# A Kullback-Leiber Divergence Filter for Anomaly Detection in Non-Destructive Pipeline Inspection

by

Ruikun Zhou

Thesis submitted to the University of Ottawa
In partial fulfillment of the requirements
For the M.A.Sc. degree in
Mechanical Engineering

Department of Mechanical Engineering
Faculty of Engineering
University of Ottawa

**Abstract**

Anomaly detection generally refers to algorithmic procedures aimed at identifying relatively rare events in data sets that differ substantially from the majority of the data set to which they belong. In the context of data series generated by sensors mounted on mobile devices for non-destructive inspection and monitoring, anomalies typically identify defects to be detected, therefore defining the main task of this class of devices. In this case, a useful way of operationally defining anomalies is to look at their information content with respect to the background data, which is typically noisy and therefore easily masking the relevant events if unfiltered.

In this thesis, a Kullback-Leibler (KL) Divergence filter is proposed to detect signals with relatively high information content, namely anomalies, within data series. The data is generated by using the model of a broad class of proximity sensors that apply to devices commonly used in engineering practice. This includes, for example, sensory devices mounted on mobile robotic devices for the non-destructive inspection of hazardous or other environments that may not be accessible to humans for direct inspection. The raw sensory data generated by this class of sensors is often challenging to analyze due to the prevalence of noise over the signal content that reveals the presence of relevant features, as for example damage in gas pipelines. The proposed filter is built to detect the difference of information content between the data series collected by the sensor and a baseline data series, with the advantage of not requiring the design of a threshold. Moreover, differing from the traditional filters which need the prior knowledge or distribution assumptions about the data, this KL Divergence filter is model free and suitable for all kinds of raw sensory data. Of course, it is also compatible with classical signal distribution assumptions, such as Gaussian approximation, for instance. Also, the robustness and sensitivity of the KL Divergence filter are discussed under different scenarios with various signal to noise ratios of data generated by a simulator reproducing very realistic scenarios and based on models of real sensors provided by manufacturers or widely accepted in the literature.

**Acknowledgements**

First of all, I would like to express my sincere gratitude and deep appreciation to my supervisors, Dr. Wail Gueaieb and Dr. Davide Spinello, for their guidance and timely advice. This work cannot be done without their support, dedication and inspiration.

I want to thank all the members in my lab, especially Sindhu Radhakrishnan and Mohammed Abouheaf. Thanks for giving me valuable suggestions and keeping me accompanied when pursuing the Master's degree.

I am very thankful to all my friends in Ottawa, especially my roommate, Jiazheng Han. I could not enjoy my life here when I was working on this thesis without all of you.

## Dedication

This is dedicated to my mother.

Thanks for your understanding and supporting me to pursue my dreams all the time.

# Table of Contents

# List of Tables

# List of Figures

# Nomenclature

## Abbreviations

| | |
|---|---|
| KL | Kullback-Leibler |
| KLD | Kullback-Leibler Divergence |
| NDT | Non-Destructive Testing |
| PIG | Pipeline Inspection Gauge |

# Chapter 1

# Introduction

Pipelines have been widely used for transportation of oil and gas around the world. Regular inspection to assure the good condition of the pipes is the first priority in pipeline maintenance [68]. Generally speaking, there are two different methods of pipeline inspection: external inspection and internal inspection [10]. Since normally the defects start from the internal surface of the pipe, inside pipeline testing is the most common way to conduct the inspection. Meanwhile, it provides the direct information of the anomalies. In the practice of in-pipe inspection, there are two important techniques developed by scholars and engineers. The first is known as the non-destructive testing (NDT), which is a group of inspecting techniques to examine the material without impairing the intended use of the detected object without causing damages [12]. The second is often referred to as the pipeline inspection gauge (PIG). PIG is the in-line inspection tool, which is able to work in the running system without interrupting the flow in the pipe and makes the NDT possible [44]. The NDT technologies have been developing rapidly since they were introduced in 1950s. Nowadays, there are several mature NDT methods, such as Visual Testing, Ultrasonic Testing, Thermal Infrared Testing, etc. Among all the methods, the electromagnetic-based techniques, including Eddy Current Testing and Magnetic Particle Testing, have developed as the main pipeline inspection approaches. This is thanks to their efficiency, convenience, and relatively low-cost [40].

Recently, with the development of modern plastic pipeline industry, plastic pipes, such as, polyvinyl chloride (PVC), exhibit several advantages, compared to traditional metal pipes. They are corrosion- and welding-free, lighter wight, easier to manipulate and connect, and have higher energy conservation and safety ratings [84]. Many countries are replacing the metal pipe infrastructure by PVC pipes or other kinds of plastic pipes. Under this situation, many electromagnetic-based NDT approaches are no longer applicable. As a result, it has become necessary to devise broader methods to conduct in-pipe testing. Considering defects on the pipelines are the changes in wall thickness or volume of the materials in a certain area or spatial position, distance sensors (also known as proximity sensors) have emerged as a better alternative. To accommodate such type of sensors, the data-processing algorithms must be altered. They are required to be more advanced and accurate because the one dimensional distance values provide less information about the defective parts in the pipe. What's more, in practice, there may be several anomalies with

different attributes (size, depth, severity, etc.) in a functional pipe. Some of them may be severe, whereas others may be incipient or insignificant such that no repair action is needed. As such, in addition to detecting anomalies, the ability to assess their attributes is also important to assure the proper course and timely action is taken by the maintenance crew.

## 1.1 Motivation

Considering the anomalies are regarded as the rare events in the probabilistic sense when processing data, the raw data from the proximity sensors can be processed with statistical and information-theoretic analysis to reveal the information of abnormal parts. For instance, if we do the cluster analysis for the whole raw dataset, the samples of anomalies may be partitioned into small populations or located in a sparse area outside the main clusters, with the comparison of the clusters of the normal portions. Or, the sample points may be recognized as rare events with low probabilities if a histogram technique used [81]. By such statistical processing or information measure analysis with high sensitivity, the NDT testing with proximity sensors is able to reveal the features of abnormal data points, which makes such a method quite appealing for the inspection of both plastic and metal pipes.

Moreover, different anomalies with different types and sizes have different analytic distributions and characteristics. The sizes of anomalies can be assessed from data processing [20]. Meanwhile, since the information-theoretic measures do not rely on the prior knowledge of the dataset, differing from the traditional signal processing using electromagnetic sensors, the distribution of the sample points can be built accordingly without prior assumptions. More importantly, the distributions of the anomaly database do not usually follow specific known distributions. For example, we know that a constant distance signal with Gaussian noise can approximate the Gaussian distribution. However, the distance values of holes on a wall with arbitrary shapes do not form a normal distribution or any other known distributions. In this regard, the information-theoretic measures are good candidates for data-processing with NDT.

Among the information-theoretic measures, the Kullback-Leiber (KL) Divergence, a member of $f - divergence$, has been well-known for its utility in anomaly detection in information theory and communication [80]. The anomaly detection in pipeline inspection may be modelled similarly, which motivates the adoption of this measure in this domain. In this context, the simplest form of anomaly detection can be reduced to distinguishing the abnormal from the normal distribution, while KL divergence is applied to measures the difference, or the "distance", between two distributions. Signals without anomalies are close to each other from a KL Divergence point of view. Nevertheless, the abnormal ones are distant from the normal clusters. If we choose the dataset from the good parts as the baseline, the KL Divergence of all parts can be calculated. As such, the anomalies can be identified, such that the further an anomaly is from the baseline, the "more abnormal" it is. This concept will be more formally developed later in the thesis.

## 1.2 Objectives and Contributions

To tackle the aforementioned problems, a Kullback-Leiber Divergence (KLD) filter is proposed herein for anomaly detection in data series generated by a broad class of proximity sensors, including the ones used in some NDT devices employed in the pipe inspection industry. In this work, the sensory data is generated by high-fidelity models of real sensors, often provided by the sensors' manufacturers, and therefore providing highly realistic testing scenarios for the filtering algorithm. The KLD filter maps the sensory data into KLD values measuring the difference between the information content associated to the sensor signal and the information content associated to a baseline noisy signal representative of the sensor, identifying anomaly as a significant divergence between the two information content. If we take the inline pipeline inspection as an example, this work tackles the in-pipe anomaly detection problem with a KLD filter to find the locations of anomalies and their sizes with a sensing robotic platform. Meanwhile, to avoid the invalidation of detection in non-metallic pipes with electromagnetic sensors, we collect the data with general proximity sensors and process it with a KLD filter. The design and development of the filter with the following properties represent the thesis' main contributions.

**Model-free filter**: The KLD based algorithm proposed here is not predicated on assumptions about the distribution of the data, and therefore it can be implemented in a model free framework. Namely, it can be adapted to include distribution models whenever they are reliably available for a specific system or sensor. Previously, KLD based tests have shown its effectiveness as a promising alternative to more traditional statistical tests for incipient fault detection [36]. Since the KLD intrinsically measures the discrepancy in the information content between two data sets, it is very robust in anomaly detection as it requires neither a priori knowledge about the nature of the anomaly nor an accurate statistical characterization [6, 74]. As a matter of fact, this is not the first attempt of applying the KLD to model free fault detection. It was tested in [82] with different simulated scenarios, where it proved to be quite effective in detecting the location of cracks from an eddy current sensor data. However, to the best of the author's knowledge, this is the first time a KLD filter is customized and applied as a model-free approach in the domain of pipeline anomaly detection.

**Size Recognition**: Unlike other approaches reported in the literature, the proposed algorithm is not only able to find the locations of the anomalies, but it is capable of identifying their sizes as well. To maximize the size accuracy, the whole volume (depth and surface area) of the defect is taken into consideration. The designer has the flexibility to assign different preference weights to each of the depth and surface area, depending on the signal to noise ratio, for example. Many researchers proposed numerous methods to distinguish abnormal from normal data series. For example, the authors in [82] were successful in identifying and locating cracks by applying KLD analysis on eddy current sensor data. However, the severity and the sizes of the anomalies were not identified.

**More accurate and sensitive filter**: Among the sate-of-the-art techniques in the field is the entropy filter proposed in [71], which was developed in collaboration with a Toronto-based pipe inspection company. The entropy filter therein considers eddy current sensory data series spatially sampling a cylindrical surface of a pipeline. With a similar

mapping process to the entropy filter, the KLD filtering algorithm presented in this thesis is built by mapping each raw data point into a local KLD value calculated on a subset of the overall data centered around the data point. A comparison study between the two filters showed that the KLD filter is more accurate and sensitive than the entropy filter and possesses more salient features, such as the ability to operate within a model free context and determine the size of the defect.

## 1.3 Thesis Outline

The structure of the thesis is organized as follows. In Chapter 2, the review of anomaly detection and the information-theoretic measures for anomaly detection are presented and discussed. The chapter also emphatically exhibits how the entropy and KLD metrics are applied in anomaly detection. To help to address this specific problem, the non-destructive pipeline inspection and pipeline inspection gauge technologies are also reviewed. Chapter 3 presents the preliminaries on entropy and KLD, and how the model free KLD filter algorithm is designed based on the theoretical concepts by building the histograms and generating the probability mass functions. Moreover, as one of the special cases of the model-free algorithm where the majority of the data can be modelled as Gaussian distribution, the KLD filter with Gaussian approximation is also proposed and shows the effectiveness for the pipeline anomaly detection scenario. Next, a case study of pipeline inspection with a class of proximity sensor models is presented in Chapter 4. In Chapter 5, the application of two different KLD filters on the case is shown and the influence of the main parameters of the KLD filter is studied with insights about the relationship between the filter performance and the signal to noise ratio, as well as the advantages of the KLD filter with respect to the entropy filter. Conclusions and future work are presented in Chapter 6.

# Chapter 2

# Literature Review

This chapter contains a brief review of the literature relevant to anomaly detection, information-theoretic measures and pipeline anomaly detection.

In section 2.1, the definition of anomaly detection as well as various technologies in the field are presented, with a focus on six commonly used techniques. We also explain how to select the efficient and effective technique according to the attributes of the testing data set based on the characteristics of different anomaly detection methods. In addition, two kinds of industrial anomaly detection scenarios are introduced.

The implementation of information-theoretic measures in anomaly detection is discussed in section 2.2. The section also addresses the relationships between the different measures. Special attention is given to the applications of two of the most commonly used information-theoretic measures, namely the entropy and the KLD metrics.

In section 2.3, the history and development of NDT and PIG methods are introduced, while focusing on these two technologies in pipeline inspection. In the last part of the section, the traditional electromagnetic methods for pipeline structural anomaly detection are presented, followed by some more recent techniques, including machine learning and image processing.

## 2.1 Anomaly Detection

Anomaly detection is very important in diverse research areas and applications in the real world. If the defects can not be found accurately and in time, significant problems might occur which might lead to devastated failures and losses. Generally speaking, anomaly detection refers to the recognition of the non-conforming behaviours in a system. Anomalies are also called outliers in data science and pattern recognition.

### 2.1.1 Anomaly Detection Techniques

Over the past half century, various anomaly detection techniques have been developed. Some of them are generic and can be used in all the subdomains of anomaly detection,

such as the information-theoretic measures derived from the statistical and information content perspectives. However, other technologies may only serve specific applications, such as the data processing methods for electromagnetic sensors, for instance, which are heavily dependent on the characteristic properties of the sensors.

As for the generic approaches, Chandola et al. [20] provide an extensive survey of anomaly detection techniques for various application domains, including classification based, nearest neighbor based, clustering based, statistical, and information theoretic methods, among others. The comparison and discussion of the first four techniques are made in [72]. The main six different kinds of anomaly detection techniques are summarized in Table 2.1. The output the of these techniques can be classified into two kinds:

**Labels:** This is when anomaly detection algorithms serve as a classifiers by assigning "normal" or "anomalous" labels to the data instances or signals. They usually take the form of threshold-based binary hypothesis tests.

**Scores:** This kind of techniques is usually based on a defined measurement to output a score that is related to the characters of each test instance. It is the theoretical fundamental of multi-class classifiers and multinary hypothesis testing.

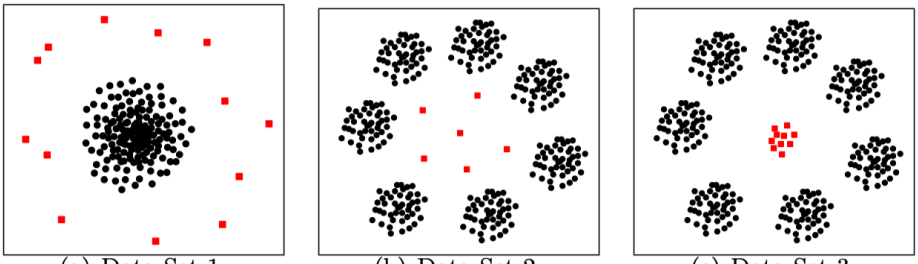
From the summary table, we can easily draw a conclusion that the anomaly detection techniques must be chosen according to the nature of the dataset to assure that the assumptions are satisfied. Otherwise, the anomalies may not be successfully detected. For instance, three different 2-D data sets are shown in Figure 2.1, where black circles stand for normal data instances and red squares represent anomalies. In the first dataset, the normal data are in a tight cluster, while the anomalous data instances are scattered around the clustering circle, but far from the center. Based on the distribution, all the assumptions of the techniques listed in Table 2.1 are met and so are applicable in this scenario. Differing from the first dataset, there are six tight clusters for the normal instances in the second dataset, instead of one, where the abnormal data instances are located among these clusters sparsely. Most techniques, such as the clustering based and the statistical methods, work perfectly for this case as well. The anomalous data instances in the third dataset are located at the center with a dense neighborhood. As a result, the clustering based methods would perform very poorly in detect the anomalies in this case. Therefore, other approaches ought to be adopted in such scenarios. In most pipeline inspection cases, the data are distributed as in the first case.

### 2.1.2 Industrial Anomaly Detection

As stated in the previous section, there are various application scenarios that one may encounter. Pipeline anomaly detection belongs to the "industrial damage detection" category, which refers to industrial anomaly detection, at large. This encompasses fault detection in the performance of machines or systems, and detecting structural defects in objects. Normally, these anomalies are caused by continuous usage, unexpected forces or pressures, and wrong operations. In practice, engineers use extensive sensors to collect the relevant

Table 2.1: Summary of six commonly used anomaly detection techniques (adopted from [20]).

| | | |
|---|---|---|
| Classification based | Description | The classifier learns from some training data with different classes and then classify the testing data into one of the classes. |
| | Assumptions /Limitations | The normal and anomalous parts can be distinguished according to their features by the classifier. |
| | Examples | Replicator Neural Network [38], Support Vector Machines [25], Deep Learning based [19] |
| Nearest neighbor based | Description | The sample points which occur far from their closet neighbors or are located in sparse neighborhood are identified as anomalies. |
| | Assumptions /Limitations | Anomalous data are located far away from their closet neighbors, while the normal ones are in a dense neighborhood. |
| | Examples | Local Outlier Factor [16], Peer Group Analysis [15] |
| Clustering based | Description | The similar data instances are grouped into clusters, the anomalous instances have different features in the cluster or do not belong to any cluster. |
| | Assumptions /Limitations | The corresponding features of anomalies can be revealed after clustering. |
| | Examples | Cluster-Based Local Outlier Factor [39], SNN clustering [30], Self-Organizing Maps (SOM) [46] based |
| Statistical | Description | Anomalies in stochastic model are regarded as rare events that occur in the low probability regions among the dataset of which the large majority is normal, and they can be recognized by building some statistical models. |
| | Assumptions /Limitations | Normal data instances are the events with high probabilities in a stochastic process, but the abnormal ones occur rarely with relatively low probabilities. |
| | Examples | Gaussian Model based-Maximum normed residual test [33], Histogram based [41, 55–57] |
| Information Theoretic | Description | The information content of the dataset is analysed with information-theoretic measures, such as, Kolmogorov complexity, entropy and KL Divergence, and the anomalies are recognized by comparing their numeric value with the majority. |
| | Assumptions /Limitations | The anomalous data instances cause irregularities in the contextual information content. |
| | Examples | Kolmogorov complexity [9], Entropy [73] |
| Spectral | Description | The significant differences in the data are captured by approximation within the subspace of the dataset, where the anomalous data instances are extinguished. |
| | Assumptions /Limitations | The dataset can be mapped into lower dimension space, in which the anomalous data instances show large difference from the normal ones. |
| | Examples | Principal Component Analysis (PCA) [29] |



(a) Data Set 1 (b) Data Set 2 (c) Data Set 3

Figure 2.1: Three 2-D data sets with different features (red squares and black dots represent anomalous and normal data instances, respectively).

data first before analyzing it using the aforementioned anomaly detection techniques to locate the anomalies and prevent further losses [20].

### Fault Detection in Condition Monitoring

In the field of condition monitoring and diagnostic engineering, the sensors monitor the performance of the machinery or one of the components of the industrial systems, such as the gas flow in a transmission pipe, for instance. The anomalies are revealed by analysing the observations while preventive measures are applied afterwards.

Extensive anomaly detection techniques have been implemented in this domain over the past decades. Davies et al. [24] reviewed the commonly used monitoring and inspection approaches for machinery monitoring, including Visual Inspection, Vibration Monitoring and Wear Debris Analysis. Recently, several up-to-date technologies have also been used in this domain. An Artificial Neural Network classifier is proposed for fault diagnosis in the condition monitoring of rolling element bearing in [79]. A model-free approach using one of spectral anomaly detection methods, the Kernel Feature Space, has been proposed to detect anomalies in the spacecraft industry [32]. It is applied on a time-series data based on the hypothesis that the anomalous instances show unexpected changes.

### Structural Defect Detection

Structural defects and damages occur due to the absence of components or some functional parts of the machinery, such as cracks in pipelines, for instance. In most cases, all the six anomaly detection techniques mentioned in Table 2.1 proved to be effective in finding the defects in this domain. The data sets for structural inspection are usually able to be processed off-line and have spatial correlations according to the detected objects.

Taking neural network based technique in as an example of the general approach in Table 2.1, which can be used in this kind of anomaly detection problems, Angeli et al. [7] proposed a Probabilistic Neural Network (PNN) based method to detect the anomalies in steel using remote eddy current sensors. A more recent work [42] extends the PNN idea

by integrating PNN and data fusion to get a better damage identification result for civil structures. Apart from these generic data processing techniques which suit almost all types of sensory data, there are some particular approaches for a certain category of sensory data, which can identify the defects with the raw data directly. However, they heavily depend on the characteristics of this raw data, specifically the signals derived from electromagnetic sensors. For instance, if an eddy current sensor is used to collect the signals for pipeline anomaly detection, the traditional magnitude and phase analysis is effective to identify the anomalies based on the difference between the magnitude and phase of each sample and a reference from the healthy part. Further, to strengthen the accuracy and effectiveness of the detection, image processing techniques are often implemented as a post-processing approach after a rough labelling of the defects with other types of sensors. For example, image processing techniques are coupled with magnetic flux leakage (MFL) testing in [43] and with remote field eddy current testing (RFECT) in [54] to conduct pipeline damage identification tasks, respectively, where the outlier detection techniques in image processing are implemented to locate the defects based on the figure of leakage and the phase change respectively. Of course, some image processing techniques are able to detect the anomalies by collecting the visual signals of the surface or inside part of the machinery. But, it is not an efficient way due to the high background noise in practice and the resolution limitation of vision sensors. In addition, these image processing approaches with/without traditional electromagnetic sensors are time-consuming and computationally expensive, compared to the generic and direct methods using the features of the raw electromagnetic data.

For regular pipeline anomaly detection, most aforementioned methodologies are applicable. But when we choose the best fit for a specific anomaly detection scenario, the nature/distribution of the collected data may need to be evaluated by the technician before decision making.

## 2.2 Information-theoretic Measures in Anomaly Detection

In section 2.1, we stated that information theoretic technique is one of the most important approaches in anomaly detection. Its theoretical basis is founded on information content and probability theory. Anomaly detection generally refers to algorithmic procedures aimed at identifying relatively rare events in data sets that differ substantially from the majority of the data set to which they belong. In the context of raw sensory data series, a useful way of operationally defining anomalies is to look at their information content with respect to the background data, which is typically noisy and therefore easily masking the relevant events if unfiltered. To process/filter the data and identify the anomalies, at least one information-theoretic measure is needed. In the following, we review the commonly used information-theoretic measures and how they can be implemented in anomaly detection algorithms.

