UMEÅ UNIVERSITY

IDENTIFYING SYMPTOMS OF FAULT IN DISTRICT HEATING SUBSTATIONS

An investigation in how a predictive heat load software can help with fault detection

Tobias Bergentz

Master thesis, 30 hp

MSc programme in Energy Engineering, 300 hp

Department of Applied Physics and Electronics. Spring term 2020

Master's thesis 2020

Identifying symptoms of fault in District Heating Substations

A investigation in how a predictive heat load software can help with fault detection

Tobias Bergentz

Department of Applied Physics and Electronics
Umeå University
Umeå, Sweden 2020

Identifying symptoms of fault in District Heating Substations
An investigation in how a predictive heat load software can help with fault detection
Tobias Bergentz

Supervisors: Johan Kensby, Utilifeed AB
Ramtin Rabiee, Department of Applied Physics and Electronics
Anjan Rao Puttige, Department of Applied Physics and Electronics

Master's Thesis 2020
Department of Applied Physics and Electronics
Umeå University

Typeset in LaTeX

Identifying symptoms of fault in District Heating Substations
An investigation in how a predictive heat load software can help with fault detection
TOBIAS BERGENTZ
Department of Applied Physics and Electronics
Umeå University

Abstract

District heating delivers more than 70% of the energy used for heating and domestic hot water in Swedish buildings. To stay competitive, district heating needs to reduce its losses and increase capabilities to utilise low grade heat. Finding faulty substations is one way to allow reductions in supply temperatures in district heating networks, which in turn can help reduce the losses. In this work three suggested symptoms of faults: abnormal quantization, drifting and anomalous values, are investigated with the help of hourly meter data of: heat load, volume flow, supply and return temperatures from district heating substations. To identify abnormal quantization, a method is proposed based on Shannon's entropy, where lower entropy suggests higher risk of abnormal quantization. The majority of the substations identified as having abnormal quantization with the proposed method has a meter resolution lower than the majority of the substations in the investigated district heating network. This lower resolution is likely responsible for identifying these substation, suggesting the method is limited by the meter resolution of the available data. To improve result from the method higher resolution and sampling frequency is likely needed.

For identifying drift and anomalous values two methods are proposed, one for each symptom. Both methods utilize a software for predicting hourly heat load, volume flow, supply and return temperatures in individual district heating substations.

The method suggested for identifying drift uses the mean value of each predicted and measured quantity during the investigated period. The mean of the prediction is compared to the mean of the measured values and a large difference would suggest risk of drift. However this method has not been evaluated due to difficulties in finding a suitable validation method.

The proposed method for detecting anomalous values is based on finding anomalous residuals when comparing the prediction from the prediction software to the measured values. To find the anomalous residuals the method uses an anomaly detection algorithm called IsolationForest. The method produces rankable lists in which substations with risk of anomalies are ranked higher in the lists. Four different lists where evaluated by an experts. For the two best preforming lists approximately half of the top 15 substations where classified to contain anomalies by the expert group. The proposed method for detecting anomalous values shows promising result especially considering how easily the method could be added to a district heating network. Future work will focus on reducing the number of false positives. Suggestions for lowering the false positive rate include, alternations or checks on the prediction models used.

Keywords: [District Heating, Substations, Fault detection, Anomaly detection]

Acknowledgements

This master thesis constitutes the end of my education at the Master of Science program in Energy Engineering at Umeå University. The thesis was preformed in collaboration with Utilifeed AB. I would like to thank Utilifeed and all of its employees for giving me this chance and for answering all of my questions. A special thanks to Johan Kensby my supervisor at Utilifeed. I would also like to thank my supervisors from the university Anjan Rao Puttige and Ramtin Rabiee and a special thanks to Oskar Finnerman. Lastly i would like to thank my family for supporting me through this process

Tobias Bergentz, Umeå, May 2020

Contents

List of Figures

List of Tables

1
Introduction

District heating delivers 71% of all energy used in Sweden for space heating and domestic hot water [1]. To stay competitive compared to other energy sources, improvements in the district heating system is needed. As an example the overall distribution losses in Swedish district heating systems are today 13% [1]. A well known way to reduce these losses is to lower the distribution temperature in district heating networks. A lowered distribution temperature is also beneficial for improved combined heat and power plants efficiency and upgraded possibilities to utilize surplus heat or improve efficiency of low temperature solar heating system [2]. By eliminating current faults in district heating system a reduction in temperature levels can be achieved [3]. Faults that can occur in district heating substation are for example, stuck valves, drifting of sensors, excess noise in measurements and constant readings[4].

Utilifeed is a company developing software for district heating companies. In an effort to increase the features available for district heating companies Utilifeed wants to give the district heating companies a tool to preform fault detection within their platform. Utilifeed has a software predicting hourly heat load, volume flow, supply and return temperatures for individual substations called EnergyPredict and Utilifeed wants to investigate how EnergyPredict could be used for detecting faults in substations. This thesis aims to find symptoms of faults in district heating substations primarily with the help of prediction from EnergyPredict and hourly meter readings from substations.

1.1 Aim of this thesis

The aim of this thesis is to develop methods to detect symptoms of faults in district heating substations. Measured data from substations will be used for this, additionally the thesis also aim to utilize the predictions from the load prediction software EnergyPredict in the detection methods. Three symptoms of faults is to be considered: drifting of measured quantities, constant values of measured quantities and anomalous values of quantities.

1.2 Previous work in this field

There is an abundance of work within the field of anomaly detection, Chandola et.al [4] gives a good introduction to the field of anomaly detection. So does the two part review article series of Katipamula and Brambley [5] [6] which focuses on fault detection and diagnostics and prognostics within the HVAC field.

There have also been a few articles in the literature regarding fault detection in district heating substations but this sub field is less explored. Seem[7] shows that methods for detecting abnormal energy usage in buildings is possible with the help of intelligent data analysis and daily reading of the energy consumption. The early work of Sandin et.al [8] gives a good introduction to fault detection in district heating. It lists and explains basic methods for outlier detection and limit checking. The report gives some examples of how abnormal quantization errors such as stuck valves etc. can be detected with the concept of Shannon entropy. Månsson et.al [9] show that by predicting a parameter using Gradient Boosting Regressor on hourly meter data a machine learning prediction model could be used for fault detection. De Nadai and van Someren [10] uses a short term forecasting model based on a mixture of ARIMA and ANN to predict gas consumption in a building. The model is trained using historical data. When the difference between the measured and predicted value exceed a predefined threshold the data point is classified as anomalous. During the progress of this thesis an interesting report by Farouq et.al [11] was published in which a reference-group approach is used to detect outliers in district heating substations. Where similarly behaving substations are grouped, and outliers where defined based on if their behaviour changed from the substations reference group.

Using predictions to find anomalies or faults is a common strategy within the fields of fault and anomaly detection but to our best knowledge this is the first study to, use a commercial software that predicts heat load, volume flow, supply and return temperatures. This is also the first study to our knowledge that uses more than one predicted quantity to detect anomalies.

1.3 Demarcation

This thesis does not consider data from the non heating periods of the year. Furthermore, the considered data are only from multi-dwellings buildings, commercial buildings, industrial demands, health and social service buildings and public administrations buildings. No single family homes where part of the data set.

2
Theory

In this section all relevant information for understanding the methods used are presented. There is also a general explanation of district heating and substations.

2.1 District heating, and district heating substations

This section provides a summary of District heating (DH) and common practices in Sweden. A thorough explanation on the subject can be found in [12]. DH is the technology of delivering heat from a central heat production plant and then distribute that heat through some medium through a city, part of a city or a building complex. District heating systems is said to have different generation, characterised by some key aspects. The first generation DH networks used steam as the heat carrier medium. These first generation DH system are still in use in some parts of the world such as Manhattan, have a few drawbacks such as large heat losses and unfortunately also safety concerns. Within systems of the the first generation of DH there have unfortunately been a few cases steam explosions, some with fatal outcome [13]. The second generation DH networks used pressurized water as heat carrier, often of temperatures above 100 °C. These systems were the dominant type of DH systems created from the 1930s to the 1970s [13]. The third generation is sometimes referred to as "Scandinavian district heating technology", due to the fact that many of the DH component manufacturers are Scandinavian based. For the third generation temperatures are often below 100 °C however the heat carrying medium is still pressurised water.

2.1.1 District heating substations

A district heating substation is the interface between the DH grid and the energy consumers. There are some major differences nationally how substations are configured, this text will be focused on District heating substations (DHS) in a Swedish context. There is a wide variety of different substation types and connection schemes, the choice of which scheme is chosen is often based on national or local standards. A main distinction between directly connected and hydraulic separated systems can be made. In a directly connected system the same water that is heated at the heating plant is directly circulated in the consumers heating system and sometimes even directly used as domestic hot water (DHW). For instance in Denmark and Germany it is common to have a direct connection for the heating system but have a hydraulic separated system for DHW. The hydraulic separation in a DHS is performed by heat exchangers. In Sweden it is most common to have hydraulic separation of both the DHW and the heating system, the most common connection type in Sweden is the parallel connection [14]. Even though there is a large variety of different connections and connection types, most DHS can be divided into three main sections: metering, the space heating system and the DHW system. In addition to these sections there is also a controller which sends signals to the actuator responsible for adjusting the valves in the DHS. An example of a parallel connected DHS is found in fig(2.1), and in the following section a brief explanation of a parallel connected DHS will be given. The DHW section in a parallel connected DHS is composed of a heat exchanger, a control valve and an actuator on the supply side of the heat exchanger controlling the flow of DH water passing through the heat exchanger. In most larger buildings a circulation pump circulates the DHW water on the secondary side. This is done so that the temperature in the hot water pipes never drops below 50°C which prevents the growth of legionella bacteria. A control valve on the secondary side controls the flow addition of cold water to the system. The space heating system is quite similar to the DHW system and has a valve controlling the flow of DH water through the heat exchanger on the primary side, and on the secondary side there is a circulation pump responsible for circulating the water through e.g. the radiators or floor heating. The controller, can be of various complexity, but usually there are a few temperature sensors helping with controlling the space heating, and an actuator controlling the control valve on the primary side. Likewise there is a temperature sensor measuring the temperature on the secondary side of the DHW system.

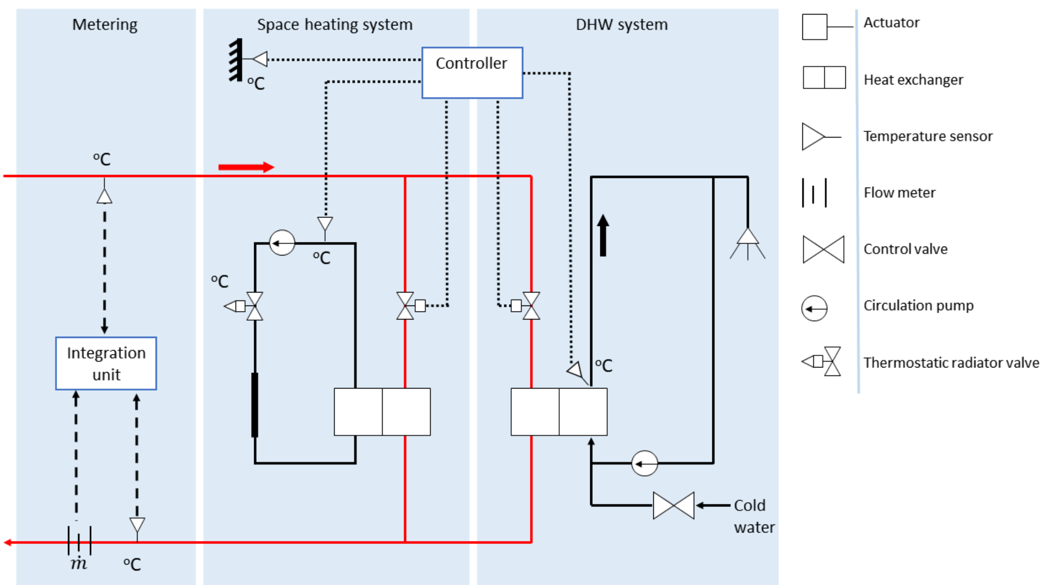

Figure 2.1: A parallel connected substation, recreated from (Månsson et al. 2019)

