MASTER'S THESIS 2020

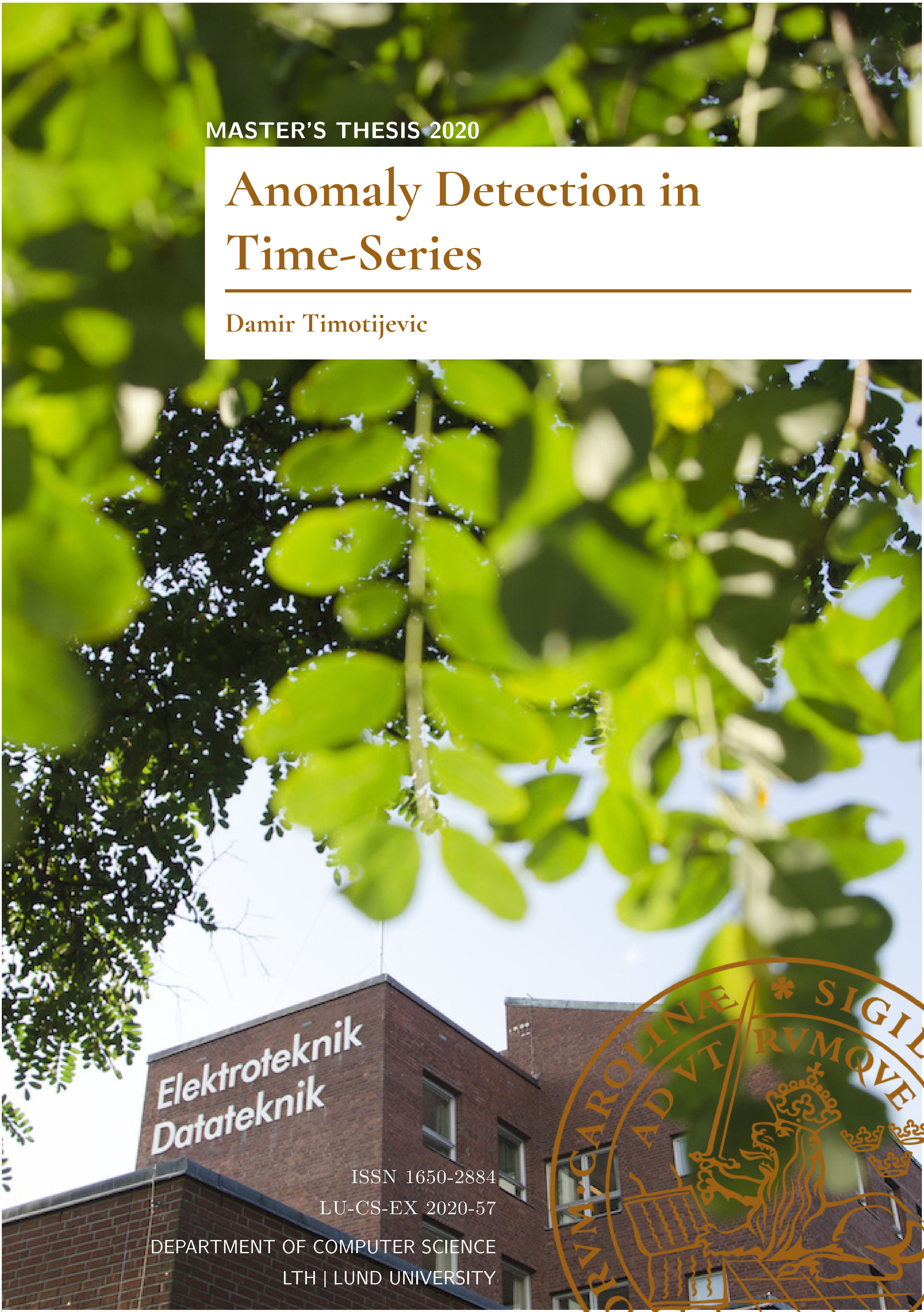
Anomaly Detection in Time-Series

Damir Timotijevic

ISSN 1650-2884

LU-CS-EX 2020-57

DEPARTMENT OF COMPUTER SCIENCE
LTH | LUND UNIVERSITY



EXAMENSARBETE
Datavetenskap

LU-CS-EX 2020-57

Anomaly Detection in Time-Series

Damir Timotijevic

Anomaly Detection in Time-Series

(An Application to Robot Joint Maintenance)

Damir Timotijevic
d.timotijevic@hotmail.com

September 10, 2020

Master's thesis work carried out at
the Department of Computer Science, Lund University.

Supervisors: Bernhard Wullt, bernhard.wullt@abb.se.com
Pierre Nugues, pierre.nugues@cs.lth.se

Examiner: Jacek Malec, jacek.malec@cs.lth.se

Abstract

Industrial robots are ubiquitous in the manufacturing industry and save human labor in highly monotonous and dangerous tasks. They are designed to operate for long time periods but over time their performance and behaviour eventually deteriorates due to wear. Early detection of abnormal behaviour is therefore of crucial importance when ensuring high availability and low operating costs in robotic systems.

In this thesis I propose a data-driven method for detecting anomalies caused by excessive wear levels in joints of industrial robots. The method uses data that is collected during a repeated diagnostic routine. It is assumed that executions of the routine where no fault is present will deviate from those in which a fault is developing. Significant deviations that are indicative of a developing fault are detected by comparing data from a fault-free execution of the routine to all succeeding executions in the distribution domain. The proposed method uses kernel density estimation to transform the data to the distribution domain and Kullback-Leibler divergence to quantify the deviation between probability distributions. To reduce the sensitivity to disturbances, the method uses an algorithm, which combines low-pass filtering and an accumulated sum of deviations (CUSUM).

The effectiveness of the proposed method is demonstrated by evaluation on real world datasets. The method is shown to outperform a benchmark algorithm in metrics such as hit rate, precision and false alarm rate.

Keywords: anomaly detection, condition monitoring, industrial robots, kernel density estimation, Kullback-Leibler divergence

Acknowledgements

I would like to thank Pierre Nugues for his guidance throughout this thesis work.

A special thanks goes out to my future colleague and former supervisor Bernhard Wullt who always provided encouragement and ideas with unwavering enthusiasm and excitement during these uncertain times of a pandemic.

Also, I would like to thank my beloved girlfriend, for all her support, motivation and poorly hidden disinterest to the contents of this thesis.

Contents

Chapter 1

Introduction

A central feature of every industrial revolution is the adoption of new technologies, ranging from the steam-powered machines in the *First industrial revolution* to microelectronics technology in the *Third industrial revolution*. Lately the concept of *Industry 4.0* has emerged as an umbrella term for a range of technologies, including the connected sensors and devices that constitute the Internet of Things (IoT), big data, and cloud computing approaches (Pereira and Romero, 2017). The emergence of these technologies has sparked an initiative in many industries to digitize products and service offerings (Geissbauer et al., 2016). One industry that stands to gain from such initiatives is the automated manufacturing industry, where industrial robots are used to ensure qualitative and quantitative requirements. By collecting and evauluating relevant machine data maintenance can be planned for and unexpected failures, which can cause significant down-times and stops to entire production lines, can be avoided.

A common approach to ensure high availability in industrial robots is by performing *preventive scheduled maintenance*. The scheduled maintenance is often based on rough estimations of the robot's components durability (Bittencourt and Axelsson, 2014). While this approach ensures high availability, it has a downside of high maintenance costs as potentially needless scheduled on-site inspections and maintenance actions are performed by service engineers. Another, more contemporary approach, is to remotely monitor the state of the system with embedded sensors, prescribing maintenance actions based on indications of impending equipment failure or declining performance. This approach, if correctly implemented, has the potential to decrease the amount of on-site inspections and unnecessary maintenance actions performed by service engineers. It thus leads to a reduction in the maintenance costs. However, it also comes with the challenge to accurately assess the condition of the system. Since not all faults are predictable, a practical implementation would be limited to faults that evolve gradually, meaning that they can be detected and serviced before causing complete breakdown.

Wear is typically a process that has a gradual evolution in industrial robots. In robot joints wear is observed between interacting components such as gearboxes and motor shafts, resulting in progressive loss of material due to relative motion. The increase of wear debris in the lubricant layer between the components is often a good indicator of wear but demands on-site visits by service technicians. Another approach is to monitor the friction due to its well known relation with wear. However, the construction of accurate friction models demands time-consuming experiments that can take years to conduct. This will thus not be the focus of this thesis. Instead, the aim is to use a data-driven approach, as opposed to a model-based approach, and indirectly infer about wear by monitoring changes to available sensory information.

1.1 What is anomaly detection?

We can define the task of anomaly detection as a classification problem, where the objective is to determine whether collected observations differ significantly from normal behaviour of a system (Pimentel et al., 2014). Almost synonymous to anomaly detection is the term *outlier detection*, which has a slightly broader definition. *Outliers* are often categorized as being either noise (weak outliers) or anomalies (strong outliers). Noise is therefore a weak form of outlier that doesn't have a large enough deviation from normal behaviour to be considered significant. As an example, points at the edge of a cluster can be considered as noise. (Aggarwal, 2016). In this thesis, we will investigate sensory information from robot joints and attempt to detect significant deviations caused by wear-related processes. Throughout the thesis, we will denote such deviations as anomalies.

In the domain of classification problems, anomaly detection can be seen as a single-class classification problem (Pimentel et al., 2014), where a single class (positive instance indicating normal behaviour) needs to be distinguished from all other possibilities. The typical setting that distinguishes anomaly detection from other classification problems is a highly imbalanced dataset with a well-sampled positive class and few or none present anomalies. This highly imbalanced dataset is often a consequence of

1. Costly and time-consuming to produce anomalous behaviours;

2. Measurements of normal behaviour are inexpensive and easily obtainable.

Because of this highly imbalanced dataset, the aim is to construct models describing a system's normal behaviour. As new observations are collected, they are compared to the model to produce an anomaly score, also called a *test quantity*. If the test quantity exceeds a predefined threshold, the observation is classified as an anomaly (Aggarwal, 2016).

1.2 Problem framing

The overarching goal of this thesis is to find or construct a method for anomaly detection and evaluate its applicability in detecting critical changes due to wear-related phenomena in joints of industrial robots. Usage of real world data is a requirement in evaluating the applicability.

A more formal and mathematical problem definition follows. Let $\mathcal{D}$ be a sequence of K batches with multiple time series collected from sensors during a repeated diagnostic routine, also called a *test cycle*, in a robotic system where

$$\mathcal{D} = \{X^{(1)}, X^{(2)}, \ldots, X^{(K)}\}$$

in which $X^{(k)} \in \mathbb{R}^{M \times N}$ is the k-th batch consisting of a sample with size M of N different system signals.

We will use the shorthand notation $\mathbf{x}_n^{(k)} = X_{*,n}^{(k)} = \left[x_{1,n}^{(k)}, \quad x_{2,n}^{(k)}, \quad \ldots, x_{M,n}^{(k)}\right]^T$ to denote the n-th system signal of the k-th batch. Here $\mathbf{x}_n^{(k)}$ could refer to a time series of the motor torque or some other system signal that may be related to the wear process in a robot joint.

Assuming that we are given one or multiple non-faulty test cycle measurements of the system signal $\mathbf{x}_n^{(0)}$, our goal is to find an approach to derive features, according to $f(\mathbf{x}_n^{(k)}) \colon \mathbb{R} \mapsto \mathbb{H}$, that can be useful in detection of anomalies. The derived features will be used to produce an anomaly score z, also called a *test quantity*, by comparing measurements $\mathbf{x}_n^{(k)}$ to the measurement $\mathbf{x}_n^{(0)}$, for which it is assumed that the system was in normal condition, in the $\mathbb{H}$-domain according to:

$$z(\mathbf{x}_n^{(k)}) \triangleq d(f(\mathbf{x}_n^{(0)}), f(\mathbf{x}_n^{(k)}))$$

where d denotes a distance function and $z \in \mathbb{R}^+$.

For example, suppose that $\mathbb{H}$ is the distribution domain, then a distance function d could be the Kullback-Leibler divergence $D_{\text{KL}}\left(p_n^{(0)} \| p_n^{(k)}\right)$. Further, the goal is to improve detection performance of wear-related changes while reducing the sensitivity to other disturbances that have an effect on the sensory information $\mathbf{x}_n^{(k)}$.

Chapter 2
Theory and Background

2.1 Industrial robots

An industrial robot is mechanically constructed from *links* which are connected by various type of *joints*. All joints connect exactly two links and are categorized by how they constrain the relative motion of the links. Robots are often categorized according to their *degrees of freedom* (DOF). Typically this refers to the number of joints that allow for rotational motion about the joint axis, also known as *revolute joints*. In industrial robots, the most common configuration is 6 DOF with links that are serially connected through revolute joints, which is often referred to as an 6 DOF elbow manipulator. Usually industrial manipulators are equipped with a hand or a gripper, called an *end-effector*, which makes interaction with the surrounding environment possible. The first three axes, denoted as the *main axes*, are used to position the end-effector and the remaining three, *wrist axes*, to achieve the desired orientation (Lynch and Park, 2017).

2.1.1 Actuation, gearing and sensors

A robot relies on *actuators* to provide the desired forces and torques for motion. Typically, these are geared electric DC motors located at each joint. A power supply converts AC-voltage to DC-voltage that is modulated by a power amplifier associated with each motor. A controller box that takes user input in the form of a desired trajectory along with positional feedback given from encoders located at each joint calculates what torque is required from each motor. The voltage is continually adjusted across the motor with the help of a current sensor to achieve the requested torque. The torque produced by a DC motor is given by

$$\tau = k_t I, \tag{2.1}$$

where k_t is the torque constant (Nm/A) and I is the current that goes through the windings of the motor's stator or rotor (Lynch and Park, 2017).

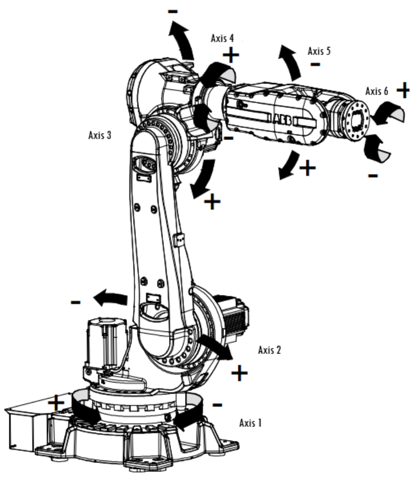


Figure 2.1: ABB IRB 6620, 6 DOF elbow manipulator (ABB, 2019).

A *rotary encoder*, also called a shaft encoder, is placed at the extending shaft of the motor to measure the angular position of the joint. The high accuracy of the angular measurements allows for differentiation and thus estimates of speed and acceleration. At the output of the motor, a *gearbox transmission* provides the high torque that is needed in robotic applications while also reducing speed. A common type is the *rotary vector reducer* which is usually found in the main axes of the manipulator due to its high reduction ratio and compactness. In the wrist axes where accuracy is of key importance, harmonic drive gearboxes, also called strain-wave gearing, are more common due to their high precision and zero backlash (Lynch and Park, 2017).

2.1.2 Friction and wear in robot joints

Friction is the tangential reactive force that emerges between surfaces in relative motion to each other. Unlike gravity, friction is not considered to be a fundamental force but arises from interaction between surfaces on a nanoscale level and depends among other things on the topology of the contact surfaces, torque levels, temperature, lubrication properties and velocity. Due to the inherent complexity of modelling these dependencies from first principles, friction models are often based on empirical observations (Al-Bender and Swevers, 2009).