### 2.2.1 Information-theoretic Measures

Generally speaking, information-theoretic measures include entropy, relative entropy (also known as KLD), conditional entropy, relative conditional entropy, information gain and information cost in the domain of anomaly detection. These measures have been extensively used in this domain. The effectiveness of the information-theoretic measures has been discussed in [80] with data sets from the University of New Mexico sendmail system call data. More recently, Callegari et al. [17] conducted a comparison study between the performances of the mainly used information-theoretic measures for anomaly identification in network traffic.

Each anomaly detection measure has its own strengths and weaknesses. The selection of the information-theoretic measure to adopt should be made according to the characteristics of the dataset. Usually, the features of the target data are studied first. Then, the corresponding measures are selected to best match the characteristics therein. In this section, we review most of these measures, with special focus on the two main approaches: the entropy and the relative entropy metrics.

### 2.2.2 Entropy Metric Application

#### Entropy

Shannon proposed the concept of information entropy in his work about a mathematical theory of communication [70]. Let $X$ be a discrete random variable mapping to a finite dimensional event space $\{x_1, x_2, \ldots, x_N\}$ of cardinality $N$. Let $P(x_i)$ be a probability mass function over the set $X$ with $\sum_{i=1}^{N} P(x_i) = 1$. The information entropy $H$, or Shannon entropy, is the expected value of the information content $-\log P(X)$ of the random variable $X$. Formally, it is defined as

$$H(X) := -\sum_{i=1}^{N} P(x_i) \log P(x_i) \tag{2.1}$$

$$:= \mathbf{E}_P[\log\left(\frac{1}{P(X)}\right)] \tag{2.2}$$

where $\mathbf{E}$ denotes the expectation. Shannon entropy is mainly used to compute the average uncertainty in a random variable. If there is an irregularity in the attributes of a certain data series, the uncertainty or diversity of the dataset increases. This concept constitutes the theoretical background of the entropy metric.

Various applications of this metric have been proposed in the past four decades to study the attributes of data and identify anomalies. In [51], the entropy metric is used to study the data dependency in a relation database. It shows the capacity of entropy to formalize the relationship in the data. Moreover, an entropy filter has been proposed in [73] to capture the anomaly in magnetic field. An extension of the entropy filter is proposed in [76] to detect anomalies/defects on the surface of gas pipelines from raw sensory data collected by

an eddy current sensor mounted on a commercial snake-like robot. This filter maps the raw data to a local information entropy space by assigning to each spatial location the entropy calculated from a neighborhood data set. A binary hypothesis testing is used to partition the space between anomaly and background, with the necessity of introducing a threshold for the hypothesis testing. The aforementioned researches failed to take advantage of the score output of the entropy metric and extend the binary output to a multi hypothesis test result.

### Rényi Entropy

Based on the concept of entropy, a more generalized information entropy, called Rényi entropy, is formulated in [67]. That is where the notation of the order of an entropy is first introduced. An entropy of order $\alpha$, also referred to as Rényi entropy, is defined as

$$H_\alpha(X) := \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{N} P(x_i)^\alpha\right) \tag{2.3}$$

$$:= \frac{1}{1-\alpha} \log\big(\mathbf{E}[P(X)^{\alpha-1}]\big) \tag{2.4}$$

where $\alpha \neq 1$. It can be shown that $H_\alpha$ converges to Shannon entrop as $\alpha \to 1$. As $\alpha \to 0$, the Rényi entropy converges to Hartley entropy, which is the logarithm of the cardinality of $X$. When $\alpha \to \infty$, Rényi entropy converges to the min-entropy $H_\infty$. The collision entropy $H_2$ is defined as $H_\alpha\big|_{\alpha=2}$. It is sometimes casually referred to as "Rényi entropy" as well.

By introducing the parameter $\alpha$ as the entropy degree, Rényi entropy metric provides a broader method for anomaly identification. Algorithms with different entropy degrees show different sensitivity to certain types of data instances. If we consider the entropy degree $\alpha$ as a variable of the Rényi entropy, the members of Rényi Entropy have different performances when calculating the entropy over the same random variable. The greater the value of $\alpha$, the greater the weight of the higher probability event account for in the sum of the entropy value, while the events with relatively lower probabilities can not be revealed. When $\alpha < 1$, the contributions of lower probability events are more reflected in the entropy. But when $\alpha$ approaches 0, the Rényi entropy tends to treat all possible events equally [47]. Statistically, anomalies can be regarded as rare events in industrial anomaly detection. In such a case, the value of $\alpha$ should be chosen between 0 and 1. Vice versa, if we want to reveal more information about the normal data instances in the raw data, $\alpha$ needs to be greater than 1. The collision entropy is a good candidate for such an example.

Based on this fact, Sheikhi et al. [71] extended their entropy filter proposed in [76] to a Rényi entropy metric. They also discussed the influence of $\alpha$ on anomaly detection in gas pipes, where the Neyman-Person criterion was applied as a decision rule mechanism to label normal and anomalous sensor samples. In a similar manner, the influence of $\alpha$ on the sensitivity of the Rényi entropy to different types of data in network monitoring is studied in [47].

### 2.2.3 KL Divergence Metric Application

#### KL Divergence

Based on the concept of entropy, the relative entropy, or Kullback-Leibler Divergence (KLD) is proposed by [49] in 1951. KLD is the expected value of the logarithmic difference (or information divergence) between two distributions $P$ and $Q$ defined over the same discrete random variable $X$ [48]. The KLD with respect to $P$ , denoted by $D_{\mathrm{KL}}(P||Q)$ is defined by

$$D_{\mathrm{KL}}(P||Q) := \sum_{i=1}^{N} P(x_i) \log\left(\frac{P(x_i)}{Q(x_i)}\right) \tag{2.5}$$

$$:= \mathbf{E}_P\left[\log\left(\frac{P(X)}{Q(X)}\right)\right] \tag{2.6}$$

The KLD is a non-negative real number. It evaluates to zero if and only if $P = Q$. It is important to notice that the KLD is not symmetric since $D_{\mathrm{KL}}(P||Q) \neq D_{\mathrm{KL}}(Q||P)$. Therefore, it has to be intended as a measure of discrepancy between the two distributions rather than a metric. Nevertheless, in most cases, it is usually informally referred to as a "distance" between two distributions or data sets. In fact, KLD is a member of a larger family of $f$-divergence measures, which are widely used distance measures in signal processing and pattern recognition.

In the field of information theory, the KLD $D_{\mathrm{KL}}(P||Q)$ is a measure of the inefficiency of a posterior probability distribution $P$ relative to a prior probability distribution $Q$. This is for instance if we know the true distribution $P$ of observations or data $X$ and the modelled distribution or the approximation of the distribution is $Q$.

#### Anomaly Detection Applications

By definition, KLD is a measure of information entropy relative to a reference set. If we regard the normal data instances as the reference in computing the KLD of some testing data set, then the KLD values corresponding to normal data would be relatively small, while the anomalous data instances would correspond to high KLD values due to the relatively large distance from the baseline data. Therefore, the KLD metric is an inviting candidate to categorize data without the need of a threshold or any prior knowledge about the distribution of the data.

A number of studies have examined the effectiveness of the KLD metric in anomaly detection. In the work of [2], KLD was proposed to capture the signal features for periodic signals generated from cognitive radio, where unexpected changes in the wireless signal were successfully identified as abnormalities. Based on this study, some more systemic and advanced methods for anomaly detection in periodic signals are reported in [1, 3, 4] with KLD measure, including the implementation of differential KLD equation for various systems. However, these studies are limited to discrete random variables and the algorithms are not applicable for continuous cases. At the same time, the differential entropy, which

is a powerful tool to derive the continuous KL divergence from the discrete, is also not been mentioned. As many signals in practice can be approximated by Gaussian distributions, the gap between the discrete and continuous cases is bridged by using Gaussian approximation for continuous random variables, such as in [83] where KLD was shown to be useful in monitoring statistics. Based on this approximation, KLD was also used to find the incipient fault in multivariate processes [37] by taking advantage of its high sensitivity to small changes in signals. Meanwhile, the fault magnitude was also estimated accurately by the relative KLD in this study. Moreover, the relationship between the performance of the KLD metric, the signal-to-noise ratio and the fault-to-noise ratio was exploited by applying the KLD to crack detection in eddy current testing [82]. However, some of these studies have neglected the issue of applicability of Gaussian approximation. Considering that the majority of anomalous data is not Gaussian, this kind of approximation may fail to detect some incipient defects in high noisy circumstances. Fortunately, some papers among them address this problem by implementing the probability density function approximation techniques to obtain the true non-Gaussian distributions or using the density ratio instead of the density function itself, such as kernel density estimation in [82] and density ratio estimation in [83]. Based on these knowledge of KL divergence, the later work of [34] addressed the fault detection in large scale systems with this metric based on non-parametric approximation with a modified version of KLD, in which the $\alpha$-relative density ratio is used, instead of the densities themselves. Recent work by [21] extended this technique to the analysis of online data with off-line references under a multivariate statistical analysis framework. Besides the anomaly detection applications in electric signals and communication contents, the application of KLD in mechanical-type fault diagnosis in bearing balls was explored in [58].

Despite this large volume of research, most of the applications of the KLD and entropy metrics only aimed at differentiating anomalous from normal data. That is, they were used to perform binary hypothesis testing. No study was conducted to demonstrate the ability of KLD in recognizing the sizes of the anomalies in a continuous manner for structural defect detection. Actually, information-theoretic measures are numeric tools, which means that the anomalies are associated with scores, rather than only labels. Taking this into consideration, of the goals of this thesis is to prove the applicability of KLD in assessing the size of the anomalies in a continuous manner for pipe inspection. Such a breakthrough would make the maintenance of pipes much more efficient.

### 2.2.4 Other Information-theoretic Measures

#### Rényi Divergence

Similar to Rényi entropy, Rényi also defined a spectrum of divergence measures, called Rényi divergence, in his work [67]. The definition is as follows.

$$D_\alpha(P||Q) = \frac{1}{1-\alpha}\log\left(\sum_{i=1}^{N}\frac{P(x_i)^\alpha}{Q(x_i)^{\alpha-1}}\right) \tag{2.7}$$

where $0 < \alpha < \infty$ and $\alpha \neq 1$. Rényi divergence is non-decreasing with respect to its order $\alpha$. As with Rényi entropy, some special cases arise when $\alpha$ takes some special values. These cases are summarized in Table 2.3, where the special values of 0, 1 and $\infty$ are obtained by taking the limits [78].

Table 2.3: Special cases of Rényi divergence.

| $\alpha$ | **Equation** | **Meaning** |
|---|---|---|
| $\alpha \to 0$ | $D_0(P\|\|Q) = -\log Q(i : P_i > 0)$ | minus the log value of Q when $P_i$ is not 0. |
| $1/2$ | $D_{1/2}(P\|\|Q) = -2\log\left(\sum_{i=1}^{N}\sqrt{P(x_i)Q(x_i)}\right)$ | minus twice of Bhattacharyya distance when natural logarithm is used. |
| $\alpha \to 1$ | $D_1(P\|\|Q) = \sum_{i=1}^{N} P(x_i)\log\frac{P(x_i)}{Q(x_i)}$ | KL Divergence |
| 2 | $D_2(P\|\|Q) = \log(1 + \mathcal{X}^2(P,Q))$ | the logarithm of 1 plus $\mathcal{X}^2$ – *divergence*, where $\mathcal{X}^2$ – *divergence* denote: $\mathcal{X}^2(P,Q) = \sum_{i=1}^{N}\frac{(P(x_i)-Q(x_i))^2}{Q(x_i)}$ |
| $\alpha \to \infty$ | $D_\infty(P\|\|Q) = \log\sup_i(\frac{P(x)}{Q(x)})$ | the logarithm of the maximum ratio of the probabilities |

Within the family of Rényi divergence, the KLD is the most commonly used divergence measure. It proved to be quite effective in many applications, as stated above. One application of Rényi divergence in unsupervised anomaly detection is presented in [66]. The broader use of this divergence family is left for future investigation.

### Conditional Entropy

Consider two discrete random variable $X$ and $Y$ with finite dimensional event spaces $\{x_1, x_2, \ldots, x_N\}$ and $\{y_1, y_2, \ldots, y_M\}$, respectively. Then, the conditional entropy of $X$ given $Y$ is defined as follows [47]:

$$H(X|Y) := -\sum_{i=1}^{N}\sum_{j=1}^{M} P(x_i, y_j)\log P(x_i|y_j) \tag{2.8}$$

$$:= \mathbf{E}[\log(P(X|Y))] \tag{2.9}$$

The conditional entropy $H(X|Y)$ measures the regularity of sequential dependencies. $H(X|Y) = 0$ if and only if the first random variable $X$ can be completely determined by the latter one $Y$. On the contrary, if these two random variables are independent, $H(X|Y) = H(X)$, which makes it equal to the Shannon entropy of $X$.

In the context of anomaly detection, if $X$ and $Y$ are two dependent random variables, then the conditional entropy measures the amount of randomness or uncertainty reduced in the random variable $X$ by observing $Y$. In other words, the mutual information between two random variables can be known after knowing $Y$, and the uncertainty of $X$ is reduced.

If we consider random variables $X$ and $Y$ from the testing and normal data instances, respectively, then the smaller the conditional entropy, the better. Namely, $X$ is similar to $Y$ if they have similar attributes. A threshold can then be set to label the anomalous data instances if the conditional entropy exceeds the threshold value. Furthermore, as conditional entropy reflects the dependencies of two random variables and the relationship between the sequential states in time series, a distributed and scalable anomaly detection algorithm is developed to provide more dynamic and reliable approach for wireless network in [8]. From another perspective, irregularity corresponds to unpredictability in the conditional entropy. We can therefore conclude that an anomaly exists if $X$ cannot be predicted by the corresponding sequences relative to baseline $Y$. This process can repeated on a case by case basis where several baselines can be made in advance. A suitable baseline is determined by the first few states in the random variable, which leads to a more reliable and flexible method. An application of this idea is presented in [62] for graph-based anomaly detection. On the other hand, if $X$ and $Y$ are from the same data set, the conditional entropy shows the regularity in this data series if they are drawn following a certain rule. The outcome in this case is analogous to Shannon entropy, but with a baseline. Taking advantage of the aforementioned salient features, a measurement on the regularity of a system using the conditional probability of the $n$th data series over the previous $(n-1)$ is presented in [80]. However, the technique has not been applied on anomaly detection.

### Conditional Relative Entropy

With the conditional probabilities in conditional entropy, a conditional version of relative entropy is defined. The conditional relative entropy of two conditional probability mass functions, $P(X|Y)$ and $Q(X|Y)$, is given by [23]

$$D_{\mathrm{KL}}(P(X|Y)||Q(X|Y)) := \sum_{i=1}^{N} P(x_i) \sum_{j=1}^{M} P(x_i|y_j) \log \frac{P(x_i|y_j)}{Q(x_i|y_j)} \tag{2.10}$$

$$:= \mathbf{E}_{P(x,y)}[\log\left(\frac{P(X|Y)}{Q(X|Y)}\right)] \tag{2.11}$$

where $X$ and $Y$ have the same definitions as in the conditional entropy. Along the same spirit of the latter, the conditional relative entropy can also be used to measure the discrepancy between two datasets. The classifying scheme in the application of anomaly detection is the same as in the conditional entropy and the relative entropy. The smaller the value, the better, in the sense that anomalies are less likely in this case. Wenke et la. [80] discussed how the conditional relative entropy can be used to reflect the difference between the training dataset and the testing dataset. Conditional relative entropy is a promising candidate for anomaly detection in complex systems and situations where the direct relationships between states cannot reveal the anomalous parts. However, the effectiveness of this metric has not been discussed in the paper.

### Mutual Information/Information Gain

Mutual information is a measurement of the amount of information that one random variable contains about another random variable. More intuitively, mutual information measures the information shared by two distributions. If we reuse the notations of $X$ and $Y$ defined above, the mutual information $I(X;Y)$ is the relative entropy between the joint distribution of $X$ and $Y$, $P(X,Y)$, and the product distribution $P(X)P(Y)$. Formally, this is defined by

$$I(X;Y) := \sum_{i=1}^{N}\sum_{j=1}^{M} P(x_i,y_j)\log\left(\frac{P(x_i,y_j)}{P(x_i)P(y_j)}\right) \tag{2.12}$$

$$:= D_{\mathrm{KL}}(P(X,Y)||P(X)P(Y)) \tag{2.13}$$

$$:= \mathbf{E}_{P_{(X,Y)}}\left[\log\left(\frac{P(X,Y)}{P(X)P(Y)}\right)\right] \tag{2.14}$$

Differing from the KLD, mutual information is symmetric. In other words, $I(X;Y) = I(Y;X)$. The relationship of entropy, conditional entropy and mutual information can be represented by the Venn diagram in Fig. 2.2. As shown in the figure, $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$. Clearly, the mutual information refers to the intersection of the self-information $H(X)$ of $X$ and the self-information $H(Y)$ of $Y$ [23]. Intuitively, mutual information is an efficient metric to detect anomalies. This is because normal data instances typically share the same attributes, while the anomalous data have different distributions. Furthermore, the mutual information metric has the inverse numerical value of the conditional entropy metric for a same anomaly detection problem. This is due to the fact that the summation of mutual information and conditional entropy is a constant, the Shannon entropy of the random variable.

An enormous amount of work has been carried out in anomaly detection with mutual information metric. In order to boost the sensitivity of this metric, Kopylova et la. [47] combined the mutual information and Rényi entropy to detect anomalous data instances as an offline computation method. In a similar manner, a modified mutual information based algorithm is proposed in [5] to capture the features for intrusion detection systems. In the context of anomaly detection, if we look at the problem from a classification perspective and use the relationship between the current state and the history data of the system as the classifier, then the classification performance can be evaluated by mutual information. Moreover, if we let $X$ and $Y$ represent the current data and the history data series, respectively, then the smaller is the conditional entropy, or the higher is the information gain, which translates in an unlike anomaly. Based on this property, mutual information is explored as an approach to evaluate the anomaly classifier for a training data set in [80]. More importantly, this work also investigates how different information-theoretic measures can be used to describe different characteristics of an audit dataset, which provides more dependable and accurate information about how to construct algorithms and evaluation rules for anomaly detection applications.

Mutual information is also known as information gain in machine learning. It measures the reduction in entropy by splitting a dataset according to a given value of an attribute.

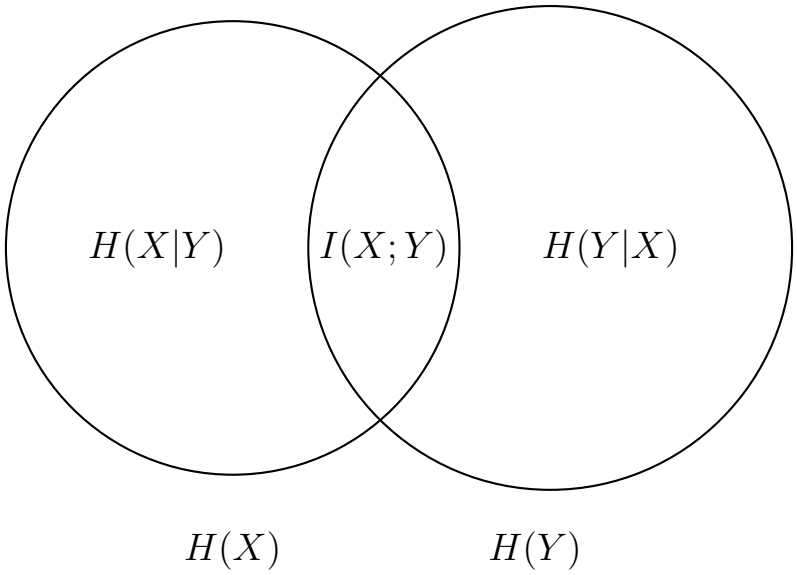



Figure 2.2: Graphical representation of the relationship between conditional entropies $H(X|Y)$ and $H(Y|X)$, and the mutual information $I(X,Y)$, in terms of entropies $H(X)$ and $H(Y)$.

The information gain of an attribute $A$ on a random variable $X$ is defined as

$$\text{Gain}(X, A) := H(X) - \sum_{v \in V(A)} \frac{|X_v|}{|X|} H(X_v) \tag{2.15}$$

where $V(A)$ is the set of all states of $A$ and $X_v$ is a subset of $X$ where $A$ has values [80]. The information gain is mainly used in decision tree to evaluate the quality of the attribute. Typically, the goal of machine learning algorithms is to maximize the information gain, which minimizes the conditional entropy.

Even though information gain shares the same quantity with mutual information, the calculation process is different. As a result, different information can be revealed by investigating the data set from different perspectives. For example, a modified version of information gain is developed in [63] to identify the anomalies by observing the ratio of the information gain of the known normal behaviours of the random variable to the entropy of the entire set of the random variable. If there are few anomalous signals in the data set, the ratio is close to one. As a result how many anomalies in the data set can be evaluated by defining thresholds or setting reference based on historic data sets.

### Jensen-Shannon Divergence

Based on the notion of KLD, a symmetric and smooth version of divergence, known as Jensen-Shannon Divergence (JSD), is proposed with an intermediate random variable. This distance measure is defined as

$$D_{\text{JS}}(P||Q) := \frac{1}{2} D_{\text{KL}}(P||M) + \frac{1}{2} D_{\text{KL}}(Q||M) \tag{2.16}$$

where $M = \frac{1}{2}(P + Q)$. JSD is also known as information radius with fixed numeric boundaries. If we use the base 2 logarithm, then $0 \leq D_{\mathrm{JS}}(P||Q) \leq 1$. However, if the natural logarithm is used, then $0 \leq D_{\mathrm{JS}}(P||Q) \leq \ln(2)$ [17]. JSD solves the problem of encountering a zero denominator in the logarithm when calculating the KLD. On the downside, its sensitivity is half that of KLD. KLD measures the distance between the two distributions, while JSD measure the distance from the two distributions to their average.

