substation has no flow and is not in operation.

4.3.2.2 Category 2: office buildings

For the next category of substations which included only office buildings, it was not possible to find a reference substation which was well performing, this was because the delta T for every building was either too high or too low throughout the operational period. The upper and lower thresholds were highly skewed for the data. Out of seven buildings analysed within this category, only a few substations had normal delta T (between 10-15 °C). Due to the nature of the signatures for this customer category, identifying a reference group which was well performing was not possible and therefore it proves difficult to identify the exact faults in the buildings. Another reason as to why it is not possible to identify a well performing substation is because there are few substations within the category and this narrows down the analysis.

4.3.2.3 Category 3: schools and public administration buildings

Another category investigated was school buildings and public administration buildings. Only a small number of buildings were identified under this category. These building types are not exclusive to the defined category because each building also has either offices or restaurants along with the said building types. The upper and lower thresholds for the data was found to be 16.9 °C and 5.3 °C, respectively. The average delta T for the subgroup was calculated to be 8.7 °C which is not too far away from the design theoretical delta T (10 °C).

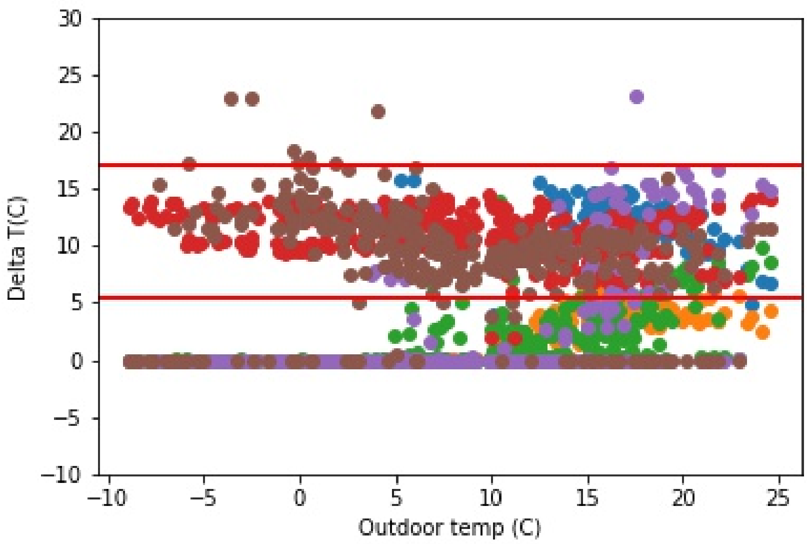


Figure 4.14: Delta T of substations in category 2: schools and public administration

4.3.2.4 Category 4: Healthcare and hospital buildings

The last category of substations analysed was hospitals and healthcare buildings. The substations which fall under this category are exclusively of one building type unlike the ones studied in other categories. This makes the findings and analysis interesting. Around 16 substations were studied in this category, the upper and lower thresholds were calculated in the same way as before. We can observe from the figure below that few substations have high delta T for high outdoor temperatures.

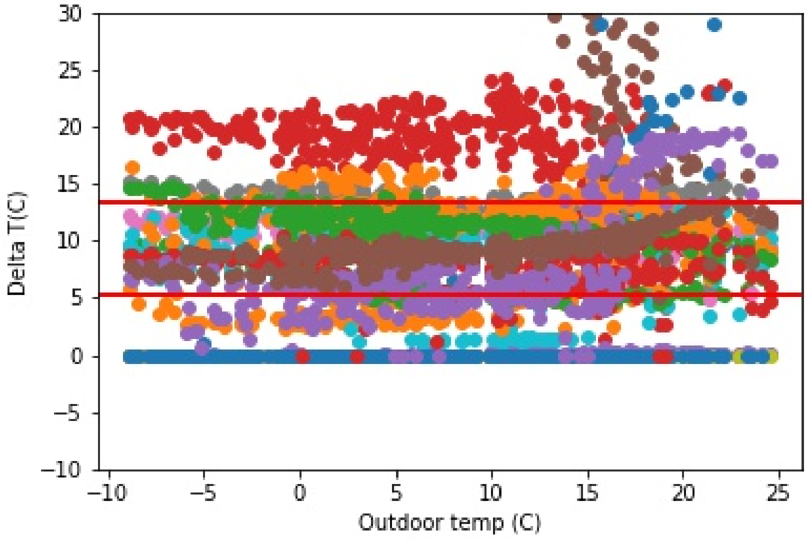


Figure 4.15: Category 3: Healthcare and hospitals.

The results from different customer categories and their average delta T are summarized in the table below:

Table 4.1: Customer categories with their average delta T for each subgroup

Customer categories	No. of buildings	Average delta T (°C)
Offices with restaurants	11	11.2
Offices	7	5.9
Schools and public administration buildings	6	8.7
Healthcare and hospitals	16	9.3

From the analysis of the delta T signatures and the table above, we see that office buildings have very low delta T compared to the other categories. Out of 41 buildings studied under the different customer categories, the delta T for office type buildings was found to be highly varied with low delta T's in many buildings indicating some sort of faults in the substation. There are also several buildings in other categories where this issue is prevalent and since it was not possible to categorise the remaining substations to carry out performance analysis using delta T signatures, it stands to reason that there are substations where low delta T syndrome occurs and this must be addressed. The reference based approach is suitable for identifying and dealing with low delta T in buildings. However, this approach is only possible when the customer categories of the substations are known and the knowledge of reference groups within the customer category. For a large number of substations, this reference group based approach requires information about the customer categories.