The interest for friction in the context of industrial robotics stems mainly from two fields: motion control and fault detection. Accurate friction models can lead to better control stability and accuracy, as in Bona and Indri (2006) and Kim et al. (2009). Since friction in robot joints relates to wear, it is also studied in the field of fault detection (Bittencourt and Axelsson, 2014).

The combined effect of friction, wear and lubrication is studied in the interdisciplinary field of *tribology*. In the field of tribology, friction is often described as a function of speed in what is known as a *Stribeck curve*. It shows that lubricated friction is a nonlinear phenomenon that that can be divided into three different regions: boundary lubrication (BL), mixed lubrication (ML) and hydrodynamic lubrication (HL). For low speeds, a high friction is observed

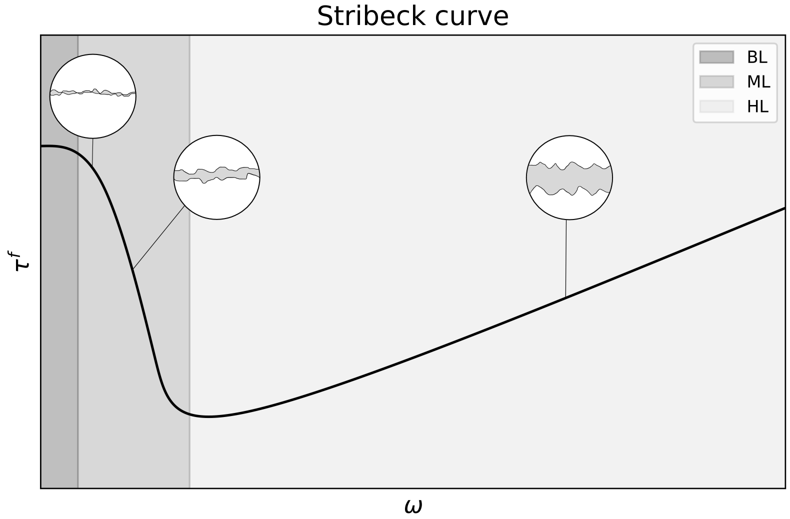


Figure 2.2: A typical Stribeck curve for a robot joint that shows friction torque τ^f as a function of the angular speed ω.

due to the roughness of the operating surfaces that are in direct contact (BL). As the speed increases, the thickness of the lubrication layer between the surfaces increases, resulting in a decrease in contact friction (ML). When the thickness of the lubrication layer reaches its maximum, the surfaces are completely separated by the lubrication layer and thus the friction becomes proportional to the force necessary to shear to lubrication layer, which depends on lubrication properties such as viscosity (HL). Wear mostly occurs in the BL and ML regions because of the contact friction between the surfaces, although in robot joints, it can also occur in the EHL region due to the high reduction ratios of the gearboxes (Bittencourt and Gunnarsson, 2012).

The following factors have been identified in experiments to have an effect on friction in robot joints; see Bittencourt (2012) for details:

Load: Increased load results in a general increase of friction in all regions of the friction curve due to the increased contact pressure between the surfaces, but a relatively larger increase can be observed in for lower speeds in the BL region.

Lubricant: Lubrication properties such as thickness and viscosity affect the magnitude of the forces needed to shear the lubrication layer. Higher viscosity results in higher friction levels in the HL-region.

Temperature: Temperature affects the viscosity of the lubricant with a higher temperature leading to a lower viscosity and thus a lower friction in the HL-region. In the BL and ML regions, higher temperatures result in higher friction which is believed to be caused by dilation in the components.

Wear: Increased friction results in increased wear debris in the form of higher concentration of metallic particles in the lubricant. This is noticed as a generalized increase in the friction curve for all regions.

Backlash: Backlash is the amount of free space or clearance caused by gaps between gear teeth. Increased backlash results in lower contact pressure between mating gears and thus a lower friction.

To detect excessive wear levels in robot joints, it is thus possible to monitor friction since increased wear can lead to increased friction. Further, as the applied torque τ from the motor of a given joint must overcome the increased friction, it is possible to indirectly infer the wear level by monitoring the torque τ, which can be estimated from the current I in the motor. However, Bittencourt (2012) shows that variations in load torques and lubricant temperatures can have a comparable effect on friction to that of wear which makes the task of accurately determining the wear level more challenging. As excessive wear levels or faults often appear as amplitude changes, a common approach is to monitor for changes in the distribution domain. See for example Bittencourt et al. (2014) and Liu et al. (2017).

2.2 Density estimation

A central topic in statistics and a way to derive features in the distribution domain is by estimation of a *probability density function* p, which is the problem of recovering the underlying probability density that generated the observations $x_1, x_2, \dots x_n$. In non-parametric density estimation, the goal is to estimate p with none or as few assumptions as possible. The estimator is often denoted as $\hat{p}$. To get a sense of the quality of $\hat{p}$, i.e. how well it estimates p, we need to define what is meant by a *good* estimate. One possible approach to quantify it is by calculating or estimating the *mean squared error* (MSE), which is defined as:

$$\text{MSE}(\hat{p}(x)) = \mathbb{E}[(p(x) - \hat{p}(x))^2] \tag{2.2}$$

Note that MSE of $\hat{p}(x)$ is a function of x. A global error measure is then given by the *integrated mean square error* (IMSE):

$$\begin{aligned} \text{IMSE}(\hat{p}) &= \mathbb{E} \int_{\mathbb{R}} [p(x) - \hat{p}(x)]^2 dx = \int_{\mathbb{R}} \text{MSE}(\hat{p}(x)) dx \\ &= \int_{\mathbb{R}} \text{Bias}^2[\hat{p}(x)] dx + \int_{\mathbb{R}} \text{Var}[\hat{p}(x)] dx \end{aligned} \tag{2.3}$$

which is sometimes referred to as the L_2-risk function. Thus a "good" estimate is one that has a overall small average squared deviation from p (Wasserman, 2006).

2.2.1 Histogram

In this section, we follow the notation outlined in Wasserman (2006). The simplest non-parametric approach for density estimation is the *histogram*. For simplicity's sake and without loss of generality, we assume that $p(x)$ is only non-zero within the interval $[0, 1]$, i.e. $x_i \in [0, 1]$, $i = 1, \dots, n$. Further we assume that $p(x)$ is continuous and that the derivative $p'(x)$ is bounded, meaning that $|p'(x)| \leq L$. The histogram is then constructed by partitioning the interval $[0, 1]$ into M non-overlapping and evenly spaced **bins**

$$B_1 = \left[0, \frac{1}{M}\right), B_2 = \left[\frac{1}{M}, \frac{2}{M}\right), \dots, B_{M-1} = \left[\frac{M-2}{M}, \frac{M-1}{M}\right), B_M = \left[\frac{M-1}{M}, 1\right]$$

If we define the **bin width** as $h = 1/M$ the histogram estimate for a given point $x \in B_k$ is given by

$$\hat{p}(x) = \frac{1}{nh}\sum_{i=1}^{n} \mathbf{1}\{x_i \in B_k\} = \frac{n_k}{nh}, \quad t_k \leq x < t_{k+1} \tag{2.4}$$

where n_k is the **bin count** and t_k and t_{k+1} is the left- and right-hand endpoints of the bin B_k. The random variable n_k is distributed according to the binomial distribution, i.e. $n_k \sim \text{Binomial}(n, p_k)$ where the **bin probability** $p_k(x)$ for a fixed point $x \in B_k$ is given by

$$p_k(x) = \int_{t_k}^{t_{k+1}} p(y)dy \tag{2.5}$$

The expected value of the binomially distributed variable n_k is np_k and its variance is given by $np_k(1 - p_k)$.

Analysis of the histogram

Using a Taylor expansion at x, we get for all points $y \in B_k$ that

$$p(y) = p(x) + p'(x)(y - x) + O(h^2), \tag{2.6}$$

where we use the fact that $|x - y| \leq h$ to get $O(h^2)$. Using Equation 2.6, we can rewrite Equation 2.5 as

$$\begin{aligned} p_k(x) &= \int_{t_k}^{t_{k+1}} (p(x) + p'(x)(y - x) + O(h^2))dy \\ &= hp(x) + h^2\left(\frac{1}{2} - \frac{x - t_k}{h}\right)p'(x) + O(h^3) \end{aligned} \tag{2.7}$$

where the factor $(1/2 - (x - t_k)/h) \in \left(\frac{-1}{2}, \frac{1}{2}\right)$. The expectation of $\hat{p}_k$ is then given by

$$\mathbb{E}[\hat{p}(x)] = \frac{p_k}{h} = p(x) + \frac{1}{2}hp'(x) - p'(x)(x - t_k) + O(h^2) \tag{2.8}$$

and thus the bias is

$$\text{Bias}[\hat{p}(x)] = \mathbb{E}[\hat{p}(x)] - p(x) = h\left(\frac{1}{2} - \frac{x - t_k}{h}\right)p'(x) + O(h^2) \tag{2.9}$$

Here we see that there is less bias if we increase the number of bins $M = 1/h$. Intuitively this is reasonable since more bins lead to a higher resolution and thus we are able to approximate $p(x)$ better. The variance of the estimator at x is given by:

$$\begin{aligned} \text{Var}[\hat{p}(x)] &= \frac{\cancel{n}p_k(1 - np_k)}{n^{\cancel{2}}h^2} = \frac{p_k(1 - np_k)}{nh^2} \\ &= \frac{\Big(hp(x) + O(h^2))(1 - O(h)\Big)}{nh^2} \\ &= p(x)/(nh) + O(1/n) \end{aligned} \tag{2.10}$$

which shows an increase in variance as the number of bins grows. Since the MSE can be expressed as a sum of variance and squared bias, we get that

$$\text{MSE} = \frac{p(x)}{nh} + \frac{1}{4}h^2 p'(x)^2 + p'(x)^2(x - t_k)^2 - hp'(x)^2(x - t_k) + O(1/n + h^3) \tag{2.11}$$

It follows that MSE $\to 0$ if $h \to 0$ as $n \to \infty$. Another implication of the Equations 2.9, 2.10 and 2.11 is that when h is large, the variance is small and the bias is large and the other way around. This means that the optimally chosen bin width h^* needs to balance the bias and variance terms. In order to achieve the fastest rate of convergence, these terms need to approach zero at the same speed, otherwise the term with the slowest speed would dominate.

To achieve the same speed, h^* needs to fulfill the condition $O(h^3) = O(1/n)$ resulting in $h^*_{MSE} = O(1/n)^{1/3}$. It can be shown, see Scott (1979) for details, that numerical integration of Equation 2.11 over $\mathbb{R}$ yields:

$$\text{IMSE} = \frac{1}{nh} + \frac{h^2}{12}\int_{-\infty}^{\infty} p'(x)^2 dx + O\left(1/n + h^3\right) \tag{2.12}$$

To find the optimal bin width h^* we differentiate Equation 2.12 with respect to h and set the result equal to zero which yields

$$h^* = \left(\frac{6}{\displaystyle\int_{-\infty}^{\infty} p'(x)^2 dx} \right)^{1/3} n^{-1/3} \tag{2.13}$$

which is the asymptotically optimal choice of bin width. We can conclude that the bins should be narrow if p has rapid changes, i.e. $\int p'(x)^2 dx$ is large. But nonetheless we see that h should shrink as $n^{-1/3}$ in order to achieve the fastest convergence rate. Plugging h^* into Equation 2.12 yields

$$\text{IMSE} = O\left(n^{-2/3}\right) \tag{2.14}$$

In the next section, we will show that we can improve on this convergence while also getting a smooth estimate of $p(x)$ by using what is called a *kernel density estimator*.

2.2.2 Kernel density estimator

The kernel density estimator $\hat{p}(x)$ is formally defined as:

$$\hat{p}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h} K\left(\frac{x - x_i}{h}\right) \tag{2.15}$$

in which K is a *kernel function* and h is the *bandwidth* of the kernel. Generally all functions fulfilling the following conditions

$$\text{(i) } K(x) \geq 0 \qquad \text{(ii) } \int_{\mathbb{R}} K(x)dx = 1 \qquad \text{(iii) } \int_{\mathbb{R}} xK(x)dx = 0 \qquad \text{(iv) } \int_{\mathbb{R}} x^2K(x)dx < \infty$$

can be used as kernel functions. Here condition (iii) means that $K(x)$ is symmetric, i.e.

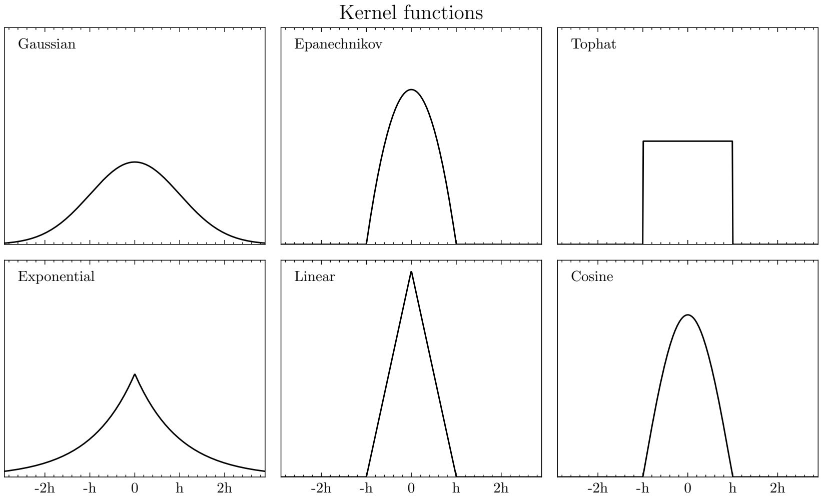


Figure 2.3: A selection of commonly used kernel functions $K(x)$.

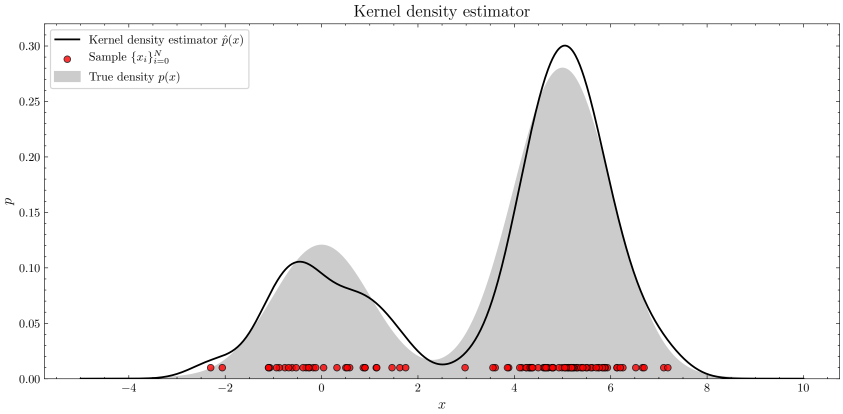


Figure 2.4: A kernel density estimation $\hat{p}(x)$ formed on two samples of size $N_1 = 30$ and $N_2 = 70$ from two Gaussian distributions $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$ where $\mu_1 = 0, \mu_2 = 5$ and $\sigma = 1.0$.