The metering in a DHS is primarily for billing purposes, since 2015 all meters in Sweden should be hourly and have a communication interface, responsible for sending the information of the measured data to the utility company [15]. The meter in an DHS is composed of four main components, two temperature sensors, a flow meter, and a integration unit. The temperature sensors are usually resistance temperature detectors, Pt1000 or Pt500 standards are the most common types [12] and they measure the temperature of the supply and return water. Flow meters comes in a large variety, the most common types in DHS are ultrasonic flow meters, inductive flow meters and velocity flow meters [16] the flow meter measures the flow of water through the primary side of the DHS. Three parameters are measured in a DHS, supply temperature °C, return temperature °C and volume flow m^3/s, they are used to calculate the heat power, using eq(2.1).

$$P_{heat} = \Delta T C_p \dot{V} \rho \tag{2.1}$$

where $\dot{V}$ is volume flow in m^3/s, ρ is the density of the fluid kg/m^3, C_p is the specific heat capacity in J/(kg K) and ΔT is the temperature difference between supply and return temperature ° C. The measurements from the sensors are sent to the integration unit and by taking the integral of the above equations the integration unit keep track of the cumulative heat use and flow that has passed through the DHS. Different meters can have different resolution and sampling frequency. It is common to have a resolution of 1 kWh for the heat use and a sampling frequency of 1 hour. For the flow and temperature sensors the accuracy varies more. If a DHS has a resolution of 1 kWh and the use is lower than 1 kWh but more than zero for an hour the reported energy will be zero until the cumulative use has exceeded the 1 kWh resolution threshold. The flow meter works in a similar fashion. The temperature sensors however most commonly reports the temperature reading at the exact moment of the sampling.

2.2 Faults in district heating substations

There are several types of faults that can occur in district heating substations. This section has a description of faults, and their influence on meter reading data. Faults in DHS can according to Gadd and Werner[15] be divided into three categories, these are component faults, construction faults, and operational faults. Since most substations today are delivered prefabricated, there is probably a decreasing number of construction faults. Component faults are simply a component that breaks. Unfortunately there is no publicly available up to date statistics of which fault that are most common in DH systems, [17], [18], [3] and [8] gives examples of typical faults in DHS, a summary of their findings is presented in the list below.

- Heat exchanger are subject to leakages, blockages.
- Valves can be over dimensioned, they can get stuck or start leaking
- Actuators shaft seizure, failure to open or close.
- Sensor can be subjected to bias, drift or poor location.
- Faults in the data transfer between meters and databases.
- Pipes can be clogged, leak or be subjected to faulty insulation.
- Energy meters can be misidentified in management system.
- High average return temperature.

- Faulty control system.

Some of these faults will create an abrupt change in the measured data, typically leakages and stuck valves, some will lead to abnormal quantization typically oversized valves [8], and some will lead to biased measurements or drift of measurements. Månsson et al. [19] categorize faults that can appear in costumers installation into five categories depending on where in the system they appear, in the: heat exchangers, control system and controller, actuators, control valves, and internal heating system of the customer. In the following subsections faults that can be categorized in to the the categories: heat exchangers, control system and controller, actuators, control valves are briefly being explained. Leaving faults in the internal heating system of the customer out with a comment that fault in this category can be of a huge variety of causes due to the often complex heating systems used in today's building. There is also a section briefly explaining possible errors in the meter reading process and lastly there will be a section describing possible symptoms of faults that can be detected with the help of meter readings.

2.2.1 Faults in heat exchanger

In the literature, common faults in the heat exchanger include, fouling, leakages and faulty installation [18] [19]. Fouling may occur when the water in the district heating system is hard, the fouling layer is then mostly composed of calcium. However other types of fouling occurs [18]. Leakage is another problem that may arise in heat exchangers especially in shell and tube exchangers [19]. There might also be some heat exchangers that are installed wrongly, with co-current flow instead of counter flow.

2.2.2 Control system and controller

The control system and the controller is responsible for managing the DHS and a fault in the control system might have large consequences. Faults in this category include faults in any of the temperature sensors, the controller or any connections between the sensors, the actuators and the controller [19].

There are a few temperature sensors helping the control system, for example sensors measuring the temperature in the building, the water on the secondary side of the space heating system and on the secondary side of the DHW system. Any of these sensors might suffer from drift causing the sensors to report higher or lower values then they did before the fault occurred.Drifting is the concept of a meter reading gradually increasing its meter error and is something all sensors experience to some extent. The sensors might also be wrongly installed from the beginning or completely lost it's connection to the controller.

2.2.3 Actuators

The actuators are responsible of regulating the control valves such that the right flow pass through the valves. When choosing actuators it is important that they are suitable for their task. Things to consider are stroke time and driving force [20].If the driving force is to small, there might be a problem opening or closing the control valve, as desired. And if the stroke time is to long, then the DHW system might be unresponsive, however actuators controlling the space heating does not need to have short stroke time [20]. Other faults in actuators includes connection problem or complete failure of the actuator, both resulting in constant or nearly constant flows through the heat exchanger controlled by the actuator.

2.2.4 Control valves

When sizing the control valves it's important to not over or undersize the valve. Oversizing will result in an inability to control small flows and a undersizing will restrict the flow too much. A solution can be to use two valves in parallel [21]. Valves are as most components, subject to wear and tear e.g. due to corrosion, erosion and water hammering they may start to leak or seize.

2.2.5 Errors in meter readings

Even if wrong measurements from the heat meter might be a problem for both the consumer and the district heating utility resulting in over or under billing, they are not a fault in the heat delivering system. These errors might include wrongfully placed temperature sensors, drifting sensors or connection loss between the integration unit and the sensors. Flow meter might also have been incorrectly calibrated or blockage might cause constant or no flow readings. There might also be a problem with connections between the integration unit and the utility company.

2.2.6 Symptoms of faults

The faults and errors described in previous sections can result in different symptoms in the meter data. Sandin et al. [8] defines the following symptoms:

- Drifting of measured quantities
- Constant values of measured quantities
- Anomalous values of quantities

The detection of any of these symptoms might indicate that there is a fault in the substation. Drifting is the concept of a meter reading gradually changing. This is something that most sensors experience, at least to some extent. There is also a possibility that changed user patterns or fouling on a heat exchanger can result in drifting like symptoms. It is hard to distinguish between drifting and changed user patterns or fouling. Sandin et al [8] suggests that, constant or nearly constant values of a quantity might suggest that for example a valve is stuck, over or undersized or that there is some problem with the actuator. Anomalous values of a quantity can be a symptom of for example a meter error, a stuck valve that for some time increases or reduces the volume flow significantly or an error in the control system. It might also just be the result of an unusual hot or cold day.

2.3 Decision trees

Decision trees are methods in statistical learning that can be used for both regression and classification. For a thorough introduction to tree based methods for statistical learning the readers are referred to [22] from which all information in this section is taken, unless otherwise stated. A brief explanation of tree based methods for regression will be given in this section. Decision trees are a segmentation of the predictors to different simple regions. The rules that are used for segmenting the predictor space can be described in tree like structure, hence the name. An example of how a predictor space can look after it has been segmented into three region, by two cuts is shown in fig(2.2).

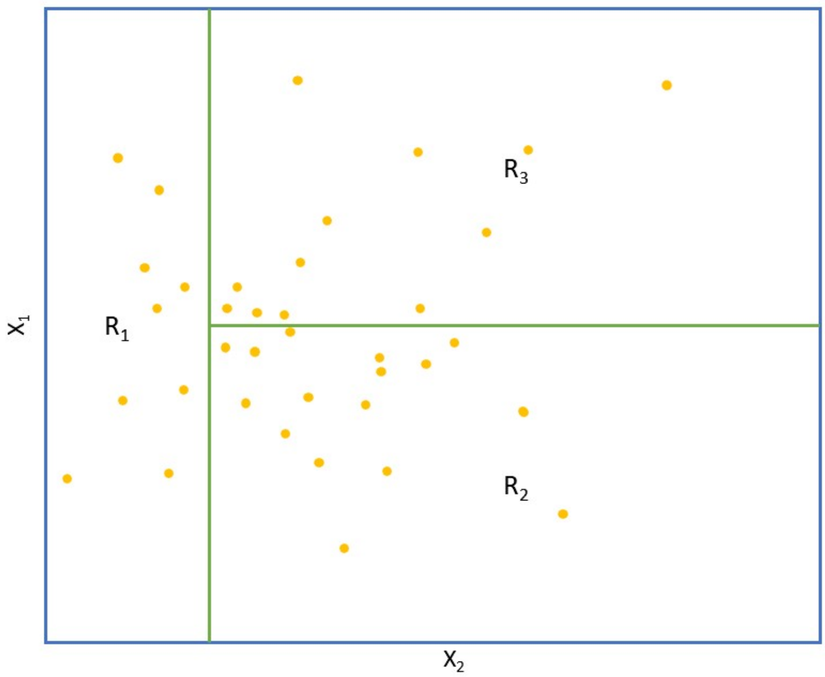

Figure 2.2: An example of what a 2D predictor space can look like after two cuts

Building a decision tree constructs of two basic steps. Given a predictor space $X_1, X_2, ...X_p$, the first step is to divide the predictor space into J non-overlapping and distinct regions, denoted $R_1, R_2, ..., R_J$. The second step is that for every observation in R_j we make the same prediction, which is usually the mean of all observation in R_j. The construction of the sub spaces $R_1, R_2, ..., R_J$ can be performed in many ways and take many forms, for simplicity we will only consider multidimensional rectangles. Since the goal is to find a good prediction as possible one wants to minimise the residual sum of squares RSS, given by

$$RSS = \sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 \tag{2.2}$$

where $\hat{y}_{R_j}$ is the mean response of all training observation in subspace R_j. One way to construct these regions would be to consider every possible combination for dividing the feature space into J regions. It would however be computational infeasible. Instead a method known as recursive binary splitting is used, which is a greedy top-down approach. Top down because it starts with one split and work it way down, greedy because in each split it tries to minimize RSS. That is find j and s such that it minimizes

$$RSS = \sum_{i:x_i \in R_1(j,s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:x_i \in R_2(j,s)} (y_i - \hat{y}_{R_2})^2 \tag{2.3}$$

where

$$R_1(j,s) = x|X_j < s \text{ and } R_2(j,s) = x|X_j \geq s \tag{2.4}$$

Repeat until J regions are available. A large tree with many nodes will fit the training data better, in fact with a large enough tree, one can perfectly predict the training data. However due to the variance bias trade-off, one usually builds a large tree and use cost complexity pruning to prune back the tree and create sub trees, details of how this is done can be found in [22], by applying cross-validation find the best sub tree.

2.3.1 Using regression trees for predicting Substation behavior

By using: heat energy, volume flow, supply and return temperature as response, weather parameters, calendar parameters as features or predictors a regression tree can be grown for each response. By feeding new predictor into this regression tree a model capable of predicting the behavior of a district heating substation can be created and by aggregating the response from several grown trees, the result can be more precise. Examples of weather parameters are, outdoor temperature and precipitation, calendar parameters includes among others, what week day it is and what time it is. This forms the basis of the prediction software EnergyPredict, used in this thesis.

2.4 Anomaly detection

In this section a brief introduction to the theory of anomaly detection is presented. Anomalies are patterns that deviate from what is considered normal behavior. The method of finding such anomalies is referred to anomaly detection. Anomalies and outliers are terms sometimes used interchangeable, hence the term outlier detection is also used for the same process. Other terms used are discordant observations, exceptions, aberrations [4]. It is worth noting that the presence of anomalies does not automatically mean there is an error or fault present nor, does the absence of anomalies indicates that the system is working as intended.

2.4.1 Different types of anomalies

In this section three types of anomalies are presented as in [4] and it's extended version [23].

- **Point Anomalies**: This is where one single data point is behaving abnormal. Usual examples are one data point that is much greater or smaller than most of the data points in the data set.