Due to the existence of a number of robust algorithms to calculate the KLD, JSD has not been witnessed as a promising approach in anomaly detection, compared to the original KLD. Nonetheless, JSD did find its way to a few applications in the domain of anomaly detection. Salem et al. [69] proposed a sequential approach to detect anomalous data in high speed networks with the help of Sketch. A similar work was also carried out in [65] to solve Change Point Detection problems in smart grids. More recently, a systematic study on the performance of the information-theoretic measures is conducted in [17]. A comparison between Shannon entropy, Tsallis entropy, Rényi entropy, KLD and JSD was made using a network traffic dataset. On top of KLD and JSD, a number of other divergence measures were proposed, as summarized in [53]. For instance the K-directed divergence takes the first half of JSD and multiplies it by two to form a new divergence measure. These divergence measures provide more options for anomaly detection in practice. An appropriate approach can be determined according to the attributes of the underlying testing database. It is worth mentioning that there has been no study that evaluated the different features of these measures in anomaly detection.

## 2.3 Non-destructive Pipeline Anomaly Detection

In this section, we present a review of non-destructive testing techniques, especially those applied to pipeline anomaly detection.

### 2.3.1 Non-destructive Testing

The non-destructive testing (NDT) for vital infrastructures, especially those related to water and energy sources, is playing a fundamental role in the world economy and in increasing the standard of living of human beings. A significant part of NDT industry is related to the network of pipelines currently in place around the world to transport water, oil and other energy sources. For these pipes, regular inspection to assure a good condition is usually the first priority in pipeline maintenance [68]. To reduce the impact of service interruption and detect the anomalies as early and effectively as possible, NDT has been vastly used in pipeline structural health and condition monitoring [14]. Several sensing technologies have been proposed in the last four decades. Among them are, eddy current, radiographic, and acoustic emission, to name a few [12, 40, 75]. Meanwhile, research in data processing and analysis has been advancing to extract relevant information in terms of defect identification and localization. A visual internal inspection with image processing techniques addressed this problem in [64]. In the work of [45] and [18], machine learning and neural networks were applied to magnetic flux leakage signals collected in NDT.

### 2.3.2 Inspection Mechanisms

The inspection mechanisms applied in inside pipeline NDT are also known as pipeline inspection gauges (PIG). They perform various maintenance operations without interrupting the basic service provided by the pipelines. The industrial term of applying PIG in practice is "pigging". Various inspection pigs have been developed to gather pipeline information by conducting measurements on the inner surface of the pipes, including recording temperature, pressure, corrosion and metal loss, diameter, bends, and curvature [35].

With the recent advances of intelligence technologies, more smart PIGs have become available, such as sophisticated robots, for instance [61]. When conducting pipeline inspection and maintenance, the in-line robotic devices are conveniently deployed inside pipelines that are typically hazardous and/or inaccessible to humans. In addition, robots might more agile and redundant than humans in certain cases. Taking advantage of this superiority, a number of robotic devices have been developed for this purpose. Some of them were deployed inside water or gas pipelines, including wheeled, legged, inchworm, snake and screw robots [59]. Snake-like robots are probably the most popular in this domain. They reached an advanced level of sophistication that allows them to operate in environments which might be too complex for conventional PIGs. Some of these robots were designed to even perform other tasks as well, such as fire fighting, and search and rescue after an earthquake [52]. A preliminary 2D parallel structure of a snake-like robot was proposed in [27, 28] for autonomous confined environment exploration. The modules of the robot are hinged with passive revolute joints, where each module is propelled by four revolute joints: two on the shoulders and two on the wheels which are in contact with the pipe. Such an architecture allows the robot to negotiate pipes of changing diameters. A similar architecture was also proposed in [77].

### 2.3.3 Pipeline Structural Anomaly Detection

The main purpose of pipeline inspection is to clear the pipes and detect anomalies, including the structural damages and abnormal flows. Structural defects identification and repair is the most important task in pipeline maintenance to prevent costly losses or environmental pollution. Robot-enabled sensing mechanisms (PIGs) have already been established among the most convenient and efficient methods to conduct NDT testing for gas or liquid pipes. Below, we will review some of the the techniques developed for pipeline anomaly detection with NDT and PIGs.

A good overview of the different types of pipeline defects and the mainly used inspection methods in pipeline anomaly detection is provided in [68]. These methods include Acoustic Emission, Electromagnetic Method, Hammer Sounding, Visual Inspection with closed circuit television and so on. In conjunction, various signal and data processing approaches have been developed to locate the defects among the raw noisy signals collected by such sensing technologies. The work in [14] summarized the methods of identifying anomalies from electromagnetic sensors, including eddy current, magnetic flux leakage (MFL), and microwave. With the exception of data analysis techniques that are based on the features

of electronic signals, other anomaly identification algorithms, in general, exploit statistical and probabilistic methods. Spinello et al. [76] employed an entropy metric on an eddy current sensor data series. Image processing-based techniques are also emerging rapidly in the pipe inspection business. An automatic defect detection algorithm was suggested in [43] based on processing images of magnetic flux leakage. A recent work by Fang et la. [31] proposed an unsupervised machine learning algorithm to detect the anomalies in sewer pipelines based on only video sequences.

## 2.4 Summary

In this chapter, the most widely used general anomaly detection techniques are surveyed, along with how information-theoretic based methods can be implemented in anomaly detection applications. Moreover, the literature of two of the most important techniques in pipeline anomaly detection, NDT and PIG, is also reviewed.

In this thesis, an algorithm for pipeline anomaly detection belonging to the class based on information theoretical tools, specifically on the KL divergence metric, is proposed and its implementation results are analyzed.

# Chapter 3

# The KL Divergence Filter

## 3.1 Preliminaries on Information Theory

In this thesis we focus on data series produced by sensors operating on mobile devices, and therefore statistical constructs are defined for discrete random variables. In addition, this allows to adopt a model free approach that is very robust since it can accommodate for signals that substantially depart from the usually assumed Gaussian distribution, therefore allowing a broader scope and applicability of the proposed method. Let $X$ be a discrete random variable mapping to a finite dimensional event space $\{x_1, x_2, \ldots, x_N\}$ of cardinality $N$. Let $P(x_i)$ be a probability mass function over the set $X$ with $\sum_{i=1}^{N} P(x_i) = 1$. The information entropy $H$, or Shannon entropy, is the expected value of the information content $-\log P(X)$ of the random variable $X$ [70]

$$H(X) := -\sum_{i=1}^{N} P(x_i) \log P(x_i) \tag{3.1}$$

The relative entropy, or Kullback-Leibler (KL) divergence $D_{\mathrm{KL}}(P||Q)$, is the expected value of the logarithmic difference (or information divergence) between two distributions $P$ and $Q$ defined over the same random variable $X$ [49], with the expectation taken with respect to $P$

$$D_{\mathrm{KL}}(P||Q) := \sum_{i=1}^{N} P(x_i) \log \frac{P(x_i)}{Q(x_i)} \tag{3.2}$$

Again, it is worth mentioning that the KL divergence is not symmetric since $D_{\mathrm{KL}}(P||Q) \neq D_{\mathrm{KL}}(Q||P)$, and therefore it has to be intended as a measure of discrepancy between the two distributions rather than a metric. Also, the KL divergence is a non-negative real number for specific distributions and it evaluates to zero if and only if $P = Q$ [48]. In the calculation, we have the convention that $0 \log \frac{0}{Q} = 0$ when $P(x_i) = 0$ and $P \log \frac{P}{0} = \infty$ when $Q(x_i) = 0$, as well $0 \log \frac{0}{0} = 0$ [23].

## 3.2 Anomaly Detection Algorithm

The KL divergence filter in this work maps each raw data point to a local KL divergence value. The sense in which the KL divergence is local will be explained soon. As a measure of information discrepancy, the KL divergence requires a baseline signal, which for anomaly detection should be representative of the noise generated by the same sensor, or at least by a sensor within the same class as the one used to collect the data. Notice that this approach is very versatile since it characterize the sensor through characteristic noise, which is often an information provided by the manufacturer.

To illustrate the main idea behind the algorithm, consider a spatial array of sensors moving along parallel trajectories, as for example in the case of radially distributed sensors carried by snake-like robots for non-destructive pipeline inspection. Each sensor produces a data series, and therefore the array can be used to sense a scalar field on a surface. A schematic is shown in Fig. 3.1, where a local portion of the sensed surface is represented in Cartesian coordinates; however this would also apply to the cylindrical inner surface of a pipeline, which can be developed on a plane without deformation. Each data point on the surface, representing a measurement of the scalar field $\phi$, is identified by two integer indexes $i = 1, \dots, N_\ell$ and $j = 1, \dots, N_w$, respectively spanning the horizontal and the vertical direction. Therefore a raw measurement of the field $\phi$ is denoted by $\phi_{ij}$. The data set comprised of all measurement data is denoted by $\mathbb{S}$.

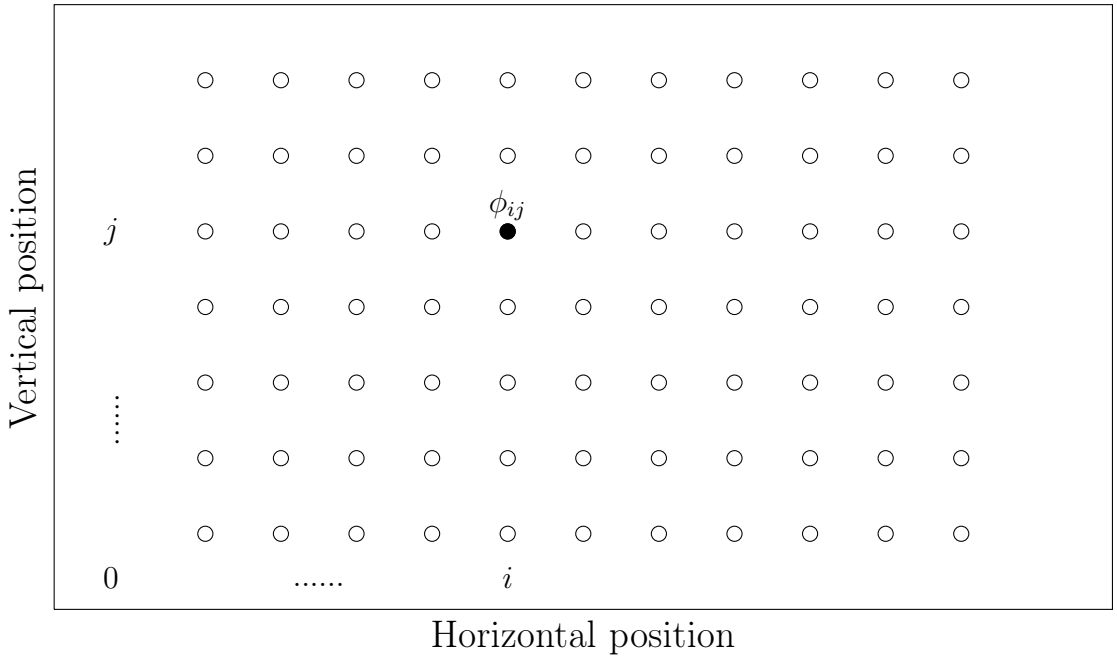



Figure 3.1: Schematic of the data points on the surface of a pipeline.

The KL divergence as expressed in (3.2) can be intended as a weighted likelihood test, where the hypothesis of a data point $\phi$ being an anomaly is tested against the null hypothesis of $\phi$ being an expected measurement. In this sense, the distribution $Q$ of the

baseline expected signal can be regarded as the prior, whereas the distribution $P$ of the sensor measurements can be regarded as the posterior. Then the anomalies can be identified by the discrepancy between these two distributions. Meanwhile, a well known problem in the calculation of the KL divergence for discrete data series may arise when the denominator in (3.2) is zero, namely $Q(X) = 0$ while the numerator $P(X) \neq 0$. This is an occurrence to be expected if $Q$ is built from noisy data substantially distinguished from anomalies. To overcome this problem, we build the histogram and probability mass function $Q$ from the entire data set. If the anomalies are relatively rare, the distribution of $Q$ is still dominated by the noise around the expected measurement, therefore providing a suitable approximation that does not suffer from empty bins. A measurement point corresponding to an anomaly evaluates to a small probability $Q(\phi) \neq 0$, therefore overcoming the issue.

### 3.2.1 Generating the baseline mass function Q

To illustrate the steps involved in building the KL filter, we consider a one dimensional data series that can be considered as the output of one of the sensor's channels around the circumferential direction. For a physical distance of 1 m, we consider the data series plotted in Fig. 3.2. The data series includes only one anomaly, which is a rare event in the data dominated by the background noise.

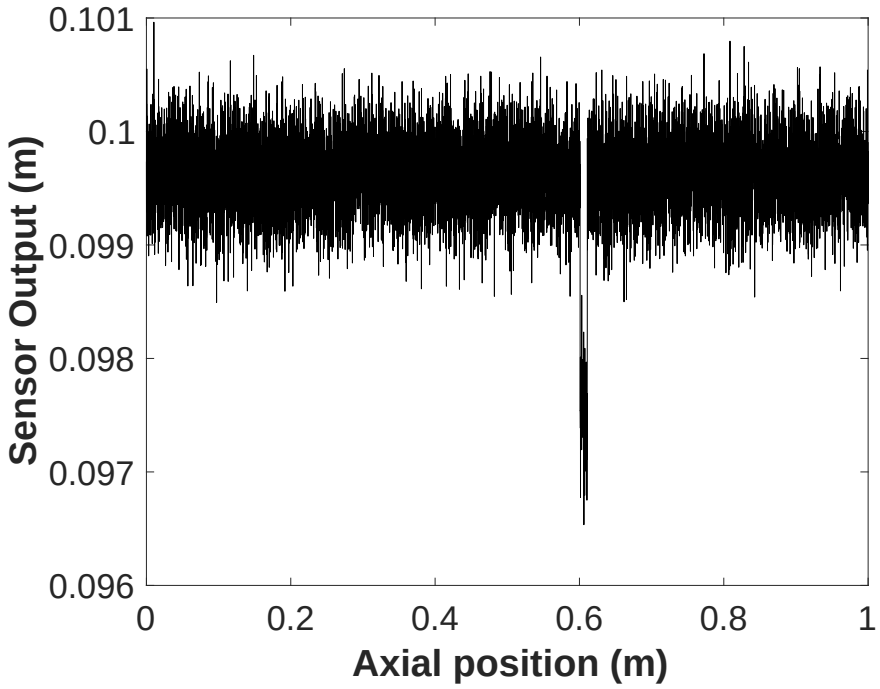



Figure 3.2: A raw sensory data series with only one anomaly.

The mass function $Q$ can be obtained by building a histogram of the all data set $\mathbb{S}$ with $k$ bins. By denoting with $a = \min \mathbb{S}$ and $b = \max \mathbb{S}$ respectively the minimum and the

maximum, and by setting the bins of equal size, the bin size is given by

$$h = \frac{b-a}{k} \tag{3.3}$$

We denote the bins boundary points as $B_n = \{a+(n-1)h, a+nh\}$, with $n = \{1, 2, ..., k\}$. Fig. 3.3 shows an example of the histogram with 20 bins. Based on the histogram, the corresponding probability mass distribution is obtained by normalizing the frequencies in every bin with respect to the cardinality of the data-set, as shown in Fig. 3.4. We denote it with $Q(\phi)$ and observe that the discrete values taken by $Q$ represent the probability that a measurement $\phi$ belongs to the corresponding bin of the histogram used to generate $Q$. Therefore, the mass function can also be used to evaluate probabilities of the original data set. The probability mass function shows the majority of the events happening around the expected measurement of the noisy sensor signal, with a tail generated by the anomaly, as expected considering that it is a rare event. We do have some zeros in $Q(\phi)$, but at that point the corresponding probabilities in $P(\phi)$ are zero as well, which returns zero value terms in the summation for the calculation of the KL divergence, according to the convention $0 \log \frac{0}{0} = 0$.

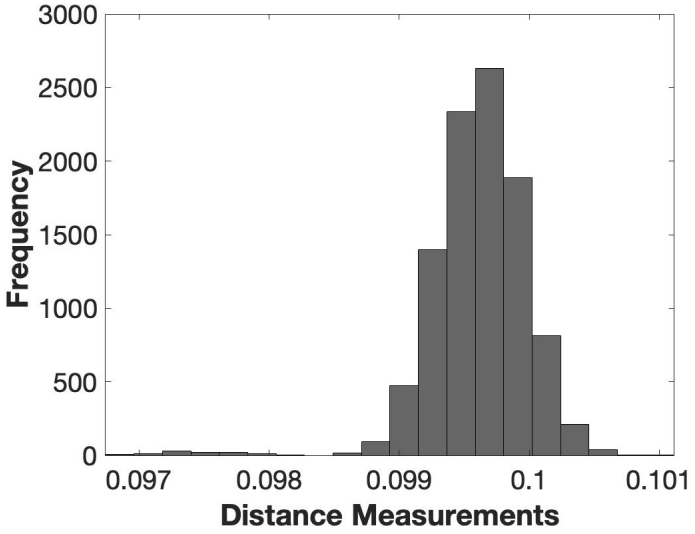



Figure 3.3: Histogram of all the data in the data series in Fig. 3.2, which is used to build the baseline $Q$.

### 3.2.2 Local KL Divergence Map

To map each raw data point $\phi_{ij}$ into a local KL divergence value, we consider two integers $l$ and $w$ and a window size $(2l+1) \times (2w+1)$ centered around the location $(i, j)$ as illustrated in Fig. 3.5, where the window is shown as the red dashed line delimiting a rectangular region enclosing a subset of $\mathbb{S}$ of cardinality $(2l+1)(2w+1)$, namely, $\Phi(i, j) \subsetneq \mathbb{S}$. From each data subset $\Phi(i, j)$ associated to the raw data point $\phi_{ij}$, a probability mass function is built by

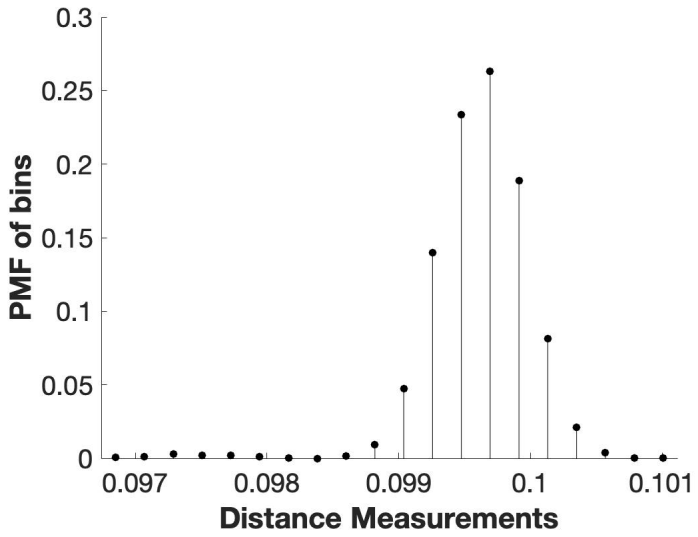



Figure 3.4: The corresponding PMF of the 20 bins in Fig. 3.3, which is denoted as $Q(\phi)$.

considering the samples in $\Phi(i,j)$ as a random variable $\Phi_{ij}$ whose realizations are the data points in the window centered at $\phi_{ij}$. This allows to build a probability mass function for each data point, so that the map into KL divergence has a local nature, allowing to detect spatial variations of the data [71].

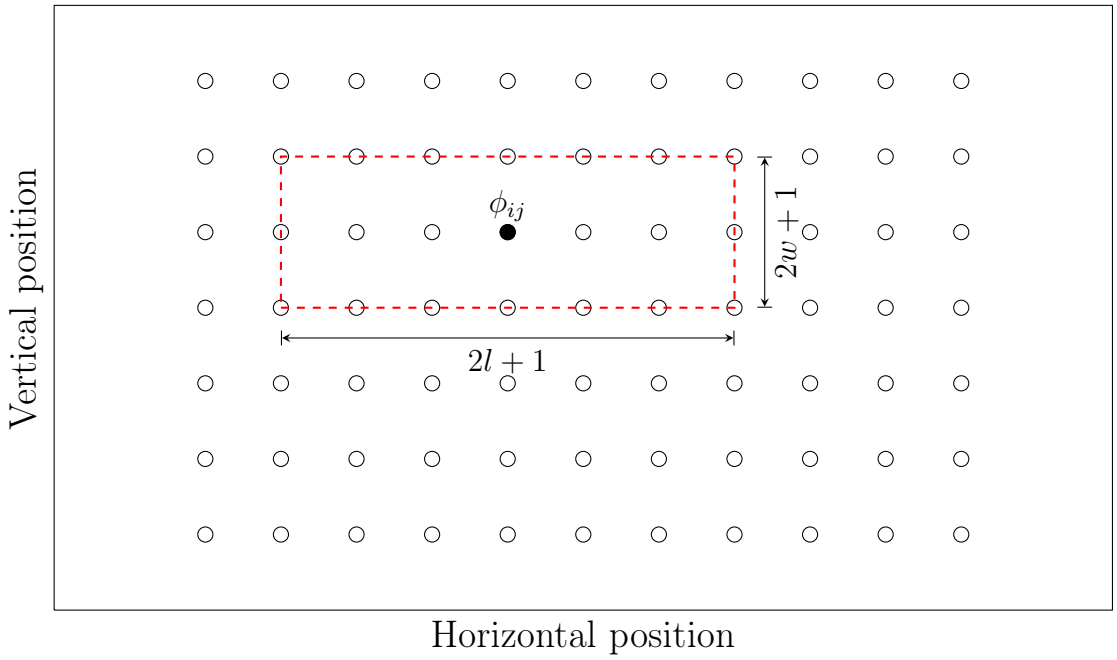



Figure 3.5: Schematic of a rectangular window centered at the phase datum $\phi_{ij}$ .