$K(x) = K(-x)$ and (iv) that it has finite variance. Typically K is chosen to be a smooth function such as the Gaussian and the amount of smoothing can be controlled by the bandwidth parameter h. To form an estimate of $p(x)$, we center a kernel $K(x)$ at each point x_i and then sum their contributions to obtain a function that can be evaluated for all $x \in \mathbb{R}$. Since Equation 2.15 only consists of additive operations, the estimate $\hat{p}(x)$ retains the continuity and differentiability properties of K, meaning that if we define K to be a smooth function the overall estimate $\hat{p}$ also becomes smooth.

Analysis of the kernel density estimator

In this section, we follow the notation outlined in Li and Racine (2006). As in the histogram, we start out by studying the bias of the estimator in a specific point x_0. Let us assume that we are given a i.i.d. sample $X_1, X_2, \ldots X_n$ from the unknown probability density $p(x)$. The expected value of $\hat{p}(x_0)$ is then given by:

$$\begin{aligned}
\mathbb{E}\left[\hat{p}(x_0)\right] &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\frac{1}{h}K\left(\frac{x_0 - X_i}{h}\right)\right] \\
&= \mathbb{E}\left[\frac{1}{h}K\left(\frac{x_0 - X}{h}\right)\right] \\
&= \int_{-\infty}^{\infty}\frac{1}{h}p(x)K\left(\frac{x_0 - x}{h}\right)dx \\
&= \int_{-\infty}^{\infty}\frac{1}{h}p(x)K\left(\frac{x - x_0}{h}\right)dx
\end{aligned} \tag{2.16}$$

where the last equality used the fact that $K(x) = K(-x)$. Now we perform a change of variable $u = (x - x_0)/h$ and $du = dx/h$ which yields

$$\int_{-\infty}^{\infty} K(u)p(x_0 + hu)du \tag{2.17}$$

Using a Taylor expansion at x_0, we get that $p(x_0 + hu)$ can be expressed as

$$p(x_0 + hu) = p(x_0) + hup'(x_0) + \frac{h^2u^2}{2}p''(x_0) + \frac{h^3u^3}{3!}p'''(x_0) + \dots \tag{2.18}$$

Plugging Equation 2.18 into Equation 2.17 results in:

$$\begin{aligned}
\mathbb{E}\left[\hat{p}(x_0)\right] &= \int_{-\infty}^{\infty} K(u)\left(p(x_0) + hup'(x_0) + \frac{h^2u^2}{2}p''(x_0) + \frac{h^3u^3}{3!}p'''(x_0) + \dots\right)du \\
&= p(x_0) + \frac{h^2p''(x_0)}{2}\int_{-\infty}^{\infty}u^2K(u)du + \frac{h^3p'''(x_0)}{3!}\int_{-\infty}^{\infty}u^3K(u)du + \dots
\end{aligned} \tag{2.19}$$

where the last equality was attained by using the conditions (ii) and (iii). To reduce cluttering, we introduce the notation $\kappa_2 = \int_{-\infty}^{\infty} u^2K(u)du$. Thus we get that the bias of the estimator at x_0 is

$$\text{Bias}\left[\hat{p}(x_0)\right] = \frac{h^2\kappa_2p''(x_0)}{2} + O(h^3), \tag{2.20}$$

which means that the overall dominating term is $O(h^2)$ as $h \to 0$.

Another interesting observation is that the bias is larger in points where there is a high degree of curvature in the estimated density, i.e. $p''(x_0)$ is large. The effect of larger bias due to curvature can be observed in the estimation of the peaks in Figure 2.4.

Next we analyze the variance which is given by:

$$\begin{aligned}
\mathrm{Var}\left[\hat{p}(x_0)\right] &= \mathrm{Var}\left[\frac{1}{n}\sum_{i=0}^{n}\frac{1}{h}K\left(\frac{x_0 - X_i}{h}\right)\right] \\
&= \frac{1}{n^2h^2}\left(\sum_{i=1}^{n}\mathrm{Var}\left[K\left(\frac{x_0 - X_i}{h}\right)\right] + 0\right) \quad \text{(by independence)} \\
&= \frac{1}{nh^2}\mathrm{Var}\left[K\left(\frac{x_0 - X}{h}\right)\right] \quad \text{(by identical distribution)} \qquad (2.21) \\
&= \frac{1}{nh^2}\left(\mathbb{E}\left[K^2\left(\frac{x_0 - X}{h}\right)\right] - \mathbb{E}^2\left[K\left(\frac{x_0 - X}{h}\right)\right]\right) \\
&= \frac{1}{nh^2}\left(\int_{-\infty}^{\infty}K^2\left(\frac{x_0 - x}{h}\right)p(x)dx - \left(\int_{-\infty}^{\infty}K\left(\frac{x_0 - x}{h}\right)p(x)dx\right)^2\right)
\end{aligned}$$

Again, a change of variable such that $u = (x - x_0)/h$ and $hdu = dx$, results in

$$\mathrm{Var}\left[\hat{p}(x_0)\right] = \frac{1}{nh^2}\left(h\int_{-\infty}^{\infty}K^2(u)p(x_0 + hu)du - \left(h\int_{-\infty}^{\infty}K(u)p(x_0 + hu)du\right)^2\right) \qquad (2.22)$$

A Taylor expansion of $p(x_0 + hu)$ as x_0 yields

$$\begin{aligned}
\mathrm{Var}\left[\hat{p}(x_0)\right] &= \frac{1}{nh^2}\left(h\int_{-\infty}^{\infty}K^2(u)\left[p(x_0) + hup'(x_0)\right] - O(h^2)\right) \\
&= \frac{p(x_0)}{nh}\int_{-\infty}^{\infty}K^2(u)du + O\left(1/n\right)
\end{aligned} \qquad (2.23)$$

where the last equality was attained by condition (iii). Equation 2.23 goes to zero if $h \to 0$ as $n \to \infty$. Thus we get that the MSE is

$$\mathrm{MSE}\left(x_0\right) = \frac{h^4\kappa_2^2(p''(x_0))^2}{4} + \frac{\kappa p(x_0)}{nh} + O(h^5) + O\left(1/n\right) \qquad (2.24)$$

where we've introduced the notation $\kappa = \int_{-\infty}^{\infty}K^2(u)du$. To find the bandwidth h that minimizes Equation 2.24, we differentiate the asymptotically dominating terms (the two first terms) with respect to h and set the result equal to zero to get

$$h^* = \left(\frac{\kappa p(x_0)}{\kappa_2^2(p''(x_0))^2}\right)^{1/5} n^{-1/5} \qquad (2.25)$$

With the asymptotically optimal bandwidth h^* plugged into Equation 2.24, we get

$$\mathrm{MSE}^*(x_0) = O(n^{-4/5})$$

which is already faster than the optimal MSE for the histogram.

However, since MSE is a pointwise property, the optimal bandwidth for a point located in a mode of a distribution isn't necessarily the optimal for a point in the tail of the distribution. Thus, we are more interested in the IMSE which is given by

$$\begin{aligned}\text{IMSE}\left[\hat{p}(x)\right] &= \int_{-\infty}^{\infty} \mathbb{E}\left[(\hat{p}(x) - p(x))^2\, dx\right] \\ &= \frac{h^4\kappa_2^2}{4}\int_{-\infty}^{\infty} [p''(x)]^2\, dx + \frac{\kappa}{nh}\underbrace{\int_{-\infty}^{\infty} p(x)dx}_{=1} + O\left(h^5\right) + O\left(1/n\right) \\ &= \frac{h^4\kappa_2^2}{4}\int_{-\infty}^{\infty} (p''(x))^2\, dx + \frac{\kappa}{nh} + O\left(h^5\right) + O\left(1/n\right)\end{aligned} \tag{2.26}$$

Differentiating the dominating terms of Equation 2.26 with respect to h and setting the result equal to zero yields

$$h^* = \frac{\kappa^{1/5}}{\kappa_2^{2/5}}\left(\int_{-\infty}^{\infty} (p''(x))^2\, dx\right)^{-1/5} n^{-1/5} \tag{2.27}$$

Plugging h^* into Equation 2.26 results in

$$\text{IMSE}^*\left[\hat{p}(x)\right] = O\left(n^{-4/5}\right) \tag{2.28}$$

which converges faster than the histogram IMSE in Equation 2.14. However, the expression for the optimal bandwidth can't be used in practice since it requires calculation of the unknown quantity $\int_{-\infty}^{\infty} (p''(x))^2\, dx$. Therefore most practical methods for bandwidth selection either try to estimate the unknown quantity (often referred to as *plug-in methods*) or utilize the method of cross-validation applied on the *integrated square error* (Gramacki, 2017).

2.3 Change detection

In statistical analysis, the process of *change detection* refers to the identification of whether or not a change has occurred and the time of any such change in the characteristics of a time series. Such a process consists of mainly two tasks:

1. Generation of *test quantities*; these are quantities that are designed to aid in the detection of possible changes. For example, in the distribution domain a test quantity could be the distance between peaks of two probability densities. A desirable property of these quantities is that they are close to zero when no change occurs and show a clear deviation from zero otherwise.

2. Design of decision rules; based on the generated test quantities, the goal of a decision rule is to raise alarms with minimal delay when a change has occurred while also keeping a low false alarm rate.

2.3.1 Test quantity

There are multiple ways to define a test quantity in the distribution domain. One approach for example, proposed in Rzeszucinski (2012) for wear monitoring in gearboxes, is to study the mode of a normally distributed residual vibration signal. It is motivated by the empirical observation that as wear develops in a gearbox, the maximal value of the normal probability density function (the mode) starts to decrease, resulting in a wider distribution. Thus, the proposed test quantity is defined as $d = 1 - p(\mu) = 1 - \frac{1}{\sqrt{2\pi}\sigma}$. This quantity is easily computable since it only demands the standard deviation but its drawbacks is the assumption of a unimodal distribution and that the signal has a normal distribution.

For our purposes, we are interested in a test quantity that is sensitive to amplitude changes, is non-parametric, and is easily computable. One such test quantity is the *Kullback-Leibler divergence* which is defined for two continuous probability distributions P and Q as

$$D_{KL}(P \parallel Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx \tag{2.29}$$

where $p(x)$ and $q(x)$ are the probability densities of P and Q. It has the desirable properties that $D_{KL}(P \parallel P) = 0$ and $D_{KL}(P \parallel Q) \geq 0$. The test quantity in Equation 2.29 is not a distance per definition since generally $D_{KL}(P \parallel Q) \neq D_{KL}(Q \parallel P)$, i.e. it is not symmetric. However, symmetry can be achieved by calculating

$$\frac{1}{2} \left(D_{KL}(P \parallel Q) + D_{KL}(Q \parallel P)\right) \tag{2.30}$$

which is referred to as the *Kullback-Leibler distance*.

2.3.2 Change detection algorithm

For the general description of change detection algorithms, we follow the terminology and notation of Gustafsson (2001). In the design of a change detection algorithm, we consider a sequence of observations of the test quantity y_k that consists of

$$y_k = \theta_k + e_k, \tag{2.31}$$

where θ_k is a deterministic component and e_k is white nose, i.e. $e_k \in N\left(0, \sigma^2\right)$. Then the goal of a change detection algorithm is two-fold; to estimate the deterministic component θ_k from y_k, and to detect rapid changes in θ_k. The deterministic component of the observation y_k is typically be separated from the noise with a low-pass filter

$$\hat{\theta}_k = H(q) y_k \tag{2.32}$$

Alternatively, low-pass filtering can also be achieved by weighting the observations by

$$\hat{\theta}_k = \sum_{k=0}^{\infty} \omega_k y_{t-k} \tag{2.33}$$

with weights ω_i that satisfy $\sum_i \omega_i = 1$. If the weights are chosen to be exponential, namely

$$\omega_k = (1 - \alpha)\alpha^k, \quad 0 < \alpha \leq 1 \tag{2.34}$$

where α is a *forgetting factor*, we obtain an estimate of θ_k that is an *exponential weighted moving average* (EWMA)

$$\hat{\theta}_k = \sum_{i=0}^{\infty} (1 - \alpha)\alpha^i y_{k-i} \tag{2.35}$$

Equation 2.35 can be rewritten in a recursive form as:

$$\begin{aligned}\hat{\theta}_k &= \alpha\hat{\theta}_{k-1} + (1-\alpha)y_k \\ &= \alpha\hat{\theta}_{k-1} + (1-\alpha)y_k - (1-\alpha)\,\hat{\theta}_{k-1} + (1-\alpha)\,\hat{\theta}_{k-1} \\ &= \hat{\theta}_{k-1} + (1-\alpha)\epsilon_k\end{aligned} \tag{2.36}$$

where $\epsilon_k = y_k - \hat{\theta}_{k-1}$ is referred to as the *prediction error*. Once an estimate is obtained, we define a *stopping rule* with the purpose to raise alarms when θ_k has exceeded a certain threshold $\hbar$. The input to a stopping rule is a distance measure s_k. In this case, we will use the prediction error as a distance measure, i.e. $s_k = \epsilon_k$, which is useful for detecting changes in the mean. If one wants to detect changes in variance, it is more suitable to choose $s_k = \epsilon_k^2$.

For the purpose of accumulating the prediction errors s_k, we define an auxiliary *test statistic* g_k which will be used to raise alarms when its accumulated sum exceeds a certain threshold $\hbar$. However, as the prediction errors are assumed to be normally distributed, they will have a natural fluctuation. Thereby the accumulated sum will have a drifting behaviour similar to that of a random walk. To prevent positive drift, we introduce a drift parameter ν which is subtracted from test statistic g_k. For a decreasing test statistic, negative drift is prevented by using a reset level a such that $g_k = \max(g_k, a)$.

In summary, we obtain the change detection algorithm presented in Algorithm 1. When the reset level is chosen to be $a = 0$, this algorithm is referred to as the cumulative sum recursive least squares filter (CUSUM RLS filter) (Gustafsson, 2001). The drift-parameter ν can be tuned as follows: If the false alarm rate is too high, ν should increase; if faster detection is needed, then ν should decrease. If more robustness is needed one can consider keeping a *run test* of g_k, i.e. not raising an alarm until g_k has exceeded the threshold $\hbar$ multiple times. For our purposes it is sufficient to have a one-sided test of the test statistic g_k since the test quantity $D_{KL}(P \parallel Q) \geq 0$, but it is also possible to extend the algorithm to use a two-sided test.