- **Contextual Anomalies**: Anomalies that due to their context can be described as anomalous are called contextual anomalies and are defined by the following two properties, contextual and behavioural. Contextual properties are properties that describe the context or neighborhood of the data point. In time series the time is the contextual property that describe the neighbouring data points. In spatial data that property might be longitude and latitude. Behavioral properties are for example the temperature in a data set that describes temperature around a city. Basically a contextual anomaly is a data point that based on only its behavioural property is not abnormal, but in the given context it can be considered as an anomaly. If you record the temperatures for a year, a sudden spike in temperature in the winter might be a contextual anomaly even if that spike temperature is not abnormal based on all recorded temperatures. It's abnormal because it occurred during a winter month.

- **Collective Anomalies**: Are a group of data points that based on individual data points is not necessarily anomalous by themselves but as a group they are anomalous. A requirement for collective anomalies are that there is some relationship between the data points, e.g. sequential or spatial. Since this thesis includes time series data only, the only form of collective anomalies encountered will be sequential anomalies, henceforth the term sequential anomalies will be used.

2.4.2 The basics of anomaly detection

An anomaly is at its core a deviation from what is normal. A basic method of anomaly detection is therefor defining what is normal and everything outside the region decided to be normal is then considered as an anomaly, however this simple method comes with several great challenges. Such as defining what is normal, the line between what is abnormal and normal is often fuzzy. The normal behavior of a process might also keep changing over time, e.g. user patterns for a credit card might change as the user earns more. Also what is normal and anomalous changes from field of application, hence migrating a method developed for one field to another is often hard. Noisy data can in some instances be quite similar to actual anomalies, and hence be hard to distinguish [23].

There are several ways to categorize and group anomaly detection methods presented in the literature. Chandola et al. [4] stated that most of anomaly detection techniques can be categorized into one of the following main technique groups: classification based, nearest neighbor based, clustering based, statistical techniques or into techniques that sprung from information theory and spectral theory. Another given way of categorize different anomaly detection methods are by the type of data used in the training period. Different anomaly detection techniques can train in three different ways: unsupervised, supervised and semi-supervised depending on the type of data that are available during the training of the anomaly detection model. There are two types of data, labeled and unlabeled. Labeled data is data that are classified as anomalous or normal, for example by use of a fault log or experts domain knowledge. Unlabeled data is data that does not have this kind of labeling. Generally labeled data is harder to obtain than unlabeled data. Unsupervised anomaly detection uses unlabeled data in its training period. Supervised anomaly detection instead uses labeled data in the training period. Lastly semi-supervised anomaly detection are methods that only uses one type of data in the training period, for example only normal behaving data points[23]. A third way of grouping different methods or algorithms is by how they approach the problem. Emmott et al. [24] which have given a benchmark of eighth algorithms, define four approaches to which they group the tested algorithms: density-based approaches, model-based approaches, nearest neighbors-based approaches, projection-based approaches a short explanation of these approaches will be given below.

Density-based approaches are a straightforward approach where one defines a probability distribution for the data points, and the more unlikely the data point is in the given density the more likely it is an outlier. Gaussian Mixture Models is one density-based method.

Model based approaches builds on the assumption that if the majority of the data points are normal, it is possible to construct a model for the data and from that model construct a decision boundary. The One-Class SVM algorithm is one model based approach.

Nearest neighbors-based approaches or distance-based approaches, uses the information from the surrounding neighbors to calculate an anomaly score. Local outlier Factor algorithm is one of the most well known algorithms in this class.

Lastly projection-based approaches are presented, here Emmott et al. [24] grouped two quite different methods, IsolationForest and Lightweight Online Detector Of Anomalies (LODA). IsolationForest builds on the principle that a data point that can easily be isolated by random axis parallel splits are anomalous, the fewer splits the more anomalous. LODA on the other hand works by creating several weak anomaly detectors from several random projections of the data. By computing the density estimation histogram for every projection and by taking the mean negative log-likelihood for each histogram, LODA produces a outlier score.

2.4.3 IsolationForrest

When introduced in 2008 IsolationForest was a whole new way of approaching outlier detection[25]. Instead of formulating what is normal it isolates anomalies. To do so it uses the key attributes of anomalies, that they are few and that they are different from normal data points. IsolationForest works by creating a chosen number of so called IsolationTrees for each data set, and by finding the average path lengths it takes to isolate a data points and ranking shorter path lengths as more anomalous. IsolationTrees are true binary trees and are generated as follows. Given that T are nodes in the tree and $X = X_1, ...X_n$ is the data. Randomly choose an attribute and a split value p and then perform the test $q < p$ where q is the attributes value, dividing the data points into T_l and T_r. Continue until one of the three conditions are met: the tree reaches a predefined height limit, there is only one data point left in the data set $|X| = 1$ or all data in X have the same values. Here height means longest path in the tree. After several trees are grown the average path length of all trees are calculated. In isolation forest path length $h(x)$ is defined as the number of edges there is from the root of the tree to the node in question. The path length is useful when comparing data points in the same data set but since the maximum height of an isolation tree grows proportional with n and the average height grows with $\log n$, it is hard to normalize the path length to give an anomaly score for comparing different data sets. However by borrowing their analysis from Binary Searching trees the authors of IsolationForest define the average of $h(x)$ as

$$c(n) = 2H(n-1) - (2(n-1)/n) \tag{2.5}$$

where $H(i)$ is the harmonic number and can be estiamted by $\log_e(i)+0.5772156649$. To normalize $h(x)$ the authors used $c(n)$. Given that $E(h(x))$ is defined as the average of several isolation trees an anomaly score can be defined as

$$s(x,n) = 2^{\frac{-2E(h(x)}{c(n)}} \tag{2.6}$$

IsolationForest has a linear time complexity and has low memory requirements [25].

2.5 Feature extraction for anomaly detection

An inherent property of time series data is its relation to other data points in the data set. This property makes it possible to extract, features that might be used in detecting anomalies. Here two features are explained, of which on is trivial.

- **Variance of sub-sequence**. A time series that has a larger than normal variance in a sub sequence might indicate that there is something unusual going on in that sub-sequence. In measured data it might indicate a process that is not in control. In predicted data a large variance might indicate that the prediction is wrong. Variance for a sub sequence given the time series X

  $$X = x_t | t = 1, .., n \tag{2.7}$$

  where t is the time index and n the total number of observation in the series. A consecutive sub-sequence, of window size $w < n$ z_p can be given as

  $$z_p = X_p, ...X_{p+w-1} \tag{2.8}$$

  for $1 \leq p \leq n - w + 1$, the variance of a sub-sequence is then given by

  $$z = \frac{1}{w-1} \sum_{i=1}^{w} (z_i - \bar{z})^2 \tag{2.9}$$

- **Raw data**. The other feature is the trivial raw data which might be useful for detecting global anomalies or extreme values.

Apart from the two above features, residual, between a predicted and an observed value, can be used in identifying anomalies. Residuals are defined in as

$$r_i = y_i - \hat{y}_i \tag{2.10}$$

where y_i is the observed value and $\hat{y}_i$ is the prediction.

2.6 Shannon entropy

Shannon entropy is within information theory a way to measure the amount of uncertainty contained in variable [26]. If a variable has n possible outcomes, and $P(X_i)$ is the probability of outcome i then the entropy $H(X)$ is given by,

$$H(x) = -\sum_{i=1}^{n} P(x_i) \log_b P(x_i) \tag{2.11}$$

where common choices for b are 2, e, 10. A simple coin toss example gives a good intuition of entropy. Assume a fair coin is tossed. The probability for head or tail is 0.5 for both, this gives the entropy.

$$H(x) = -(0.5 \log_2(0.5) + 0.5 \log_2(0.5) = 1.0 \tag{2.12}$$

and for a coin with two heads, the entropy will become 0 since the $\log(1) = 0$. The higher the entropy the more uncertainty there is in a variable.

3
Methods

In this chapter a description of the proposed method is outlined. The chapter starts with an general description of the methods proposed for finding symptoms of faults after which each process is described in more detail. From studying the literature in the subject, three symptoms of faults have been identified. These are as mentioned in section 2.2.6: drifting, abnormal quantization and anomalous values. The project was therefore divided into three main blocks aimed towards finding each of the mentioned symptom. Two of the proposed methods, those suggested for drift detection and anomaly detection, are based on comparing the difference between predictions made by EnergyPredict and the measured data from the observed substation. The proposed method for identifying abnormal quantization only uses measured data. In fig(3.1) an outline of the work done in this thesis can be found with the three main blocks displayed. In the following sections the data used is first presented, then the proposed method for abnormal quantization detection is presented. Later in the chapter the suggested method for drift detection is presented after which the suggested method for anomaly detection followed by the method of validation are presented.

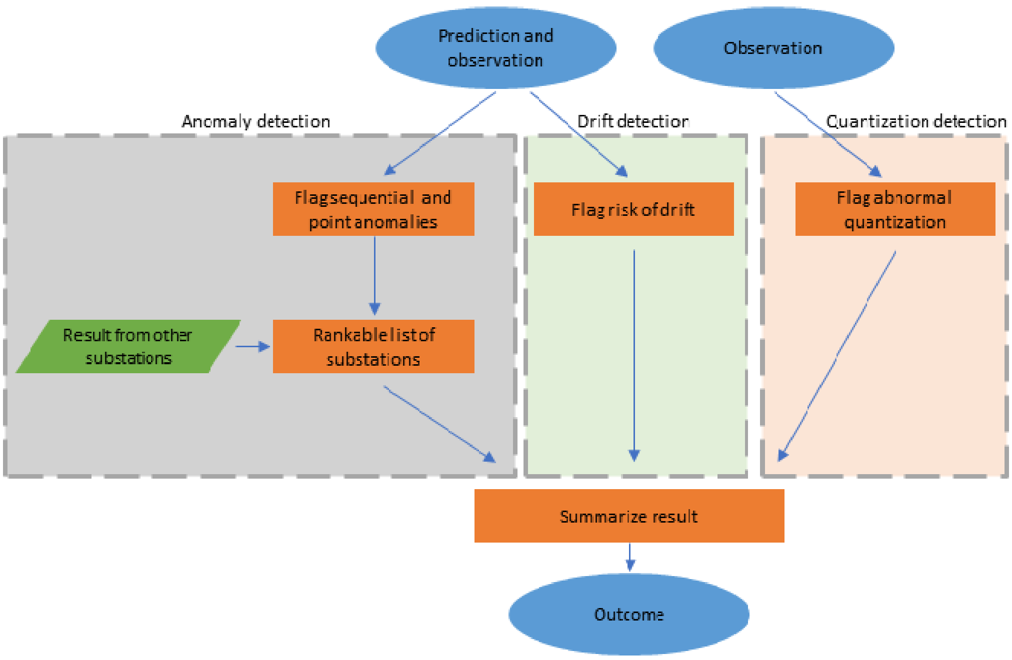

Figure 3.1: A flowchart of this thesis showing the basics of the three main blocks.

3.1 Used data set

For this investigation a data set composed of data from 672 substations from a district heating network in southern Sweden was used. The 672 substations where composed of Multi-dwellings buildings, commercial buildings, Industrial demands, Health and social service buildings and Public administrations buildings. However no single family homes where investigated.

The prediction model obtained from the prediction software EnergyPredict was trained on 355 days from the beginning of January to the end of the year, and the predicted period was 79 succeeding days from first of January. Fig(3.2) shows an overview of how the predictions were created for each substation. This data was composed of hourly predictions of heat energy, volume flow, supply and return temperatures, for each individual DHS in the data set here after this data will be referred to as the predictions. For the same period hourly measured data of heat energy, volume flow, supply and return temperatures was obtained from the meters in the DHS's. This data will hereafter be referred to as the measured data. As mentioned earlier both the prediction and the measured data are hourly, giving approximately 1 896 observations for each of the 672 substations in the data set. All observations (hours) that lacked any of the predicted or observed variables in an observation were removed totally from the data set, that is both prediction and measured data are completely removed for that observation, giving a slightly smaller data set. There was a choice made to only considering the heating season, were space heating is in use. This choice was made due to the fact that in the change between heating and non heating period, fluctuations in energy use can be quite significant from one day to an other.

