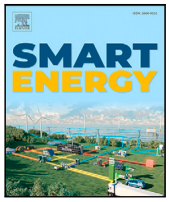
Smart Energy 16 (2024) 100159

Contents lists available at ScienceDirect

Smart Energy

journal homepage: www.journals.elsevier.com/smart-energy



Fault detection for district heating substations: Beyond three-sigma approaches

Chris Hermans *, Jad Al Koussa, Tijs Van Oevelen, Dirk Vanhoudt

Flemish Institute for Technological Research (VITO), Boeretang 200, 2400 Mol, Belgium
EnergyVille, Thor Park 8130, 3600 Genk, Belgium

ARTICLE INFO

Keywords:
4GDH
District heating
Smart thermal grids
Smart energy systems
Fault detection
Anomaly detection
Quantile regression

ABSTRACT

The topic of this paper is fault detection for district heating substations, which is an important enabler for the transition towards fourth-generation district heating systems. Classical fault detection approaches are often based on anomaly detection, commonly making the implicit assumption that the errors between the measurements and the predictions made by the baseline model are i.i.d. and following an underlying Gaussian distribution. Our analysis shows that this does not hold up in the field, showing clear seasonality in the error over time. We propose to replace the Gaussian error model by a quantile regression model in order to provide a more nuanced fault threshold, conditioned on time and other input variables. Additionally, we observed that properly training the baseline model comes with its own challenges due to this time dependency, which we propose to resolve by employing an ensemble of models, trained on different periods of time. We demonstrate our method on unlabelled operational data obtained from a Swedish district heating operator to illustrate its use in the field. In addition, we validate it on labelled data from our residential lab setup, testing a variety of common faults.

1. Introduction

District heating (DH) networks play an important role in decarbonizing heat demand and contributing to the global energy transition. These networks operate by centralizing heat production and subsequently distributing it through a network of pipes to end consumers. The heat serves multiple purposes, including space heating, domestic hot water (DHW) production, and various thermal processes. Notably, this system exhibits the capacity to harness local fuel or heat resources that would otherwise go to waste, thereby enhancing overall energy efficiency and sustainability [1,2].

Suboptimal or faulty performance of district heating substations or components on the secondary side of the substation contributes to an increase in the return temperature of the DH network [3]. It should be noted that we refer to faults as an indication that the system is not performing as it should, which might lead to malfunctions or breakdowns. Such faults can be technical issues affecting the hardware and software at the substation level and the secondary side, wrongly set settings in components and controllers or wrongly designed systems. Examples of faults include fouling of heat exchangers, under or over dimensioned valves, broken actuators of automated valves and distorted measurements from temperature sensors. Also, the heat curves can be set too high or the secondary system could be hydraulically unbalanced. In the work of Månsson [4], a detailed overview of the different faults is given. Several faults do not affect the comfort at the customer, so they are not directly detected. However, as they contribute to an increase in the return temperature of the DH network, their repercussions are multifaceted and directly impact the overall efficiency of the network: heat losses in the network increase, the efficiency of production units such as geothermal plants or gas condensing boiler decreases, the power-to-heat ratio in combined heat and power (CHP) plants is reduced, the utilization of low-grade heat such as industrial excess heat or solar thermal heat becomes limited, and the overall volume flow in the DH network increases [2,3,5]. Additionally, it becomes more difficult to reduce the DH supply temperature due to capacity limitations in the network [3]. This would lead to, among other effects, a decrease in the efficiency of heat pumps, limiting the integration of DH networks in Smart Energy Systems in which DH systems can provide short-term storage for intermittent renewable energy [6,7]. With the trend of moving towards fourth-generation DH systems [8], the importance of lower return temperatures has become increasingly relevant, since the margin for error decreases with lower network temperatures [9]. In fact, reducing return temperatures is an enabler for the transition towards fourth-generation DH systems. These findings underscore the critical importance of maintaining optimal performance at DH substations to ensure energy efficiency and sustainability.

* Corresponding author at: Flemish Institute for Technological Research (VITO), Boeretang 200, 2400 Mol, Belgium.
E-mail address: chris.hermans@vito.be (C. Hermans).

https://doi.org/10.1016/j.segy.2024.100159
Received 1 February 2024; Received in revised form 23 September 2024; Accepted 24 September 2024
Available online 30 September 2024

The critical role of return temperature in district heating networks has prompted proactive measures by network operators. These initiatives aim to curtail return temperatures within the network through innovative customer engagement strategies. For instance, by introducing motivational tariffs to their customers with penalties for high return temperatures and benefits for low return temperatures, periodical audits of building installations, counselling services, closer relationships with the customers and training of installers [4,9,10].

Previous research indicated that between 43% and 75% of analysed substations in some DH networks operate suboptimally [3,11]. District heating operators and technology providers have also been reporting about these issues. Hofor, the district heating operator in Copenhagen, estimates that 50% of the substations in that network runs suboptimally due to faults [12]. Also, the company Samson mentions that due to faults, a large substation can cost up to 100,000 euros/year of unnecessary network pump power [13].

District heating utilities have mainly been relying on time-consuming, costly, and error-prone manual inspection and analysis in order to determine substations with suboptimal or faulty operation [2,3,14,15]. The scale of inspections can be large: for instance, Wien Energie, the DH operator of the network in Vienna reports that more than 23,000 trouble-shootings and 12,000 repairs have to be performed per year on their network [13]. The progressive digitalization of district heating (DH) networks, mandated by legal frameworks in some countries [16], yields an increase in available data. This creates the potential for using data-driven methods for automatic fault detection and diagnosis, or for assisting operators to do so. Machine learning (ML) can process large amounts of data automatically and identify hidden patterns and fault signatures that may not be detected by manual methods, helping the DH network operator to reduce down times, costs and the impact on the network [15,17]. For example, the company Samson states that 30%–80% fewer service trips are taken due to the introduction of automated fault detection and diagnosis techniques in a network [13].

Despite the increment in available data and improved logging, manual inspection and classification of faults still requires a lot of time and effort. As such, labelled data is hard to come by, and even when it is available it is too scarce to apply classical supervised learning methods. This lack of labelled data is an intrinsic problem of district heating system monitoring, and in order to deal with this, a number of different methods have been developed over the years. Neumayer et al. [15] categorize them into traditional methods such as manual analysis, thresholds, and physical models, and machine learning methods such as classification, regression, and clustering. The review shows that a variety of methods is used, mostly focusing on unsupervised anomaly detection (AD) rather than fault detection, due to a lack of labelled datasets. Van Dreven et al. [17] groups the methods used for fault detection differently, by separating between data mining and knowledge discovery (DMKD) techniques and AD techniques. On one hand, DMKD is the process of identifying patterns and insights in large complex datasets. It includes clustering and regression to identify patterns, relationships, and trends within the data, to make informed decisions and predictions. On the other hand, AD is the process of identifying anomalies, which consist of data deviating significantly from normal or expected behaviour.