Algorithm 1: Change detection algorithm

```
  Parameters; $\alpha$, $\nu$, a
  Initialization; $\hat{\theta}_0 = y_0$, $g_0 = 0$
  Output; Alarm if $g_k \geq \hbar$

        $\hat{\theta}_k = \alpha\hat{\theta}_{k-1} + (1-\alpha)y_k$
        $\epsilon_k = y_k - \hat{\theta}_{k-1}$
        $s_k = \epsilon_k$
        $g_k = \max(g_{k-1} + s_k - \nu, a)$
```

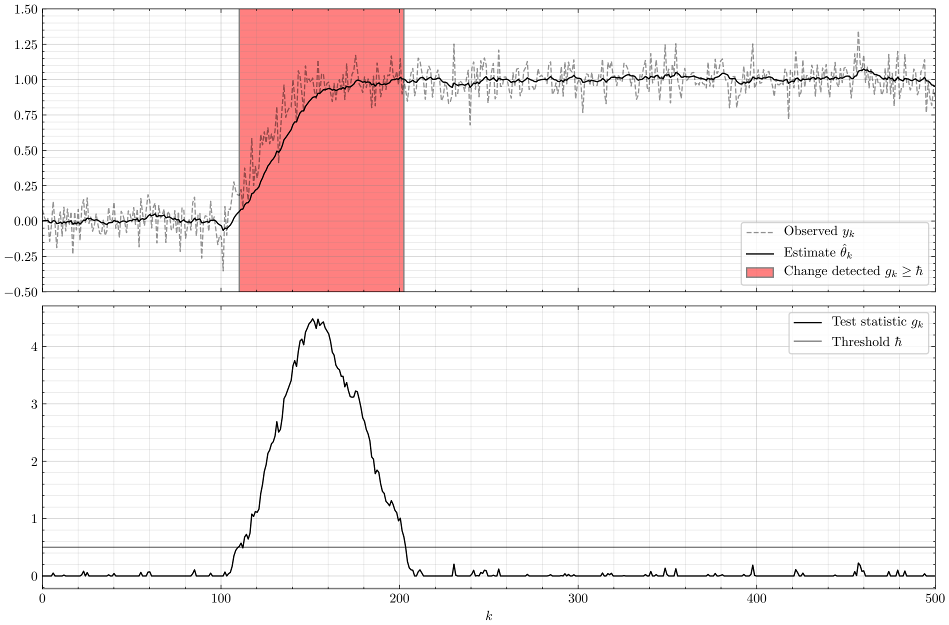


Figure 2.5: An illustrative example to demonstrate the change detection algorithm. A signal θ_k is perturbed by white noise $e_k \in N(0, \sigma^2)$, $\sigma = 0.1$. The estimate $\hat{\theta}_k$ is obtained by Equation 2.35. In the interval $100 < k < 150$ the signal undergoes a linear change such that $\theta_k = 1.0$ for $k \geq 150$. During this interval an increase in the test statistic g_k can be observed. Here the reset level a is chosen to be $a = 0$.

Chapter 3
Evaluation

The following sections give a brief overview of the collected data, the systems involved, and finally, results from the proposed method are presented and evaluated against a baseline algorithm.

3.1 Data collection

The collected datasets consist of measurements that are used in a system for *repeatability analysis*. Each measurement is collected during a diagnostic routine, also referred to as a *test cycle*, that is unique for each system and designed to achieve high friction and to expose weaknesses in the structure such as backlash. During the test cycle, only one axis is moving at a time to achieve a better fault isolation. The routine consists of moving $\pm 5^{\circ}$ repeatedly three times while sampling system signals such as torque, feed-forward torque and velocity at a rate of 2 kHz. For our purposes, the data is resampled to 248 Hz, yielding a sample period of $T = 4.032$ms. The system signals are synchronized by maximal autocorrelation of feed-forward torque, which is a calculated signal without any measurement noise.

The datasets can be divided into two parts, one labeled and one without labels. The labeled dataset consists of known failure cases, which we will abbreviate as FC. In the known failure cases, the existence of information in the form of lubricant analyses and service reports allowed for annotation on a case-by-base basis under consultation of an expert. The main purpose of this dataset it to compare performance of the proposed method against a benchmark algorithm in metrics such as hit rate and false alarm rate.

The unlabeled dataset is a much larger and diverse dataset with measurements collected from robots that are used in a production environment. The population spans robots of many different types with a large variation in the test cycles. The purpose of the unlabeled dataset is to evaluate the proposed method against the benchmark in a production environment and to quantify metrics such as alarm rate, consistency and the isolability of the alarms, which are of key importance in condition monitoring systems.

3.1.1 Benchmark algorithm

The benchmark algorithm is part of a conditon monitoring system. Its purpose is to estimate the condition of a manipulator by keeping fault indicators for each joint. For every new test cycle, the fault indicators are updated and an alarm is raised if the indicator for any joint passes over a predefined threshold. The threshold is chosen so that exceeding it should signal an impending failure. Fault indicators are calculated by accumulating four different metrics that quantify deviation from the nominal, i.e. the first test cycle, for which it is assumed that the manipulator's condition was non-faulty.

The exact implementation of how these metrics are calculated is unknown and can be considered to be a black box. The algorithm outputs three different levels for the condition of an axis; OK, warning, and error. In the evaluation, warnings and errors will be considered to be anomalies.

3.2 Known failure cases

Table 3.1: A failure case summary.

Failure case	Robot type	Axis	Component	Type of failure	Environment	Dataset
1	IRB 6700	2	Gearbox	Wear	Stage	$\mathcal{D}^{\text{train}}$
2	IRB 6600	6	Motor	Wear	Production	$\mathcal{D}^{\text{test}}$
2	IRB 6600	2	Gearbox	Wear	Production	$\mathcal{D}^{\text{test}}$
3	IRB 4600	1	Gearbox	Wear	Stage	$\mathcal{D}^{\text{test}}$
4	IRB 6640	3	Bearings	Wear	Stage	$\mathcal{D}^{\text{test}}$
5	IRB 7600	5	Gearbox	Wear	Stage	$\mathcal{D}^{\text{test}}$
6	IRB 6640	4	Motor	Oil leakage	Production	$\mathcal{D}^{\text{test}}$
8	IRB 6700	3	Gearbox	Wear	Production	$\mathcal{D}^{\text{test}}$
10	IRB 6700	2	Gearbox	Wear	Stage	$\mathcal{D}^{\text{train}}$
11	IRB 6640	2	Gearbox	Wear	Stage	$\mathcal{D}^{\text{test}}$

The known FCs consist of both robots that are used in a staged environment, of which some were undergoing accelerated wear tests, and robots used in different production environments. During accelerated wear tests, a robot is run with a constant high load in a controlled environment for several months, or even years, until a breakdown occurs and maintenance is needed. At mostly regular time intervals during the accelerated wear tests and less frequently in a production environment, a diagnostic test cycle is executed yielding the dataset:

$$\mathcal{D}^{FC} = \{X^{(1)}, X^{(2)}, \ldots, X^{(K)}\},$$

where K is the total number of executed test cycles. FC1 and FC10 will be used as a training dataset for parameter tuning and the rest of the FCs will be used for evaluation.

In FC10, an accelerated wear test is conducted in a staged environment on a 6-axis robot in the ABB IRB 6700 robot family. The test cycle is executed every 6th hour during a period of 2 years with occasional interruptions. During this period, measurements of the iron concentration in the lubricant of the gearbox in axis 2 are also collected. In Figure 3.1, excessive

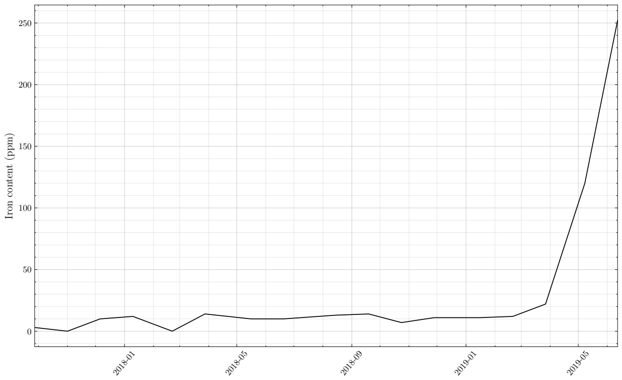


Figure 3.1: Measurements of iron concentration (ppm) in lubricant of a gearbox during accelerated wear test (FC10, axis 2). Notice how the wear develops rapidly in the matter of two months.

levels of wear can be detected by the increased iron concentration in the lubricant of the gearbox in axis 2.

In FC1, an accelerated wear test is conducted in a staged environment on a 6-axis robot in the ABB IRB 6700 robot family. The test cycle is executed less frequently than in FC10 (once a day) and without measurements of iron content. The failure occurs in axis 2 due to a worn out gearbox.

The system signal of particular interest which will be used to estimate the wear level is the motor torque τ. The rationale behind this, as mentioned before in Sect. 2.1.2, is that excessive levels of wear cause increased friction. To overcome the friction torques, the applied motor torque needs to increase. Thereby, indirectly, one can reason about wear levels from τ. However, it is important to note that friction is also affected by the lubricant temperature and variations in load. A selection of torque data from multiple test cycles in FC10 are shown in Figure 3.2. In Figure 3.2 one can observe that excessive wear causes changes in amplitude of the applied motor torque τ which makes it suitable to study changes in the distribution domain.

3.2.1 Results

A Gaussian kernel was chosen for the estimation due to its smoothing behaviour. Optimal bandwidths $\hbar$ of the kernels $\kappa_{\hbar}$ were found by conducting a grid search in the parameter space of $\hbar \in [0.1, 2]$ and chosen according to the best cross-validation performance.

Figure 3.3 shows an example of the difference between torque data of two test cycles, one for which the manipulator is in normal condition and one when the wear level is high. In this example, it is easier to distinguish the increased wear when the torque data is transformed to the distribution domain.

Test quantity

In the next step, the KDEs $\hat{p}^{(k)},\ k = 1 \dots K$ were compared to the nominal KDE $\hat{p}^{(0)}$ using the Kullback-Leibler divergence defined in Equation 2.29. The resulting distances for FC10

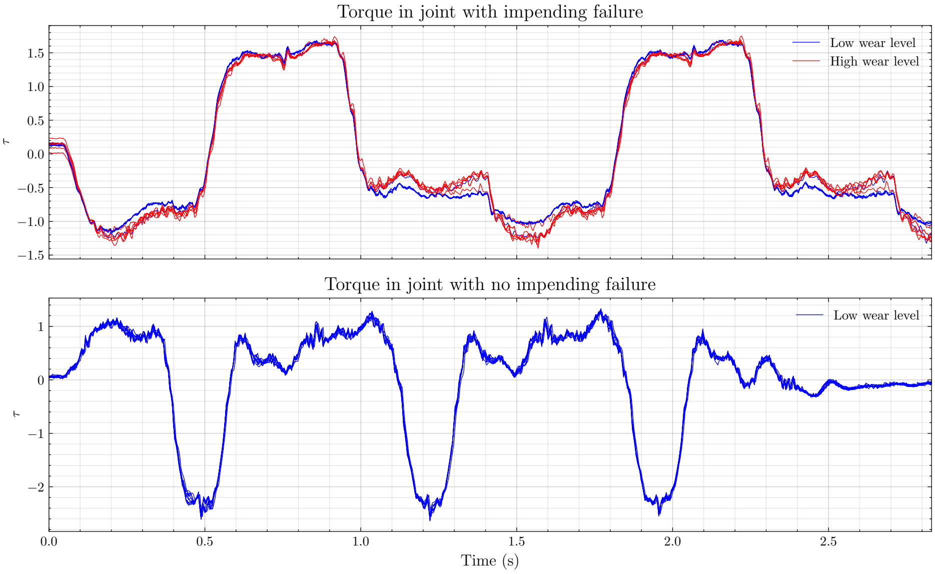


Figure 3.2: Standardized torque τ in a selection of multiple test cycles (FC10, axis 2 and 5). Measurements in normal and anomalous operation are represented by blue and red color respectively. Note that excessive wear appears as changes in amplitude of the applied motor torque τ.

and FC1 are shown in Figures 3.4 and 3.5 respectively. In both cases, it is observed that the divergence increases as the manipulator gradually deviates from normal condition.

In FC10, we expect to see an increase in the test quantity D_{KL} along with the increased iron concentration in the lubricant of the gearbox in axis 2 (see Figure 3.1). The effect of excessive wear is clearly visible in the trend of increasing divergence in the test quantity of axis 2. However, sharp peaks in the divergence that are not related to wear can also be observed. It was discovered that these peaks coincide with test cycles performed directly after the manipulator had been non-operational for a longer time due to lubricant analyses. This had caused the lubricant to cool off. Similar effect can be observed for the same dates in all axes but with slightly smaller amplitude.

As discussed in Sect. 2.1.2, temperature variations can have a considerable effect on friction in robot joints which would explain the appearance of the peaks for low wear levels in Figure 3.4. An increase of D_{KL} can also be seen in axis 3 which is probably due to its codependency with axis 2.

In FC1, an increase in the test quantity is expected along with the increased wear level that occurred in axis 2 between 2016-05-09 and 2016-06-16. In 2016-05-09, it was reported that axis 2 started making noise, indicating unusual behaviour. Wear in the gearbox of axis 2 was confirmed in 2016-05-12 with a lubricant analysis, which showed a high concentration of metallic particles in the gearbox. In 2016-06-16, the gearbox of axis 2 was replaced, before any failure had occurred. During the period of high wear level, a clear response can be observed in the test quantity D_{KL} of axis 2 in Figure 3.5. It can also be seen that the test quantity is smaller for all other axis.

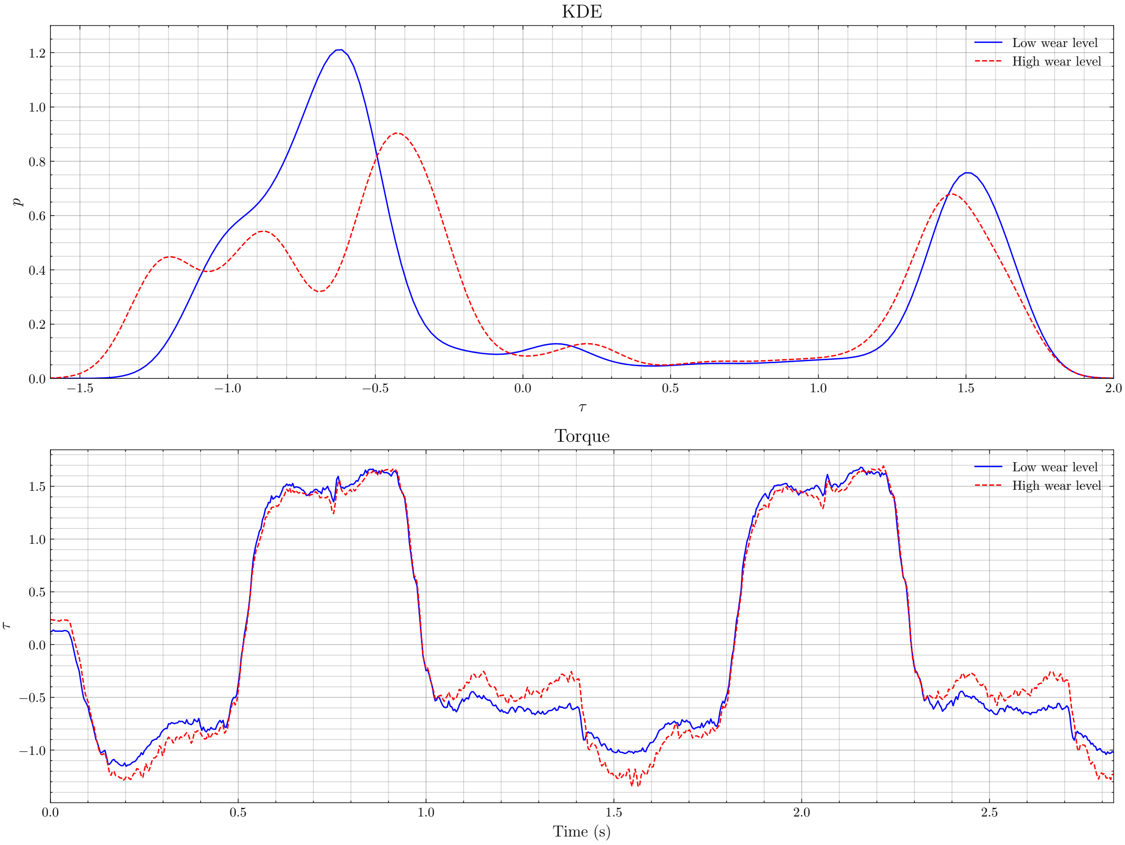


Figure 3.3: Comparison of KDEs and torque measurements for a non-anomalous and anomalous test cycle. It seems that an excessive wear level is easier to distinguish in the distribution domain.