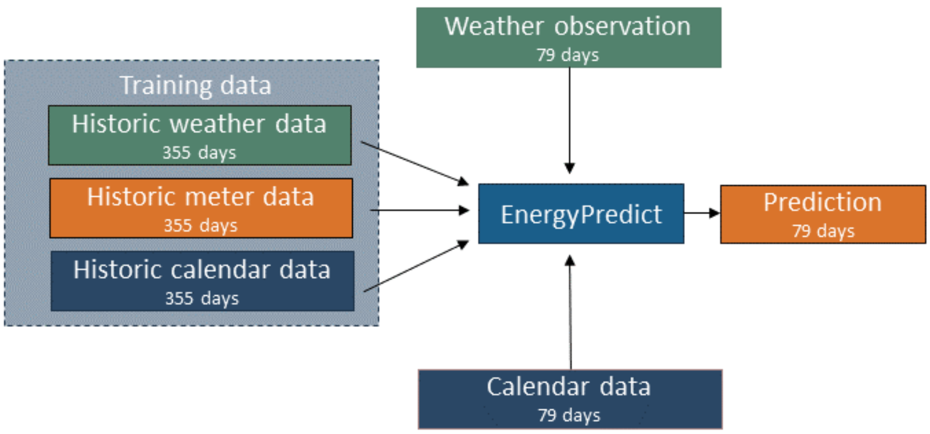

Figure 3.2: A overview of how EnergyPredict created the predictions used in each substation.

3.2 Detecting abnormal quantization

The method in this sections is based on a method presented by Sandin et al. [27]. The idea is that substations with over/undersized valves or poorly working actuators might have a reduced Shanon's entropy in their measured data. By eq(2.11) the Shannon entropy can be calculated for each quantity, were the possible outcomes n is equal to all the values the quantities has taken during the period. And the probability $P(x_i)$ is based on the times each value is observed during the considered period. Substations with low entropies should be investigated. The features considered are heat energy and volume flow, the choice to only consider these two features are based on the fact that a stuck, over/undersized valve or faulty actuator will mainly affect heat energy and volume flow.

It was observed that substations with low entropy's for heat energy and volume flow are usually substations with low flow and energy use in general. Which is naturally, for example substations with many recorded zeros, will natural report lower entropy's. A proposed method is therefore to set two thresholds a and b. Such that if a substations is in the a percent of substations with the lowest entropy, but not b percent with the lowest variance, one would flag for further investigation. The reasoning behind this is that, if one only ranked on entropy substations with lots of recorded zeros or with generally low flow or measured heat energy will be favoured. Setting a constraint so that one only consider substation with the lowest a percent of the substations ranked on entropy and the highest b percent in variance will overcome this problem. The thresholds suggested are a=5% and b=80%.

3.3 Drift detection

The proposed method for drift detection is quite basic. Given a prediction model trained on historical data, if one compares the mean of the prediction with the measured data, the difference should be relatively small. If not, this could be a sign that something has happened to the substation such as fouling in the heat exchanger, drift in any of the sensors or a change in the user patterns.

The method proposed is implemented by simply giving a score 1-5 based on the the absolute value of the relative error $|\delta x|$ between the mean of the measured and predicted variable. The thresholds used are given below.

- $|\delta x| < 0.1$ gives score 1
- $0.1 \leq |\delta x| < 0.2$ gives score 2
- $0.2 \leq |\delta x| < 0.4$ gives score 3
- $0.4 \leq |\delta x| < 0.6$ gives score 4
- $|\delta x| \geq 0.6$ gives score 5

3.4 Detecting anomalies

In this section and the following subsection a proposed method for detecting sequence and point anomalies is being presented. The idea is that by observing the residuals between a good prediction, here provided by EnergyPredict, and measured values one can identify abnormal residual.

Instead of observing where residuals are above or below a given threshold, which might be hard to define for all substations, the idea is to observe where the residuals are abnormal. The reasoning behind this is that a prediction that does have some accuracy issues, for example a prediction that has a mean shift compared to the measured data could still be used to identify abnormal data points, an example of a mean shift can be seen in fig(3.3). Where by observing the residuals between the prediction and the measured data, the used method can identify a period around 4th of March which seems anomalous. Even though there seems to be a slight shift between measured values and the predictions from the beginning of the showed period .

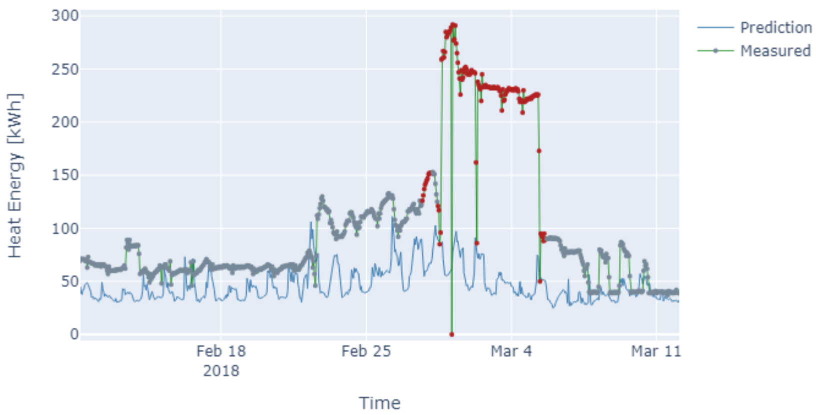

Figure 3.3: By identifying abnormal residuals rather than using a threshold on the residuals a prediction that is not perfect can still be used for identifying anomalies

The method for identifying sequential and point anomalies are similar and will here be presented together. Where they differ it will be stated by the use of the terms sequential anomalies and point anomalies, otherwise when the term anomalies are used we mean both sequential and point anomalies. The outline for this method is, cleaning the prediction models from now on noted as prediction cleaning, find most anomalous data points/sequential data points in each substation from now on noted as anomaly detection, compare results between substation and rank according to largest errors from now on noted as ranking. The outline for this method can be observed in fig(3.4), as observed in the flowchart the procedure for handling point and sequential anomalies only differ in the way that for point anomalies there is no aggregation of data points .

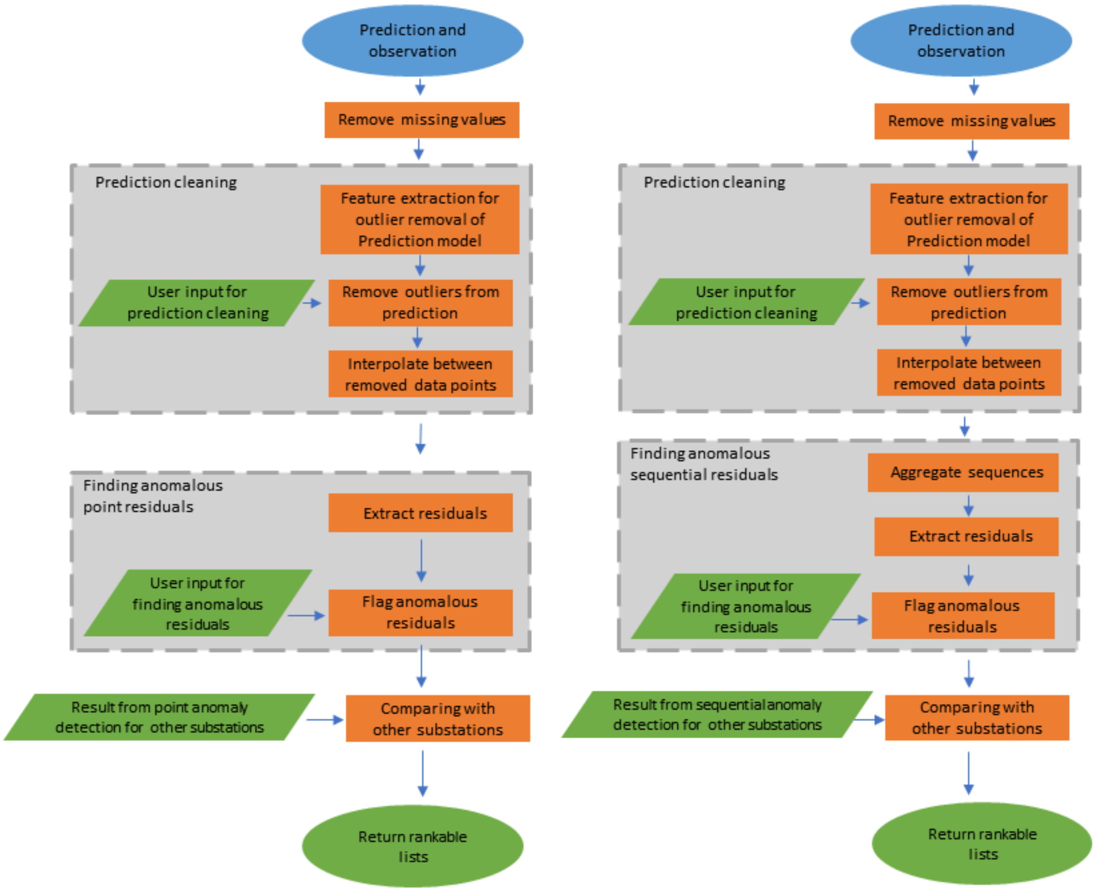

Figure 3.4: A flow chart describing the outline of the proposed anomaly detection method, the procedure are quite similar for both handling point and sequential anomalies.

The absolutely first step in the proposed method is to remove any non complete data entries from the data set. As an example if there is a missing observation of volume flow for a particular hour for the specific data set, all observations and predictions for that substation is removed.

3.4.1 Remove suspicious data points in the prediction

The prediction method that Utilifeed uses to model the relation between measured data in the substation and external predictors is a tree based method. The use case intended when designed was primary prediction, and not fault detection. In the data set used in this study, oddities in the form of a zigzag fashion appear in the prediction data for some substations, see example fig(3.5). These zigzag patterns and sudden drops and spikes in the graph, could be an indication of that the tree based method has found correlations between oddities in the metering data and predictors used for training the model. It is likely that improving the cleaning process for the metering data used in training the prediction model and/or redefining the set of predictors would eliminate this problem. Such measures are being worked on but they are out of the scope for this master thesis project. The proposed way to reduce the effects of these oddities in this project is therefore to simply remove the data points where these oddities appear and interpolate between the two non removed closest data points. The motivation behind this strategy is that we want as few false positives as possible and are therefore prepared to sacrifice some data points that are removed incorrectly to lower the risk of many false positives.

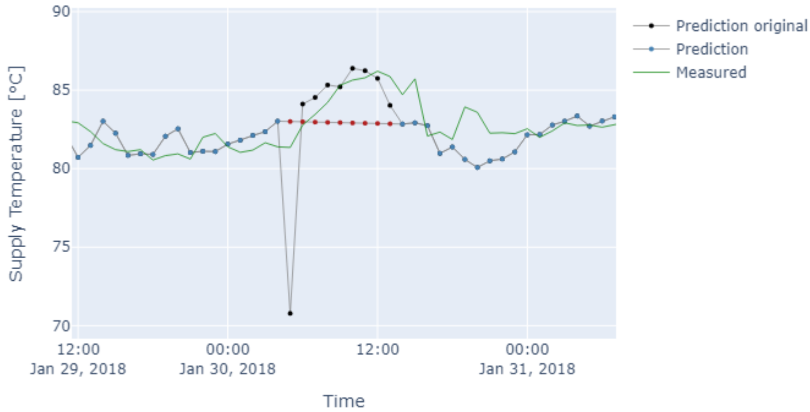

Figure 3.5: An example of unwanted behavior in the prediction model, Here blue dots represent prediction model, red dots are the linear interpolation, and black dots are the dots we have removed and replaced with linear interpolation.

The process of removing abnormal data points starts with a feature extraction. These features are then used in IsolationForest to identify the most abnormal data points. The prediction data points that are flagged as abnormal are removed and a interpolation is performed on the removed data points. The decision to use an anomaly detection algorithm to identify anomalous prediction data points rather than using global thresholds, was made because of the vast variations of substations and data patterns in the observed data set.