Within the AD methods, a mix of expert knowledge and statistical parameters play a pivotal role to determine operational limits and expected behaviour, which serve as critical decision points for automatically identifying anomalies. Sandin et al. [18] present a number of methods for fault detection in substations: they suggest the use of limit checking with linear thresholds, and using statistical notions such as the Z-score, which is an indicator of the normalized distance between measurement and mean, normalized by the standard deviation sigma. In the work of Månsson et al. [11], the 3-sigma rule is used to automatically detect and rank poorly performing substations. This rule expresses that nearly all values of a normal distribution fall within three standard deviations from the mean. Values falling outside that range are considered as anomalies. In Yliniemi et al. [19], the authors introduce a method to detect temperature sensor faults in substations. The authors calculate the variance of the measurements of the temperature sensor, and consider that measurements within 2-sigma indicate faults. In Gadd et al. [3], a manually-set threshold based on expert knowledge is used to identify substations with low primary temperature difference.

When using regression or deep learning, assumptions need to be considered to choose the corresponding limits. Zhang et al. [20], who proposed a method for anomaly detection of DH substations using a simplified physical model consisting of long short-term memory (LSTM), combined with a variational auto-encoder, state that anomaly thresholds were decided upon using trial-and-error. They use thresholds of 93% and 99.5% on the reconstruction error to detect the anomalies. In the work of Calikus et al. [21], the authors suggest a method to rank abnormal substations based on their power signature using robust regression, making use of the 3-sigma rule for anomaly detection as well.

The work presented in this paper takes a closer look at a common assumption made in many of the papers above: more specifically that the observed errors follow a Gaussian distribution, are unbiased (*i.e.* can be characterized by a mean of zero, with no overestimation or underestimation taking place), and demonstrate homoscedasticity (there is only one single finite standard deviation σ that is uniform across all levels of the independent variables). In particular the heteroscedasticity of the error distribution poses a problem for the reliability of some of the techniques above, as thresholds defined in terms of percentiles break down when the underlying error model is no longer valid. To remedy this, we propose a simple adjustment to the classical AD approach: replacing the Gaussian error model by a linear/non-linear quantile based regression model.

Quantile regression is a versatile statistical technique that extends traditional linear regression by allowing for the estimation of conditional quantiles of the response variable, rather than just the mean. This makes it particularly useful for analysing the impact of predictors across different points of the distribution of the dependent variable. Because it is such a versatile technique, it is used across various domains such as finance [22], healthcare [23] and climate studies [24]. Additionally, quantile regression has been used in anomaly detection across different fields in recent years. However, we do not believe it has been introduced in the field of district heating systems, with bearing fault detection for wind turbines [25] as the application closest to our work. The majority of these techniques employ tailored quantile neural networks, whose complexity and data requirements are a liability rather than an asset for our particular application. We have opted for the more interpretable classical linear quantile regression, where the contribution of each input variable is directly reflected by its associated weight, and a more time efficient non-linear variant based on the transformation of the input space using Gaussian basis functions.

Additionally, we have observed that the classical AD approach can fail if the training period of the baseline model is chosen poorly, resulting in a non-representative model for what constitutes normal behaviour. In order to deal with this, we propose the use of an ensemble of models, trained on data from different periods of time. This will increase robustness, and even allow for an early differentiation between fault types. However, the way these models are employed differs significantly from how ensembles are commonly used in literature [26]. In classical ensemble learning, the trained models each contribute a single prediction, and the result is a single weighted sum. Our method performs no such aggregation, but analyses the difference between predictions to draw conclusions. As such, it should not be confused with classical ensemble techniques.

The paper is structured as follows. Section 2 will cover our methodological approach. We will start by introducing the classical approach commonly used in fault detection by anomaly detection, followed by an analysis of some of the problems encountered in this approach,

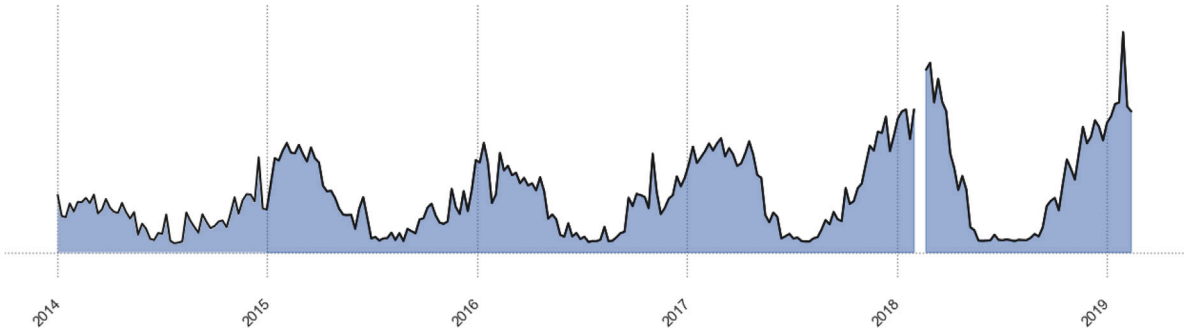


Fig. 1. Five years of data from a substation in a Swedish DH network. The data was collected between 2014 and 2019, and is visualized at a resolution of a single point per week. Each point represents the root mean square error (RMSE) between the predicted volumetric flow rate, and the actual volumetric flow rate. As the data is normalized, the values have no actual meaning, so they were left out of the figure. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and our proposed adjustments as a consequence of these observations. Section 3 will demonstrate our method on unlabelled operational data obtained from a district heating operator, as well as labelled data from our residential lab experiments. Finally, we conclude our paper in Section 4.

2. Methodology

2.1. Classical approach

In this paper, we propose to follow a classical AD approach, consisting of three steps: (1) modelling the baseline behaviour, (2) modelling the error distribution, and (3) determining anomalies in new observations.

2.1.1. Model the baseline behaviour

The first step consists of determining what normal (baseline) behaviour looks like. This requires access to a dataset that was recorded when the operation of the substation was not faulty, in order to train a representative model on it. In literature, there exists a large variety in the choice of models, from physics-based white-box models to black-box models based on deep learning. In principle, a very high level of accuracy is not a strict requirement, assuming that errors induced by faults are larger than the difference in errors between a moderate and high accuracy model. What is essential is that the model enables proper distinction between normal and anomalous behaviour. As such, some form of linear model is often more than sufficient. In our work we have opted for a polynomial ridge regressor, a linear regression algorithm where the input features are extended with polynomial terms, and the mean squared error (MSE) loss is supplemented with an L_2 regularization term over the weights. The degree of the polynomial and the weight of the regularization term are determined by hyperparameter optimization, more specifically Bayesian optimization (BO) using Tree-based Parzen Estimators [27]. This choice of regression model strikes a good balance between high training speed and sufficiently high accuracy.