Simple threshold

In the results of Figure 3.4 and 3.5, it is observed that the chosen test quantity is sensitive to both changes in wear and temperature. Based on those results, it seems that a threshold of the test quantity D_{KL} chosen as $\hbar = 0.04$ would suffice for early detection of increased wear levels in both training cases, see Figure 3.6.

In the training cases, this threshold would yield a hit rate of 56.1% and a false alarm rate of 0.01% as seen in Table 3.2. In the same table, we see that the same threshold yields similar results on the test dataset $\mathcal{D}^{\mathrm{test}}$, which suggests that it adapts properly to new and previously unseen data. Note that a true positive (TP) refers a test cycle for a given axis for which it was assessed, under consultation of an expert, that signs of wear were detectable. Thus, the hit rate measures the ability to detect signs of an impending failure in the torque sequences of a test cycle.

Change detection algorithm

To increase the method's robustness, the refinement presented in Sect. 2.3.2 was implemented. Although a simple threshold on the test quantity D_{KL} at $\hbar = 0.04$ yields a relatively low false alarm rate, the method can be made less sensitive to noise or other disturbances

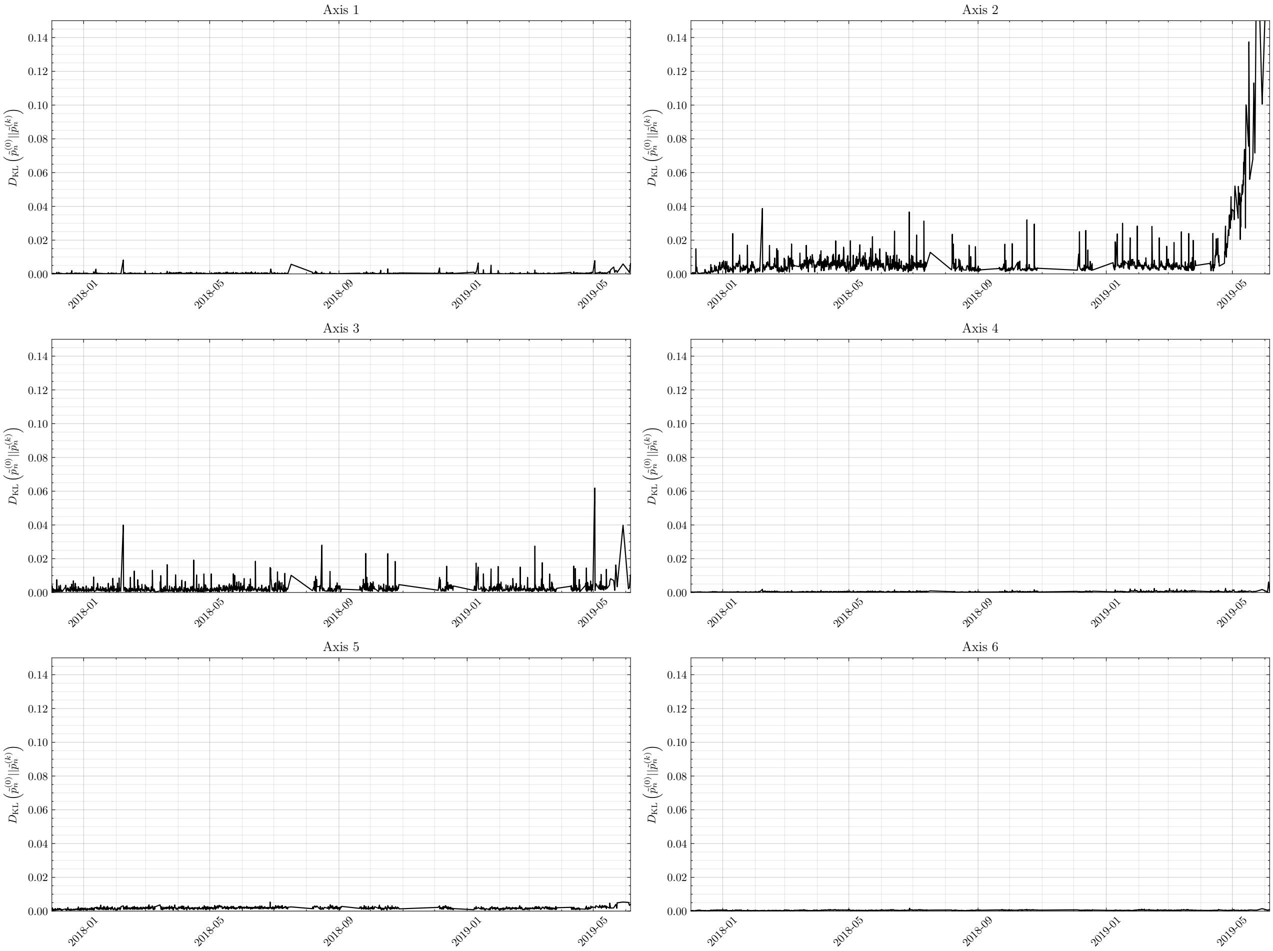


Figure 3.4: Test quantities $D_{\mathrm{KL}}\left(\hat{p}_n^{(0)}\|\hat{p}_n^{(k)}\right)$, $k = 1 \dots K$ of each axis in the accelerated wear test of FC10. There is a noticeable effect in the divergence of axis 2 with the increased iron concentration in the lubricant, see Figure 3.1. However, there are also are a few peaks that are not related to high wear levels. Notice also the codependency between axis 2 and 3 and that the divergences for remaining axes are considerably lower.

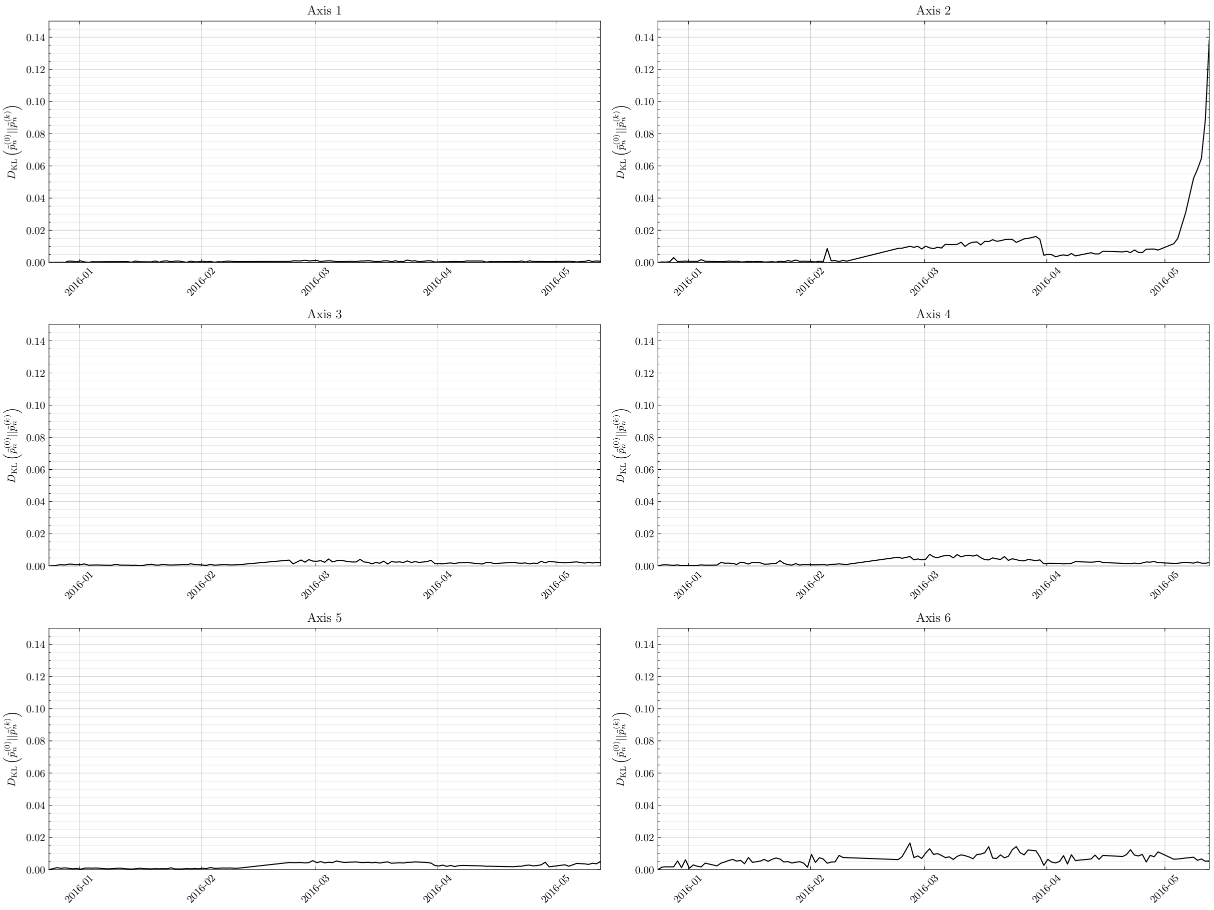


Figure 3.5: Test quantities $D_{\mathrm{KL}}\left(\hat{p}_n^{(0)} \| \hat{p}_n^{(k)}\right),\ k = 1 \ldots K$ of each axis in the accelerated wear test of FC1. In this case it was reported that axis 2 started making noise 2016-05-09, indicating unusual behaviour. A lubricant analysis was performed 2016-05-12 that showed a high concentration of metallic particles in the gearbox of axis 2. In 2016-06-16 the gearbox of axis 2 was replaced before any failure occurred. During this period a clear response can be seen in the test quantity of axis 2. Notice that the test quantity is considerably smaller in all other axes.

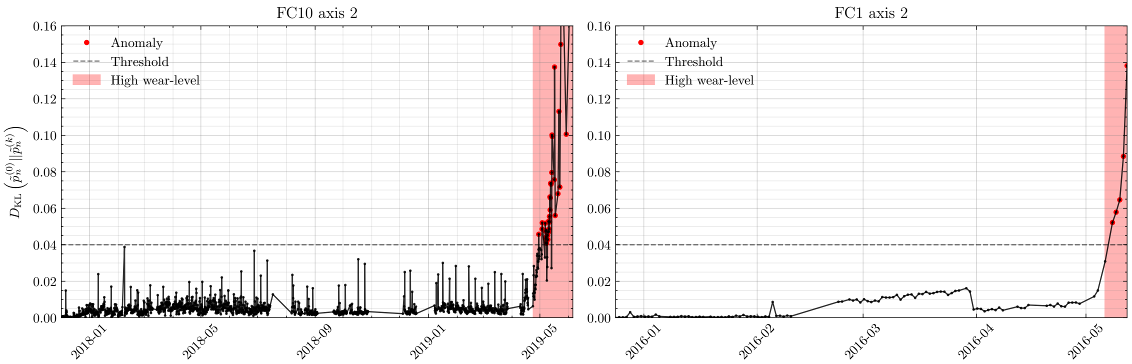


Figure 3.6: Detected true anomalies, marked as red points, in the test quantity of axis 2 in FC10 (left) and FC1 (right) with a threshold set at $\hbar = 0.04$. The period for which it was assessed that signs of increased wear were detectable is shown with a red background filling. Note that the component with increased wear level was replaced shortly after the last test cycle in both cases, before any failure had occurred.

Table 3.2: Obtained results on training and test datasets, $\mathcal{D}^{\text{train}}$ and $\mathcal{D}^{\text{test}}$, when a threshold of the test quantity D_{KL} is chosen as $\hbar = 0.04$. TP = True positive, FN = False negative, etc.

FC	TP	FP	TN	FN	Accuracy	Hit rate	False alarm rate	F1-score
1	5	0	666	1	0.999	0.833	0.0	0.909
2	52	46	5603	10	0.990	0.839	0.008	0.65
3	8	1	344	1	0.994	0.889	0.003	0.889
4	3	0	106	1	0.991	0.750	0.000	0.857
5	3	1	48	2	0.944	0.6	0.020	0.667
6	5	0	358	1	0.997	0.883	0.000	0.909
8	4	10	234	1	0.956	0.800	0.041	0.421
10	41	1	8965	35	0.996	0.539	0.0001	0.695
11	13	5	1152	0	0.996	1.0	0.004	0.839
$\mathcal{D}^{\text{train}}$	46	1	9631	36	0.996	0.561	0.0001	0.713
$\mathcal{D}^{\text{test}}$	88	63	7845	16	0.990	0.846	0.008	0.690

Table 3.3: Results for change detection algorithm with a threshold set at $\hbar = 0.142$ on training and test datasets, $\mathcal{D}^{\text{train}}$ and $\mathcal{D}^{\text{test}}$.

FC	TP	FP	TN	FN	Accuracy	Hit rate	False alarm rate	F1-score
1	4	0	666	2	0.997	0.667	0.0	0.800
2	60	11	5638	2	0.998	0.968	0.002	0.902
3	7	0	345	2	0.994	0.778	0.0	0.875
4	1	0	106	3	0.972	0.250	0.0	0.400
5	3	0	49	2	0.963	0.6	0.0	0.75
6	5	0	358	1	0.997	0.883	0.0	0.909
8	5	13	231	0	0.948	1.0	0.053	0.435
10	54	0	8966	22	0.998	0.712	0.0	0.831
11	9	0	1157	4	0.997	0.692	0.0	0.818
$\mathcal{D}^{\text{train}}$	58	0	9632	24	0.998	0.707	0.0	0.829
$\mathcal{D}^{\text{test}}$	90	24	7884	14	0.995	0.865	0.003	0.826

that are not related to excessive wear levels, such as temperature variations in the lubricant, as was seen in FC10. The reasoning behind this approach is that wear is a process that develops gradually while large temperature variations often occur in isolation in conjunction with maintenance.

By estimating $\hat{\theta} = \mu_{KL}$ from the raw test quantities D_{KL} as described in Equation 2.35, the signal is separated from the noise. On top of that, an auxiliary test statistic g_k, which holds an accumulated sum of the prediction errors, is used to determine if there has been a change in the estimated mean μ_{KL}. Note that it is now the test statistic g_k that is being evaluated against the threshold to detect anomalies.

In Figure 3.7, we see that the weighted exponential moving average and drift parameter ν make the test statistic g_k less sensitive to noise and temperature variations compared to the raw test quantity D_{KL} seen in Figure 3.6. The results obtained in Table 3.3 show that this refinement significantly increases the hit rate while also reducing the false alarm rate.

In Figure 3.8, which shows the results for axis 2 and 6 in $\mathcal{D}^{\text{FC2}} \in \mathcal{D}^{\text{test}}$, we observe that the chosen threshold adapts properly to new previously unseen data. Here we also observe that the test statistic g_k has a clear response to the increased wear level and that it returns to normal levels after a component has been replaced. In this case, most false positives are caused by a slow reduction of the test statistic after replacement of component.