5 Conclusion

In this chapter the main conclusions and findings from the thesis are summarised. The aim of this thesis was to employ a data driven approach to analyse the performance of district cooling substations and detect faults in the district cooling substations. In this thesis, different approaches are employed taking inspiration from the studies carried out in district heating systems and the results are presented. The energy and delta T signatures were the main focus of this study. Data was taken from 180 different substations in Gothenburg and analysed. The analysis of the energy signatures of the different substations showed that ranking substations based on the regression score and using the R^2 score to determine the dispersion can be used to detect faults and abnormal behaviour in substations. Several factors determine the cooling energy in a building and the mode of operation.

Out of 180 substations analysed, 53 substations were found to have normal operation with a high R^2 score. 39 substations had very low R^2 scores which indicates faults in the substations and poor performance. Other methods such as clustering which have been previously done on meter readings are inconclusive when applied on signatures, they are more suitable for time series data where mode of operation is available and investigated. The categorization of substations as well performing and poorly performing takes into account several factors such as building type and location, period of operation and mode of operation. It is rather subjective and the authors choice to consider these and many other factors while analysing the performance of substaions.

The delta T signatures which were studied for the 41 buildings within different categories required a different approach to detect faults in the substations. It was found that office buildings have the most variations when it comes to delta T signatures. This could be due to lack of maintenance or high operational hours leading to faults in the equipment. It is important to note that, despite implementing a data driven approach, a significant amount of domain knowledge is required while dealing with faults. A recurring issue while carrying out this study has been how one would treat these faults from a domain knowledge point of view.
The results from this study can be used as source of inspiration for future work in the fault detection domain.

6

Scope for future work

In this thesis, a data driven approach was implemented to study the behaviour of substations on a large scale. As mentioned, there are different methods to choose from when dealing with such a large set of data and the choice clearly depends upon the expected outcome from the work.

This project studied the operational data of substations as a function of outdoor temperature to detect faults and analyse the performance. There are however, other methods to identify and detect faults. Working with time series models to determine the mode of operation and different time of operation can be done to recognise the frequency of faults in the buildings. Seasonal variation and performance of substations for different seasons is also an interesting approach for performance diagnosis.

Identifying substations which have abnormal operation has been done in this study. Future research can include a way to categorize and group well performing and poorly performing substations based on their operational signatures. An approach to detect faults in substations is presented in this thesis but strategies to deal with these faults and resolve them must be done. Return temperature signatures which were considered in this thesis were analysed for 180 substations but further analysis requires more domain knowledge. The relation between these three different signatures can prove an important tool in performance analysis for district cooling systems.

Bibliography

[1] Iea. Energy efficiency 2020 – analysis.

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

[3] Abrar Inayat and Mohsin Raza. District cooling system via renewable energy sources: A review. *Renewable and Sustainable Energy Reviews*, 107:360–373, 2019.

[4] Maria Jangsten. *Gothenburg District Cooling System-an Evaluation of the System Performance Based on Operational Data.* Chalmers Tekniska Hogskola (Sweden), 2020.

[5] Sara Månsson, Kristin Davidsson, Patrick Lauenburg, and Marcus Thern. Automated statistical methods for fault detection in district heating customer installations. *Energies*, 12(1):113, 2018.

[6] Ece Calikus, Sławomir Nowaczyk, Anita Sant'Anna, and Stefan Byttner. Ranking abnormal substations by power signature dispersion. *Energy Procedia*, 149:345–353, 2018.

[7] Alaa A Olama. *District cooling: Theory and practice.* CRC Press, 2016.

[8] Svend Frederiksen and Sven Werner. District heating and cooling, 2013.

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

[10] Mingkun Dai, Xing Lu, and Peng Xu. Causes of low delta-t syndrome for chilled water systems in buildings. *Journal of Building Engineering*, 33:101499, 2021.

[11] S Tredinnick and G Phetteplace. District cooling, current status and future trends. *Advanced district heating and cooling (DHC) systems*, pages 167–188, 2016.

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

[13] Shiraz Farouq, Stefan Byttner, Mohamed-Rafik Bouguelia, Natasa Nord, and Henrik Gadd. Large-scale monitoring of operationally diverse district heating substations: A reference-group based approach. *Engineering Applications of Artificial Intelligence*, 90:103492, 2020.

[14] Ece Calikus, Sławomir Nowaczyk, Anita Sant'Anna, Henrik Gadd, and Sven Werner. A data-driven approach for discovering heat load patterns in district heating. *Applied Energy*, 252:113409, 2019.

[15] Panagiota Gianniou, Xiufeng Liu, Alfred Heller, Per Sieverts Nielsen, and Carsten Rode. Clustering-based analysis for residential district heating data. *Energy conversion and management*, 165:840–850, 2018.

[16] Stoyan Danov, J Carbonell, Jordi Cipriano, and Jaime Martí-Herrero. Approaches to evaluate building energy performance from daily consumption data considering dynamic and solar gain effects. *Energy and Buildings*, 57:110–118, 2013.

[17] Michel Noussan, Matteo Jarre, and Alberto Poggio. Real operation data analysis on district heating load patterns. *Energy*, 129:70–78, 2017.

[18] Alexander Martin Tureczek, Per Sieverts Nielsen, Henrik Madsen, and Adam Brun. Clustering district heat exchange stations using smart meter consumption data. *Energy and buildings*, 182:144–158, 2019.

[19] Maria Jangsten, Torbjörn Lindholm, and Jan-Olof Dalenbäck. Analysis of operational data from a district cooling system and its connected buildings. *Energy*, 203:117844, 2020.

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

AVANCEZ
1829

CHALMERS
UNIVERSITY OF TECHNOLOGY