While data availability in the field differs between district heating operators, we assume that at least the following data is available:

- $\dot{V}$: volumetric flow rate
- T_{out}: outdoor temperature
- $\bar{T}_{out}$: mean outdoor temperature over the last 24 h
- T_r: return temperature
- T_s: supply temperature

During our experiments, we restricted ourselves to the use of these variables only. As we are interested in modelling patterns in the *relationship* between these variables, the choice of labelling some as inputs and others as outputs is somewhat arbitrary. We have adopted a model that forecasts the current volumetric flow rate $\dot{V}$, based on a combination of the variables above, with different time offsets. In addition, we have included temporal features to enrich the model (weekday, time of day, day of year). For the remainder of this work, we will refer to the baseline model f, the input variables $\mathbf{x}$, the measured flow rate as y, and the predicted flow rate as $\hat{y} = f(\mathbf{x})$.

2.1.2. Model the distribution of the prediction error

The next step consists of establishing an error model. This step is commonly performed implicitly, simply assuming that the error $\epsilon = \hat{y} - y$ follows a Gaussian distribution $\epsilon \sim \mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2)$, with mean $\mu_\epsilon = 0$ and a standard deviation σ_ϵ, which is derived from the observed errors.

2.1.3. Observe new measurements to determine anomalies

The final step consists of determining a threshold to specify when a measurement can be classified as an anomaly. Common choices here are the 2-sigma or 3-sigma threshold, which theoretically correspond to a 95% or 99.7% probability respectively of falling within the data distribution. In practice, the exact thresholds are sometimes chosen empirically, based on how they hold up in the field.

In a live monitoring scenario, sensors produce new measurements constantly, which can be compared to the predictions made by the baseline model. If the difference exceeds the aforementioned threshold, the measurement is labelled as an anomaly. If sufficient anomalies within a certain period of time are reported, the substation is flagged for inspection. The exact number of anomalies required to report a fault needs to be decided by the end user, and is dependent on factors such as baseline model accuracy and temporal resolution of the data.

2.2. Data analysis

In this section, we will take a closer look at 5 years of data from a substation in a Swedish DH network, collected over a period between 2014 and 2019. In Fig. 1, we take a look of the prediction error as it evolves through time, visualized at a resolution of a single point per week to improve readability. Each point represents the normalized root mean square error (RMSE) between the predicted volumetric flow rate, and the actual volumetric flow rate. As the data is normalized, the values have no actual meaning, so they were left out of the figure. The first thing to notice is that the error is not independent and identically distributed (i.i.d.). There is clear seasonality present in the data, which means that an error value that otherwise would seem perfectly viable in winter could be a clear sign of an anomaly during summer. As such, the error is clearly conditioned on the time of the year, which indicates that it is not identically distributed.

In Fig. 2, we take a closer look at the actual probability distribution of the errors within a single year. The first distribution represents the period Jan 2014–Dec 2014, while the second represents Jan 2015–Dec 2015. Both distributions, represented by the blue histogram, are

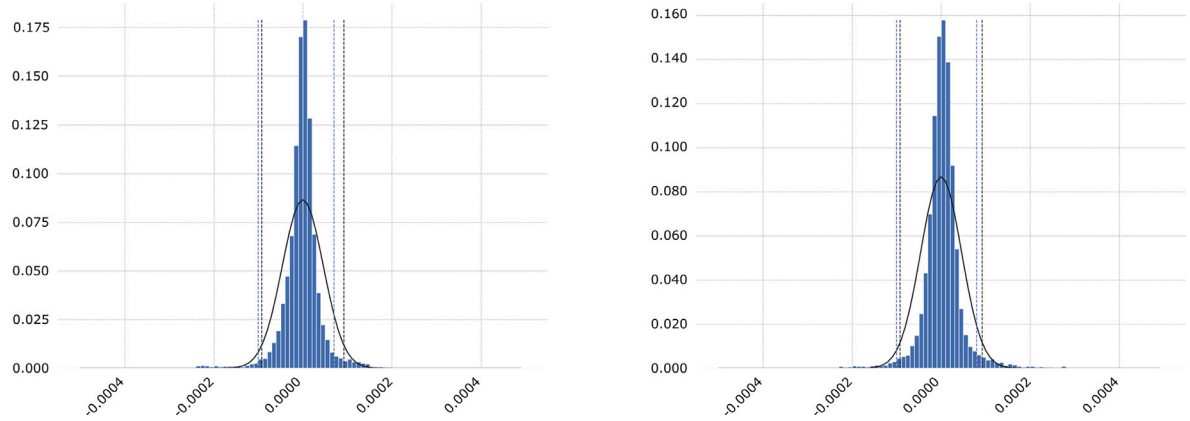


Fig. 2. Two examples of probability distributions of the error between observed measurements and baseline model predictions, superimposed by the computed corresponding Gaussian distribution $\mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2)$ computed directly from the same data, plotted at the same resolution. Even though the distribution has a typical bell shape, using the standard deviation to find the corresponding 2.5th and 97.5th percentile (holding 95% of the data), produces error thresholds (black dashes) that differ significantly from those computed directly from the data (blue dashes). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

superimposed by the Gaussian $\mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2)$ computed directly from the data. Despite the typical bell shape, simply computing the standard deviation will not produce proper results when trying to establish threshold boundaries for the underlying data. Assuming the Gaussian distribution is a correct representation, the standard deviation should allow us to determine the thresholds for the 2.5th and 97.5th percentile (corresponding to the lowest 95% absolute errors in the distribution), illustrated by the black dashed lines). However, if we compare these thresholds to the actual thresholds computed directly from the data, depicted by the blue dashed lines, we can see a significant difference.

This is just an example from one single substation, but discrepancies such as these are common. While the sources of these irregularities may vary, if we desire a system that is robust regardless of the specifics of a single substation, we need to accommodate for this. As such, we want to: (1) compensate for the fact that measurement error is not i.i.d., e.g. by conditioning the error model on time and possibly other features, and (2) define thresholds in terms of the likelihood that the sample falls inside the data distribution (e.g. a threshold explicitly of 95% instead of 2-sigma), requiring an alternative error model.

2.3. Training the baseline behaviour model

As we have mentioned in Section 2.1.1, in order to train a representative model for the baseline behaviour, access to data that was recorded when the operation of the substation was not faulty is required. However, as we are unaware of the trustworthiness of the data, or how much we actually need for a representative training sample, this poses additional practical challenges by itself.