3.2.2 Comparative evaluation

In the results presented in Table 3.4, we observe that the benchmark algorithm has a predictive power with a hit rate of 58.8% and a false alarm rate of 16.1% on the test dataset. By comparing the performance metrics on the test dataset of the different approaches in Table 3.5, we observe that the benchmark algorithm has a significantly higher false alarm rate compared to the proposed method, which suggests that it has a low signal-to-noise ratio. The hit rate, which is often referred to as the sensitivity or recall and measures the rate of correctly

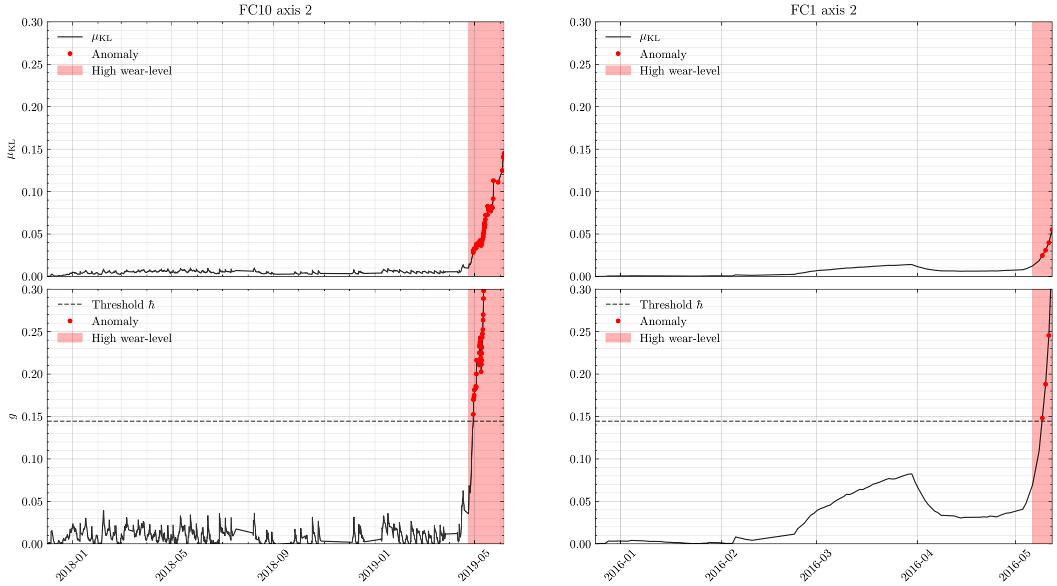


Figure 3.7: Change detection algorithm applied on training dataset $\mathcal{D}^{\text{train}}$ with a threshold for g set at $\hbar = 0.14$. The upper plots show the smoothing effect of the EWMA μ_{KL} for FC1 and FC10. The lower plots show the thresholded test statistic g_k. It can be observed that μ_{KL} significantly reduces the influence of noise and temperature variations in FC10 (compare to Figure 3.6). The robot in FC1 was restarted 2016-04 due to unknown reason. An increase in g_k can be observed during the period of high wear-level in both FC1 and FC10.

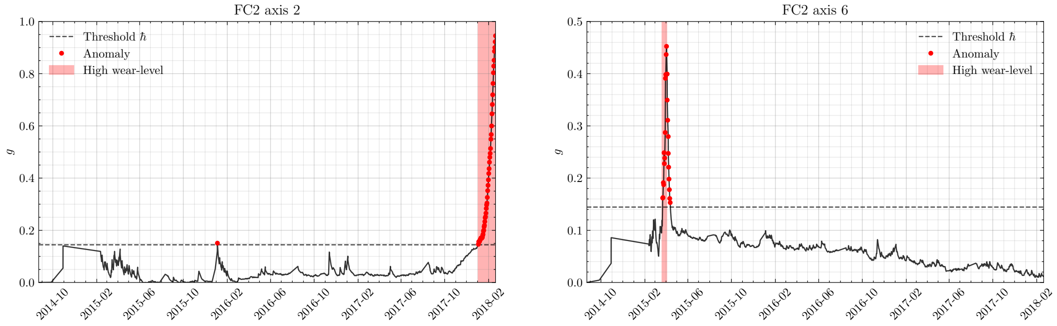


Figure 3.8: Change detection algorithm applied on $\mathcal{D}^{\text{FC2}} \in \mathcal{D}^{\text{test}}$ with a threshold for the test statistic g set at $\hbar = 0.14$. In this case the robot was used in a production environment where both axis 2 och 6 had periods of increased wear level. In the motor of axis 6 the wear level was noticed in time, leading to a replacement in 2015-02-16, before total breakdown. We can observe that g of axis 6 (right plot) increases significantly during the period of high wear and decreases steadily as the motor is replaced. Three years later wear was noticed in the gearbox of axis 2 with a following replacement in 2018-02-16. During the replacement the iron concentration in the lubricant was measured to be 627 ppm, six times the limit of recommended concentration. Note that most of the false positives originate from g not decreasing fast enough after the replacement of the motor in axis 6.

Table 3.4: Results for benchmark-algorithm on training and test datasets, $\mathcal{D}^{\text{train}}$ and $\mathcal{D}^{\text{test}}$.

FC	TP	FP	TN	FN	Accuracy	Hit rate	False alarm rate	F1-score
1	6	0	660	0	1.0	1.0	0.0	1.0
2*	-	-	-	-	-	-	-	-
3*	-	-	-	-	-	-	-	-
4	3	0	106	1	0.999	0.75	0.0	0.857
5	3	14	38	2	0.719	0.6	0.269	0.273
6	5	108	258	1	0.707	0.833	0.295	0.084
8	3	122	148	3	0.547	0.5	0.452	0.046
10	29	482	8373	44	0.941	0.397	0.055	0.100
11	6	62	1047	7	0.939	0.462	0.06	0.148
$\mathcal{D}^{\text{train}}$	35	482	9033	44	0.945	0.443	0.051	0.117
$\mathcal{D}^{\text{test}}$	20	306	1597	14	0.816	0.588	0.161	0.111

* Results could not be obtained.

Table 3.5: Comparison of performance metrics on $\mathcal{D}^{\text{test}}$.

Algorithm	Accuracy	Hit rate	False alarm rate	Precision	F1-score
Threshold	0.990	0.846	0.008	0.583	0.690
Change detection	**0.995**	**0.865**	**0.003**	**0.789**	**0.826**
Benchmark	0.816	0.588	0.161	0.061	0.111

identified positives, is also seen to be significantly higher in the proposed method.

3.3 Unlabeled dataset

The unlabeled dataset consists of approximately 157,000 measurements from 374 different robot systems. This dataset is collected with the purpose of evaluating alarm rate, consistency, and isolability of both algorithms. Isolability and consistency are meaningful metrics in the context of decision-making in maintenance actions, giving an answer to what extent it is possible to localize an impending failure.

3.3.1 General analysis

Alarm rate

The axial alarm rate can provide insights of how sensitive the algorithms are to noise. In Figure 3.9, we see that the alarm rate is significantly higher for the benchmark algorithm in all axes. In the wrist axes (4-6), consensus can be found in a higher alarm rate for axis 6 for both algorithms.

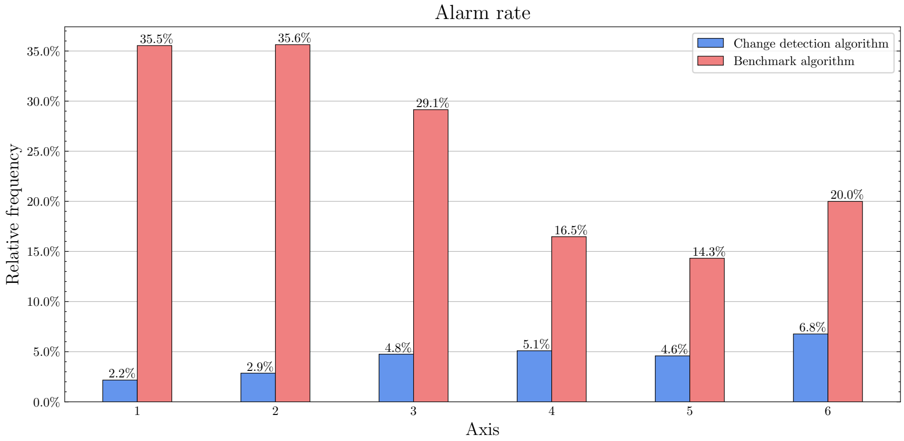


Figure 3.9: The alarm rate of each axis.

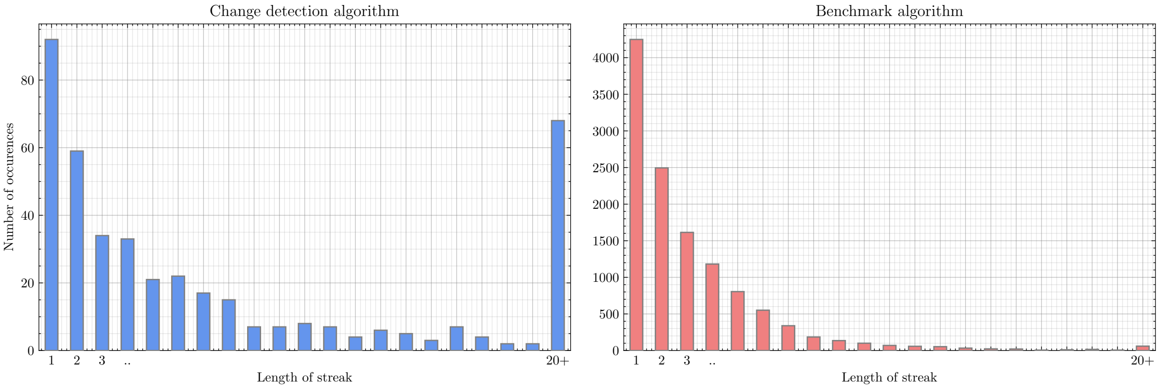


Figure 3.10: The distribution of alarm streaks in the proposed- and the benchmark algorithm. Streaks of length 20+ are aggregated into a single bar. Notice the different scales in the number of occurrences.

To explore the consistency of the methods and if the large difference in alarm rate is a result of wear-related phenomena or noise, we study streaks of consecutive alarms. In Figure 3.1 and the known failure cases in Section 3.2, we observed that wear evolves gradually, thus we expect that multiple alarms occur consecutively in wear-related cases.

In Figure 3.10, we see that single alarms dominate in both algorithms with a fading tail for multiple consecutive alarms. However, the concentration of single alarms is much larger in the benchmark algorithm which suggests that is has a low signal-to-noise ratio. This is also suggested by the resemblance between its distribution and the distribution of coin toss streaks seen in Figure 3.11. In white noise, positive and negative outcomes are equally likely. Thereby the distribution of streaks in accumulated white noise and coin tosses should decrease as $O(2^{-n})$ where n is the streak length. This behaviour is however not observed for streaks in the proposed method as for example streaks of length 3 and 4 are equally likely, which suggests that the proposed algorithm is less sensitive to noise.

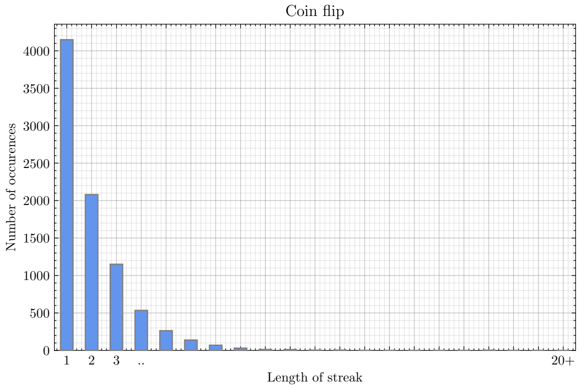


Figure 3.11: The distribution of streaks in 34 000 coin tosses with equally likely outcome of heads and tails.

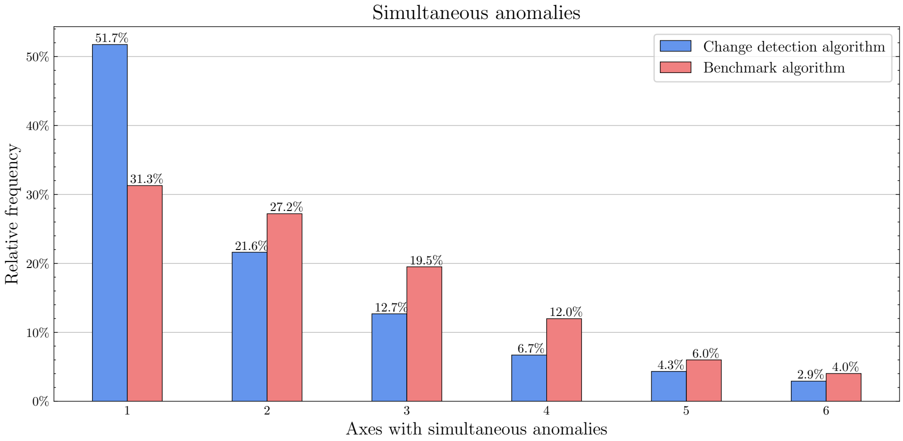


Figure 3.12: Number of axes with simultaneous anomalies during a test cycle.

Isolability

As a first attempt to evaluate the isolability of the algorithms, we investigate in how many axes alarms are raised in simultaneously during one test cycle in case of any detected anomaly. In Figure 3.12, we observe that the alarms of the proposed algorithm are in general more isolated compared to the benchmark algorithm. In a majority of the test, cycles alarms are isolated to one axis which suggests a strong fault isolability.

However, the statistic in Figure 3.12 can be somewhat misleading since it doesn't take into account the distribution of alarms in the succeeding test cycle. For example, a case where an alarm is raised in axis 6 while the succeeding test cycle has an alarm in axis 2, would suggest weak fault isolability.

In Figure 3.13 we observe that in a majority of cases predictions remain the same in the succeeding test cycle when an anomaly has been predicted in one or multiple axes by the proposed algorithm. For the benchmark algorithm, we observe that most predictions change in the succeeding test cycle, which again suggests a weak fault isolability.

For test cycles with multiple predicted anomalies, it is of interest to study what axis combinations are the most common for further evaluation of isolability. If multiple anomalies

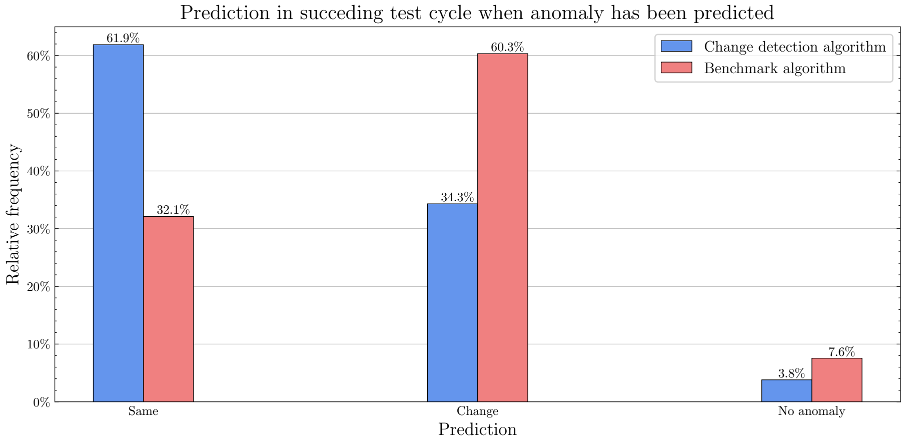


Figure 3.13: Prediction in succeeding test cycle when an anomaly has been predicted in one or multiple axes.

are predicted, there is a better isolation capability if these are contained in the main axes (1-3) or equally the wrist axes (4-6) compared to if they were predicted to be in axes that are far from each other.