One of the advantages of this algorithm is that no information about the anomalies is needed. Since the prior distribution $Q$ is built from the entire data set $\mathbb{S}$, which includes

each subset $\Phi(i,j)$, then for every sample $\phi \in \Phi_{ij}$ the probability $Q(\phi) \neq 0$, and the calculation of the KL divergence is well posed. We stress one more time that as long as the anomalies are rare, $Q$ is expected to be almost fully characterized by noise around expected measurement, and therefore we can avoid the indeterminacy caused by $Q(\phi) = 0$, while the local KL divergence would measure the local discrepancy between a data point and the overall data set which closely resembles noise. The size of the window characterizes the resolution of the filter. At the limit, a window as large as the entire data set would not capture any variation since $P \equiv Q$ for all $\phi \in \mathbb{S}$. An approach adopted when there are few bins with zero frequency is to simply ignore the empty bins and implicitly set the corresponding terms to zero, namely, ignore these terms in the summation without significantly affecting the result if the number of empty bins for $Q$ is small as compared to the total number of bins [60]. Another possibility is to fictitiously add the same constant small frequency to all bins, so that no empty bins exist [1]. If the number of empty bins for $Q$ is significant, a more rigorous approach is to adopt Jenson-Shannon divergence [17].

If we set the window size as 121×3 and take a data point located in 0.602 m on the axial direction in Fig. 3.2 as an example, the histogram of the corresponding data set $\Phi(i,j)$ is shown in Fig. 3.6 with 20 bins. By following the same procedure as the one used for $Q$, a posterior probability mass function $P_{ij}$ is associated to the random variable $\Phi_{ij}$ by normalizing the histogram with respect to the cardinality of the data set $\Phi(i,j)$, with an example illustrated in Fig. 3.7. As before, the probability of the random variable $\phi \in \Phi(i,j)$ is calculated as the mass of the corresponding bin's probability mass function.

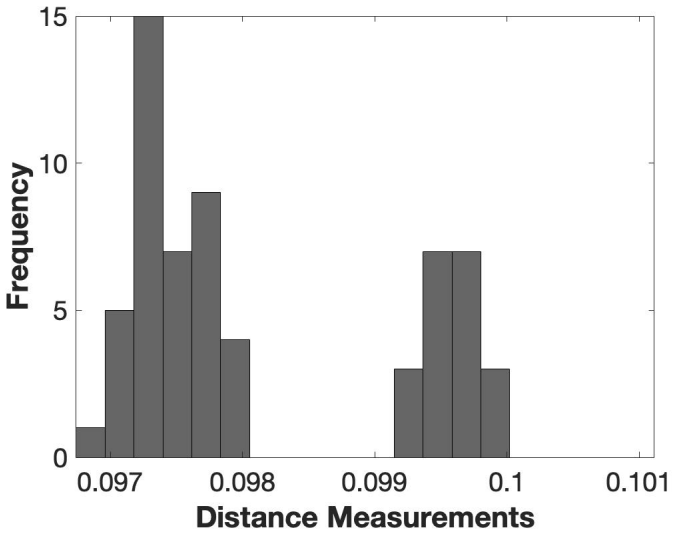



Figure 3.6: Histogram of $\Phi(i,j)$ when $\phi_{ij}$ is a data point from the series in Fig. 3.2 located on the anomaly with $k = 20$ bins.

Moreover, by comparison of the two histograms, we can conclude that the histogram of $P_{ij}$ significantly differs from $Q$'s, since for $P_{ij}$ most events fall into the first few bins due to the fact that the anomaly corresponds to smaller distance measurement. As a result, it appears with high frequency due to the fact the window to build the histogram captures a

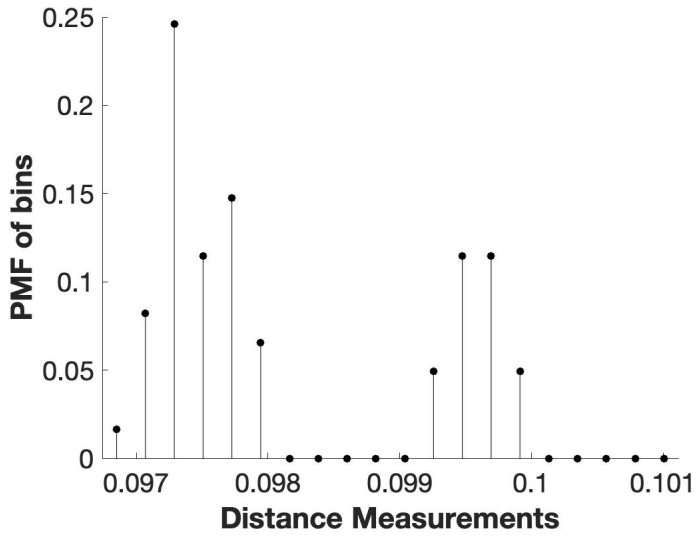



Figure 3.7: The corresponding PMF of the 20 bins in Fig. 3.6, which is denoted as $P_{ij}(\Phi_{ij})$.

significant amount of data related to anomaly measurements in the first few intervals. On the other hand, the histogram in Fig. 3.3 is dominated by the higher values of distance measurements that are representative of the output of the sensor when no anomaly is detected.

The last step is to measure the local discrepancy between the prior distribution $Q$ and the posterior $P_{ij}$ for all $i, j$, which results into mapping each raw data point $\phi_{ij}$ into a local KL divergence

$$\begin{aligned} D_{\mathrm{KL}}(P_{ij}||Q) &= \sum_{\phi \in \Phi_{ij}} P_{ij}(\phi) \log \frac{P_{ij}(\phi)}{Q(\phi)} \\ &= \sum_{n=1}^{K} P_{ij}(\phi \in B_n) \log \frac{P_{ij}(\phi \in B_n)}{Q(\phi)} \end{aligned} \tag{3.4}$$

where $B_n$ is the $n$-th bin in the posterior $P_{ij}$, with total number of bins $K$. The expression (3.4) greatly simplifies the calculation of the KL divergence, since the summation extends over the bins rather than the entire data samples in $\Phi(i, j)$. However, we need to clarify the meaning of the expression $Q(\phi)$, since the bins used to build the histogram that generates $Q$ can in general be different than the ones used for the histogram that generates $P_{ij}$. Let $\delta_{ij}$ be the uniform bin size of the histogram of $\Phi(i, j)$, and $h$ be the bin size for the histogram of $\mathbb{S}$ as defined above. We define the discrete probability density functions $p_{ij}$ and $q$ by normalizing each value of the corresponding mass functions by the bin sizes (in this case uniform), so that

$$P_{ij}(\phi \in B_n) = p_{ij}(\phi_n)\delta_{ij}, \qquad Q(\phi) = q(\phi_n)h \tag{3.5}$$

where $\phi_n \in B_n$ is any element in the bin $B_n$, which can be chosen arbitrarily within the bin since by construction all the data points in a bin have the same probability. Note that

the bin size for the histogram generating $Q$ is in general different, that is $k \neq K$, but $q(\phi_n)$ accounts for this since the mass $Q$ is normalized by the bin size. Substituting into (3.4)

$$D_{\mathrm{KL}}(P_{ij}||Q) = \sum_{n=1}^{K} p_{ij}(\phi_n)\delta_{ij} \log \frac{p_{ij}(\phi_n)\delta_{ij}}{q(\phi_n)h} \tag{3.6}$$

This expression allows to use different number of bins and different bin sizes for data sets that may have different structures. Since the summation is extended only to the bins for the prior $P_{ij}$, and since $\Phi(i,j) \in \mathbb{S}$, this calculation is equivalent to

$$D_{\mathrm{KL}}(P_{ij}||Q) = \sum_{m=1}^{k} p_{ij}(\phi_m)\delta_{ij} \log \frac{p_{ij}(\phi_m)\delta_{ij}}{q(\phi_m)h} \tag{3.7}$$

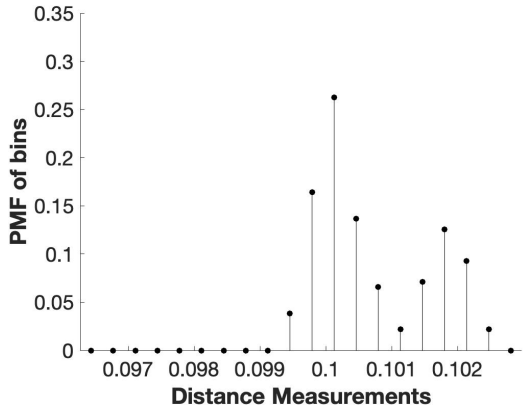
since the summation would include several zero terms of the type $0 \log \frac{0}{q(\phi_m)h}$, corresponding to the empty bins of the histogram for $P_{ij}$. The only non-zero terms are therefore the ones in (3.6).

By repeating the process of step one and two, each data point $\phi_{ij}$ can be mapped to a local KLD value. Given the local nature of the KL divergence map as conferred by the window size around each original data point, the algorithm is able to spatially differentiate among different data points. The three probability distributions in Fig. 3.8 refer to two points on different data sets from the same proximity sensor and their common baseline, in which anomalies of different size were present. The PMF of the data point on the larger anomaly is more narrowly distributed around the sample point representing the anomaly value, which results in a larger discrepancy from the baseline distribution $Q$ in (c). As we will numerically verify in Chapter 5, larger anomalies correspond to larger values of the KL divergence, establishing therefore an immediate intuitive interpretation of the KLD filtered data.

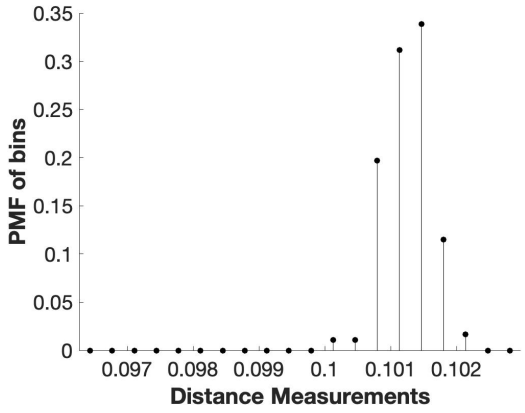
## 3.3 Anomaly Detection with Gaussian approximation

The procedure for the KL divergence filter outlined above does not require any assumption on the distributions $P_{ij}$ and $Q$, which are built from data series. The advantage of not depending on specific assumptions on the data distribution is offset by the need of binning of the histograms, and the calculation of the KL divergence through a summation that may involve several terms. Therefore the robustness of the method is offset by possible relative algorithmic complexity.

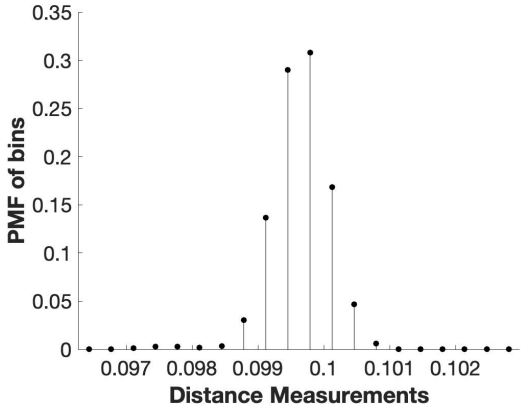
The data characterizing the noise is expected to be well represented by a Gaussian distribution. If such approximation can be acceptable also for the anomaly data, then the Gaussian approximation allows to obtain analytical expressions for the KL divergence filter that greatly simplify it. Clearly the choice has to be done based on trade-offs dictated by how acceptable the simplifying assumptions are with respect to the specific data. In this




(a) PMF of smaller anomaly data




(b) PMF of larger anomaly data



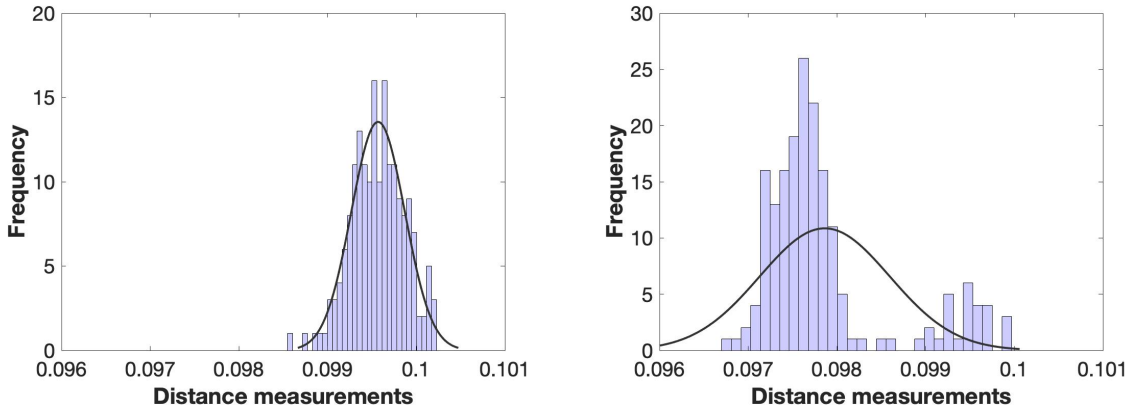

(c) PMF of the common baseline Q

Figure 3.8: Comparison between the probability mass functions of two anomalies with different sizes, the PMF of the larger anomaly in (b) shows less similarity than the smaller one in (a) in shape with the baseline in (c).

section we present the KL divergence filter derived under this set of assumptions, namely Gaussian approximations for noise and anomaly data. To briefly illustrate this idea, an example of the Gaussian approximation of pure noisy data and anomalous data from the data series in Fig. 3.2 with histograms is given in Fig. 3.9.

### 3.3.1 KLD filter with Gaussian approximation

Shannon proposed the concept of differential entropy in [70], which extends the idea of information entropy, initially defined for discrete variables, to continuous random variables. For a continuous random variable $\phi$ with probability density function $p_{ij}(\phi)$, the differential




(a) Noisy data (b) Anomalous data

Figure 3.9: The Gaussian approximations using histogram for the noisy raw sensory data (axial position = 0.1 m) and anomalous raw sensory data (axial position = 0.605 m) from data series in Fig. 3.2, showing (a) the Gaussian distribution approximation of the baseline data set representing noisy expected measurements; and (b) the data with anomalies for which the Gaussian approximation is less accurate. Black lines are the approximating Gaussian distributions.

entropy is defined as

$$H(p_{ij}) := -\int_A p_{ij}(\phi) \log p_{ij}(\phi) \mathrm{d}\phi \tag{3.8}$$

where $A$ is the support set of $\phi$. Based on the concept of differential entropy, the KL Divergence between two absolutely continuous distributions $p_{ij}(\phi)$ and $q(\phi)$ of a continuous random variable $\phi$ is defined as

$$D_{\mathrm{KL}}(p_{ij}||q) := \int_A p_{ij}(\phi) \log \frac{p_{ij}(\phi)}{q(\phi)} \mathrm{d}\phi \tag{3.9}$$

where $p_{ij}$ and $q$ are the analogous of $P_{ij}$ and $Q$ in the discrete case. To approximate $P_{ij}$ and $Q$ by Gaussian distributions, we calculate the first two statistical moments of $\Phi(i,j)$ as

$$\mu_{ij} = \frac{1}{|\Phi(i,j)|} \sum_{\phi \in \Phi(i,j)} \phi \tag{3.10a}$$

$$\sigma_{ij}^2 = \frac{1}{|\Phi(i,j)|} \sum_{\phi \in \Phi(i,j)} (\phi - \mu_{ij})^2 \tag{3.10b}$$

where $|\Phi(i,j)|$ denotes the cardinality of $\Phi(i,j)$. Given the entire data set $\mathbb{S}$, the corresponding first two statistical moments to approximate $Q$ can be computed in the same way, and will be denoted by $\mu_2$ and $\sigma_2$. If the distributions for the entire data set and for

the actual signal are assumed to be Gaussian, then

$$p_{ij}(\phi) \sim N(\mu_{ij}, \sigma_{ij}^2) \tag{3.11a}$$

$$q(\phi) \sim N(\mu_2, \sigma_2^2) \tag{3.11b}$$

It is worth mentioning that if there are few anomalies in the data set, the noise dominated signals can be well approximated by Gaussian distributions. However, when adopting the Gaussian approximation, $q$ can also be obtained from purely noisy signals without incurring the indeterminacy affecting the discrete case, since the support set of the Gaussian function is $[-\infty, +\infty]$. From the definition of Gaussian distribution, the expressions of the PDFs can be explicitly written as

$$p_{ij}(\phi) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left(-\frac{1}{2}\left(\frac{\phi - \mu_{ij}}{\sigma_{ij}}\right)^2\right) \tag{3.12a}$$

$$q(\phi) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{1}{2}\left(\frac{\phi - \mu_2}{\sigma_2}\right)^2\right) \tag{3.12b}$$

By substituting (3.12) into (3.8), the differential entropy of a continuous Gaussian distributed random variable can be written in closed form as [50]:

$$\begin{aligned} H(\Phi_{ij}) &= \mathrm{E}\left[-\ln p(\phi)\right] \\ &= -\int_{-\infty}^{+\infty} p_{ij}(\phi) \ln p_{ij}(\phi) \mathrm{d}\phi \\ &= \int_{-\infty}^{+\infty} p_{ij}(\phi) \left(\frac{1}{2}\ln\left(2\pi\sigma_{ij}^2\right) + \frac{(\phi - \mu_{ij})^2}{2\sigma_{ij}^2}\right) \mathrm{d}\phi \\ &= \int_{-\infty}^{+\infty} p_{ij}(\phi) \left(\frac{1}{2}\ln\left(2\pi\sigma_{ij}^2\right)\right) \mathrm{d}\phi + \int_{-\infty}^{+\infty} p_{ij}(\phi) \frac{(\phi - \mu_{ij})^2}{2\sigma_{ij}^2} \mathrm{d}\phi \\ &= \frac{1}{2}\left(\ln\left(2\pi\sigma_{ij}^2\right) + 1\right) \end{aligned} \tag{3.13}$$

Similarly, it is possible to obtain a closed form expression for the KL divergence [11]

$$\begin{aligned} D_{\mathrm{KL}}(p_{ij}||q) &= \mathrm{E}_{p_{ij}}\left[\ln \frac{p_{ij}(\phi)}{q(\phi)}\right] \\ &= -\int_{-\infty}^{+\infty} p_{ij}(\phi) \ln \frac{p_{ij}(\phi)}{q(\phi)} \mathrm{d}\phi \\ &= \int_{-\infty}^{+\infty} p_{ij}(\phi) \ln \left(\frac{\frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{\frac{-(\phi-\mu_{ij})^2}{2\sigma_{ij}^2}}}{\frac{1}{\sqrt{2\pi\sigma_2^2}} e^{\frac{-(\phi-\mu_2)^2}{2\sigma_2^2}}}\right) \mathrm{d}\phi \\ &= \int_{-\infty}^{+\infty} p_{ij}(\phi) \ln \left(\frac{\sigma_2}{\sigma_{ij}} \frac{e^{\frac{-(\phi-\mu_{ij})^2}{2\sigma_{ij}^2}}}{e^{\frac{-(\phi-\mu_2)^2}{2\sigma_2^2}}}\right) \mathrm{d}\phi \end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{+\infty} p_{ij}(\phi) \ln \frac{\sigma_2}{\sigma_{ij}} \mathrm{d}\phi + \int_{-\infty}^{+\infty} p_{ij}(\phi) \ln \left( \frac{e^{\frac{-(\phi-\mu_{ij})^2}{2\sigma_{ij}^2}}}{e^{\frac{-(\phi-\mu_2)^2}{2\sigma_2^2}}} \right) \mathrm{d}\phi \\
&= \frac{1}{2} \ln \left( \frac{\sigma_2^2}{\sigma_{ij}^2} \right) + \int_{-\infty}^{+\infty} p_{ij}(\phi) \left( \frac{-(\phi-\mu_{ij})^2}{2\sigma_{ij}^2} + \frac{(\phi-\mu_2)^2}{2\sigma_2^2} \right) \mathrm{d}\phi \\
&= \frac{1}{2} \ln \left( \frac{\sigma_2^2}{\sigma_{ij}^2} \right) - \frac{1}{2\sigma_{ij}^2} \int_{-\infty}^{+\infty} p_{ij}(\phi)(\phi-\mu_{ij})^2 \mathrm{d}\phi + \frac{1}{2\sigma_2^2} \int_{-\infty}^{+\infty} p_{ij}(\phi)(\phi-\mu_2)^2 \mathrm{d}\phi \\
&= \frac{1}{2} \ln \left( \frac{\sigma_2^2}{\sigma_{ij}^2} \right) - \frac{1}{2} + \frac{1}{2\sigma_2^2} \int_{-\infty}^{+\infty} p_{ij}(\phi)(\phi-\mu_{ij}+\mu_{ij}-\mu_2)^2 \mathrm{d}\phi \\
&= \frac{1}{2} \ln \left( \frac{\sigma_2^2}{\sigma_{ij}^2} \right) - \frac{1}{2} + \frac{1}{2\sigma_2^2} \left( \int_{-\infty}^{+\infty} p_{ij}(\phi)(\phi-\mu_{ij})^2 \mathrm{d}\phi + \int_{-\infty}^{+\infty} (\mu_{ij}-\mu_2)^2 \mathrm{d}\phi \right. \\
&\qquad \left. +(\mu_{ij}-\mu_2) \int_{-\infty}^{+\infty} 2p_{ij}(\phi)(\phi-\mu_{ij}) \mathrm{d}\phi \right) \\
&= \frac{1}{2} \ln \left( \frac{\sigma_2^2}{\sigma_{ij}^2} \right) - \frac{1}{2} + \frac{1}{2\sigma_2^2} \left( \sigma_{ij}^2 + (\mu_{ij}-\mu_2)^2 + 0 \right) \\
&= \frac{1}{2} \left( \ln \frac{\sigma_2^2}{\sigma_{ij}^2} + \frac{\sigma_{ij}^2}{\sigma_2^2} + \frac{(\mu_{ij}-\mu_2)^2}{\sigma_2^2} - 1 \right)
\end{aligned}
\tag{3.14}
$$

where $E_{p_{ij}}$ is the expectation with respect to $p_{ij}$. With this equation, each data point on the plane can be associated with a local KLD value in the same manner as the model-free KLD filter, evaluating zero when the two distributions are identical in terms of the first two moments.