2.4. Proposed adjustments

In order to deal with the limitations of the classical approach described above, we propose two adjustments: (1) the use of a drop-in replacement for the Gaussian error model, but one more capable of representing the actual distribution: *quantile regression models*, and (2) the use of an ensemble of models trained at different moments of time, which increases robustness of the method and allows for rudimentary fault classification.

2.4.1. Quantile regression models

A quantile regression model $Q_\tau(\epsilon \mid \mathbf{x}')$ represents the τ-th quantile of the conditional distribution of the response variable $\epsilon = |\hat{y} - y|$, where ϵ is the measurement error whose distribution we are trying to model, given the input variables $\mathbf{x}' = (1, x'_1, \ldots, x'_n)$. Unlike in the Gaussian case, we chose to model the absolute error instead, as it is easier to interpret and aggregate these values. The parameter τ is a value between 0 and 1, representing the desired quantile level. In other words, if we want to determine the threshold for which there is a 95% probability the measured error ϵ is smaller than said value, we need to compute a quantile regressor with a value of $\tau = 0.95$. The proposed method can be adapted to work with the signed error. Thus, assigning a 95% probability threshold would correspond to the measurement error falling between the 0.025 and 0.975 quantile, which is a bit harder to interpret and visualize, especially when comparing to aggregated error values such as the MAE and RMSE. This will become more clear in Section 3, while discussing the results of our experiments.

Depending on the selected input variables $\mathbf{x}'$, a corresponding value ϵ_τ can then be computed using the same formula from classical linear regression:

$$\epsilon_\tau = q_\tau(\mathbf{x}') = \mathbf{w}\mathbf{x}' = w_0 + \sum_{k=1}^{n} w_k x'_k \tag{1}$$

The difference lies in the employed loss function to determine $\mathbf{w}$. Whereas classical linear regression minimizes the mean squared error (MSE) a.k.a. the L_2 loss $L_2(\hat{y}_i, y_i) = (\hat{y}_i - y_i)^2$ summed over all training points $y_i = f(x'_i)$, quantile regressors minimize the pinball loss instead. For reference, the standard formula for the pinball loss is as follows:

$$L_\tau(y, \hat{y}) = \max\{(\tau - I(y \le \hat{y})) \cdot (y - \hat{y}), (\tau - I(y > \hat{y})) \cdot (\hat{y} - y)\} \tag{2}$$

or when written in terms of measurement error ϵ and error threshold ϵ_τ:

$$L_\tau(\epsilon, \epsilon_\tau) = \max\{(\tau - I(\epsilon \le \epsilon_\tau)) \cdot (\epsilon - \epsilon_\tau), (\tau - I(\epsilon > \epsilon_\tau)) \cdot (\epsilon_\tau - \epsilon)\} \tag{3}$$

where $I(\cdot)$ is the indicator function that equals 1 if the condition inside is true and 0 otherwise. However, most practical implementations of quantile regression include an additional L_2 regularization term $\lambda \mathbf{w}^\top \mathbf{w}$ in the loss function, where λ is a tuneable hyperparameter.

Because $Q_\tau(\epsilon \mid \mathbf{x}')$ is conditioned on the input variables $\mathbf{x}'$, we can account for seasonality by including temporal parameters such as the day of the year. In a similar fashion, we can include other dependencies such as the supply temperature T_s. Note that the inputs $\mathbf{x}'$ for the error model are different from the input variables $\mathbf{x}$ of the baseline model. In addition, in case of the presence of an L_2 regularization term, the corresponding hyperparameter λ can be tuned using Bayesian optimization, just like in the baseline models. The only caveat is that for the error model, the pinball loss as defined in Eq. (3) has to be used during hyperparameter optimization instead.

Linear regression models can be extended to capture non-linearity in the data through the use of basis functions. E.g., our baseline behaviour

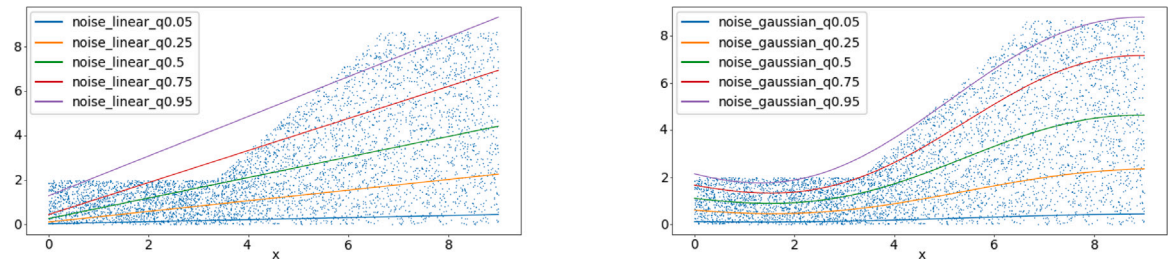


Fig. 3. Linear quantile regressors $Q_\tau(y \mid x)$ vs. non-linear quantile regressors $Q_\tau(y \mid \boldsymbol{\Phi}(x))$, where the response variable y is conditioned on a single input variable x. All data points (x, y) were sampled under a piece-wise linear curve g, such that $x \in [0, 1], y \in [0, g(x)]$. The Gaussian basis functions $\boldsymbol{\Phi}$ applied to x introduces a very continuous non-linearity in the thresholds produced by the quantile regressor, more closely resembling the shape of g.

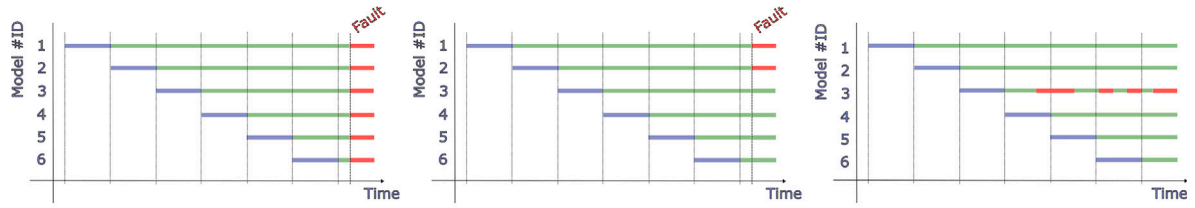


Fig. 4. Three possible scenarios, using an ensemble of models for fault detection. The x-axis shows progress through time, the y-axis indicates model number. The blue area labels the model training data, the green area shows when the model thinks the substation is performing normally, while red indicates the model warns for a fault. (a) An immediate fault occurred, e.g. a valve broke; (b) A case of gradual decline/drift, e.g. due to aging; (c) Detecting a faulty sub-model at model ID #3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

model is a polynomial ridge regressor, a linear model where the original input space x is transformed to a set of polynomial inputs using a polynomial basis function $\phi_i(x) = x^i$:

$$y = w_0 + \sum_{i=0}^{d}\sum_{k=1}^{n} w_{ik}\phi_i(x_k) = w_0 + \sum_{i=0}^{d}\sum_{k=1}^{n} w_{ik}x_k^i \tag{4}$$