3.4.1.1 Feature extraction for prediction cleaning

The features extracted are variances of overlapping segments. Window size is chosen to be $n = 6$ data points and the step size is chosen to be $m = 3$. The algorithm for defining variance of segment is to be found in algorithm(1) in appendix A. The choices of n and m are arbitrary, but after trails they seemed reasonable, however no formal investigation of the best choice of n and m where done due to the lack of labeled data.

It should be noted that a large increase in a quantity also gives a large variance. However the argument can be made that interpolating those data points, have likely little effect since when interpolating a large increase or decrease in a variable the change from the original data is not relatively large, for the most part fig(3.6) shows an example of this.

3.4.1.2 Finding abnormally high variances of segments and interpolate

The next step is to find segments with abnormally high variances. To do this the decision was made to look at three of the four variables: heat energy, volume flow,

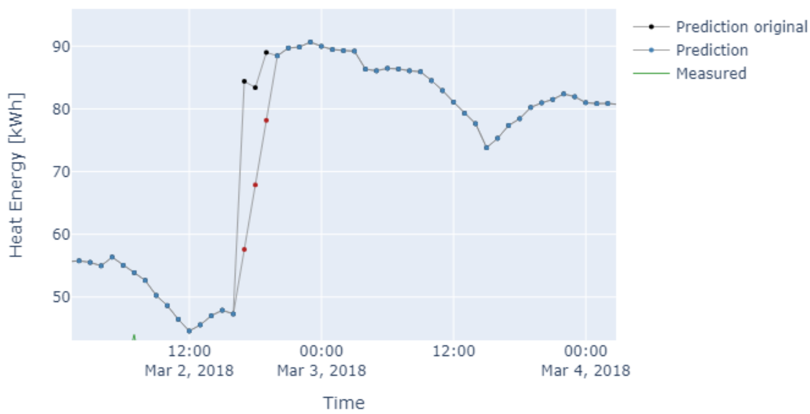

Figure 3.6: Linear interpolation where abnormally high variance of segment might as in this case interpolate between data points when a large increas in a variable is detected

supply and return temperature, simultaneously. The variables chosen are the ones with the highest Kurtosis. The choice of not using all quantities is based on the fact that heat energy is a linear combination of the other variables so adding that will not increase the information, rather would the curse of dimensionality make it harder for IsolationForest. The choice of using the quantity with the highest Kurtosis is based on a statement from the original paper of IsolationForest [25]. Saying that "Kurtosis is sensitive to the presence of anomalies and hence it is a good attribute selector for anomaly detection" [25]. After the three variables were chosen an implementation of IsolationForest from the Python library scikit-learn was used to perform the outlier detection on the three series of variance of segment (VOS) [28]. All parameters in IsolationForest where chosen to their default value, except the threshold of 0.6 in the anomaly score which is slightly more conservative than the default threshold, the anomaly score is calculated with in IsolationForest with the use of eq(2.6) and eq(2.5). The threshold was determined to be a reasonable threshold after observing a few substations at different threshold. However no formal hyper parameter study was performed due to the lack of labeled data. IsolationForest returns a list of indexes for where the variance of segment is abnormally high, based on the threshold of the anomaly score chosen. These data points where removed and a linear interpolation between neighbouring data points where performed.

3.4.2 Finding anomalous residual

The main idea behind the proposed method is to find the residuals that are anomalous, and these are found using IsolationForest. The residuals are found for a series as described in eq(2.10) were i are indexes of the series and $\hat{y}_i$ is the predicted value and y_i is the observed value. The residuals for the four quantities heat energy, volume flow, supply and return temperatures where calculated.

For identifying the anomalous residuals, scikit-learns implementations of IsolationForest were once again used, the thresholds and settings chosen were the same as in the prediction cleaning process, with the same motivation. This time only two features where used. Heat energy is as mentioned before a linear combination of the three other quantities measured, meaning using all four would not give any more information. After the trials with three parameters the choice was made that two features should be used. The two features used were residuals of volume flow and heat energy, with the motivation that a total energy increase/decrease would be able to be detected with the residuals of the heat energy, and a change in the temperature difference between supply and return will result in a change in the residuals of volume flow.

3.4.3 Finding anomalous sequences

To find the most abnormal sequences the approach is similar. The sequences are extracted in form of a jumping window with overlap, much like the one in algorithm(1), here window size $n = 12$ was chosen and a step size jump of $m = n/2$ was chosen. In the windows an aggregation of the data in each window is done by taking the mean of the features in each window. This gives a time series of aggregated data points. IsolationForest is then applied on the two aggregate series of residuals of heat energy and residuals of volume flow and flags the most abnormal aggregated data points. IsolationForest uses the same settings as used previously. The reason for using overlapping windows is that else non overlapping window might cut a sequence anomaly in half, making it less anomalous when aggregated and therefore not detectable. The choice of using a step size of half the window size and not less was made because the more windows there are the more similar they get to each other, making detecting outliers harder. The choice of a correct window size is important and a too large window might miss shorter anomalous sequences, but a too small will be heavily influenced by any point anomalies in the window. The choice of $n = 12$ was made after several trails, however due to lack of labeled data, a hyper parameter study was not performed.

3.4.4 Ranking

The overall goal of the developed tool is to identify substations that are faulty or have experienced faults. However since IsolationForest flags the most anomalous data points, a substations with really small residuals everywhere will get flags on data points that are likely perfectly normal, so called false positives. In fig(3.7) we can see a substations that has a precise prediction for both volume and energy over the whole plotted period, despite this IsolationForest has flagged the red dots in the graph as anomalous.

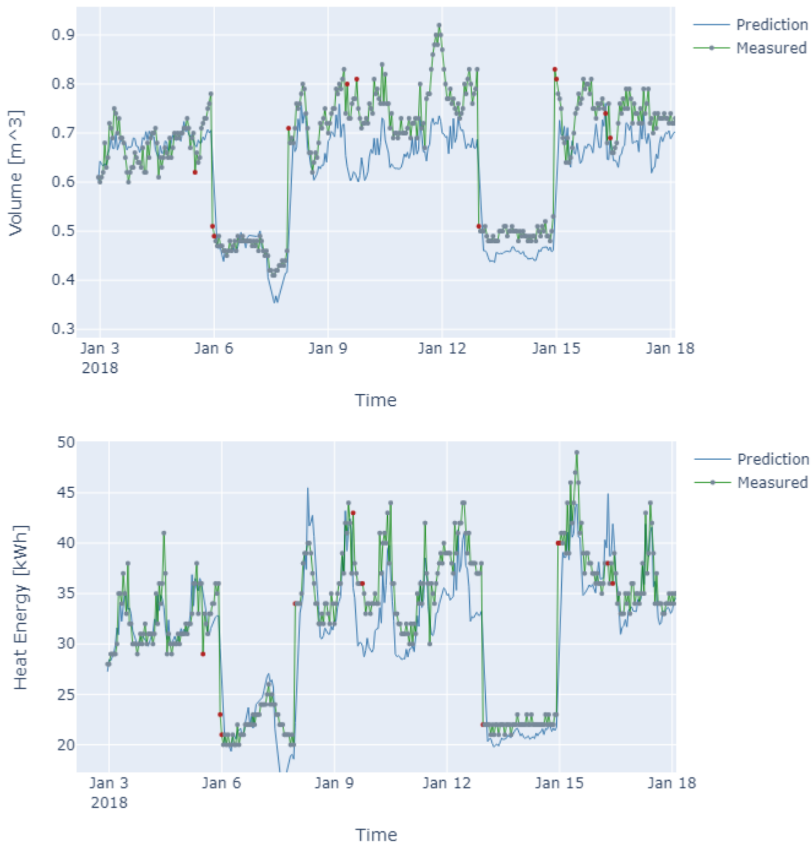

Figure 3.7: A substation for which the prediction is accurate, but where IsolationForest still finds anomalous data points (red).

Since IsolationForest is applied to the residuals for each substation it might identify anomalous residuals in substations even if there is nothing that indicates that a fault or an error has occurred there. To overcome this problem a solution is to rank the substation in a list where greater risk of anomalous data points would rank higher. Two methods for ranking are presented below and each of these can be introduced to either point or sequence anomalies and both can be applied to each variable used in IsolationForest.

The first ranking is based on the sum of the n largest residuals flagged as anomalous by IsolationForest, normalized by the average value of the measured quantity during the observed period. The idea is that for every substation find the Ranking value $RV_{avgNorm}$ as described by the equation below

$$RV_{avgNorm} = \frac{\sum_{i=1}^{n} \max_i Z}{\bar{X}} \tag{3.1}$$

Where $\max_i Z$ is the i^{th} largest residual flagged by IsolationForest as anomalous quantity in question. $\bar{X}$ is the average of the quantity X. By arranging the $RV_{avgNorm}$ number for each substation in the investigated data set the theory is that substations with larger risk of faults or errors will rank higher in the list. In a similar manner a $RV_{resNorm}$ can be given for each substation in a network, where the $RV_{resNorm}$ is defined as below.

$$RV_{resNorm} = \frac{\sum_{i=1}^{n} \max_i Z}{\sum |y|} \tag{3.2}$$

Where $\max_i Z$ is defines as previously mentioned, and $\sum |y|$ is the sum of all of the absolute values of the residuals.

As mentioned this ranking can be applied to either aggregated sequences or point anomalies and also to different quantities. In this thesis four rankings are evaluated, all with $n = 5$ and all applied on the heat energy variable the rankings are the following:

A. Ranking based on $RV_{resNorm}$ evaluated for aggregated sequence for the quantity heat Energy
B. Ranking based on $RV_{resNorm}$ evaluated for point anomalies for the quantity heat Energy
C. Ranking based on $RV_{avgNorm}$ evaluated for aggregated sequence for the quantity heat Energy
D. Ranking based on $RV_{avgNorm}$ evaluated for point anomalies for the quantity heat Energy

3.5 Validation

Within the field of fault and anomaly detection, there is often a lack of labeled data. This leads to problems when evaluating anomaly and fault detection methods. Some different strategies for overcoming this problem is presented in the literature. A common way to test anomaly detection algorithm is to test them on data sets specifically created for anomaly training and testing anomaly detection algorithms. However since many anomaly detection algorithms are specifically developed for certain processes and use cases, these standard data sets might not be usable. An other approach is to artificially induce errors in data sets. This requires that one knows that the data set which is being induced by the artificial errors is a data set without any anomalies, otherwise there is no way to evaluate the result. A third way is to use domain experts and let them evaluate the method. In the following sections the validation process for the three different anomaly detection methods, drift detection, abnormal quantization detection and anomaly detection are presented.

3.5.1 Validating abnormal quantization detection method

The method for evaluating the quantization detection method is quite straight forward. Two quantities are evaluated volume flow and heat energy. If the method flagged a substation as in risk of abnormal quantization an investigation of that substation was performed by observing the graphs of the corresponding quantity. There after a classification on whether the substation might suffer from an abnormal quantization was done. As mentioned before no labeled data was available, so the investigation was made by the author and might be biased.

3.5.2 Validating drift detection

The problem with detecting drift is that one need a label to tell if a detected change is actually a drift. The proposed method to detect drift only compares the mean of the prediction for one period with the observed value for that period. There are many reasons why these might be different. To let an expert group detect drift is almost impossible, an increase or decrease in energy use over a long period might just be an mild change in the weather. Inducing artificial drift is not possible due to the nature of the proposed method, this will be further elaborated in the discussion.

3.5.3 Anomaly detection

Since the anomaly detection method proposed is specific to detection anomalies in district heating substations and no data set with anomaly labels are available there is no labels to compare our findings to. Since the method is applied on the whole set of 672 substations and results in a ranked list, a method using artificially induced errors would require a lot of manual work to identify a large number of well performing substations to which artificially error could be induced. The method of using artificially induced errors also suffers from a bias problem. There is a risk that the faults introduced are faults that method have easier to detect, and that one miss inducing faults that the method has more difficulties to detect.