An entropy filter based on expression (3.13) together with an hypothesis testing based on a log likelihood ratio test has been presented in [71]. By comparing the expressions (3.13) and (3.14), it is clear that a test based on KL divergence has the advantage of including the means, and therefore it is expected to perform better with signals that have a similar variance with respect to the background noise but significantly different sample mean. More generally, the KL divergence filter based on Gaussian approximations of the prior and the posterior measures the discrepancies in terms of the first two statistical moments of the corresponding data sets. Although the Gaussian approximation of the anomaly data can be very inaccurate, as in the example just illustrated, the filter may still work if the discrepancy in terms of statistical moments is significant. We will show some results for which this is indeed the case.

# Chapter 4

# A Case Study: Pipeline Non-destructive Inspection

As stated in Chapter 2, inside pipeline inspection is a core application of anomaly detection technologies. In most cases, data acquisition is performed in challenging circumstances. For instance, underground oil pipes are often covered by thick sluggish oil residues, which affects the quality of the collected data. For this work, field data was not available to test with. However, high-fidelity simulated data was generated with the help of professional software along with realistic sensor models provided by the sensors' manufacturers. In this chapter, we present the details of how the data acquisition was performed before it is processed in the following chapters.

## 4.1 Pipeline with Defects

A virtual pipe is built to mimic a real pipe anomaly detection case conducted by a Toronto-based company in [71], where more than ten anomalies, represented by holes of different sizes, are scattered along two circumferential positions on two connected parts of a metal pipe. The model of the pipe is constructed with the CAD software Solidworks© before it is exported to CoppeliaSim© simulator [22]. The latter allows to construct simulated environments with data generated by high fidelity models of commercial sensors. We consider a 1 m long pipe-like geometrical cylinder as a test bed, with a pipe's internal diameter of 400 mm and a thickness of 2 mm. A total of eleven defects are placed along the pipe's surface, as schematized in the two-dimensional representation in Fig. 4.1. We embedded defects, or anomalies, of different nature. The first type is the welding strip represented by two black straight lines. Welding strips are very common and they are often found between pipe segments. Therefore it is important for a detection device to be able to distinguish them from other defects. The width and depth of the welding strip are 10 mm and 2 mm, respectively. The second type of defects considered is through holes. These are holes that go through the full thickness of the pipe surface. They are represented by black circles. The third and last type of defects is blind holes and are represented by red

circles. These are circular dents which do not fully penetrate through the pipe. The depth of the blind holes is 1 mm (half the pipe's thickness). The holes' diameters vary between three possible values: 5 mm, 10 mm, and 15 mm. The holes' circumferential positions are either 180° or 270°. The detailed diameters of the holes are given in Tables 4.1 and 4.2. The 3D model of the pipe in Solidworks is shown in Fig. 4.2.

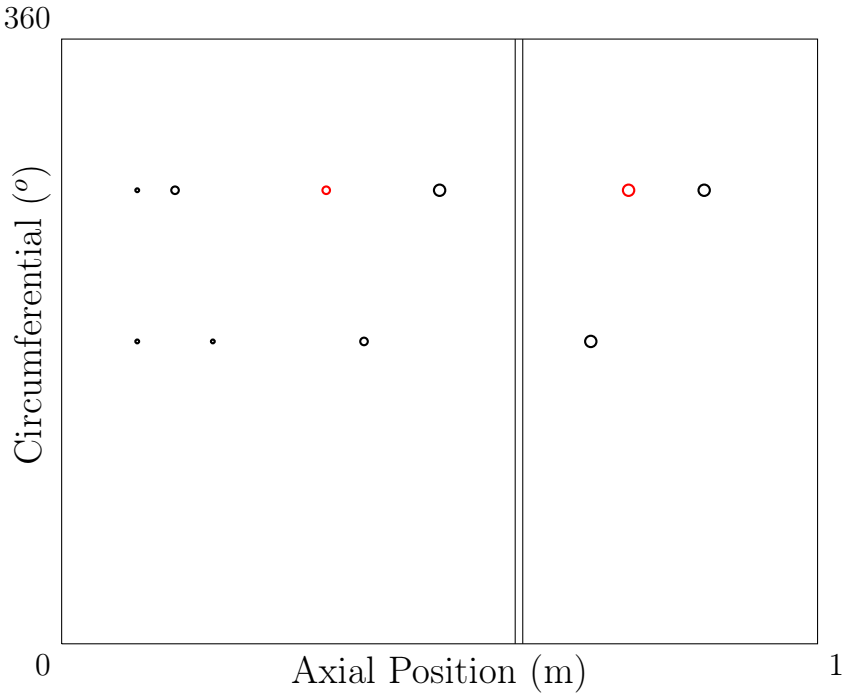



Figure 4.1: Schematic of the anomalies on the pipeline, including ten holes represented by the circles and a welding strip represented by two black lines.

Table 4.1: The diameters of the four holes along 180°

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| $\phi$5 mm | $\phi$5 mm | $\phi$10 mm | $\phi$15 mm |

Table 4.2: The diameters of the six holes along 270°

| 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| $\phi$5 mm | $\phi$10 mm | $\phi$10 mm (blind) | $\phi$15 mm | $\phi$15 mm (blind) | $\phi$15 mm |

## 4.2 Data Generation

Sensory data that reproduce that collected by a mobile robot for non-destructive pipeline inspection are generated using the simulator environment CoppeliaSim©. The workspace

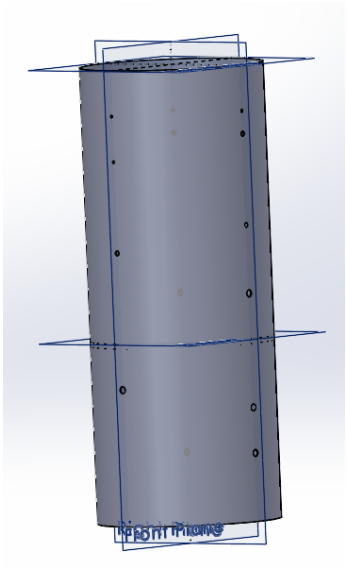


Figure 4.2: Solidworks 3D model of the pipeline shown in Fig .4.1.

in CoppeliaSim is rendered in Fig. 4.3, with a sensing device comprised of 36 proximity sensors mounted along the circumferential surface of a cylindrical device called the carrier. The cylindrical payload geometry with a circumferential array of sensors is a common arrangement for non-destructive inspection of pipelines as it allows to reconstruct the sweep of the inner surface of the pipe. It is also robust with respect to the axial rotation of the device [13].

A laser pointer is mounted as the sensing element on each of the 36 proximity sensors. This is a common arrangement in commercial proximity sensors. The 36 laser pointers are evenly attached on the carrier surface, 10° apart from each other, as illustrated in Fig. 4.4. The orange core represents the sensor carrier. The carrier can simultaneously move axially (back/forth) and radially (rotary motion) with the help of one revolute and one prismatic joint. The distance values between the proximity sensors and the inner surface of the pipe can be collected by scanning the whole pipe with proper forward and rotating angular velocities. The diameter of the carrier is 200 mm. The base distance between sensors and the pipe's inner surface when the carrier is axially centered along the pipe is 100 mm. The data series from one channel scanning (i.e., one of the 36 sensors) is shown in Fig. 4.5.

Given the 36 channels, the whole pipeline can be scanned with a proper linear and angular carrier velocities. That way, each data point can be mapped to its circumferential and axial coordinates. Let $v$ and $\omega$ be the carrier's constant linear and angular velocities, respectively. Then, the location of a specific point on the scanning path of the sensor can

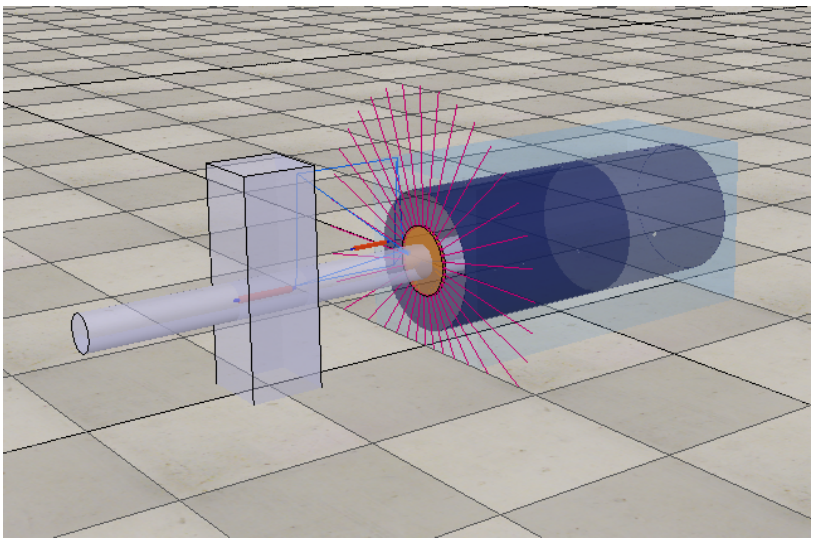


Figure 4.3: CoppeliaSim workspace showing the sensing device with 36 circumferential proximity sensors.

be found by

$$X = v\,t \tag{4.1a}$$
$$Y = \omega\,R\,t - \theta_i \tag{4.1b}$$

where $X$ and $Y$ are the point's axial and circumferential positions, respectively; $t$ is the time, $R$ is the pipe's inner radius, and $\theta_i = (i-1) \times 10°$ is the initial circumferential position corresponding to proximity sensor $i \in \{1, 2, 3, ..., 36\}$. In a discrete sampling process, $t = k\,\Delta t$, where $\Delta t$ is the sampling period and $k = 0, 1, 2, \ldots$, is the sample index number. The values used in this study are

$$v = 0.01\,\mathrm{m/s} \tag{4.2a}$$
$$\omega = 100\,°/\mathrm{s} \tag{4.2b}$$
$$\Delta t = 0.01\,\mathrm{s} \tag{4.2c}$$

Here is an example of the sampling process. At a linear speed of $v = 0.01\,\mathrm{m/s}$, the carrier needs 100 s to run the whole 1 m-pipe. The resolution of the proximity sensors is 0.1 mm, which is in line with many commercial sensors in practice. The resolution parameter determines the total of possible outcomes for the raw distance measurements within a given data set, which also provides the limits of the number of bins defined in section 3.2. For instance, if we have the range of a raw measurement data set from 0.0963 m

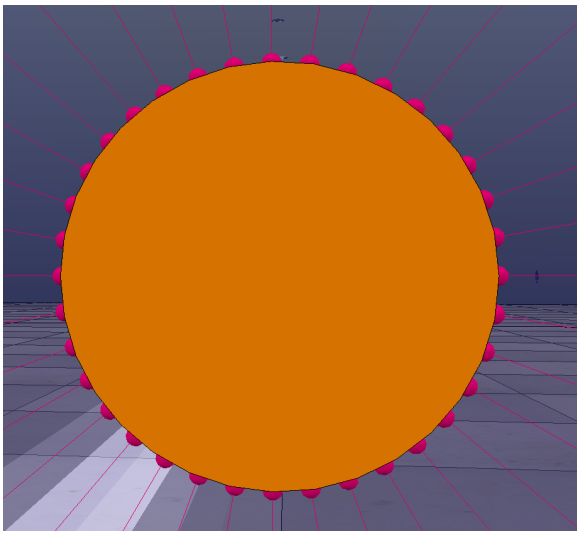


Figure 4.4: The layout of the sensor on the carrier.

to 0.1030 m. Then with every 0.1 mm, a possible outcome of the measurement is given, thus 67 in total. To assure the histogram meaningful and accurate, the number of bins must not be greater than the total outcomes of the entire raw sensory dataset.

Thus, at a sampling rate of 100 Hz, 10000 axial and 360 circumferential samples are collected. Each sample can then be mapped onto its spatial coordinates on the pipe using relation (4.1). It may happen that certain points are mapped to zero values if the sampling frequency is too low. In such a case, these points are assigned the values of their nearest neighbor along the same circumferential direction. The simulation parameters can be found in Appendix A.

So that our study is as realistic as possible, a Gaussian white noise is added to the collected data. This type of noise is typical in commercial sensors. The relationship between the detection distance and signal to noise ratio (SNR) in a typical commercial sensor is studied in [26]. The outcome is shown in Fig. 4.6. The relationship between the two is almost linear. The figure indicates that the typical SNR corresponding to a detection distance of 10 cm, which is the base distance between the sensor and the pipe surface, is 50 dB. The SNR given here is the ratio of the power of the signal to the power of the noise in dB. It is defined as

$$\text{SNR} = 10 \log_{10}(\frac{P_{signal}}{P_{noise}}) \tag{4.3}$$

where $P_{signal}$ is the average power of all the elements in the signal (meaningful input), while $P_{noise}$ is the average power of the noise, which is the variance of the Gaussian noise.

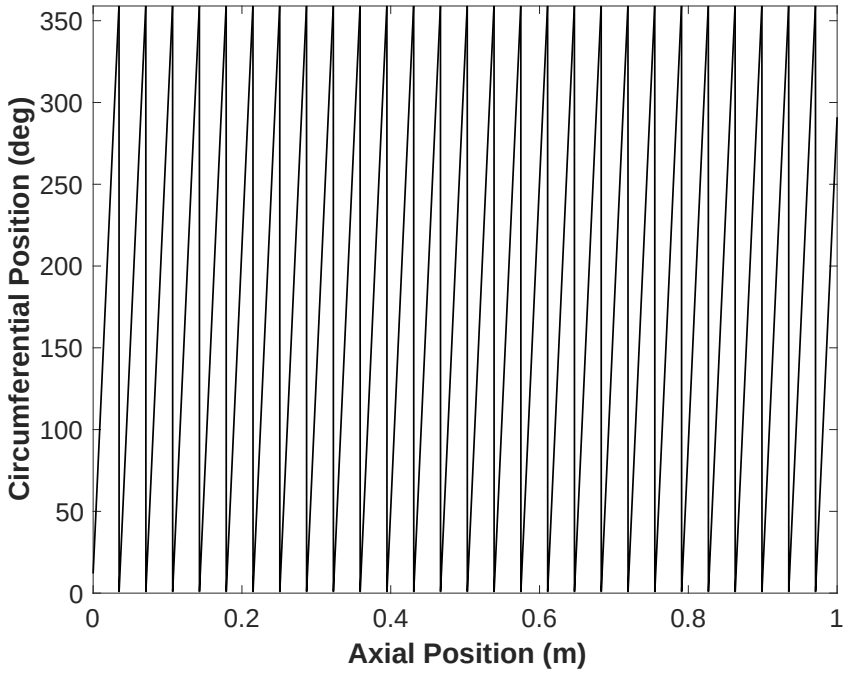



Figure 4.5: Data series from one channel scanning.

Since the raw distance values from the sensors are used in the calculation of the signal power, the SNR level varies from case to case with different initial measurement distances given a certain power of the Gaussian noise. Therefore, to reveal the actual difference between the useful signal and the noise and study the real effect of the noise, we use the Fault-to-Noise-Ratio (FNR) [82] defined as follows.

$$\mathrm{FNR} = 10\log_{10}(\frac{P_{fault}}{P_{noise}}) \tag{4.4}$$

where $P_{fault}$ is the average power of the fault signal. The fault signal is the difference between the actual sensory output and the initial measurement distance, namely, the deviation from the baseline. As a result, with the same noise power, FNR is consistent with different measurement distances for a certain inspection scenario. Similarly, the fault severity on a signal can be computed by the Signal to Fault Ratio (SFR) with the derived fault signal, defined as

$$\mathrm{SFR} = 10\log_{10}(\frac{P_{signal}}{P_{fault}}) \tag{4.5}$$

These three ratios are related by

$$\mathrm{SNR} = \mathrm{SFR} + \mathrm{FNR} \tag{4.6}$$

In this case, when SNR = 50 dB, the FNR level is 5.15 dB. Considering the SNR level of the aforementioned commercial proximity sensor under this application scenario in practice is 50 dB, the default SNR value of 50 dB (FNR = 5.15 dB) will be adopted throughout the rest of the thesis unless otherwise specified.

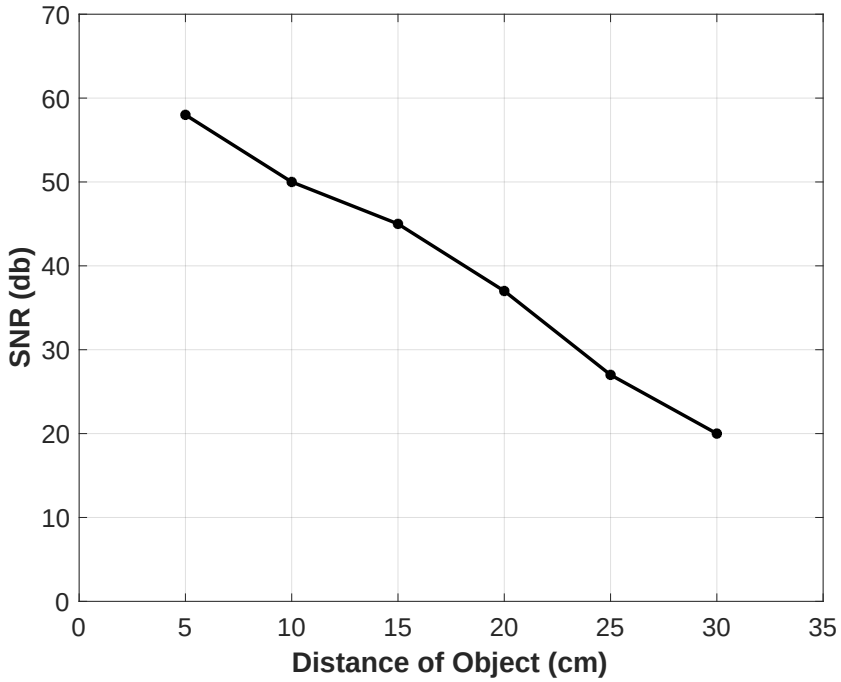



Figure 4.6: The relationship between the measurement distance and SNR in a typical commercial sensor [26].

# Chapter 5

# Results and Discussion

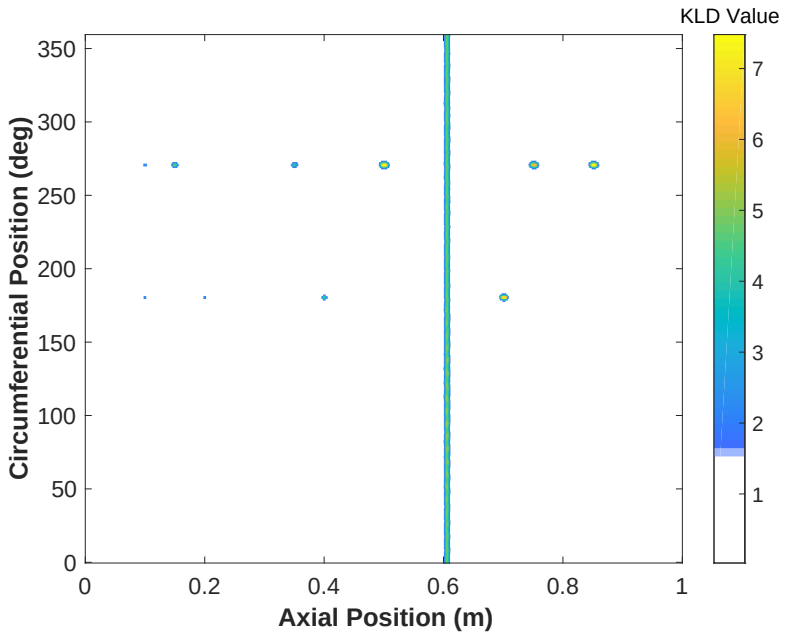
We show the implementation and application of the KLD filter by using the virtual pipeline inspection scenario described in Chapter 4 with the one meter long pipeline.

## 5.1 Results with KL Divergence Filter

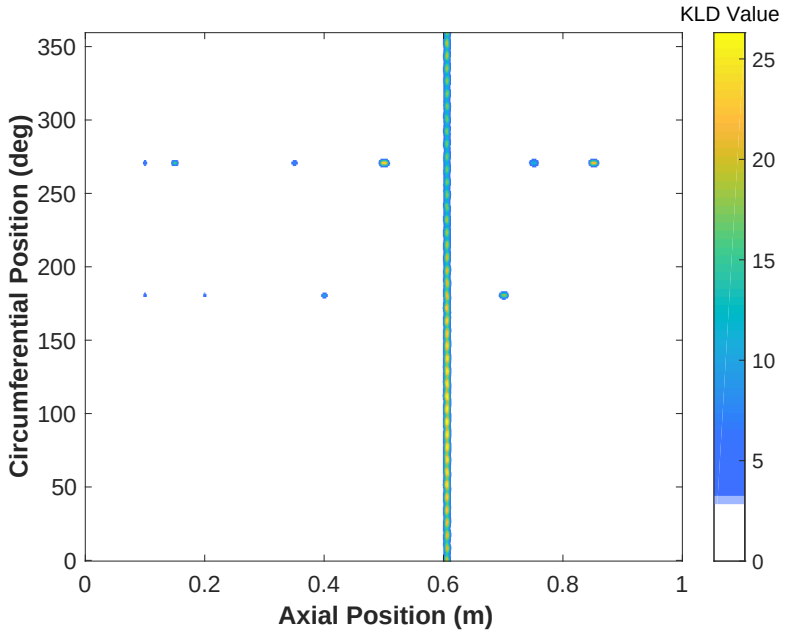
### 5.1.1 Model-Free Filter

The inner surface of the pipeline can be developed on a plane, as schematized in Fig. 4.1. Results in Fig. 5.1(a) show the output of the KL divergence filter presented in section 3.2. Since there are 67 possible outcomes in total, varying from 0.0963 m to 0.1030 m in every 0.1 mm, we set the number of bins $k$ of $Q$ to assign every outcome a probability to reveal the distribution of the entire data set. For $P_{ij}$, the number of bins $K$ is set as 60 to get a more accurate discrete probability density function, and the window size parameters are $w = 1$ and $l = 60$. Raw data is generated with SNR set as 50 dB, as described in detail in Chapter 4. Unless otherwise specified, all results have been generated with this set of parameters.