Another useful basis function is the Gaussian function $\phi_i(x) = \exp\left(-(\mu_i - x)^2/2\sigma_i\right)$, where μ_i and σ_i respectively represent its centre and standard deviation:

$$y = w_0 + \sum_{i=0}^{m}\sum_{k=1}^{n} w_{ik}\phi_i(x_k) = w_0 + \sum_{i=0}^{m}\sum_{k=1}^{n} w_{ik}\exp\left(-\frac{(\mu_i - x_k)^2}{2\sigma_i}\right) \tag{5}$$

The parameter m defines the number of basis functions, and as such the resolution at which the centres μ_i are spread out over the feature input space. More advanced formulations are possible, but the common denominator is that the original input space is transformed into an alternative feature space, followed by a standard linear regression on these newly generated features. This allows for the capture of non-linearities in the data, while remaining linear in terms of training of the model.

Similarly, we are able to extend quantile regression models to capture non-linearities in the distribution of the response variable ϵ, using these Gaussian basis functions $\boldsymbol{\Phi}$, by applying standard quantile regression to the transformed input variables: $Q_\tau(\epsilon \mid \boldsymbol{\Phi}(\mathbf{x}'))$. We illustrated this principle in Fig. 3, where we have created an artificial dataset where the range of the response variable is conditioned on the value of input variable.

Finding the optimal set of hyperparameters, such as the weight of the regularization term λ, the choice of input variables $\mathbf{x}'$, or the choice between linear and non-linear regression, can be performed using the same framework used in Section 2.1.1. The only requirement is the replacement of the MSE loss by the pinball loss in the Bayesian optimization process. Depending on the particularities of the substation data, and the accuracy of the baseline model, it could be that simpler models perform better than more complex ones. Like in any regression problem, it is always good practice to perform hyperparameter optimization to choose a model of the right complexity.

2.4.2. Ensemble of baseline behaviour models

Our second proposal consists of *creating additional baseline models* as time progresses trained on a sliding window into their recent past (e.g. trained only on data from the last year), instead of retaining only a single baseline behaviour model. This has several advantages:

- Because an ensemble has multiple models, they might *capture different aspects of the data*, giving a more complete picture of the normal behaviour. E.g., in case the complete parameter space was not covered before a fault occurred, and it occurred in an unseen part (for example in winter when all training data was from spring till fall), at least the model detect anomaly again once the measurements come back into familiar parts of the parameter space (in this example: spring). This aspect is important when we do not have a lot of data yet.
- Similarly, if for some reason training resulted in a sub-par trained model, it will disagree with the surrounding models when flagging incoming readings as an anomaly. This is illustrated in Fig. 4c.
- Because models are trained at different points in time, this allows for some *initial fault classification* to take place, despite a lack of labelled data. More precisely, it allows us to differentiate between immediate faults (such as a broken valve), and gradual decline/drift (e.g. due to aging). We illustrate this in Figs. 4a and 4b.

3. Results

In this section, we will cover two sets of experiments. The first experiment consists of a controlled lab experiment, where we have introduced several faults, resulting in a labelled dataset. The second experiment evaluates operational data from the substation that we have visited earlier in Section 2.2. This is unlabelled data, meaning that the determination of the substation state is uncertain over the entire time period.