Comparing Tables 3.6 and 3.7, we notice two things: First, multiple anomalies are in general less common for the proposed algorithm. Secondly, when multiple anomalies are predicted they are generally in axes with close proximity to each other.

3.3.2 Case follow-ups

The section describes three cases from the unlabeled dataset that were considered to be of interest and for which further inquiry made it possible to conclude that a failure had occurred. The purpose of this section is to exemplify what the statistics in the previous section results in individual cases.

Case A. In this case, a service report stated that there had been an oil leakage in the motor of axis 4 that was noticed 2019-07, leading to a replacement of the motor before any failure occurred in 2019-07-11. In Figure 3.14, it can be observed that the torque data from test cycles run before 2019 are similar while data after show a slight deviation from the nominal. This is reflected in the increasing test statistic g for test cycles after 2019. In this case, we also notice that the test cycles are run on a weekly schedule, resulting in a slow accumulation of the test statistic.

When comparing the predictions of the two algorithms, we notice that the proposed algorithm predicts no anomalies, although a clear trend that is indicative of an impending failure can be observed. In contrast, the benchmark algorithm has a high alarm rate and predicts anomalies when axis 4 of the manipulator is in both healthy and non-healthy condition. In Figure 3.15, we see that the benchmark algorithm also has a high alarm rate for the remaining axes, even though no failures where reported for these during the time of the measurements.

Table 3.6: Five most common axis combinations where more than one anomaly is predicted by the proposed algorithm. Notice that a frequency of 6 for a given combination here means that there are 6 different robots that each have test cycles, with more than one predicted anomaly, of which the the given combination is the most common. The six circles represent the different axes from axis one to axis 6. A filled circle represents a predicted anomaly.

Combination	Frequency	Relative frequency
○●●○○○	6	0.194
○○○●●●	6	0.194
○○○○●●	4	0.129
○○●○○●	3	0.097
○○○●○●	3	0.097

Table 3.7: Five most common axis combinations where more than one anomaly is predicted by the benchmark algorithm.

Combination	Frequency	Relative frequency
●●●○○○	43	0.195
●●●●●●	25	0.113
○●●○○○	20	0.090
●●●○○●	18	0.081
●●○○○○	17	0.077

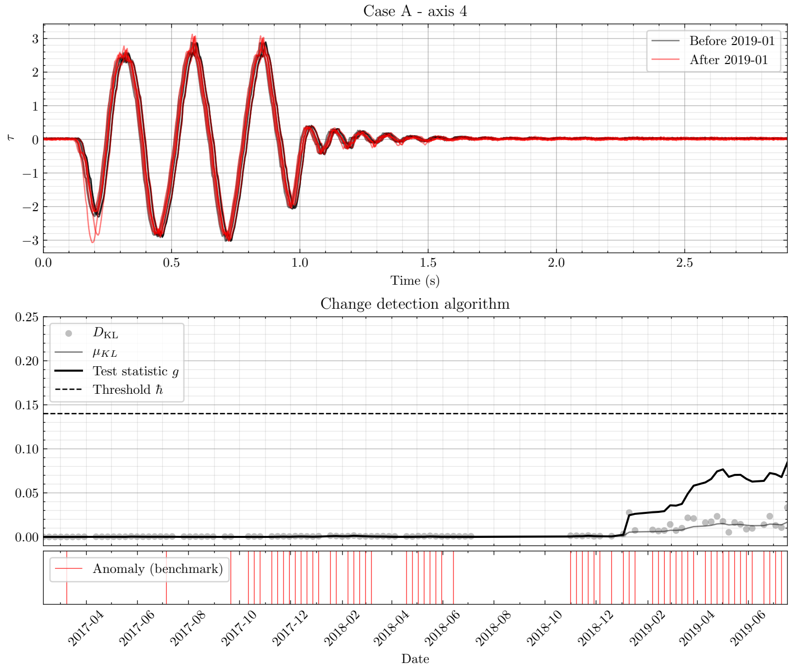


Figure 3.14: Torque data and results for change detection algorithm with a threshold set for the test statistic g at $\hbar = 0.14$. Predicted anomalies by the benchmark algorithm are shown as vertical red lines in the bottom plot. In this case it was noticed that the motor of axis 4 had an oil leakage 2019-07, leading to a replacement 2019-07-11. In the upper plot we notice that the torque in test cycles run after 2019 show a slight deviation from the nominal torque data.

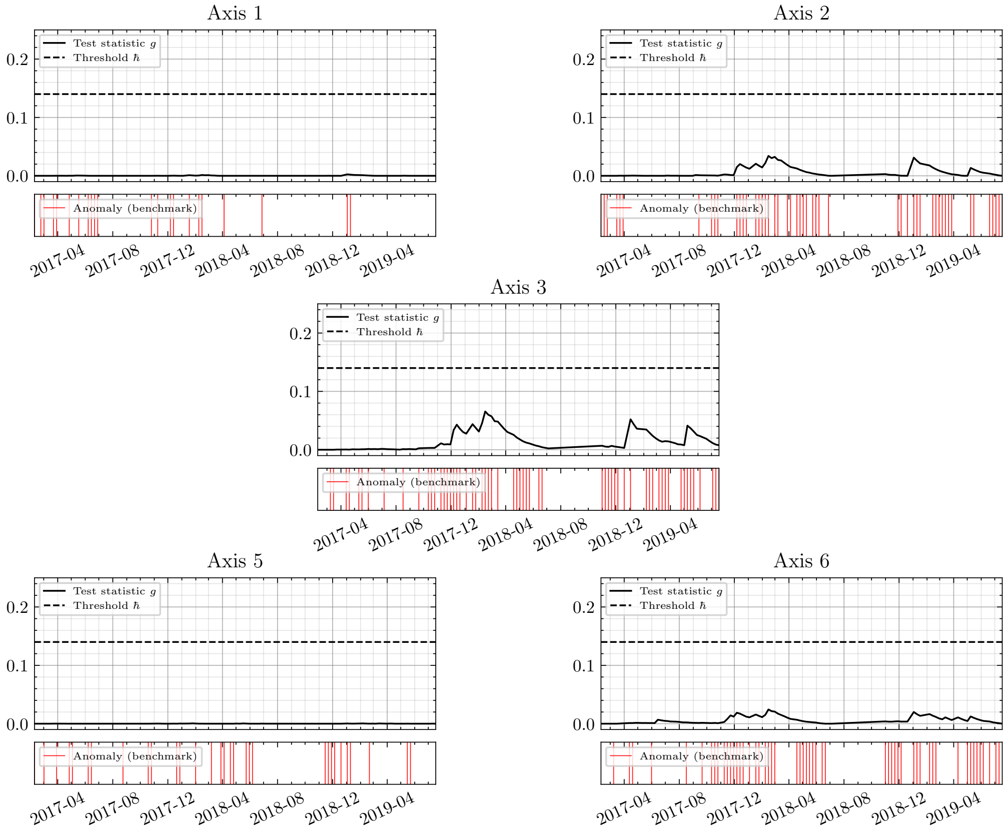


Figure 3.15: Results for change detection algorithm with a threshold set for the test statistic g at $\hbar = 0.14$ for all remaining axes in case A. Predicted anomalies by the benchmark algorithm are shown as vertical red lines in the bottom plots.

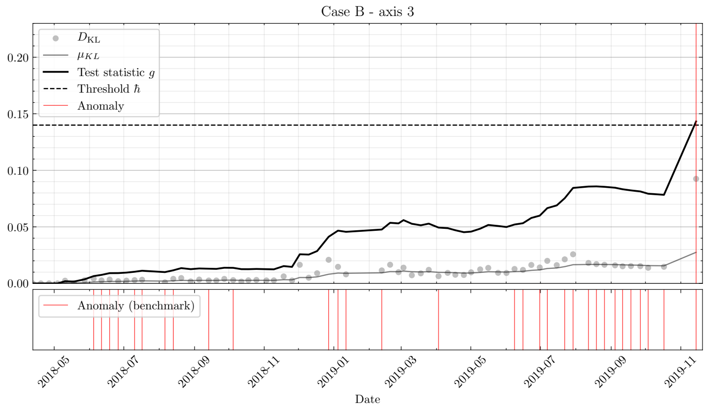


Figure 3.16: Results for change detection algorithm with a threshold set for the test statistic g at $\hbar = 0.14$. Predicted anomalies by the benchmark algorithm are shown as vertical red lines in the bottom plot. In this case it was noticed that the gearbox of axis 3 was damaged, leading to a replacement 2019-11-17.

Case B. In this case, a service report stated that the manipulator had a damaged gearbox in axis 3, resulting in a replacement before failure in 2019-11-17. In Figure 3.16, we notice a clear trend in the test statistic that indicates an impending failure. The test statistic exceeds the threshold in the last run test cycle, three days before the replacement. We also notice that the test cycles are run on a weekly basis with occasional interruptions that can last up to a month. The interruptions and the weekly schedule constrains the rate of increase in the test quantity.

In Figure 3.17, the test statistic of axis 2 also shows a clear trend that exceeds the threshold. In the same service report, it is stated that the oil of the gearbox in axis 2 was heavily worked and that there were visible filings of iron. Although these observations may be related to wear, further inquiry is needed to determine the definite cause. Further, the service report states a healthy condition for the remaining axes (1, 4, 5 and 6). Again, the benchmark algorithm shows a high alarm rate for both healthy and non-healthy axes.

Case C. In this case, a service report stated that maintenance action had been taken 2019-07-08 to replace the motor of axis 6 that was malfunctioning due to oil leakage. Furthermore, it is stated that the components of the remaining joints are in good condition. In 3.18, we observe how the torque data gradually deviates from the nominal torque data. The same deviation is also observed in the changing shape of the estimated kernel densities. In Figure 3.19, we notice a periodicity in the test statistic of axis 2 that seems to be connected to longer interruptions, possibly due to holidays.

Similar increases in the test statistic g can be seen for all other axes but smaller in magnitude which suggests that these anomalies are not wear-related. Nonetheless, there are fewer predicted anomalies compared to the benchmark algorithm and the anomalies are mostly isolated to axis 6 where the fault is located.

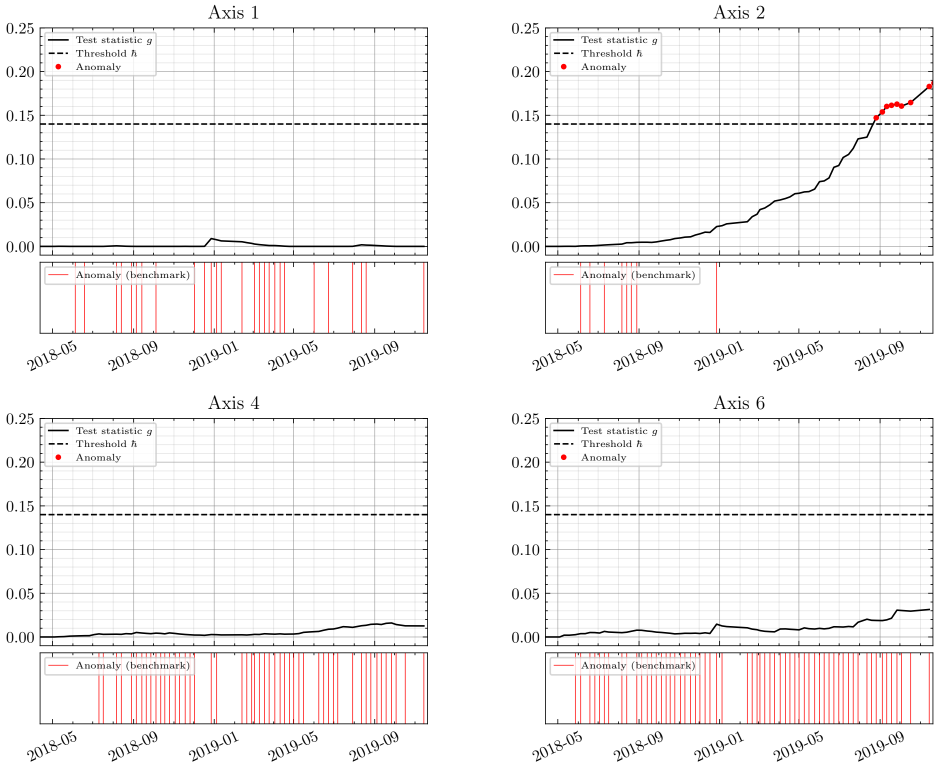


Figure 3.17: Results for change detection algorithm with a threshold set for the test statistic g at $\hbar = 0.14$ for remaining axes in case B. Predicted anomalies by the benchmark algorithm are shown as vertical red lines in the bottom plots. No torque data was found for axis 5.

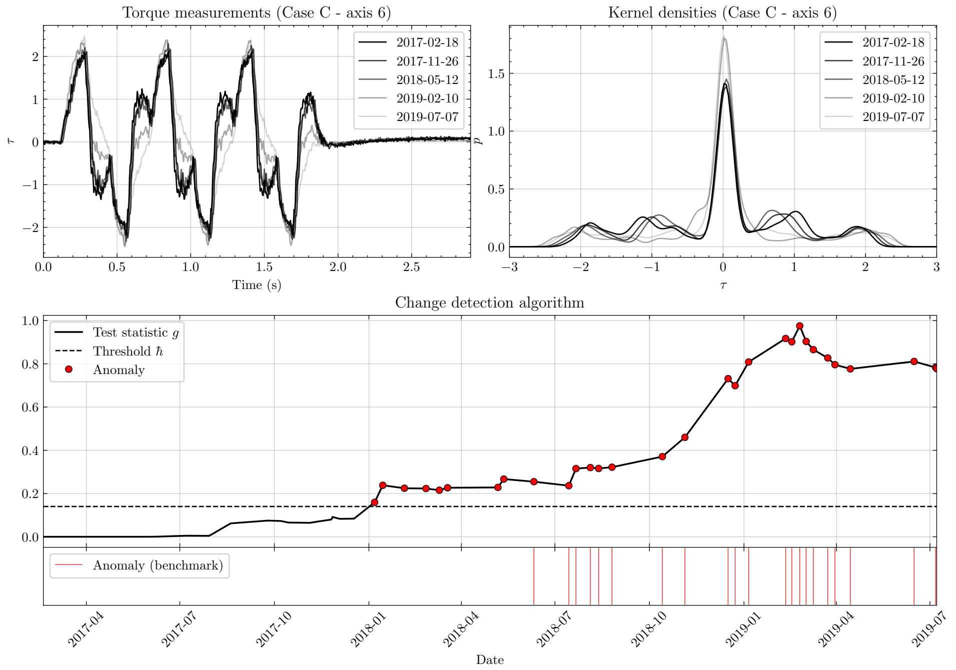


Figure 3.18: In the upper plots a sample of standardized torque data are shown with the corresponding kernel density estimations. Results for the change detection algorithm with a threshold for the test statistic g set at $\hbar = 0.14$ are shown in the middle plot. Predicted anomalies by the benchmark algorithm are shown as vertical red lines in the bottom plot.