The color map shows the level sets of the KL divergence value associated to the raw distance sensory data, with white referring to background noise. This gives an immediate visualization of detected anomalies, and the shade of color correlates with their size, which in this case is related to the whole volume, relying on both the depth along the radial dimension and the diameter of the faced area of the holes, but mainly the latter when the noise level is low. Nevertheless, the relative sizes of the holes can be clearly distinguished. Wider anomalies, for example $\phi = 15$ mm, are represented in yellow, whereas shallower and/or smaller ones, for example $\phi = 5$ mm are represented in blue. The output of the filter also clearly identifies anomalies of different nature, distinguishing between holes (circles) and a weld (line across the circumferential direction). Both through holes and blind holes are all detected.




(a)




(b)

Figure 5.1: Result from (a) the model-free KLD filter and (b) KLD filter with Gaussian approximation.

### 5.1.2 Gaussian Approximation Filter

The KL Divergence filter with Gaussian approximation has been presented in section 3.3. Results from the same data set are presented in Fig. 5.1(b), with clear detection of anomalies. The parameters are consistent with the model free algorithm. In this case the Gaussian

approximation works well, with the advantage of gained simplicity in the implementation. The scale of the KLD filter output is different due to the different structure of the equations, see Chapter 3.

As discussed in Chapter 3 and confirmed by the results presented here, the KLD filter with Gaussian approximation can be very effective even if the distribution of the posterior $P_{ij}$ is poorly approximated by a Gaussian. This is due to the structure of equation (3.14), which shows how the approximated filter can still detect and classify different distributions as long as there are non negligible difference in the first two moments. The accuracy of the actual KL divergence values may be poor, but results can be relevant if one is content with classification and detection, at least as a first approximation. The simplicity of (3.14) should definitely be considered when evaluating trade-offs.

The results presented below are all obtained with the model-free approach. For the specific data sets, the same qualitative conclusions could be drawn by using the Gaussian approximation.

## 5.2 Anomaly Size Identification

As noted above, the KLD filter can identify the location of anomalies and classify their sizes (volume of the material) relative to each other. To better illustrate this feature, consider the simulated raw data in Fig. 5.2, which is the output of one channel of a distance sensor axially spanning a portion of a cylindrical pipeline while crossing six holes (anomalies), see Chapter 4 for details on sensor data generation. The holes have three different diameters, namely $\phi = 15\,\text{mm}$, $\phi = 10\,\text{mm}$, and $\phi = 5\,\text{mm}$. Two holes (the third and the sixth peak) are blind with 1 mm depth. The fifth anomalous signal, spanning the entire circumferential direction, is due to a welding. The sign of the anomaly from welding is the opposite as the holes since the weld is a protrusion.

We implement the model-free KLD filter to the raw data. The output of the KLD filter is shown in Fig. 5.3, which is a longitudinal section, taken at the circumferential position 270°, of the results in Fig 5.1(a). The blue curve shows a qualitative characterization of the relative sizes of the holes, with the relative height of the KL divergence peaks correctly classifying the anomalies from the original data, with the peaks of the two blind holes slightly lower than the peaks associated to the through holes with the same diameter. In engineering applications, the severity of a hole is typically determined based on two parameters: the diameter and the depth. If we categorize the anomalies according to their KLD magnitudes, the first one can be regarded as a small size hole, while the second and the third ones are medium size holes, and the last three holes are of a large size. Therefore, the size recognition of the anomalies under this regular SNR level is dominated by the diameter. But the differences in depth among the holes are also revealed by the magnitude. Although the anomalies in this specific example are quite pronounced also in the raw data, the KLD filter enhances the relative difference in peaks giving additional insights.

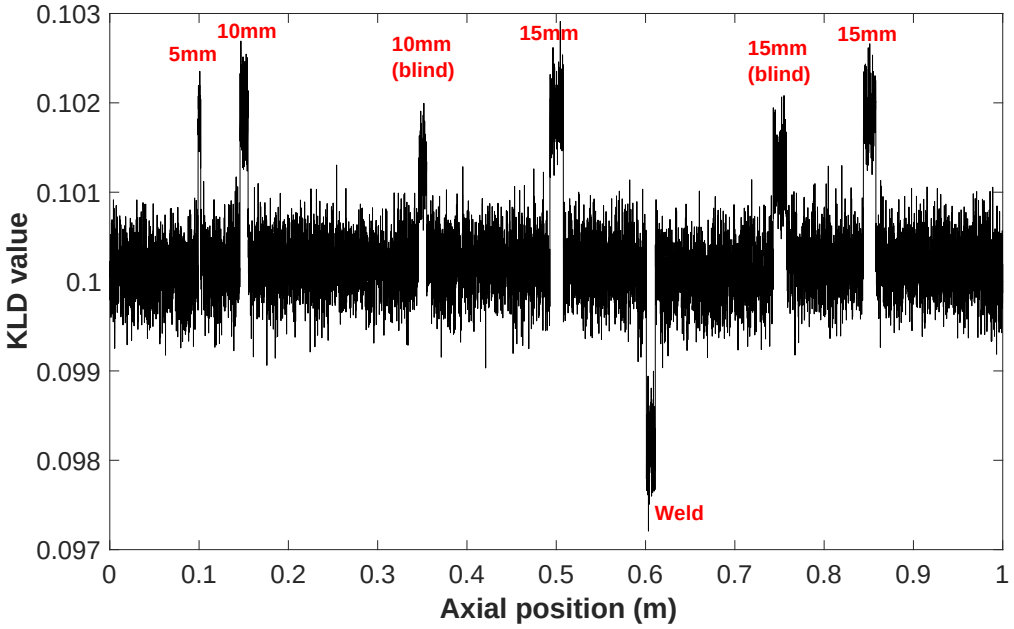



Figure 5.2: Raw sensory data of the data series along 270°.

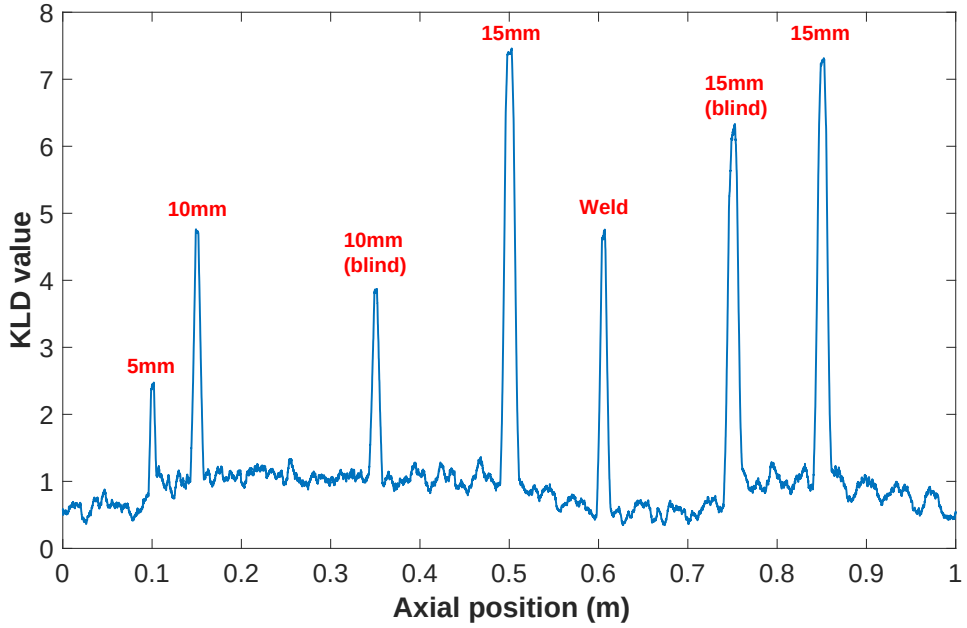



Figure 5.3: Classifications of the anomalies using the KLD filter on data from Fig. 5.2, where all the anomalies are detected and the relative size of the holes is qualitatively identified.

## 5.3 Qualitative Study of the Influence of Parameters

From the modelling in Chapter 3, we know there are four important parameters, namely the number of bins $k$ for the prior $Q$ and $K$ for the posterior $P_{ij}$, and the window size parameters $w$ and $l$. The window size parameters affect both the model free and the Gaussian approximation filters, whereas the binning parameters affect only the model free.

### 5.3.1 Number of bins

When choosing the number of bins to build a histogram from a given data set there are general considerations to account for. Clearly, if the number of bins $K$ is too small, one looses details on possible relevant trends in the frequencies of the occurrences of the data points; on the other hand, excessively atomized bins result in spurious local oscillations of the data that add unnecessary complexity without adding relevant information on the shape of the underlying distribution. At the limit, having only one bin prevents the filter to discriminate between data points since any notion of contiguity is lost in the histogram. Fig. 5.4 illustrates an output of the filter with only one bin based on the data series in Fig. 5.2, where it is clear the loss of performance with respect to cases presented above. Moreover, it is better to have more bins in most cases, but not exceed the limit determined by the resolution of the sensors. Another undesirable effect observed when adopting very few bins is that the filter cannot distinguish between very noisy signals and anomalies. An example with 2 bins is shown in Fig. 5.5 for the entire date set, where tremendous healthy data are recognized as anomalies. A way to determine an acceptable number of bins is to run the algorithm with different values of the parameters for a data set with known anomalies, observing the performance of the filter in terms of detection. For the cases presented here the SNR is 50 dB, and the number of bins $K$ is varied from 3 to 67, where as explained in Chapter 4, 67 is the number of possible outcomes for a distance measurement $\phi$ based on the sensor's resolution and the geometry of the workspace. Clearly this number varies with the problem, but in general it can be determined based on the specific device.

### 5.3.2 Window size

Another important parameter is the window size, defined as Fig. 3.5. The window size is crucial to confer locality to the filter, and ultimately being able to spatially distinguish features. The window size needs to be sufficiently large to include enough data points around a given sample to smooth out eventual local fluctuations that may not be representative, and concurrently it has to be sufficiently small to be able to represent relevant trends.

For the class of examples considered here, the window size is also important to determine the extension of the anomaly on the plane, that is, the diameter of the holes in the pipeline, which can be reduced to the problem of spatial resolution of the filter. If the window size if too small, the local posteriors $P_{ij}$ are all very similar regardless of the extension of the anomaly due to their local nature, and therefore classification can be more difficult.

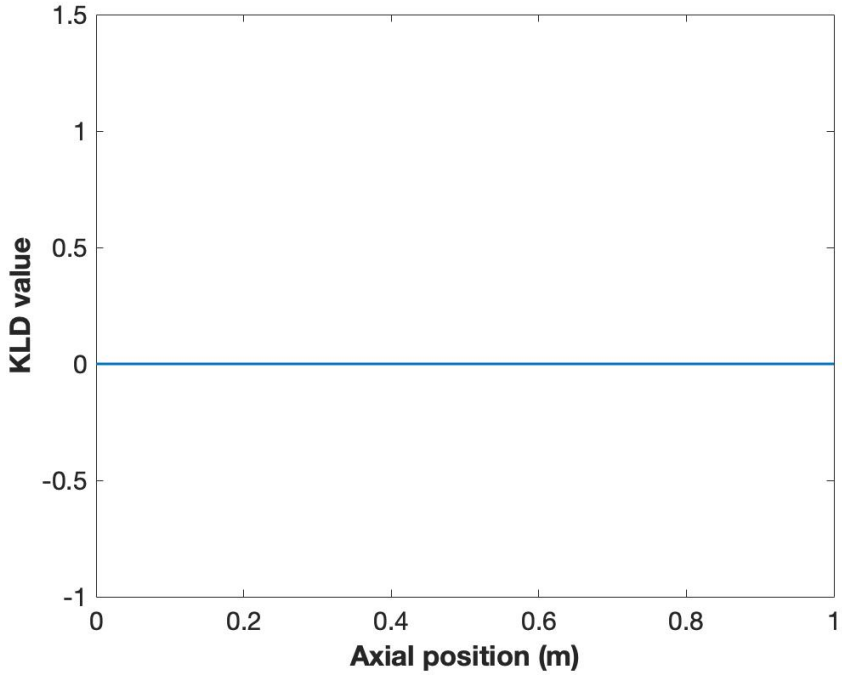



Figure 5.4: Output of the KLD filter with only one bin in building the histogram of $P_{ij}$ for the data series in Fig. 5.2, where the KLD values of all the data are close to zero and no information revealed.

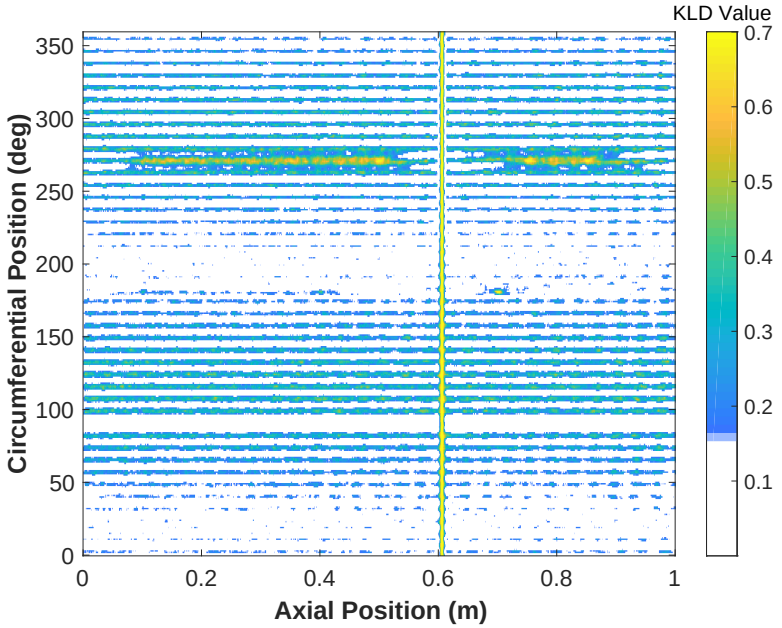



Figure 5.5: Output of the KLD filter with $K = 2$ bins in building the histogram of $P_{ij}$, which fails to detect the anomalies.

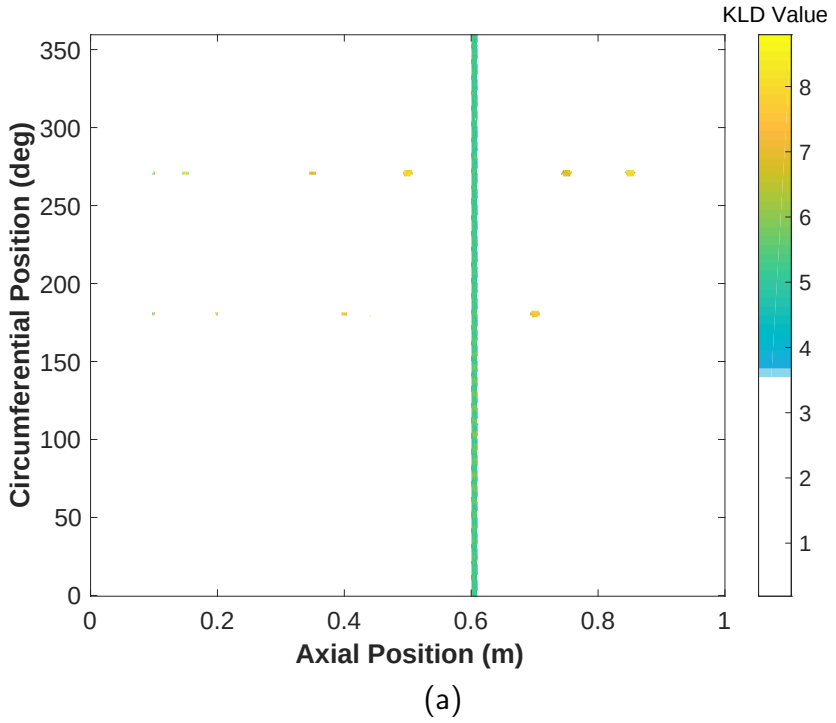
Results in Fig. 5.6 refer to two scenarios in which the performance of the filter is deteriorated by the choice of the window size. In (a), a narrow window size ($w = 1$, no cross channel correlation) results in all anomalies are classified as having the same size, and in (b) a large window size results in identifying some of the regions contiguous to the holes as anomalies, due to spurious spatial correlations.

For data sets with unknown anomalies, the algorithm should be calibrated on data sets with known anomalies that are of interest for detection, by setting the window size to be slightly smaller that the area of the anomalies of interest.

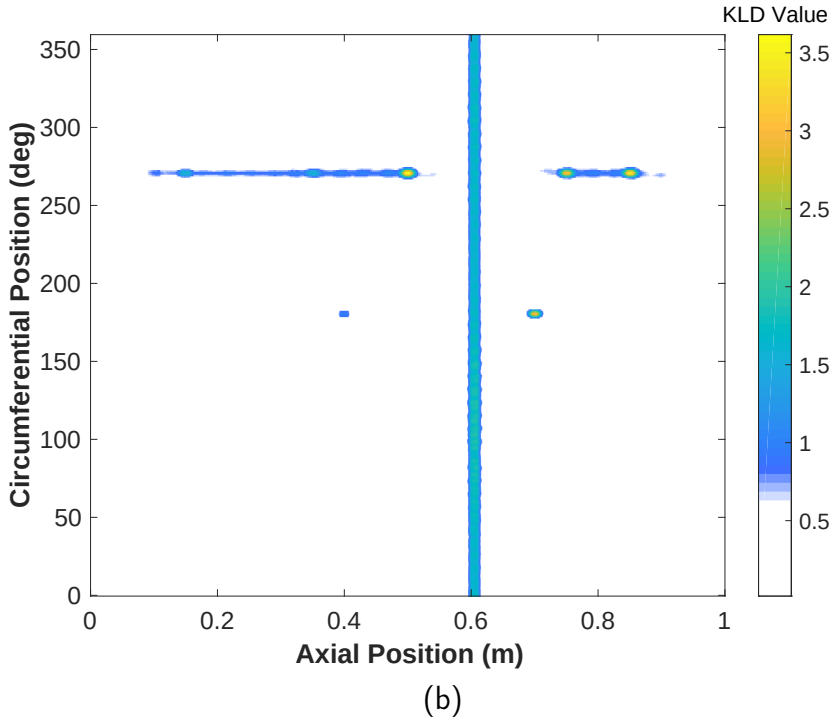
## 5.4 Signal to Noise Ratio

The level of background noise of in-pipe inspection depends on the sensors, pipeline surface finish and surface cleanliness. If the surface is clear and processed well, it is easy to detect the anomalies. However, the surface is expected to be non-ideal with defects in practical applications, for instance presenting corrosion or other contaminants that may contribute to a low SNR detection environment. As a result, the SNR influences the output of detection significantly. The relationship between SNR and the percentage of detection is plotted in Fig. 5.7 with 10 given holes as anomalies, where if a hole is not masked by the noise and can be recognized clearly, this anomaly is counted as one in the summation. The number of holes successfully detected has been lumped into the same counter, irrespective of the size. The welding part is neglected, as it spans a vast area along all the circumferential positions and it can always be detected. All the experiments in the statistics are run with the same window size ($121 \times 3$) and 60 bins in building the histogram for $P$. As shown, the percentage of detection decreases as the SNR decreases below 40 dB, and the noise masks the anomaly signals when the SNR is below 30 dB. Fig. 5.8 gives four different scenarios of the sensory data series along 270° under different SNR levels, where almost all the signals are masked by noise when SNR = 29 dB.

Besides the number of detected holes, the sizes of the anomalies are influenced by the inspection circumstance as well. Since the data points of the through holes with 2 mm depth are farther deviations from the noisy signals, the distribution of these samples has less similarity with the baseline, which contributes a higher KLD value, comparing with the blind holes with 1 mm depth. As a result, the blind holes are easier to be masked by the noise for low SNR, which is consistent with most engineering practice scenarios, as for instance when the shallower holes on gas pipes are much easier masked by imperfectly polished inner surfaces or residual stains. In the transition to low SNR, the output of the filter degrades by first presenting inaccurate holes' sizing as in Fig. 5.9, where results calculated with 40 dB show downsizing of the blind holes. The third blind $\phi$10 mm and fifth blind $\phi$15 mm holes are regarded as small size and medium size, respectively. In this case, the depth of the hole has a greater weight in terms of size recognition and the diameter is not the only dominating parameter any more, compared to the previous results in section 5.2. When the SNR goes lower, the relatively smaller holes, the first and second ones (the small size holes in the 40 dB scenario), are masked by background noise eventually as in Fig. 5.10 with results calculated for 34 dB for the same data series.




(a)




(b)

Figure 5.6: Outputs of the KLD filter with inappropriate window size, showing (a) the filter cannot identify the sizes with $l = 10$ and $w = 0$, where all ten holes are identified as the large size; (b) Some healthy signals around the holes are recognized as anomalies incorrectly and some small anomalies are missed due to the large window with $l = 200$ and $w = 3$.

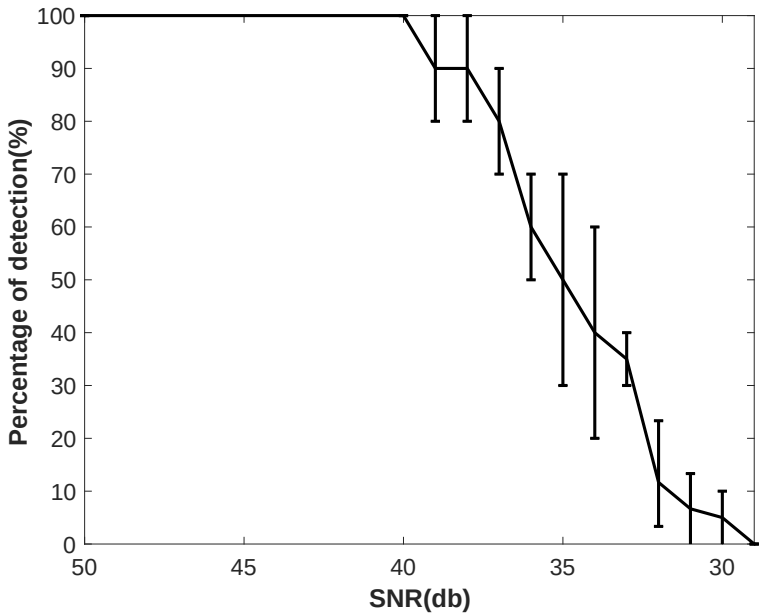



Figure 5.7: Percentage of detection of the holes with respect to SNR. The detected number of holes decreases with the SNR decreasing.