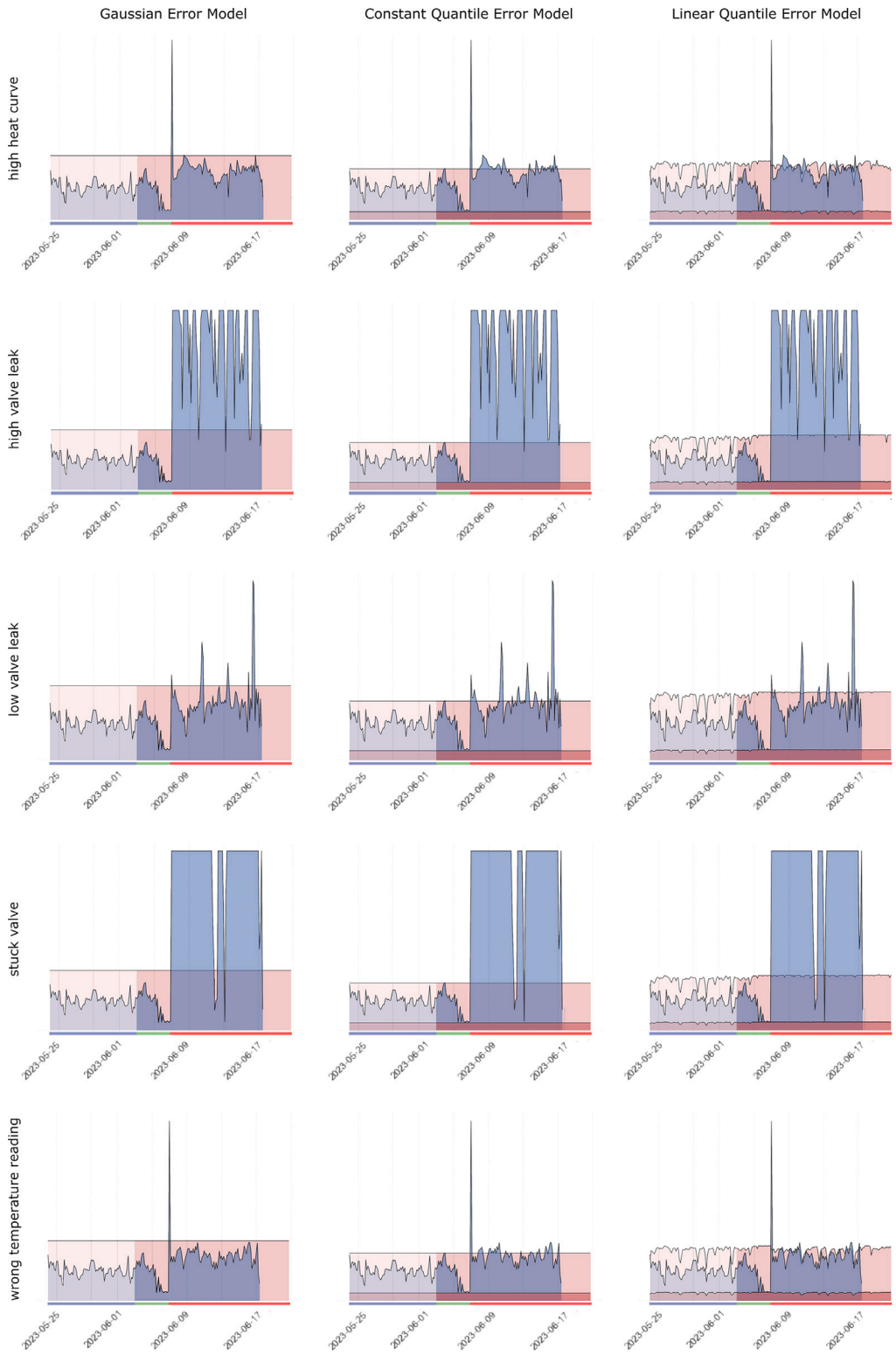
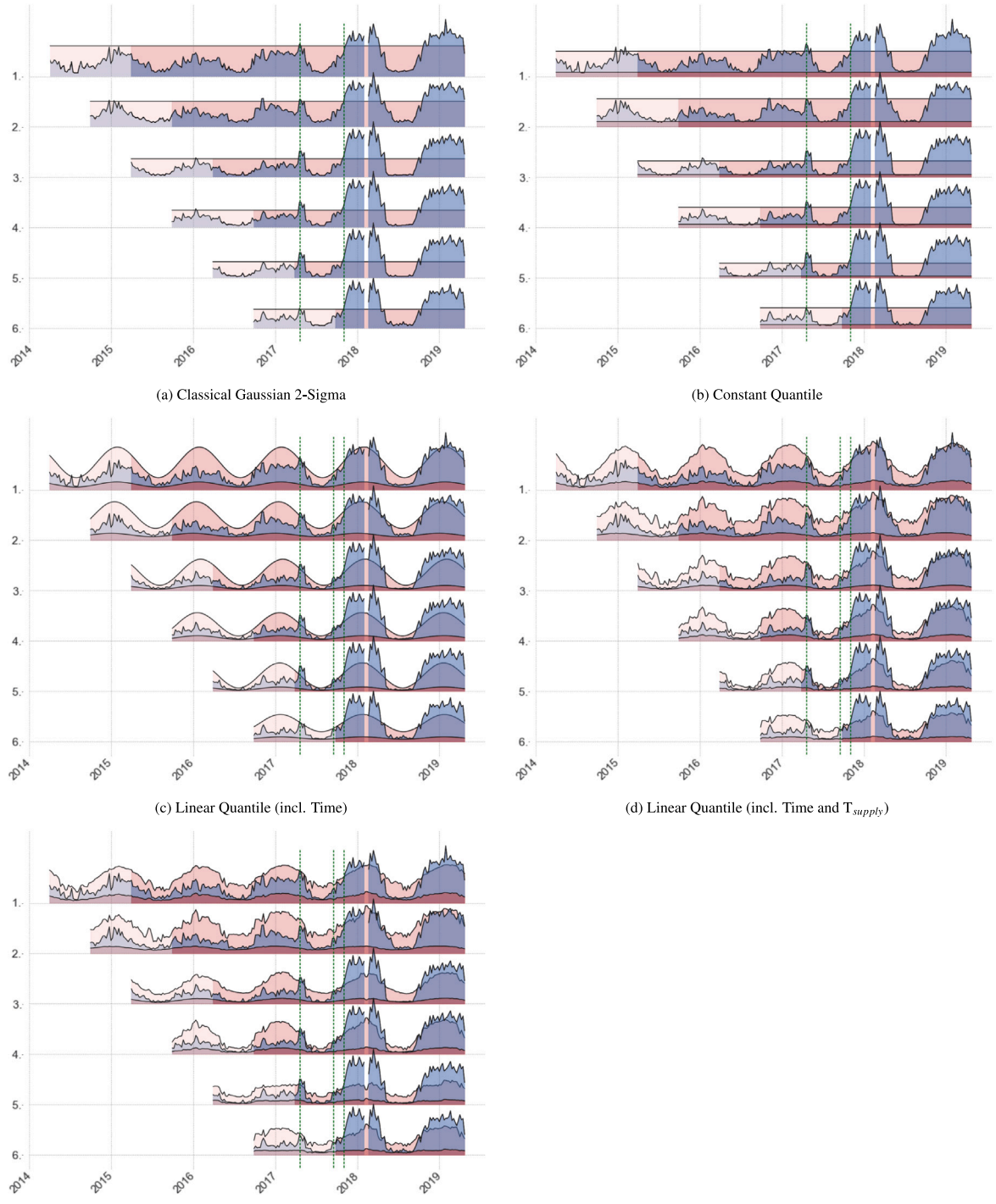


Fig. 5. Evaluating the lab scenario, with different types of labelled faults. If the blue curve (measured error) surpasses the light red (top) curve (95th percentile threshold), we suspect a fault has occurred. Each row corresponds to a different fault type, each column to a error model type. Each plot is underlined with colour information corresponding to Fig. 4, using the same colour legend: blue for the training period, green for the testing period without faults, red for the testing period with a fault present. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1. Presentation of the results

In both cases we will perform the inspection visually, using the results depicted in Figs. 5 and 6 respectively. As in Fig. 1, we have reduced the resolution of the visualization to a single point per week to improve readability. The blue curve represents the RMSE of the volumetric flow rate over a fixed period of time. Because each point constitutes an aggregated error over multiple positive values, we expect these values to lie between the 0.50 and 0.95 quantile during normal operation for the majority of the time, as the RMSE tends to be higher



(a) Classical Gaussian 2-Sigma

(b) Constant Quantile

(c) Linear Quantile (incl. Time)

(d) Linear Quantile (incl. Time and T_{supply})

(e) Non-linear Quantile (incl. Time and T_{supply})

Fig. 6. Evaluating the live scenario, with different types of error models. If the blue curve (measured error) surpasses the light red (top) curve (95th percentile threshold), we suspect a fault has occurred. See Section 3.1 for more information on how to interpret the results. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

than the mean, but dampens the impact of a single surpassed threshold. The resolution of this aggregation naturally depends on the use case, ranging from a week for a historical analysis, to maybe one to five minutes in a real-time live system.

Aside from the blue curve, two more are drawn in each plot: a lighter and a darker shaded red curve in front and behind the blue data curve, depicting the computed 0.95 and 0.50 quantile respectively. As such, the lighter curve corresponds to the classical 2σ threshold in an implicit Gaussian error model. Additionally, the training period is depicted by a lighter shade of each colour at the beginning of the sequence. In case we made use of a model ensemble, all models are shown in the order they were trained.

It should be noted that none of the figures have an explicit value on the y-axis, as the actual values were not important to the visualization. The only thing that matters is to record if the data crosses the 0.95 quantile threshold. Any labels on the y-axis refer to the model numbers, not any actual flow rate values.