For the above mentioned reasons the decision to use an expert group to evaluate the anomaly detection methods proposed in this thesis where made. The expert group were presented with the 15 highest ranking substations for each of the four rankings explained in the previous section. Overlapping between the four rankings made the total high ranking substation evaluated to 30 unique substations. A reference group for comparison, was formed by 20 unique substations. These 20 substations were chosen among the 70% lowest ranking substations for each ranking, that is 5 from each evaluated ranking.

The expert group were given a set of questions for each of the 50 evaluated substations. For each substation evaluated, the expert group got a graph of the heat energy and volume flow consumption reported for each hour during the examined period. For each substation the expert group where asked two multiple choice questions.

Question 1: Given your experience or possibly a fault log, does this substation during the given period experience anomalies in the data? With anomaly we here mean either a physical error or a meter error.

1. The data shows significant anomalies that is most likely caused by a fault (or a planned major interference in the system). If I would get alerted on this deviation we would send out a technician or inform the customers immediately.
2. Anomalies are observed that could be caused by a fault or a meter error but I'm not certain, and/or the consequence are small. I would like to get informed and keep the substation under observation and check it/inform the owner when there is a good opportunity.
3. There are minor oddities in the data but they are likely not caused by faults or meter errors. If I would get alerted on this deviation no action would be taken. Might still be good to keep a log in case the situation changes.

4. No anomalies can be observed, everything seems to be working as it should.

Question 2: Given your expertise do you think the prediction model performs reasonably well during the given period?
1. Yes
2. No

If the expert group answered 1 or 2 on the first question the substation was seen as having anomalous data for the investigated period. If the answer was 3 or 4 the substation was seen as being without anomalous data for the given period.

4 Results

The result from the anomaly detection method and the abnormal quantizitation detection method is presented in this chapter. However since no satisfactory method for validating the result from the drift detection method where found no result will be given from that method, a shorter discussion regarding the drift detection method is however given in the next chapter. First the result of the abnormal quantization detection method is presented, thereafter the result from the anomaly detection method is presented. Most of the graphs in this chapter does not show the whole 79 days investigated, but a short period. This is done for readability.

4.1 Results from the abnormal quantization detection method

The proposed method with the given thresholds identified 13 substations that according to the used thresholds of were in risk of quantization error when observing the quantity heat energy. In 11 of them the reason was likely a low meter resolution, the most common resolution in the investigated network were 1 kWh. However these 11 substations all had in common a resolution of 10 kWh see fig(4.1). Two substations however showed interesting behaviour one of which is shown in fig(4.2), the other one showed similar behaviour. When instead investigating the volume flow, all of the flagged substation had a low meter resolution, which likely was triggering the flagging.

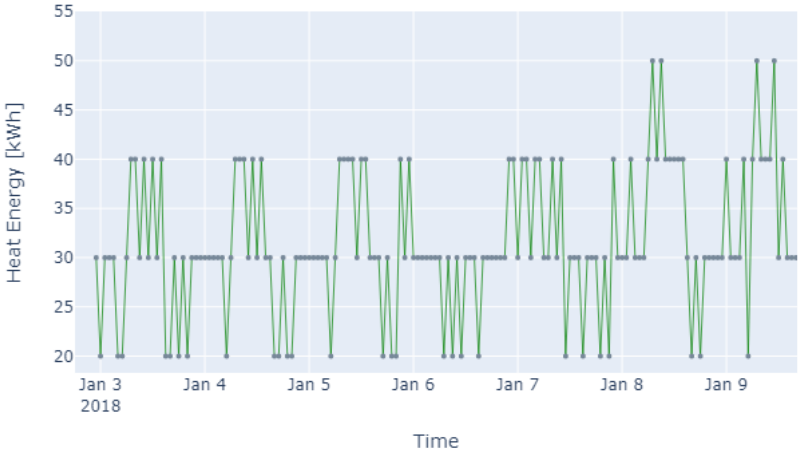

Figure 4.1: Example of a substation classified with risk of quantization error, that has a low meter resolution. Similar patterns was present for the whole investigated period.

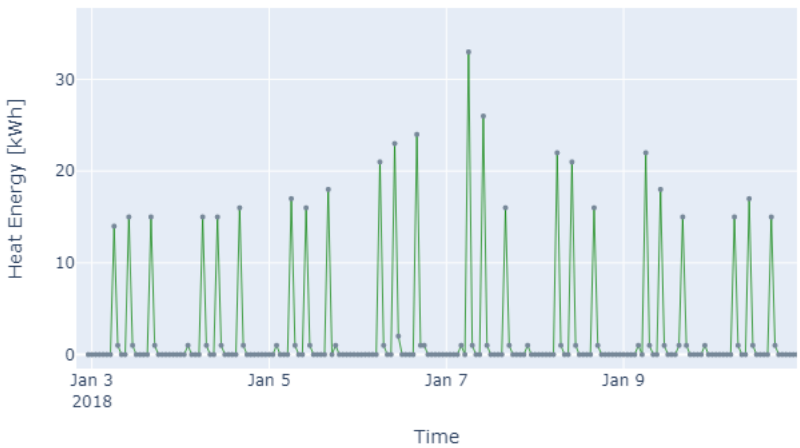

Figure 4.2: Interesting substation behaviour detected with quantization error detection. Similar behaviour was present for the whole investigated period.

4.2 Anomaly detection

In this section the result from the anomaly detection method will be presented. First comes a presentation of the result of the prediction cleaning process, after which a presentation of the result from the validation of the proposed four rankings is presented.

4.2.1 Prediction cleaning

In the cleaning of the anomalous prediction data points an average of approximately 85 data points were removed for each substation. The goal was to remove data points that were at risk of interfering with the anomaly detection strategies later. A comparison between when the prediction cleaning was in use, and when it was not in use showed larger residuals for heat energy when it was in use but only by an increase of 0.3% seen on average over the whole data set. The difference in the top 15 ranking substations for each evaluated ranking between when the prediction cleaner was in use and not, was four substation. Of those all had clearly qualified to the top ranking 15 due to obvious errors in the predictions, in fig(4.7) one can see one substation where the prediction were classified as poor by the expert groups. The blue dots are the prediction and the black dots, are the data points which were removed in the prediction cleaning process, and the red dots are the linear interpolation between the removed data points. For all four substations similar result were obtained. Even though the overall prediction does not improve with the prediction cleaning method, a reduction in large deviations due to anomalies in the predictions have likely taken place.

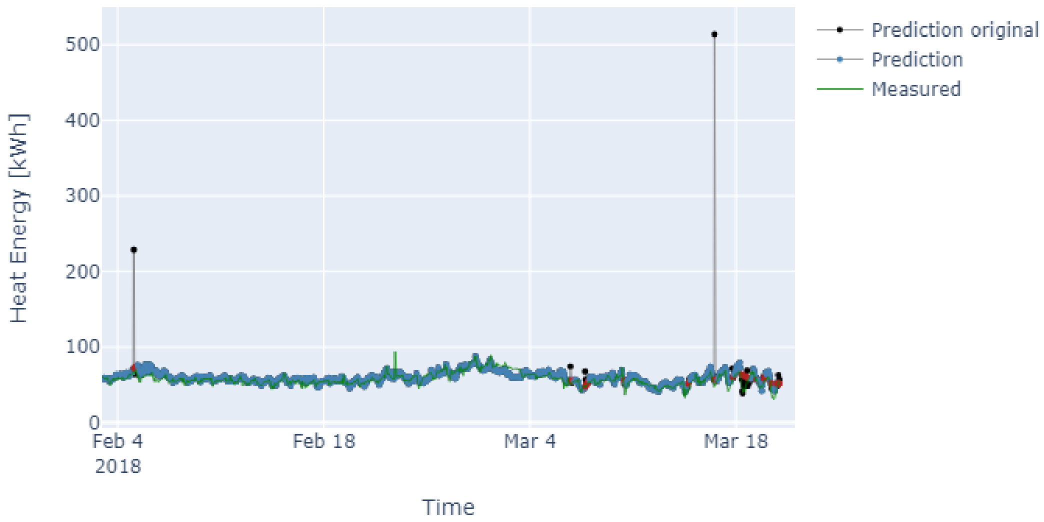

Figure 4.3: A plot over predicted and actual heat energy use in a substation. A solid line indicates mesured data, blue dot indicates the prediction. Black dots indicates prediction data points removed by the prediction cleaner, red dots indicates the interpolations that replaced the removed data points.

4.2.2 Result from the validation of the proposed rankings

In the following section the result from the validation method is presented. Each ranking evaluated is presented in its own section. Since the expert group was composed by two experts, there were some substations were the expert group answered contradicting. Substations for which both of the expert answered question 2 with that the prediction did not perform reasonably well were removed from the rest of the validation. A conservative approach to leave substation for which the expert group disagreed up on the question if the prediction model performed well was chosen. We will hereafter denote those predictions that was not classified as preforming reasonably well as poor predictions in fig(3.3) an example of a poor prediction can be seen. Based on the answers from question 1 to the expert group the substations were classified into three groups: Containing anomalies, Not containing anomalies and Ambiguously classified. Where those substations in the first group, containing anomalies all had gotten the answer 1 or 2 from the experts. The substations in the second group, not containing anomalies, got the answer 3 or 4. Each substation were the expert group disagreed were classified into the last group ambiguously classified. If a substation got a 1 from one expert and a 3 from the other, it would end up in this group.

4.2.3 Validation of ranking A

In the case of ranking A, $RV_{resNorm}$ for aggregated sequences, 5 of the 15 substations were classified with a poor performing prediction model. Of the 10 substations left for evaluation, 7 were classified into the containing anomaly group. One substation was classified into the, not containing anomaly group, and for 2 the answers were ambiguous. In fig(4.4-4.6) examples of substations that were classified into the different groups by the experts are shown.

Table 4.1: Summarized result from the evaluation of ranking A, after removing the 5 substations which were unambiguously classified as having a poor prediction model were removed.

Classified as	**Number of substation**
Containing anomalies	7
Not containing anomalies	1
Ambiguously classified	2

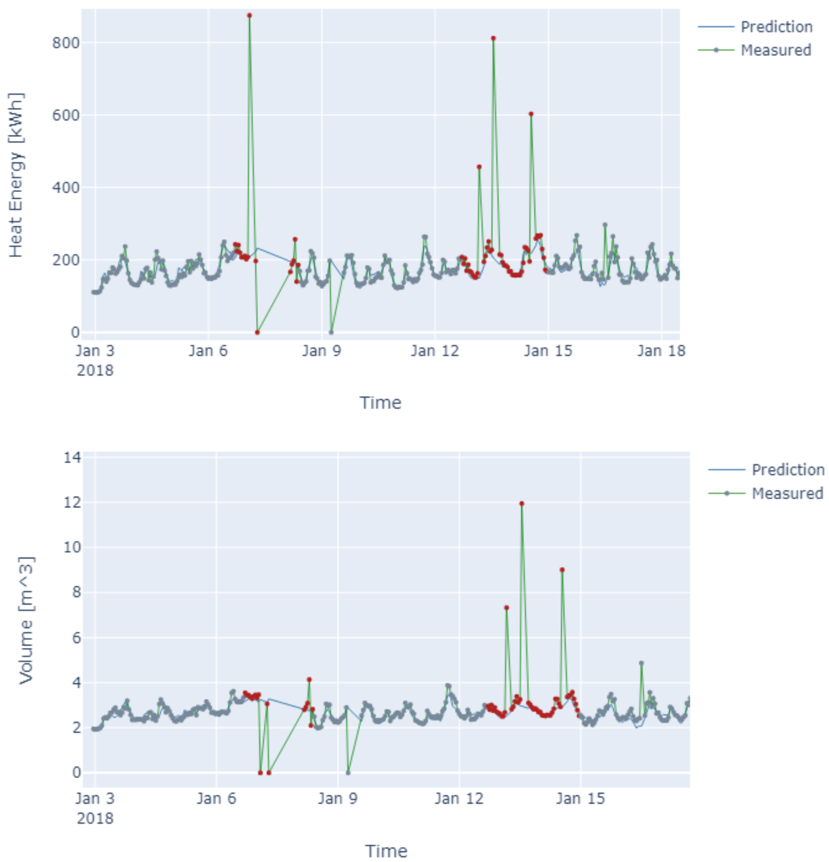

Figure 4.4: Two plots showing one of the substations were both experts classified it as having anomalies in the data. The solid line shows the prediction made with EnergyPredict, the dotted line shows the measured values, red dots indicated were anomalies were identified.