## 5.5 Comparison between Entropy Filter and KLD Filter

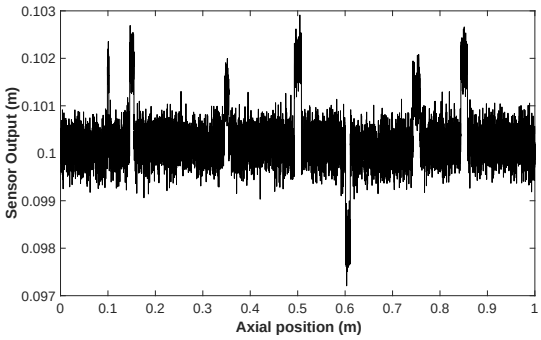
### 5.5.1 Result with Entropy Filter

A local entropy filter was proposed in [71] for the same class of problems, namely in-pipe anomaly detection with eddy current sensors. As in the KLD proposed here, the entropy filter is structured in two steps in which each raw data point is associated with a neighbor data set $\Phi(i, j)$, which is considered as the sample space for a probability distribution that in turn is used to map the initial data point into a local information entropy calculated through the formula (3.1).

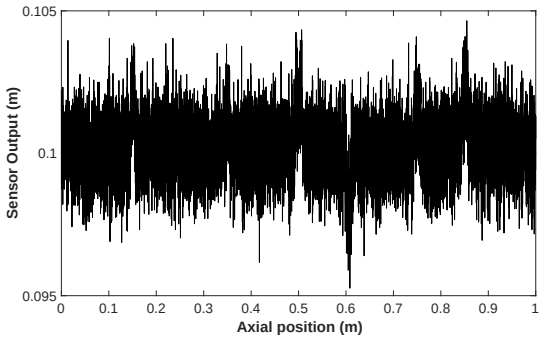
For the purpose of comparison, we implement the entropy filter on the data set with SNR $= 50\,\mathrm{dB}$, with results shown in Fig. 5.11, where the window size parameters are the same as the KLD filter. The colored areas indicate the locations and the sizes of the defects, with the difference that smaller entropy values correspond to anomalies which are generated by signals with higher information content. For the SNR considered here, both filters perform well in identifying anomalies.

### 5.5.2 Comparison with Binary Hypothesis Testing Classifier

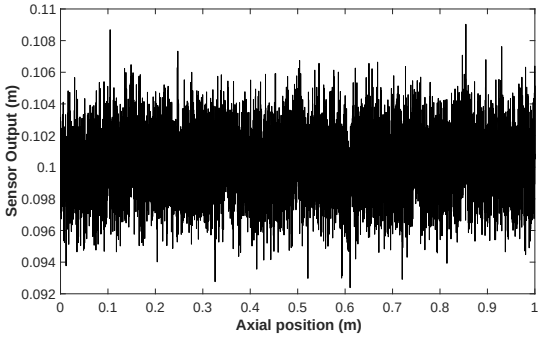
Low SNR environment testing is the key indicator for the effectiveness of a filter. The authors of the entropy filter tested the effectiveness of the algorithm with a binary hypothesis testing. Therefore, to be consistent with the original results in the paper, a comparison



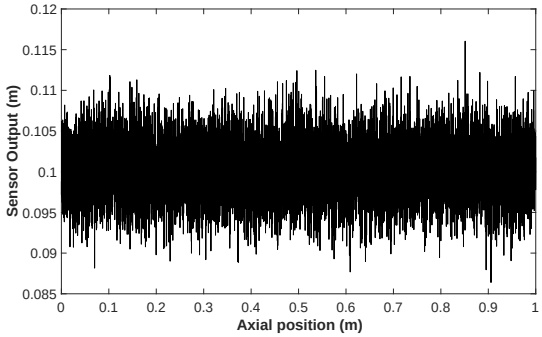

(a) SNR=50 dB, FNR=5.15 dB




(b) SNR = 40 dB, FNR=0.89 dB




(c) SNR = 34 dB, FNR=0.24 dB




(d) SNR = 29 dB, FNR=0.07 dB

Figure 5.8: Raw sensory data series along 270° in Fig. 5.2 under different SNR levels (the power of the fault signal is the same). When SNR is below 40 dB, the shapes of the anomalous signals are barely distinguishable.

between the KLD filter and the entropy filter with binary hypothesis testing classifier and SNR=39 dB environment is illustrated in Fig. 5.12. For this set of results, the entropy filter fails to detect the third blind hole at the circumferential position 270°. In this case the KLD filter is more sensitive to incipient anomalies due to the use of the baseline data set as the prior, which is an important feature of the KLD filter where the hypothesis testing classifier is embedded into the algorithm, as opposed to the entropy filter which requires an additional step since the entropy is an absolute measure of information rather than a relative one.

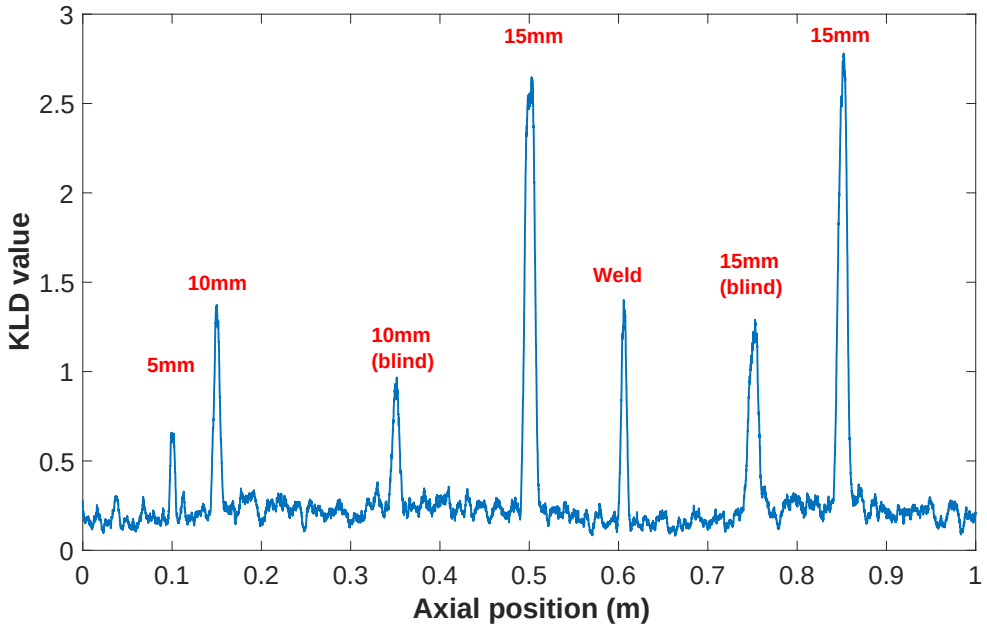



Figure 5.9: The magnitude of the blind holes is underestimated with 40 dB, as compared with the original plot in Fig. 5.3.

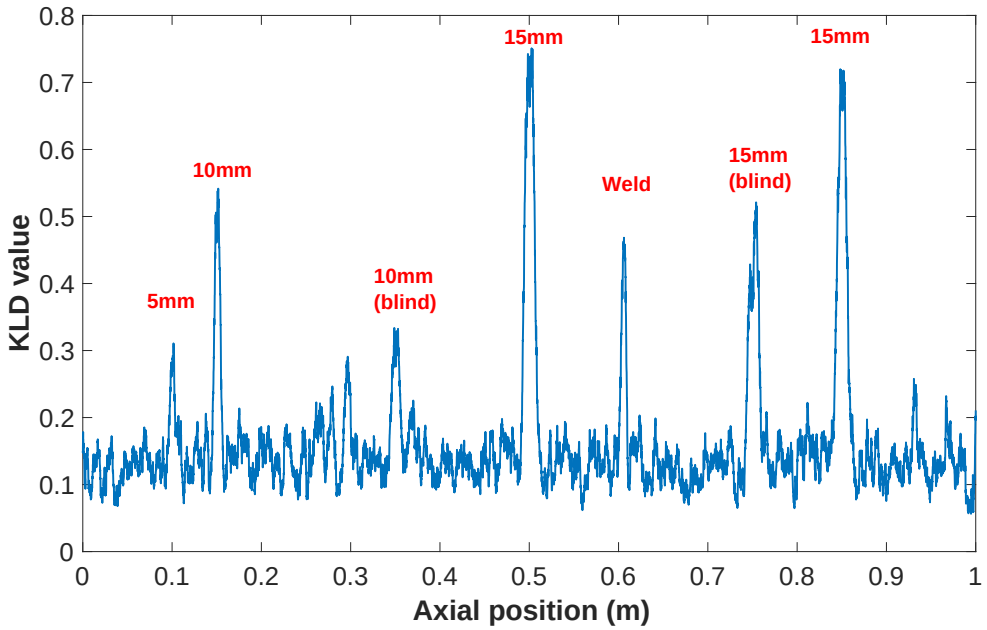



Figure 5.10: The first small size through hole and the second medium size blind hole are missed when SNR = 34 dB based on the same data series in Fig. 5.3.

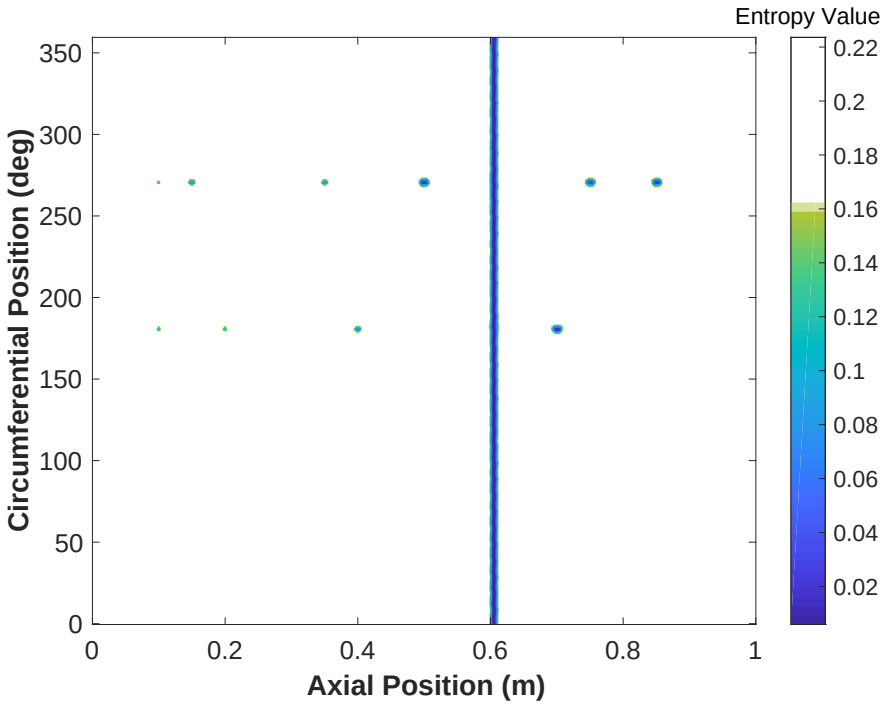
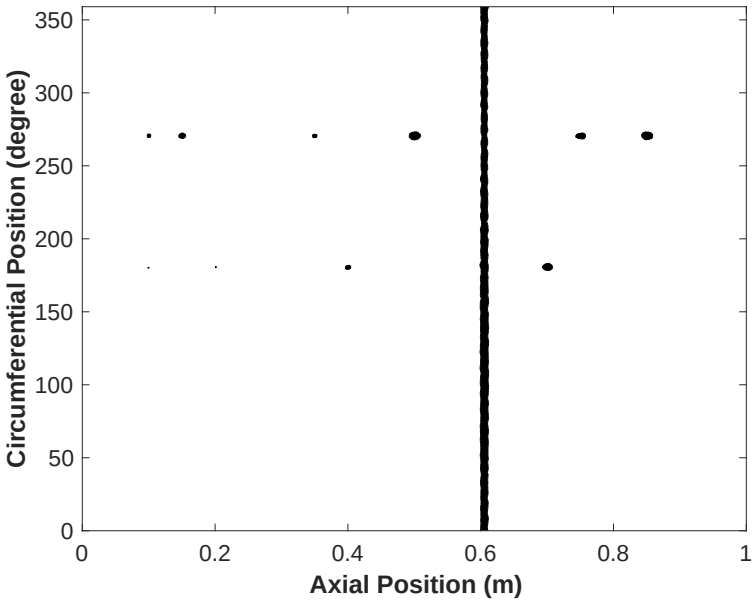
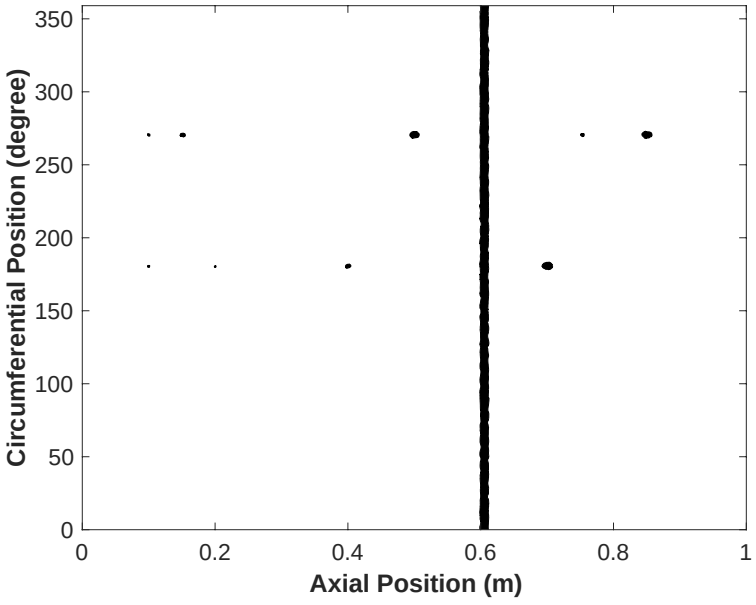



Figure 5.11: Results of the entropy filter on the same dataset of the KLD filter in Fig. 5.1.1.




(a) KLD filter




(b) Entropy filter

Figure 5.12: Comparison between the binary testing result with KLD filter and entropy filter, where the third hole along 270° direction is missed by the entropy filter when SNR = 39 db.

# Chapter 6

# Conclusions and Future Work

## 6.1 Conclusions

We proposed an anomaly detection algorithm with specific focus on data series produced by a class of sensing devices mounted on mobile robots for non destructive inspection of pipe-like environments. The anomaly detection algorithm maps the raw sensory data to a local Kullback–Leibler divergence value that measures the discrepancy between two probability distributions. Leveraging on this feature of the KL divergence, we build a baseline or prior probability distribution from the whole data set, and a posterior probability distribution form a subset of the all data centered at any given sample. The discrepancy between the two, as measured by the corresponding KL divergence, is a local measure of difference in information content between the data in the neighbor of the given sample, and the overall data set which is dominated by noise around the sensor's expected measurement. By repeating the procedure for every raw data sample point we obtain the output of the KL divergence filter.

The procedure just explained is model free in the sense that the local KL divergence is calculated on probability mass functions obtained from histograms of the data series. The advantage of this approach is that it does not rely on any assumption on the underlying distributions, with the downside of requiring to build histograms with the complexity of binning, which may vary due to the local nature of the algorithm with respect to the data set. For Gaussian probability distributions, the expression for the KL divergence simplifies into an explicit closed form that involves the first two statistical moments, therefore offering a very fast tool to identify anomalies. Although the Gaussian approximation for data with high information content (which typically characterizes anomalies in noisy signals) is not accurate, the KLD filter can still be effective if the baseline distribution and the local data distribution significantly differ in the first two moments. Comparisons between the two versions of the filter on the same data set are shown, demonstrating the effect just discussed.

The algorithm is tested on a simulated data set from a distance sensing device (proximity sensor) spanning a one meter long pipe with 36 channels along the circumferential

direction. Results show that the algorithm is able to identify the locations and qualitatively classify the sizes of a set of anomalies (holes on the pipe) with raw data generated with SNR of 50 dB, which is a typical value in experimental conditions. The diameter and depth of the defects are both detected. As expected, the performance of the algorithm deteriorates as the SNR decreases, with narrower and shallower holes more likely to be missed. Since the SNR is a crucial parameter in anomaly detection applications, its effect on the detection success ratio has been parametrically investigated, showing that below a specific threshold the detection rate is virtually unaffected by the SNR. The effect of algorithm parameters has also been briefly considered, specifically in terms of bin numbers for the histograms generating probability mass functions needed to calculate the KL divergence, and in terms of window size, that is the locality parameter of the filter. A comparison with a similar information theoretical filter based on entropy has also been drawn, showing that the KLD filter performs better with lower SNR signals.

Moreover, the algorithm has also been validated with a real experimental data set from a pipeline anomaly detection company, generated using a snake-like robot and Remote Field Eddy Current Sensors. The real metal pipe consists of two parts, with 11 and 16 anomalies, respectively, located along two circumferential positions on each part. The two components are connected by a welding ring. The application of the KLD filter results in a correct anomaly detection. However, the results cannot be shown here not to violate the company's intellectual property rights.

Overall, the proposed algorithm is promising in detecting anomalies for the class of sensory data considered here.

## 6.2 Future Work

In this work, the data set is synthetic. It is software-generated from simulation scenarios. A real experiment based on this simulation scenario can be devised to further validate this algorithm with randomized anomalies on different pipes in the future. Besides, the field of anomaly detection is very fervent given the number of possible applications that are made possible by advances in computing, sensing, image processing, and applied machine learning among others. Here we indicate two possible further research directions that focus on information theoretical tools, which belong to the same framework of the work presented in the thesis. Moreover, the algorithm in the thesis has been applied offline, but it will be interesting to consider extensions to online applications to integrate it within decision making loops in autonomous mobile sensor.

### 6.2.1 Other Information-theoretic Measures

From the perspective of divergence measures, KLD is one of the widely used divergence in the family of $f$-divergence. There are some similar divergence metrics, such as Jensen-Shannon Divergence, for instance. All these metrics have the potential to be employed in anomaly detection applications for different types of data sets. Moreover, in the family of

entropy, conditional entropy and relative conditional entropy also showed their effectiveness in this field. But, they have not been used in structural defect identification. For example, conditional entropy has been proved to be a good candidate for multi hypothesis testing with several baselines to identify the types of the anomalies. In the future, a comparison of the results between these information measure metrics can be made with different kinds of signals to figure out their advantages and disadvantages. Based on the conclusions, the main features of different information-theoretic measure metrics can be adopted for choosing the most adequate anomaly detection approach. This can also help finding out the best match for the underlying dataset. For example, if there are many zeros in the distributions when measuring the discrepancy between two distributions, then JS Divergence may be a better choice and the KLD should be avoided.

### 6.2.2 Online Detection Applications

In this work, we use the offline dataset to show the functionality of the KLD filter, which heavily relies on the knowledge of the entire data set. The prior distribution must be derived over all the data points before computing the KLD. How to obtain a reasonable prior based on pure noisy data or the signals of some unknown healthy parts of the pipeline with the historic data sets can be a fruitful research avenue. With the prior, the KLD filter may be able to be used in some online detecting applications. Furthermore, The KLD showed its strong capacity to detect incipient and small anomalies in information theory and communication application. As long as there is a slight difference in the data series, there is a large deviation showing on the curve of the KLD values. If the prior can be set appropriately, the KLD filter can be used on a running system and the data can be collected for analysis simultaneously. As such, the structural defects or the condition monitoring anomalies can be revealed timely and accurately. Of course, other information-theoretic measures may also be used in such online applications if the above requirements are met.

# APPENDICES

# Appendix A

# Simulation with CoppeliaSim

The simulation scenario is built with the robotic simulator CoppeliaSim. Expect the pipe which is imported from Solidworks, there are three main components in the sensing mechanism, proximity sensor, carrier and joints. Here we list the detailed parameters of the simulation environment.

Proximity sensor is the most important part in this simulation, which determines the type and quality of the collected dataset. We use the built-in LaserPointer sensors in the software, and the parameters are set as table A.1. And all the 36 channels are set up with the same parameters.

Table A.1: The setting parameters of proximity sensors.

| Sensor subtype | Laser |
|---|---|
| Volume type | Ray |
| Range | 0.4 m |
| Offset | 0 m |
| Detection | Front and back face detection |

Two joints provide the power and possibility for the rotational and translational movement of the mechanism in the simulation, which are functioning as the motors plus some supporting parts in practice. One is the prismatic joint, and it drives the carrier move forward. The other one is a revolute joint, which makes the rotation of the carrier possible. The parameters of these two joints can be found in table A.2.

The parameters of carrier are easy to define among there main parts of the sensing mechanism. The process is as follows. First, choose a cylinder and set the sizes as needed, for instance, 0.1 m diameter with 2 m long; second, change the property of the cylinder to 'Body is dynamic', and define an approximate value for the mass, 0.05 kg in this case; last, attach all the proximity sensors to the carrier.

Regarding other parts and setting environment, the sizes of the base are set as needed and other parameters of it are default. We have a 0.2 m wide,0.2 m long and m high cuboid as the base. Considering the dynamic setting influences the quality of collected

Table A.2: The setting parameters of the joints.

| Prismatic Joint | |
|---|---|
| Mode | Torque/force mode |
| Position range | 1 m |
| Range | 0.4 m |
| Target velocity | 0.01 m/s |
| Maximum force | 10e6 N |
| **Revolute Joint** | |
| Position is cyclic | Yes |
| Position range | 1 m |
| Target velocity | 100 deg/s |
| Maximum torque | 2.5 × 10e8 N*m |

data. to make sure the distance values are accurate, we change the Dynamic Contents as the following table A.3. And a graph in the scenario is used to collect the data from the sensors. All the data stream are shown and stored on the graph, and then they are exported as an Excel file to do the analysis.