3.2. Lab generated data

For our first experiment, illustrated in Fig. 5, we performed controlled experiments in a lab environment. We emulated the operation of a heating system in a house connected to a DH system via a substation. Radiators were placed in a climate chamber to emulate the heat emission and heat demand.

The initial two weeks consisted of recording baseline data, without any faults. For the next ten weeks, every two weeks the substation was exposed to a different kind of fault, providing us with labelled data. The faults emulated in the lab were:

- *Minor valve leak:* the primary flow control valve not fully closed at expected closed position, leading to minor primary water flow when there is no heat demand.
- *Valve leak:* the primary flow control valve not fully closed at expected closed position, but at a larger opening than above.
- *Stuck valve:* the primary flow control valve opening fixed at 50% when there is heat demand.
- *High heat curve:* the heat curve set for the controller of the substation is at a higher value than needed.
- *Temperature sensor deviation:* a bias in the temperature sensor readings used by the substation's controller measuring the secondary supply temperature.

For each of the time series associated with a type of fault, we added the baseline data up front for model training, providing us with a total of five sequences of four weeks. The data was recorded at a 10 s resolution, but all models were trained at a one minute resolution to allow for the capture of sufficient dynamics using only a limited amount of lags, retaining a small model size and the associated reduced computational cost. For visualization purposes, the RMSE between predicted and measured volumetric flow is shown at a resolution of an hour per point. Values of the RMSE have been capped at three times 2σ for improved readability.

For this simulation we did not make use of the ensemble of models, as there was no seasonality present due to a lack of seasonal heat demand from actual real human occupants. As such, each fault is evaluated on a single common training set and their own test set, with the purpose of verifying how the new error models compare to traditional sigma-based approaches. For each fault, we are comparing three error models:

- (left column) For reference, we also include the *Gaussian error model*, where we compute the quantile threshold corresponding to 2σ ($\tau = 0.95$).
- (centre column) The simplest *(constant) quantile error model* does not use any conditional input variables $\mathbf{x}'$. The quantile is determined directly based on the values of the error ϵ. This model is close to the implicit Gaussian error model, as the threshold is constant, but is based on the training data distribution itself, rather than the distribution corresponding to a Gaussian model.
- (right column) The next model adds additional information in the form of the *supply temperature* to the input variables $\mathbf{x}'$.

Analysing the results from this experiment shows that in this experiment, for each of the faults a classical two-sigma approach would have sufficed to capture the induced faults. In case of the low valve leak however, if the size of the leak would have been slightly lower, the 2σ approach would have failed to detect the initial moment the fault occurred. In contrast, the quantile regressors with their tighter bounds would have had better odds to catch it. Additionally, we notice that the quantile model conditioned on the supply temperature is also able to capture some of the expected dips in the predicted error.

3.3. Live data

In the second experiment, illustrated in Fig. 6, we analyse five years of data from a substation of a Swedish district heating operator. The goal is to determine if a fault occurred in that period and at which point in time they should have sent a maintenance engineer. All data was captured at a one hour resolution, and all models were trained at this resolution. For visualization purposes, the RMSE between predicted and measured volumetric flow is shown at a resolution of a single week per point. The simulation was performed using an ensemble of models, where every half year a new model was trained on the last year of data. We only show the first six models of the ensemble, because after the last model we can establish with a high degree of confidence that a fault has occurred.

We will compare a variety of error models, to illustrate their qualitative differences. The RMSE for each six models (depicted by the blue areas) is identical in every example, only the chosen error model differs:

- Once again, we also include the *standard Gaussian error model* (Fig. 6(a)), computing the quantile threshold corresponding to 2σ ($\tau = 0.95$).
- The *constant quantile error model* (Fig. 6(b)) without any conditional input variables $\mathbf{x}'$.
- By *introducing temporal information* such as the time of day and the day of the year to the input variables $\mathbf{x}'$, the linear quantile model (Fig. 6(c)) is able to capture seasonality, while retaining a very simple wave form.
- The next model (Fig. 6(d)) includes the *supply temperature* into the input variables $\mathbf{x}'$.
- The final model transforms the set of input variables $\mathbf{x}'$ from the previous model through Gaussian basis functions, introducing a *non-linear quantile regression error model* (Fig. 6(e)).

The classical 3σ approach (not depicted on the Figure) does not detect any anomalies. If we reduce the threshold to 2σ, a first anomaly is detected in April 2017 (Fig. 6(a), first green dashed line), where the measurements briefly exceed the threshold for a single week. If we extend the classical approach by going from a single model to an ensemble of models, the detected peak becomes more pronounced. By the end of that same year, the threshold is clearly exceeded even for the single-model classical approach (second green dashed line). Switching to a constant quantile model (Fig. 6(b)), we can see that the 0.50 and 0.95 quantile thresholds are tightened around the training data. This makes the flagging of the initial anomaly in April 2017 more pronounced. Additionally, if the quantile model is made conditional on temporal information (Fig. 6(c)), we are able to capture the seasonal pattern in the errors, introducing new fault warnings. When the first signs of a fault occurred in spring 2017, the classical and constant error models were still able to catch this fault. However, due to the wave form, the new quantile model is able to catch fault indications in the beginning of autumn 2017 (submodels 3–6, centre green dashed line) that the Gaussian and constant quantile model have missed. Adding an additional layer of information in the form of the supply temperature to the input variables, allows for an even more nuanced picture (Fig. 6(d)). At several points in the simulation, we can see that the uptake in prediction error was correctly anticipated by the quantile error model. Finally, in this particular instance the introduction non-linearities to the error model does not add much to the detection of faults. The main difference is a lowered 0.95 threshold some of the later submodels, making the faults more pronounced.

For the example above, based on a trade-off between model complexity and added value to what faults the model can capture, the most appropriate model would have been the linear quantile model conditioned on temporal features and supply temperature. However, this is of course not the general rule, and is dependent on the substation. In practice, model selection should always be performed using hyperparameter optimization based on pinball loss. In this set of experiments,

the error model type was fixed, even though hyperparameters such as the number of Gaussian basis functions were still optimized for with BO.

4. Conclusion

In this paper, we have taken a closer look at a common implicit assumption that is made by fault detection approaches for district heating substations, more specifically that the observed errors follow a Gaussian distribution, and are unbiased and homoscedastic. We have illustrated that this assumption does not hold up in practice, and made several proposals to improve robustness of the classical fault detection methods based on anomaly detection:

- The implicit Gaussian error model at the heart of these methods should by replaced by a quantile regression model. Such a model provides a better representation of the actual error distribution, and can be conditioned on a set of input variables containing additional information.
- We advocate the use of an ensemble of error models, trained on data from different periods of time. This increases the general robustness of the system through the ability to capture different aspects of the data, and the detection of sub-par baseline or error submodels, and allows for a differentiation between direct faults and faults based on drift/aging.