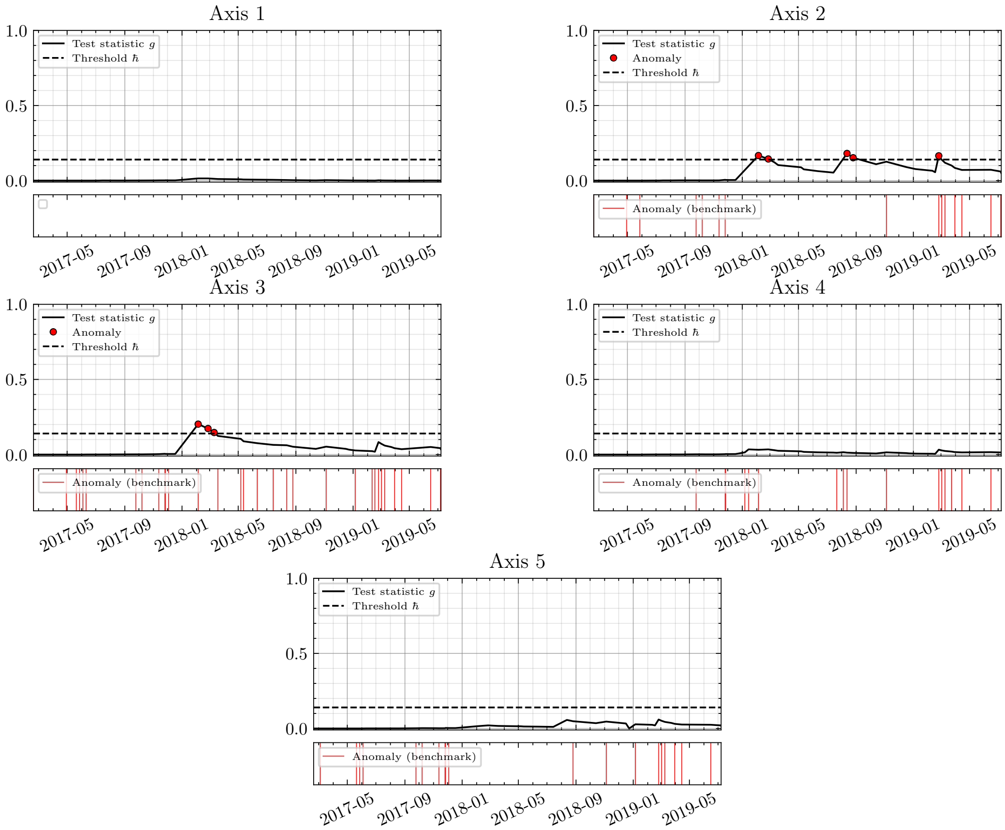


Figure 3.19: Results for the change detection algorithm with a threshold for the test statistic g set at $\hbar = 0.14$ for all remaining axes in case C. Predicted anomalies by the benchmark algorithm are shown as vertical red lines in the bottom plots. Notice the periodicity of the predicted anomalies (by the proposed algorithm) in axis 2.

3.4 Discussion

The final suggested method to monitor changes in the condition of robot joints shows promising results for early detection of impending failures. In the evaluation, we observed that the test statistic has a clear response to increased wear level but still some issues remain that need to be considered.

3.4.1 Algorithmic refinement

It was observed that temperature variations in the lubricant have a significant effect on the friction levels in a robot joint. To discriminate between temperature variations and increased wear level, we introduced refinements such as an exponential weighted moving average and a CUSUM-test statistic which resulted in a increased hit rate and false alarm rate.

However this refinement of the method assumed that the temperature variations have isolated occurrences while wear is a phenomena that develops gradually. Temperature variations that do not adhere to this assumption will remain an issue but no such variations have been observed. If this would present itself as an issue one could further exploit the fact that temperature variations have an effect on all robot joints at the same time as we observed in FC10. Another possible way to tackle this issue is by estimating the temperature with a Kalman filter but that would require a heat model and measurements of the environment temperature.

The refinements to the method also caused a slight delay as an auxiliary test statistic was introduced. When diagnostic test cycles are run daily this poses no problem. However, when test cycles are run less frequently, we observed that in some cases, such as case A and B in the unlabeled dataset, the test statistic accumulates slowly. This suggests that in a production environment the parameters of the algorithm would need to be adapted to how often the test cycle is scheduled to run. Is is also assumed that test cycles are executed in regular time intervals, but as we observed in both datasets this isn't always the case. With longer time between test cycles, the prediction becomes less reliable and as of now this uncertainty is not reflected in the prediction.

As of now, explicit predictions of when a failure will occur and fault isolation on a component level are not being made. For predictions of this kind, a better understanding of how wear develops with usage and time is required. With a labeled dataset of each failure type, it would be possible to extend the method by monitoring multiple test statistics between the density of each executed test cycle and failure type, which would allow for fault isolation on a component level.

3.4.2 Alarm rate

Another issue that presented itself in the unlabeled dataset is a higher alarm rate in axis 6 compared to the other axes, which seems to be an issue mostly related to manipulators used in a production system. The control system of the manipulator relies on correct load definitions. When incorrect load data is defined the path accuracy is impaired and there is a risk for overshooting.

In general, increased load leads to increased contact pressure between the components which results in increased friction levels. Incorrect load definitions are believed to be the major cause of the higher alarm rate in axis 6 compared to other wrist axes for both methods. Regarding the distribution of alarms per axis, it is expected that axis one should have a relatively lower rate of alarms compared to other axes because this is usually the least utilized axis.

3.4.3 Consistency

In the evaluation of the methods' consistency, we assumed that the more consistent the method is, i.e. the longer alarm streaks it produces, the more likely it is that the cause of the alarms is wear-related. It can be questioned if this is a valid assumption since incorrect load definitions could result in alarming patterns with multiple consecutive alarms.

To further discriminate between consecutive alarms caused by incorrect load definitions and those caused by wear-related phenomena, a more valid approach perhaps would be to study the trend of the test statistic when multiple consecutive alarms have been raised. In such situations, a positive trend is more suggestive of wear-related alarms rather than alarms caused by incorrect load definitions, which are more likely to affect the system abruptly with no following trend. Nonetheless, in the study of the alarm streaks we observed that single alarms are less common in the proposed method compared to the benchmark (90 compared to 4200 single alarms) which suggests that the proposed method is much less sensitive to noise and isolated disturbances such as temperature variations.

The analogy between the alarm streaks of the benchmark and the coin toss simulation could easily be misunderstood. Its purpose is not to suggest that the performance of the benchmark is similar to that of a random guessing. That statement wouldn't be true since we observed that the benchmark has predictive power in the labeled dataset. Rather the purpose was to illustrate that the distribution for shorter alarm streaks has similar characteristics to that of streaks in random walks or coin tosses where outcomes are equally likely, suggesting that the benchmark has a low signal-to-noise ratio.

3.4.4 Isolability

When evaluating the isolability of the methods, we observed that simultaneous alarms on multiple axes during a diagnostic test cycle are more likely to occur in predictions from the benchmark method. To complement the shortcoming that isolability was only evaluated in the spatial domain, statistics for isolability in the time domain were also gathered. In both cases, it was observed that the proposed method in general has a better isolability compared to the benchmark. Whether this also results in better fault isolability is harder to give a definitive answer to, although the higher hit rate and lower false alarm rate in labeled dataset and the presented follow-up cases would suggest so.

To further explore the isolability of the methods, it was of interest to study the most common axial combinations of alarms. In this section, we observed that in most cases combinations of axes with close proximity to each other were more common in the proposed method compared to the benchmark, which suggests that the proposed method has a better capability to isolate.

Since a predominance of alarms in axis 6 can be observed in most of the combinations, it is believed that these alarm patterns are caused by incorrect load definitions. Examination of the event logs of some of the involved systems showed error codes such as “Joint load too high” were common during the same period as the alarms which gave more strength to this belief. But no conclusive evidence could be presented due to unavailability of event log data for all systems of interest.

Chapter 4
Conclusions

In this thesis I have proposed a method for detecting wear in joints of manipulators that operate in a repetitive manner. The validity of the proposed method has been shown on real failure cases in both staged and production environments. To reduce the sensitivity to disturbances such as temperature variations, an algorithm that combines low-pass filtering and a CUSUM-test is suggested. In evaluation on real failure cases, the proposed algorithm has been shown to decrease the false alarm rate while increasing the hit rate and precision.

Furthermore, two comparative evaluations of the proposed method were made against a benchmark algorithm. In the first evaluation, it was shown that the proposed method has a significantly lower false alarm rate and a higher hit rate. In the second evaluation that was made on a larger unlabeled dataset, it was also shown that the proposed method raises significantly fewer alarms. In an attempt to quantify other metrics that are considered important in condition monitoring systems, such as consistency and isolability, it was shown that the proposed method raises more consistent and isolated alarms compared to the benchmark method.

However, issues that affect the performance of the method still remain. The refinement of the method which reduced sensitivity to temperature variations introduced a delay in the response to wear-related phenomena, resulting in a slowly accumulating fault indicator especially when diagnostic test cycles are run less frequently. This leads to the conclusion that the parameters of the method need to be adapted to how often the diagnostic test cycle is scheduled. Furthermore, incorrect load definition is a disturbance which still poses a problem since it increases the risk of impaired path accuracy and overshooting and thus also affects the manipulator during the diagnostic test cycle.

As future work, it would be of interest to investigate if the same method is applicable without the need to enforce diagnostic routines. Instead the diagnostic routine could consist of:

1. Searching for suitable and repeating sequences in user-provided code and

2. Sampling all sequences found in 1) to detect wear-related changes.

Such an approach would increase availability and allow for a much higher sample rate but would come with the drawback of reduced repeatability as production plans often change.

References

(2019). *IRB 6620, M2004, Product specification*. ABB. Rev. AA.

Aggarwal, C. C. (2016). *Outlier Analysis*. Springer Publishing Company, Incorporated, 2nd edition.

Al-Bender, F. and Swevers, J. (2009). Characterization of friction force dynamics. *Control Systems, IEEE*, 28:64 – 81.

Bittencourt, A. C. (2012). *Modeling and Diagnosis of Friction and Wear in Industrial Robots*. PhD thesis, Linköpings universitet.

Bittencourt, A. C. and Axelsson, P. (2014). Modeling and experiment design for identification of wear in a robot joint under load and temperature uncertainties based on friction data. *IEEE/ASME Transactions on Mechatronics*, 19(5):1694 – 1706.

Bittencourt, A. C. and Gunnarsson, S. (2012). Static friction in a robot joint - modeling and identification of load and temperature effects. *Journal of Dynamic Systems, Measurement and Control*, 134(5).

Bittencourt, A. C., K. Saarinen, S., Sander-Tavallaey, Gunnarsson, S., and Norrlöf, M. (2014). A data-driven approach to diagnostics of repetitive processes in the distribution domain – applications to gearbox diagnostics in industrial robots and rotating machines. *Mechatronics*, 24(8):1032 – 1041.

Bona, B. and Indri, M. (2006). Friction compensation in robotics: an overview. In *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference, CDC-ECC '05*, volume 2005, pages 4360–4367.

Geissbauer, R., Vedso, J., and Schrauf, S. (2016). Industry 4.0: Building the digital enterprise. Technical report, PWC.

Gramacki, A. (2017). *Nonparametric Kernel Density Estimation and Its Computational Aspects*. Studies in Big Data. Springer International Publishing.

Gustafsson, F. (2001). *Adaptive Filtering and Change Detection*. John Wiley and Sons, Ltd.

Kim, H. M., Park, S. H., and Han, S. I. (2009). Precise friction control for the nonlinear friction system using the friction state observer and sliding mode control with recurrent fuzzy neural networks. *Mechatronics*, 19(6):805 – 815.

Li, Q. and Racine, J. S. (2006). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Liu, Y., Yan, C.-Y., and Zhang, F. (2017). Fault diagnosis and health assessment for rotating machinery based on kernel density estimation and kullback-leibler divergence. In *Fault Diagnosis and Detection*, chapter 6. IntechOpen.

Lynch, K. and Park, F. (2017). *Modern Robotics: Mechanics, Planning, and Control*. Cambridge University Press.

Pereira, A. and Romero, F. (2017). A review of the meanings and the implications of the industry 4.0 concept. *Procedia Manufacturing*, 13:1206 – 1214.

Pimentel, M. A., Clifton, D. A., Clifton, L., and Tarassenko, L. (2014). A review of novelty detection. *Signal Processing*, 99:215 – 249.

Rzeszucinski, P. (2012). *Development of reliable vibration-based condition indicators and their data fusion for the robust health diagnosis of gearboxes*. PhD thesis, The University of Manchester.

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, 66(3):605–610.

Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer-Verlag, Berlin, Heidelberg.

INSTITUTIONEN FÖR DATAVETENSKAP | LUNDS TEKNISKA HÖGSKOLA | PRESENTERAD 2020-08-27

EXAMENSARBETE Anomaly Detection in Time-Series
STUDENT Damir Timotijevic
HANDLEDARE Pierre Nugues (LTH)
EXAMINATOR Jacek Malec (LTH)

Anomalidetektering i tidsserier

POPULÄRVETENSKAPLIG SAMMANFATTNING **Damir Timotijevic**

För robotsystem inom produktionsindustrin eftersträvas hög tillgänglighet tillsammans med låga driftkostnader. Drift medför dock förekomst av ett oundvikligt slitage av systemets komponenter. Detta arbete presenterar en metod för att diagnostisera tillståndet för mekaniska komponenter i ett robotsystem.

För att säkerställa hög tillgänglighet inom produktionsindustrin utförs schemalagt förebyggande underhåll i regelbunda intervaller efter bestämda tider av användning. Denna typ av underhållsstrategi minskar risken för oväntade driftsstopp men kan dock medföra ökade kostnader till följd av att många potentiellt obehövliga underhållsåtgärder måste utföras av servicetekniker på plats. En annan, modernare strategi för underhåll är att fjärrövervaka systemets tillstånd med inbäddade sensorer samt föreskriva underhållsåtgärder baserat på indikationer av överhängande risk för utrustningsfel eller minskande prestanda. Denna strategi, kallad tillståndsbaserat underhåll, kan leda till en minskad mängd av onödiga underhållsåtgärder och därmed minskade underhållskostnader. Dock medför metoden svårigheten att bedöma systemets tillstånd på ett tillförlitligt sätt.

Som ett led i detta nya sätt att tänka kring underhåll presenteras i arbetet en metod som kan nyttjas för att upptäcka ökade slitagenivåer i lederna för en industrirobot. Slitage som uppstår mellan de olika komponenterna i en robotled, t.ex. kugghjulen i växellådan, leder till ökad friktion. Denna ökning av friktion kräver en ökning av motorns vridmoment som i industrirobotar vanligtvis är en mätbar storhet. Därmed är det möjligt att vid en repeterande rörelse dra slutsatser om slitagenivån i en robotled genom att studera motorns vridmoment.

Eftersom slitage oftare är lättare att upptäcka som amplitudförändringar i vridmomentet studeras signalerna genom att transformeras till distributionsdomänen. Jämförelsen av distributionen för vridmomentet för en och samma rörelse över tiden med vridmomentets distribution då roboten varit i felfritt skick gör det möjligt att dra slutsatser om förekomsten av ökade slitagenivåer.

Resultaten visar att metoden med framgång kan tillämpas med syftet att upptäcka ökade slitagenivåer i robotleder i riktiga fall. Vidare gjordes två jämförande utvärderingar av metoden mot en referensalgoritm, i den första visades att metoden har en högre sensivitet, bättre precision samt en betydligt mindre mängd falska larm jämfört med referensalgoritmen. I den andra utvärderingen visades att den föreslagna metoden ger upphov till färre, mer konsekventa samt mer isolerade larm.