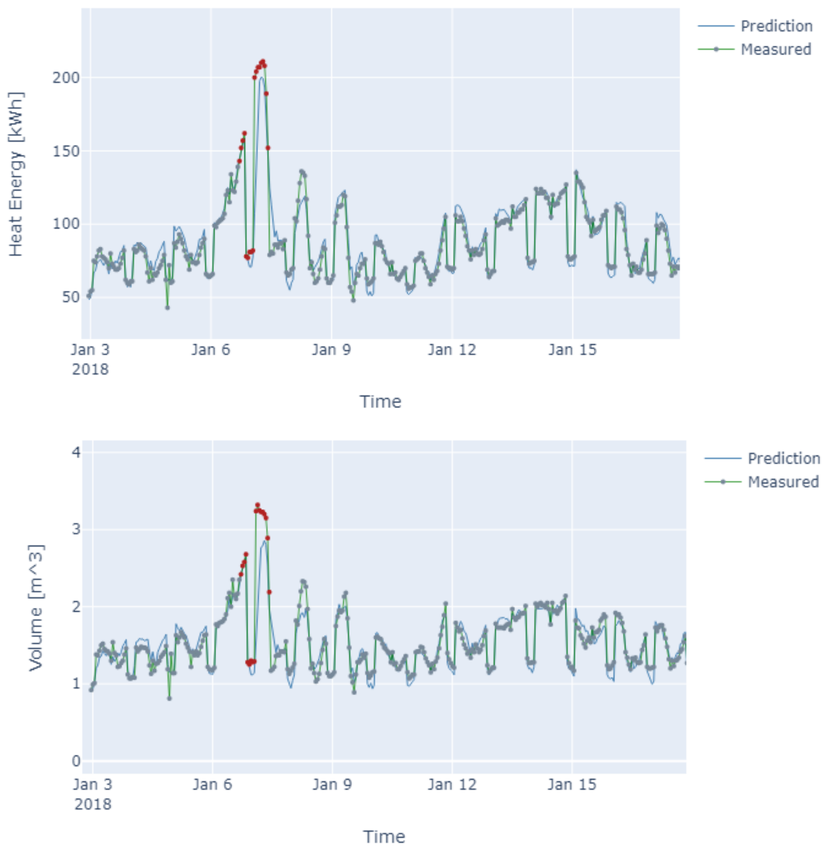

Figure 4.5: Two plots showing one of the substations were both experts classified it as not having any anomalous data points even though it ranked among the 15 highest in ranking A. The solid line shows the prediction made with EnergyPredict and the doted line shows the measured values, red dots indicated were anomalies were identified.

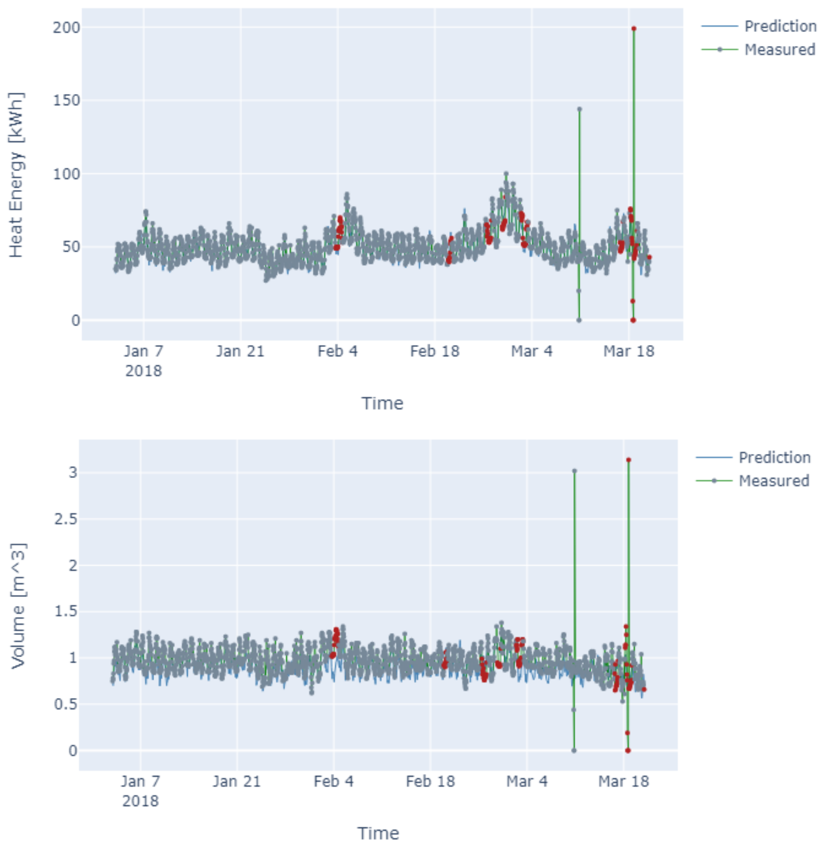

Figure 4.6: Two plots showing one of the substations were the experts disagreed to classify it as anomalous or not. The solid line shows the prediction made with EnergyPredict and the dotted line shows the measured values, red dots indicated were anomalies were identified.

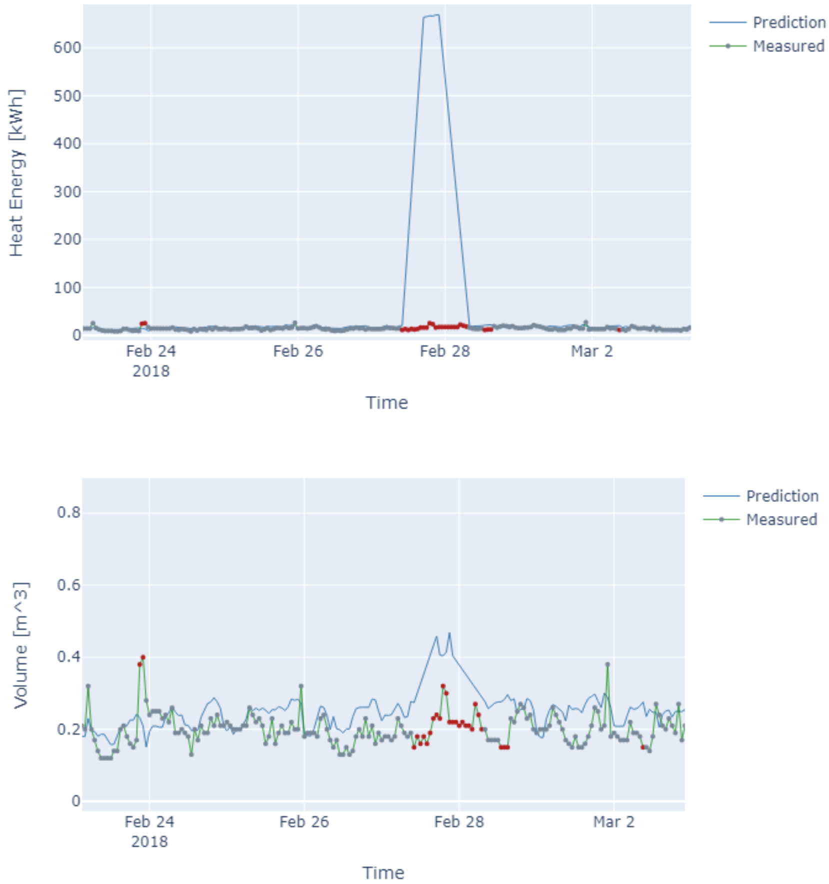

Figure 4.7: Two plots that show a substation were both experts agreed that prediction model did not perform reasonably well, and therefor were removed from further evaluation.

4.2.4 Validation of ranking B

Of the 15 highest substations ranked in ranking B, 4 substations were classified as having poor performing prediction models. Leaving 11 substations for further evaluation, of those 11, 7 were classified into the containing anomalies group. For 4 substation the answers were ambiguous, classifying them into the ambiguously classified group, none of the substations were classified of both experts as without any anomalies.

Table 4.2: Summarized result of the evaluation of ranking B, after the 4 substations that were unambiguously classified as having a poor prediction models were removed.

Classified as	Number of substation
Containing anomalies	7
Not containing anomalies	-
Ambiguously classified	4

4.2.5 Validation of ranking C

In the case of $RV_{avgNorm}$ for aggregated sequences, 8 of the 15 substations were classified as having poor prediction models. Of the remaining 7 substations 4 were classified into the containing anomalies group, 3 were classified into the ambiguously classified group and none were classified as being without anomalies.

Table 4.3: Summarized result from the evaluation of ranking C, after the 8 substations that was unambiguously classified as having a poor prediction model was removed.

Classified as	Number of substation
Containing anomalies	4
Not anomalous	-
Ambiguously classified	3

4.2.6 Validation of ranking D

Also here 8 prediction models were classified with not performing reasonably well. Of the remaining 7 substations 2 were ambiguously classified, and 5 were classified into the, containing anomalies group.

Table 4.4: Summarized result from the evaluation of ranking D, after the 8 substations that were unambiguously classified as having a poor prediction models were removed.

Classified as	Number of substation
containing anomalies	5
Not anomalous	-
Ambiguously classified	2

4.2.7 A collective evaluation of the four rankings

If one collectively evaluate the four rankings, 30 unique substations are considered. Of those 8 are classified as having poor prediction models leaving 22 for further evaluation. Of the 22 substation 14 are considered as having anomalous data points within them, only 1 substation is considered to be without anomalies and 7 are ambiguously classified.

Table 4.5: Summarized results from a collective evaluation of the rankings A, B, C, D, 30 unique substations were considered of which 8 were removed due to poor prediction models.

Classified as	Number of substation
Containing anomalies	14
Not anomalous	1
Ambiguously classified	7

4.2.8 Result from the reference group of well performing substations

From the reference group, of 20 random low ranking substations one were classified into the group, containing anomalies, 9 were classified into the, not containing anomalies group and, 10 were ambiguously classified.

Table 4.6: Summarized result of the evaluation of the reference group.

Classified as	**Number of substation**
containing anomalies	1
Not anomalous	9
Ambiguously classified	10

5 Discussion

The aim of the project was to find methods of detecting substations showing symptoms of fault, three symptoms where investigated. This turned out to be a difficult task. The first symptom of fault dealt within this thesis was abnormal quantization, an adjustment to the method suggested by Sandin et al. [27] was proposed and an evaluation was conducted. While evaluating the result from the quantization detection method it quickly became clear that the results where heavily influenced by the meter resolution in the substation. It seems like it is the meter resolution that is measured or flagged by this technique rather than any real quantization problems. It is probably not possible to develop this method any further simply due to the fact that the limiting factor is the meter resolution. To find any substations where for example an oversized valve is causing problems a higher resolution and sampling frequency is likely needed.

The initial idea for the drift detection method was to use several prediction models for the same substation each trained on a different year but predicting the same period, and then comparing the mean between these models and measured data. The idea was that a general trend or an abrupt change would indicate a drifting problem. This method would still be hard to validate but a substation with several prediction models generating similar means for all prediction but not the measured data would certainly be interesting. This method was however not possible to test due the fact that the data available did not stretch over so many years that this method would be possible. However as more data becomes available this initial idea might be possible to investigate further. As mentioned in case of drift detection the biggest problem have been to find a way to validate the method. A good way to do this is yet to be found. This is the reason why no results are being presented for this method. A suggested method have been to induce artificial error, however this is not a method that will give any reliable result. As an example if one induced an artificial drift and an artificial random error to a substation with a well performing prediction model the proposed drift detection method would detect the substation where the artificial drift affects the mean enough to be detected by the thresholds but, if a random error is induced it would not detect the error. The problem however lies in that when the method is applied to a whole DH network there is no way of knowing if the detected shift in mean value is due to a poor prediction or an actual change in the DHS. This severely limits the use of the proposed drift detection method, since it would be foolish to do any kind of intervention based on a non validated method. On the other hand the drift detection method might be useful as an indicator that a retraining of the prediction model is needed. One could of course use other metrics to create an indication of when a prediction model need retraining, for example the mean absolute scaled error. In this way the drift detection might be useful.