Based on our experiments in the lab, we have shown that the proposed method works. Due to the nature of the induced faults however, the lab experiments did not highlight the advantages of the proposed changes over the traditional approach in this context, and in the future we will be expanding our tests to include more gradual faults to illustrate the utility of the ensemble, and more subtle faults where the classical approach fails. The verification of our method on live data however has shown already that the underlying seasonality in the error is clearly captured by the newly proposed error model. In addition, we have shown how quantile regression can be extended to capture non-linearity in the underlying data distribution through the application of basis functions to the conditional inputs.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

We would like to thank Jonne van Dreven and Nico Robeyn for their support in performing the lab experiments used in this work.

References

[1] Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions: Strategy on heating and cooling. Eur Commission COM 2016;51.
[2] Frederiksen S, Werner S. District heating and cooling. Professional Publishing Svc.; 2013.
[3] Gadd H, Werner S. Fault detection in district heating substations. Appl Energy 2015;157:51–9. http://dx.doi.org/10.1016/j.apenergy.2015.07.061.
[4] Månsson S, Johansson Kallioniemi P-O, Thern M, Van Oevelen T, Sernhed K. Faults in district heating customer installations and ways to approach them: Experiences from Swedish utilities. Energy 2019;180:163–74. http://dx.doi.org/10.1016/j.energy.2019.04.220, URL https://www.sciencedirect.com/science/article/pii/S0360544219308606.
[5] Lygnerud K, Werner S. Annex TS2 implementation of low-temperature district heating systems. 2021.
[6] David A, Mathiesen BV, Averfalk H, Werner S, Lund H. Heat roadmap europe: Large-scale electric heat pumps in district heating systems. Energies 2017;10(4). http://dx.doi.org/10.3390/en10040578, URL https://www.mdpi.com/1996-1073/10/4/578.
[7] Mathiesen B, Lund H, Connolly D, Wenzel H, Østergaard P, Möller B, et al. Smart energy systems for coherent 100 solutions. Appl Energy 2015;145:139–54.
[8] Lund H, Østergaard PA, Chang M, Werner S, Svendsen S, Sorknæs P, et al. The status of 4th generation district heating: Research and results. Energy 2018;164:147–59. http://dx.doi.org/10.1016/j.energy.2018.08.206.
[9] Lund H, Thorsen JE, Jensen SS, Madsen FP. Fourth-Generation District Heating and Motivation Tariffs. ASME Open J Eng 2022;1:011002. http://dx.doi.org/10.1115/1.4053420, arXiv:https://asmedigitalcollection.asme.org/openengineering/article-pdf/doi/10.1115/1.4053420/6853903/aoje_1_011002.pdf.
[10] Leoni P, Geyer R, Schmidt R-R. Developing innovative business models for reducing return temperatures in district heating systems: Approach and first results. Energy 2020;195:116963. http://dx.doi.org/10.1016/j.energy.2020.116963, URL https://www.sciencedirect.com/science/article/pii/S0360544220300700.
[11] Månsson S, Davidsson K, Lauenburg P, Thern M. Automated statistical methods for fault detection in district heating customer installations. Energies 2019;12(11):113. http://dx.doi.org/10.3390/en12010113.
[12] Honoré K. The age of digitalization and flexibility - from consumer to flexumer in the district heating system. In: 9th International conference on smart energy systems ; conference date: 12-09-2023 through 13-09-2023. Copenhagen; 2023.
[13] Proceedings of the IEA DHC annex TS4 conference, Novemer 20-21,2023. Germany: Fraunhofer Institute for Energy Economics and Energy System Technology; 2023.
[14] Månsson S, Johansson Kallioniemi P-O, Thern M, Van Oevelen T, Sernhed K. Faults in district heating customer installations and ways to approach them: Experiences from Swedish utilities. Energy 2019;180:163–74. http://dx.doi.org/10.1016/j.energy.2019.04.220.
[15] Neumayer M, Stecher D, Grimm S, Maier A, Bücker D, Schmidt J. Fault and anomaly detection in district heating substations: A survey on methodology and data sets. Energy 2023;276:127569. http://dx.doi.org/10.1016/j.energy.2023.127569.
[16] Verordnung über di eVerbrauchserfassung und Abrechnung bei der Versorgung mit Fernwärme ode rFernkälte (FFVAV)/Ordinance on metering and billing for the supply of district heating or district cooling, URL https://www.gesetze-im-internet.de/ffvav/BJNR459110021.html.
[17] van Dreven J, Boeva V, Abghari S, Grahn H, Al Koussa J, Motoasca E. Intelligent approaches to fault detection and diagnosis in district heating: Current trends, challenges, and opportunities. Electronics 2023;12(6). http://dx.doi.org/10.3390/electronics12061448, URL https://www.mdpi.com/2079-9292/12/6/1448.
[18] Sandin F, Gustafsson J, Delsing J, Eklund R. Basic methods for automated fault detection and energy data validation in existing district heating systems. International symposium on district heating and cooling: 03/09/2012-04/09/2012. District Energy Development Center; 2012.
[19] Yliniemi K, Van Deventer J, Delsing J. Sensor fault detection in a district heating substation. In: IMEKO TC10 international conference on technical diagnostics: 09/06/2005-10/06/2005. 2005.
[20] Zhang F, Fleyeh H. Anomaly detection of heat energy usage in district heating substations using lstm based variational autoencoder combined with physical model. In: 2020 15th IEEE conference on industrial electronics and applications. IEEE; 2020, p. 153–8.
[21] Calikus E, Nowaczyk S, Sant'Anna A, Byttner S. Ranking abnormal substations by power signature dispersion. Energy Procedia 2018;149:345–53.
[22] Uribe J, Guillen M. Quantile regression for cross-sectional and time series data: Applications in energy markets using R. 2020, http://dx.doi.org/10.1007/978-3-030-44504-1.
[23] Le Cook B, Manning WG. Thinking beyond the mean: a practical guide for using quantile regression methods for health services research. Shanghai Arch Psychiatr 2013;25(1):55.
[24] Wasko C, Sharma A. Quantile regression for investigating scaling of extreme precipitation with temperature. Water Resour Res 2014;50(4):3608–14.
[25] Xu Q, Fan Z, Jia W, Jiang C. Quantile regression neural network-based fault detection scheme for wind turbines with application to monitoring a bearing. Wind Energy 2019.
[26] González S, García S, Del Ser J, Rokach L, Herrera F. A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. Inf Fusion 2020;64:205–37.
[27] Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. Adv Neural Inf Process Syst 2011;24.