Table A.3: The setting of dynamic contents in CoppeliaSim.

| | |
|---|---|
| Dynamics Engine | Newton |
| Dynamic setting | Very accurate |
| Simulation time step | dt = 10 ms |

# Appendix B

# Matlab Code for the KLD Filter

## B.1 Model-free KLD filter code

The Model-free KLD filter code from reading the data to the outputs is shown as below:

```
% read the data
d0=xlsread('10.27_data_all.xls');
d=d0(2:10001,2:37);
t=d0(2:10001,1)-0.01;
[ns1,ns2] = size(d);
if d(1,1) == 0
    d(1,:) = 0.1*ones(1,36); %36 sensors
end

%calculate the x&y
v = 0.01; w = 100; R = 0.1; l = 360;
% define the parameters according to setting in the simulation
y0 = [0:10:350];
D = zeros(ns1,360);
for i = 1:size(d,2)
     x = [1:nd1]';
     pos = x - y0(i); % x = w*t by accident
     if pos < 0
        y = pos + 360;
     else
        y = fix(pos - fix(pos/l)*l)+1;
     end

    for z=1:size(y,1)
          if y(z)<=0
             y(z)=y(z)+360;
          end
```

```
        end
D(sub2ind(size(D), x, y)) = d(:,i);
end
% map the data to corresponding postion
for n = 1:size(D,2)
        for m = 1:size(D,1)
              if D(m,n) == 0
                    if m == 1
                        D(m,n) = 0.1;
                    else
                        D(m,n) = D(m-1,n);
                    end
              end
      end
end

disp('Building discrete pdfs for KLD computation')
A=awgn(D,50,'measured'); % add Gaussian noise to the signal
RawD = roundn(A,-4); % define the resolution of the sensors
[nd1,nd2] = size(RawD);

%Obtaining the distribution of Q over the entire data set
X = [min(min(RawD)):0.0001:max(max(RawD))];
w = size(X,2); % define the number of bins k for Q
max_valueR = max(max(RawD)); min_valueR = min(min(RawD));
binQ = (max_valueR - min_valueR)./w;
edgesQ = min_valueR:binQ:max_valueR; %edges for Q
Q = histcounts(RawD,edgesQ,'Normalization', 'probability');

% assigning the number of bins and window size for P
prompt_in = {'Number of bins to build PDF',
'Number of data samples in each channel',
'Number of contigous channels to be included'};
title_entr = 'Input for KLD computation';
def_val = {'60','60','1'};
ui = inputdlg(prompt_in,title_entr,1,def_val);
K = round(str2double(ui(1,1)));
samp = round(str2double(ui(2,1)));
lc = round(str2double(ui(3,1)));

%Convert Q from k bins into K
for m = 1: K
       f = find(X>=edgesP(m) & X<edgesP(m+1));
      Q_P(m) = sum(Q(1,f));
end
```

```
% Define the boundaries of the data windows to compute the KLD
low_b = 1;
hi_b = size(RawD,1);

bH1 = (1:size(RawD,1)) - round(samp/2);
bH2 = (1:size(RawD,1)) + round(samp/2);

for i = 1:size(RawD,1)
     if (bH1(i) < low_b)
        bH1(i) = low_b;
        bH2(i) = min(hi_b , bH2(i) + (i - low_b));
     end

    if (bH2(i) > hi_b)
       bH2(i) = hi_b;
       bH1(i) = max(low_b , bH1(i) - (hi_b - i));
    end
end

%KL Divergence computing
KLD_d = zeros(size(RawD));
k = -lc:lc;
for j = 1:nd2
     currentChannels = mod(j-1+k,360)+1;
     for i = 1:nd1
          currentP = RawD(bH1(i):bH2(i), currentChannels);
          P = histcounts(currentP,edgesP,'Normalization', 'probability');
          for n = 1: size(P,2)
               if (P(n) == 0)
                  KL_n = 0;
               else
                  KL_n = P(n)*log(P(n)/Q_P(n));
               end
               KLD_d(i,j) = KLD_d(i,j) + KL_n;
          end
     end
end

% plot the 2D result of the KLD filter
figure
imagesc(XX,Y,KLD_d');
set(gca,'YDir','normal');
xlabel('Axial Position (m)','FontWeight','bold');
ylabel('Circumferential Position (deg)','FontWeight','bold');
```

```
set(gca,'FontSize',15);
colorbar
h=colorbar;
set(get(h,'Title'),'string','KLD Value');
```

## B.2 KLD filter with Gaussian approximation code

The Gaussian approximation KLD filter is a special case of the model-free algorithm. The only difference in code is the KL divergence computing part.

```
% computing the KLD values with Gaussian approximation
Mean=zeros(size(RawD));
STD=zeros(size(RawD));

%calculate mean and standard deviation for P
for j = 1:nd2
      currentChannels = mod(j-1+k,360)+1;
      for i = 1:nd1
            currentP = RawD(bH1(i):bH2(i), currentChannels);
            subset = reshape(currentP,[],1);
            Mean(i,j) = mean(subset);%compute the mean
            STD(i,j) = std(subset);% compute the standard deviation
       end
end

% compute the mean and standard deviation for Q
Q_m = mean(mean(Mean));
Q_std = mean(mean(STD));
% KL divergence equation with Gaussian approximation
KLD = 0.5*(2*log(Q_std./STD)+(STD./Q_std).^2+((Mean-Q_m)./Q_std).^2-1);

% plot the 2D result
figure
imagesc(XX,Y,KLD');
set(gca,'YDir','normal');
xlabel('Axial Position (m)','FontWeight','bold');
ylabel('Circumferential Position (deg)','FontWeight','bold');
set(gca,'FontSize',15);
colorbar
h=colorbar;
set(get(h,'Title'),'string','KLD Value');
```

# References


[1] Mostafa Afgani, Sinan Sinanovic, and Harald Haas. Anomaly detection using the Kullback-Leibler divergence metric. In *2008 First International Symposium on Applied Sciences on Biomedical and Communication Technologies*, pages 1–5, Aalborg, Denmark, October 2008. IEEE.

[2] Mostafa Afgani, Sinan Sinanovic, and Harald Haas. Information Theoretic Approach to Signal Feature Detection for Cognitive Radio. In *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, pages 1–5, New Orleans, LA, USA, 2008. IEEE.

[3] Mostafa Afgani, Sinan Sinanovic, and Harald Haas. Hardware implementation of a Kullback-Leibler Divergence based signal anomaly detector. In *2009 2nd International Symposium on Applied Sciences in Biomedical and Communication Technologies*, pages 1–6, Bratislava, November 2009. IEEE.

[4] Mostafa Z Afgani. *Exploitation of Signal Information for Mobile Speed Estimation and Anomaly Detection*. PhD thesis, The University of Edinburgh, 2011.

[5] Fatemeh Amiri, MohammadMahdi Rezaei Yousefi, Caro Lucas, Azadeh Shakery, and Nasser Yazdani. Mutual information-based feature selection for intrusion detection systems. *Journal of Network and Computer Applications*, 34(4):1184–1199, 2011.

[6] A. Anderson and H. Haas. Kullback-leibler divergence (kld) based anomaly detection and monotonic sequence analysis. In *2011 IEEE Vehicular Technology Conference (VTC Fall)*, pages 1–5, Sep. 2011.

[7] M. Angeli, P. Burrascano, E. Cardelli, S. Fiori, and S. Resteghini. Classification of eddy current NDT data by probabilistic neural networks. In *IJCNN'99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339)*, volume 6, pages 4012–4014, Washington, DC, USA, 1999. IEEE.

[8] Chrisil Arackaparambil, Sergey Bratus, Joshua Brody, and Anna Shubina. Distributed monitoring of conditional entropy for anomaly detection in streams. In *2010 IEEE International Symposium on Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW)*, pages 1–8. IEEE, 2010.

[9] Andreas Arning, Rakesh Agrawal, and Prabhakar Raghavan. A linear method for deviation detection in large databases. In *KDD*, volume 1141, pages 972–981, 1996.

[10] Yong Bai and Qiang Bai. Pipeline Inspection and Subsea Repair. In *Subsea Pipeline Integrity and Risk Management*, pages 73–99. Elsevier, 2014.

[11] Dmitry I. Belov and Ronald D. Armstrong. Distributions of the Kullback-Leibler divergence with applications: Distributions of the Kullback-Leibler divergence. *British Journal of Mathematical and Statistical Psychology*, 64(2):291–309, May 2011.

[12] Harold Berger and Symposium on Nondestructive Testing Standards, editors. *Nondestructive testing standards: present and future.* Number 1151 in STP / ASTM. ASTM International, Philadelphia, Pa, 1992.

[13] R. Bickerstaff, M. Vaughn, G. Stoker, M. Hassard, and M. Garrett. Review of sensor technologies for in-line inspection of natural gas pipelines. *Sandia National Laboratories, Albuquerque, NM*, 2002.

[14] J. Blitz. *Electrical and Magnetic Methods of Nondestructive Testing.* CRC Press, January 1991.

[15] Richard J Bolton, David J Hand, et al. Unsupervised profiling methods for fraud detection. *Credit Scoring and Credit Control VII*, pages 235–255, 2001.

[16] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104, 2000.

[17] Christian Callegari, Stefano Giordano, and Michele Pagano. An information-theoretic method for the detection of anomalies in network traffic. *Computers & Security*, 70:351–365, 2017.

[18] A.A. Carvalho, J.M.A. Rebello, L.V.S. Sagrilo, C.S. Camerini, and I.V.J. Miranda. MFL signals and artificial neural networks applied to detection and classification of pipe weld defects. *NDT & E International*, 39(8):661–667, December 2006.

[19] Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*, 2019.

[20] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3):15, 2009.

[21] Hongtian Chen, Bin Jiang, and Ningyun Lu. An improved incipient fault detection method based on Kullback-Leibler divergence. *ISA Transactions*, 79:127–136, August 2018.

[22] Coppelia Robotics. CoppeliaSim. https://www.coppeliarobotics.com/, Accessed March 4, 2020.

[23] T. M. Cover and Joy A. Thomas. *Elements of information theory.* Wiley-Interscience, Hoboken, N.J, 2nd ed edition, 2006.

[24] A Davies. *Handbook of Condition Monitoring Techniques and Methodology*. Springer Netherlands, Dordrecht, 1998.

[25] Manuel Davy and Simon Godsill. Detection of abrupt spectral changes using support vector machines an application to audio signal segmentation. In *Acoustics, Speech, and Signal Processing (ICASSP), on International Conference*, volume 2, pages II–1313. IEEE, 2002.

[26] TI Designs. *Proximity Sensing of up to 30-cm Range With > 15-dB SNR and Robust Capacitive Touch Reference Design*, 2015.

[27] Lounis Douadi, Davide Spinello, and Wail Gueaieb. Dynamics and Control of a Planar Multibody Mobile Robot for Confined Environment Inspection. *Journal of Computational and Nonlinear Dynamics*, 10(1):011005, September 2014.

[28] Lounis Douadi, Davide Spinello, Wail Gueaieb, and Hassan Sarfraz. Planar kinematics analysis of a snake-like robot. *Robotica*, 32(05):659–675, August 2014.

[29] Haimonti Dutta, Chris Giannella, Kirk Borne, and Hillol Kargupta. Distributed top-k outlier detection from astronomy catalogs using the demac system. In *Proceedings of the 2007 SIAM International Conference on Data Mining*, pages 473–478. SIAM, 2007.

[30] Levent Ertöz, Michael Steinbach, and Vipin Kumar. Finding topics in collections of documents: A shared nearest neighbor approach. In *Clustering and Information Retrieval*, pages 83–103. Springer, 2004.

[31] Xu Fang, Wenhao Guo, Qingquan Li, Jiasong Zhu, Zhipeng Chen, Jianwei Yu, Baoding Zhou, and Haokun Yang. Sewer pipeline fault identification using anomaly detection algorithms on video sequences. *IEEE Access*, 8:39574–39586, 2020.

[32] Ryohei Fujimaki, Takehisa Yairi, and Kazuo Machida. An approach to spacecraft anomaly detection problem using kernel feature space. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 401–410, 2005.

[33] Frank E Grubbs. Procedures for detecting outlying observations in samples. *Technometrics*, 11(1):1–21, 1969.

[34] Anis Hamadouche, Abdelmalek Kouadri, and Azzedine Bakdi. A modified Kullback divergence for direct fault detection in large scale systems. *Journal of Process Control*, 59:28–36, November 2017.

[35] Hyung Seok Han, Jae Jong Yu, Chan Gook Park, and Jang Gyu Lee. Development of inspection gauge system for gas pipeline. *KSME international Journal*, 18(3):370–378, 2004.

[36] Jinane Harmouche, Claude Delpha, and Demba Diallo. Incipient fault detection and diagnosis based on kullback–leibler divergence using principal component analysis: Part i. *Signal Processing*, 94:278–287, 2014.

[37] Jinane Harmouche, Claude Delpha, and Demba Diallo. Incipient fault amplitude estimation using KL divergence with a probabilistic approach. *Signal Processing*, 120:1–7, March 2016.

[38] Simon Hawkins, Hongxing He, Graham Williams, and Rohan Baxter. Outlier detection using replicator neural networks. In *International Conference on Data Warehousing and Knowledge Discovery*, pages 170–180. Springer, 2002.

[39] Zengyou He, Xiaofei Xu, and Shengchun Deng. Discovering cluster-based local outliers. *Pattern Recognition Letters*, 24(9-10):1641–1650, 2003.

[40] C. Hellier. *Handbook of nondestructive evaluation*. McGraw-Hill, 2001.

[41] Paul Helman and Jessie Bhangoo. A statistically based system for prioritizing information exploration under uncertainty. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27(4):449–466, 1997.

[42] Shao-Fei Jiang, Chun-Ming Zhang, and CG Koh. Structural damage detection by integrating data fusion and probabilistic neural network. *Advances in Structural Engineering*, 9(4):445–458, 2006.

[43] Mojtaba Rostami Kandroodi, Babak Nadjar Araabi, Majid Nili Ahmadabadi, Farshad Shirani, and Maisam Mansoob Bassiri. Detection of natural gas pipeline defects using magnetic flux leakage measurements. In *2013 21st Iranian Conference on Electrical Engineering (ICEE)*, pages 1–6, Mashhad, Iran, May 2013. IEEE.

[44] Rajesh Keshwani and Shiben Bhattacharya. Design and optimization of eddy current sensor for instrumented pipeline inspection gauge. *Sensor Review*, 28(4):321–325, September 2008.

[45] A. Khodayari-Rostamabad, J.P. Reilly, N.K. Nikolova, J.R. Hare, and S. Pasha. Machine Learning Techniques for the Analysis of Magnetic Flux Leakage Images in Pipeline Inspection. *IEEE Transactions on Magnetics*, 45(8):3073–3084, August 2009.

[46] Teuvo Kohonen. Self-organizing maps: Ophmization approaches. In *Artificial neural networks*, pages 981–990. Elsevier, 1991.

[47] Yuliya Kopylova, Duncan A Buell, Chin-Tser Huang, and Jeff Janies. Mutual information applied to anomaly detection. *Communications and Networks, Journal of*, 10(1):89–97, 2008.

[48] Salomon Kullback. *Information Theory and Statistics*. Wiley & Sons, 1959.

[49] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

[50] A Verdugo Lazo and P Rathie. On the entropy of continuous probability distributions (corresp.). *IEEE Transactions on Information Theory*, 24(1):120–122, 1978.

[51] Tony T. Lee. An infornation-theoretic analysis of relational databases—part i: Data dependencies and information metric. *IEEE Transactions on Software Engineering*, 10:1049–1061, 1987.

[52] Pål Liljebäck, Kristin Ytterstad Pettersen, Øyvind Stavdahl, and Jan Tommy Gravdahl. *Snake robots: modelling, mechatronics, and control*. Springer Science & Business Media, 2012.

[53] Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151, 1991.

[54] Qingwang Luo, Yibing Shi, Zhigang Wang, Wei Zhang, and Yun Zhang. Imaging of Local Defects of Pipes Based on Deconvolution Technology. In *19 th World Conference on Non-Destructive Testing 2016*, page 8, 2016.

[55] G Manson. Identifying damage sensitive, environment insensitive features for damage detection. In *Proceedings of IES Conference*, 2002.

[56] Graeme Manson, S Gareth Pierce, Keith Worden, Thomas Monnier, Philippe Guy, and Kathryn Atherton. Long-term stability of normal condition data for novelty detection. In *spie the international society for optical engineering*, pages 323–334. International Society for Optical Engineering; 1999, 2000.

[57] Graeme Manson, SG Pierce, and Keith Worden. On the long-term stability of normal condition for damage detection in a composite panel. *Key Engineering Materials*, 204:359–370, 2001.

[58] Zahra Mezni, Claude Delpha, Demba Diallo, and Ahmed Braham. Bearings Ball Fault Detection Using Kullback Leibler Divergence in the EMD Framework. In *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, pages 729–734, Chongqing, October 2018. IEEE.

[59] Josep M. Mirats Tur and William Garthwaite. Robotic devices for water main in-pipe inspection: A survey. *Journal of Field Robotics*, 27(4):491–508, June 2010.

[60] Kevin P. Murphy. *Machine learning: a probabilistic perspective*. Adaptive computation and machine learning series. MIT Press, Cambridge, MA, 2012.

[61] JB Nestleroth. Pipeline in-line inspection challenges to ndt. *insight-wigston then northampton*, 48(9):524, 2006.

[62] Caleb C Noble and Diane J Cook. Graph-based anomaly detection. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 631–636, 2003.

[63] Feng Pan and Weinong Wang. Anomaly detection based-on the regularity of normal behaviors. In *2006 1st International Symposium on Systems and Control in Aerospace and Astronautics*, pages 6–pp. IEEE, 2006.

[64] Claudio Piciarelli, Danilo Avola, Daniele Pannone, and Gian Luca Foresti. A Vision-Based System for Internal Pipeline Inspection. *IEEE Transactions on Industrial Informatics*, 15(6):3289–3299, June 2019.

[65] Istvan Pinter, Lorant Kovacs, Andras Olah, Rajmund Drenyovszki, David Tisza, and Kalman Tornai. Application of jensen-shannon divergence in smart grids. In *Proceedings of 5th Scientific and Expert Conference TEAM*, volume 2013, pages 287–290, 2013.

[66] Sirikarn Pukkawanna. *Unsupervised Anomaly Detection in Massive Traffic Using S-transform and Renyi Divergence*. PhD thesis, Nara Institute of Science and Technology, 2015.

[67] Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 547–561, Berkeley, CA, 1961. University of California Press.

[68] Piervincenzo Rizzo. Water and Wastewater Pipe Nondestructive Evaluation and Health Monitoring: A Review. *Advances in Civil Engineering*, 2010:1–13, 2010.

[69] Osman Salem, Farid Naït-Abdesselam, and Ahmed Mehaoua. Anomaly detection in network traffic using jensen-shannon divergence. In *2012 IEEE International Conference on Communications (ICC)*, pages 5200–5204. IEEE, 2012.

[70] C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[71] F. Sheikhi, D. Spinello, and W. Gueaieb. Rényi Entropy Filter for Anomaly Detection With Eddy Current Remote Field Sensors. *IEEE Sensors Journal*, 15(11):6399–6408, November 2015.

[72] Farid Sheikhi. Entropy filter for anomaly detection with eddy current remote field sensors., 2014.

[73] Arie Sheinker, Nizan Salomonski, Boris Ginzburg, Lev Frumkis, and Ben-Zion Kaplan. Magnetic anomaly detection using entropy filter. *Measurement science and technology*, 19(4):045205, 2008.

[74] J. Silva and S. Narayanan. Average divergence distance as a statistical discrimination measure for hidden markov models. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(3):890–906, May 2006.

[75] A Sophian, G Y Tian, D Taylor, and J Rudlin. Electromagnetic and eddy current NDT: a review. *Insight*, 43(5):5, 2001.

[76] Davide Spinello, Wail Gueaieb, and Roderick Lee. Entropy filter for anomaly detection with eddy current remote field sensors. In *2011 IEEE International Symposium on Robotic and Sensors Environments (ROSE)*, pages 95–100, Montreal, QC, Canada, September 2011. IEEE.

[77] K. Suzumori, S. Wakimoto, and M. Takata. A miniature inspection robot negotiating pipes of widely varying diameter. In *2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422)*, volume 2, pages 2735–2740. IEEE, 2003.

[78] Tim Van Erven and Peter Harremos. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

[79] G Vijay, P Srinivasa Pai, and N Sriram. Artificial neural network based condition monitoring of rolling element bearing using vibration signals'. *Advances n Mechanical Engineering*, page 449, 2010.

[80] Wenke Lee and Dong Xiang. Information-theoretic measures for anomaly detection. In *Proceedings 2001 IEEE Symposium on Security and Privacy. S&P 2001*, pages 130–143, Oakland, CA, USA, 2001. IEEE Comput. Soc.

[81] Y. Y. Yao. Information-Theoretic Measures for Knowledge Discovery and Data Mining. In *Entropy Measures, Maximum Entropy Principle and Emerging Applications*, volume 119, pages 115–136. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003. Series Title: Studies in Fuzziness and Soft Computing.

[82] Abdulrahman Youssef, Claude Delpha, and Demba Diallo. An optimal fault detection threshold for early detection using Kullback–Leibler Divergence for unknown distribution data. *Signal Processing*, 120:266–279, March 2016.

[83] Jiusun Zeng, Uwe Kruger, Jaap Geluk, Xun Wang, and Lei Xie. Detecting abnormal situations using the Kullback–Leibler divergence. *Automatica*, 50(11):2777–2786, November 2014.

[84] M H Zgoul and S M Habali. An Investigation into Plastic Pipes as Hot Water Transporters in Domestic and Industrial Applications. *JJMIE*, 2(4):11, 2008.