To detect anomalies is a difficult task, and as shown by the result of the expert group even the experts have a hard time agreeing on what is an anomaly. Approximately one third of the substations the experts disagreed. This might of course be due to different perceptions of what should be classified as an anomaly. But it highlights a problem with anomaly detection that defining what is normal and not can even be a problem for domain experts. The lack of consensus in the expert group for many substation, makes it hard to draw any solid conclusion about the performance of the anomaly detection method. However the method seems to have potential to identify anomalous substations and this work shows that the method has potential but also a few shortcomings that needs to be addressed in future work.

5.1 Method and result, of the anomaly detection method

In this discussion the choices made in the anomaly detection method will be discussed in greater detail. There will also be a short discussion regarding the result from the anomaly detection method. During the course of this work it has become evident that the prediction model used, which is based on regression trees, performs well for the vast majority of substations. However for some investigated substations it seems like over fitting is a problem leading to poor predictions in those substations. Here and through out the rest of this discussion the term poor predictions refers to predictions that for some periods greatly deviates from all neighbouring predictions or observations in such a way that one clearly can suspect that the prediction is not accurate. It is likely that meter errors are one part of this problem where the model is trained on obviously faulty data, which then leads to an over fitting problem. For example it was noted by the experts in the validation process that several substations had missing values or zeros right before or after a spike in the data, and the theory is that all those missing value or zeros have in some way been aggregated to one data point. See the error at 7th of January in fig(4.6) for an example of this phenomena. It is likely that these oddities are present in the training data, which means that EnergyPredict has trained on data where these oddities are present which likely is one of the contributing reason to why some predictions models are not performing as desired.

An inherent problem with the method chosen, that utilizes the residual between predictions and meter values is that there is no possible way to identify if it is the prediction model or the meter data that is anomalous. The method is dependent on the fact that the prediction models in the vast majority of cases in fact performs well. The choice to compare the substations in a ranking was made because the method used for identifying anomalous residuals, will find and flag anomalous data points in most substations. By ranking with relatively larger errors higher in the list. An operator using the developed method can start by investigating the substation higher in the list and work her way down the list to smaller and smaller relative errors. As mentioned there is no way for IsolationForest to know if it is the measured data or the prediction model that is anomalous. This became evident when investigating the 30 unique substations that make up the substations that ranked among the 15 highest in any of the rankings A, B, C, D. Of those substation 8 where marked as having a poor performing prediction model. While none of the substations in the reference group was classified with having a poor prediction model. This is natural when you elaborate around it, the developed method is favoring larger residuals, meaning that models that have poor performing prediction models are in risk of ranking high in any of the suggested rankings. From the result it is evident that the ranking A and B, the ranking normalized by the total sum of all residuals, are performing better in this way with 5 respectively 4 substation classified as having poor prediction model. Compared to the ranking C and D where for both rankings 8 prediction models where classified as being poor prediction models, hence including all of the poor prediction models found. The reasoning behind removing these poor performing substation before any further evaluation was done due to the fact that, since the ranking was done by taking the sum of the n =5 largest residuals which were flagged as anomalous by IsolationForest, a substation that has qualified to the top 15-list and had a poor prediction model had likely done so due to the poor prediction model and not because an oddity in the measured data. As an example, fig(4.7), shows a substation where the prediction of the heat energy makes takes on a value around February 28, several hundred times any measured value during the observed period. This error or anomaly in the prediction model is what has qualified the substation among the top 15 worst substations.

If a recommendation on which ranking is to be made, A and B gives overall lower false positives if we include substations with poor prediction models and substations classified in to the, not anomalous, and ambiguously classified groups, in the false positives category. Of the top 15 ranking substations in ranking A and B both classified 7 out of 15 where classified as having anomalies in the data. It is unclear if this rate of true positives will decrease or increase if more of the top ranking substations where to be considered e.g 25. Nevertheless it shows that the method works. And considering that this method is easily implemented it can certainly be of help when identifying substations that are at risk of beeing faulty in an DH network. If the number of false positives due to poor predictions is reduced the tool will certainly be even more useful.

5.2 Improvement and future work

For the quantization detection, it seems like further work is dependent on better meter data with higher resolution. As more and more substations are pre-manufactured, before installation it will likely become more unusual with problems of over or undersized valves maybe making the quantization error a thing of the past. In the drift detection method, future work would be to find a way to test and evaluate the proposed method alternatively to develop a whole new method.

As mentioned earlier the domain experts noticed that there were that quite a few of the substations which had missing values or zeros right before or after a spike in the data, and the theory is that all those missing value or zeros have in some way been aggregated to one data point. With this background information a test were performed where the meter data was replaced with a centred rolling mean with a window size of 6. When examining the 15 worst ranking substations using the $RV_{resNorm}$ for aggregated sequences there is still a problem with poor predictions. Seven of the 15 substation had not satisfactory prediction model or compromised data in some other way, of the remaining 8 substations, 6 shows evidence of anomalies. None of the examined substations had the pattern with missing values or zeros recognized by the experts. This alteration with a rolling mean for the meter data might therefore be a method to build future work on. In this thesis only ranking based on heat energy has been validated. A brief investigation on using volume flow instead and utilizing the rolling mean of the meter data discussed previously shows promising results. By investigating the 15 highest ranking substations when using a rolling mean with window size of 6 in the $RV_{resNorm}$ for aggregated sequences base on volume flow. A preliminary investigation shows that 7 substations have poor performing prediction models, but of the remaining 8 substation all but one shows sign of anomalies in the data. This is however results that need to be validated in future work.

In order to reduce the number of false positives a way to check if the prediction models are performing okay and not producing run-away predictions, like the one in fig(4.7), is needed. This can be achieved by either some check on the prediction models, flagging some predictions as non-reliable for fault detection or some alterations to EnergyPredict so that no or only very few of the produced models have the runaway behaviour as seen in fig(4.7). One should keep in mind that EnergyPredict is a tool developed to minimize the average error in the prediction rather than provide a platform for anomaly detection. It should also be noted that EnergyPredict performs really well for the vast majority of substations, but as discussed previously the method used finds substations with abnormal residuals, which naturally are either substations with an anomalies in the measured values or in the predictions. Therefore when this anomaly detection method is in use by an operator, and a operator finds a substation with a prediction that is not satisfactory, an option should be to retrain the prediction model so that a better prediction is obtained.

6
Conclusion

Three methods were developed for investigating symptoms of faults in district substations. The abnormal quantization method does not show potential for being useful for detection symptom of fault. The method proposed for drift detection gives, has not been evaluated properly due to difficulties in finding a suitable validation method. The developed method for identifying anomalies shows potential for being used in future anomaly detection methods, especially if measures to reduced the numbers of prediction models causing false positives to even lower levels can be taken into place. The aim of the project was to investigate if there was a possibility to use Utilifeed's prediction software EnergyPredict to detect symptom three distinct symptoms of fault. Two out of three methods proposed used EnergyPredict, one showed promising results and possibilities for further developments.

Bibliography

[1] Statens energimyndighet. *Energiläget 2020*. ET 2020:1. 2020.

[2] Hakan İbrahim Tol. *District heating in areas with low energy houses*. Thesis for the degree of doctor of philosophy, Technical University of Denmark, 2015.

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

[4] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection. *ACM Computing Surveys*, 41(3):1–58, jul 2009.

[5] Srinivas Katipamula and Michael R. Brambley. Review article: Methods for fault detection, diagnostics, and prognostics for building systems—a review, part I. *HVAC and R Research*, 11(2):169–187, 2005.

[6] Srinivas Katipamula and Michael R. Brambley. Review article: Methods for fault detection, diagnostics, and prognostics for building systems—a review, part II. *HVAC and R Research*, 11(2):169–187, 2005.

[7] John E. Seem. Using intelligent data analysis to detect abnormal energy consumption in buildings. *Energy and Buildings*, 39(1):52–58, 2007.

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

[9] Sara Månsson, Per Olof Johansson Kallioniemi, Kerstin Sernhed, and Marcus Thern. A machine learning approach to fault detection in district heating substations. *Energy Procedia*, 149:226–235, 2018.

[10] Marco De Nadai and Maarten Van Someren. Short-term anomaly detection in gas consumption through ARIMA and Artificial Neural Network forecast. *2015 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems, EESMS 2015 - Proceedings*, pages 250–255, 2015.

[11] Shiraz Farouq, Stefan Byttner, Mohamed-rafik Bouguelia, Natasa Nord, and Henrik Gadd. Engineering Applications of Artificial Intelligence Large-scale monitoring of operationally diverse district heating substations : A reference-group based approach . *Engineering Applications of Artificial Intelligence*, 90(December 2019):103492, 2020.

[12] Svend Frederiksen and Sven E Werner. *Fjärrvärme och fjärrkyla*. Studentlitteratur, Lund, 1. uppl. edition, 2014.

[13] 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.

[14] Svensk fjärrvärme. fjärrvärmecentralen - Utförande Och Installation. 2004.

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

[16] Energiföretagen. Energimätare för termisk energi Tekniska bestämmelser F:104, vers. 2. 2019.

[17] Jouko Pakanen and Juha Kuisrnin. Fault diagnosis methods for district heating substations. 1996.

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

[19] Sara Månsson, Per Olof Johansson Kallioniemi, Marcus Thern, Tijs Van Oevelen, and Kerstin Sernhed. Faults in district heating customer installations and ways to approach them: Experiences from Swedish utilities. *Energy*, 180:163–174, 2019.

[20] Sara Månsson, Kristin Davidsson, Patrick Lauenburg, and Marcus Thern. Automated statistical methods for fault detection in district heating customer installations. *Energies*, 12(1):113, 2019.

[21] Kimmo Yliniemi, Jan Van Deventer, and Jerker Delsing. Sensor fault detection in a district heating substation. In *Proceedings of the 10th IMEKO TC10 International Conference on Technical Diagnostics: Budapest, Hungary, June 2005, 2005*, 2005.

[22] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning.* Springer Series in Statistics. Springer New York, New York, NY, 2009.

[23] Arindam Banerjee and Vipin Kumar. Anomaly Detection : A Survey Technical Report Department of Computer Science TR 07-017 Anomaly Detection : A Survey Varun Chandola , Arindam Banerjee , and Vipin Kumar. (July), 2009.

[24] Andrew Emmott, Shubhomoy Das, Thomas Dietterich, Alan Fern, and Weng-Keen Wong. A Meta-Analysis of the Anomaly Detection Problem. *arXiv preprint arXiv:1503.01158*, (2015), mar 2015.

[25] Fei Tony Liu, Kai Ming Ting, and Zhi Hua Zhou. Isolation forest. *Proceedings - IEEE International Conference on Data Mining, ICDM*, pages 413–422, 2008.

[26] G Hackeling. *Mastering Machine Learning with Scikit-Learn, Second Edition.* Packt Publishing, second edition, 2017.

[27] Fredrik Sandin, Jonas Gustafsson, Robert Eklund, and Jerker Delsing. Basic methods for automated fault detection and energy data validation in existing district heating systems. *Proceedings of DHC13, the 13th International Symposium on District Heating and Cooling*, pages 182–189, 2012.

[28] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller. Scikit-learn. *Journal of Machine Learning Research*, 12(1):2825–2830, jun 2011.

A

Appendix

```
Data: Prediction time series PTS
Result: Series with variance for each segment VOS
initialization;
i=1;
start=0;
end=n;
while not at end of PTS do
    VOS[i] = variance of PTS[start:end] ;
    start=start+m ;
    end=end+m;
    i=i+1;
end
```

Algorithm 1: The algorithm for defining the variance of segment