Yang Zuo - Squat Detection in Railway Switches & Crossings Using Point Machine Vibration

Railway maintenance with AI
Squat Detection in Railway Switches & Crossings
Using Point Machine Vibration

Yang Zuo

Luleå 2022
Division of Operation, Maintenance and Acoustics
Department of Civil, Environmental and Natural Resources Engineering
Luleå University of Technology
SE- 97187 Luleå, Sweden
In God we trust
In love we believe
The research work presented in this thesis has been carried out in the Division of Operation, Maintenance and Acoustics at Luleå University of Technology with financing from Järnvägstekniskt Centrum (JVTC). I have received invaluable support and help from so many colleagues. They all have helped to contribute to this thesis in various ways.

First of all, I would like to thank my main supervisor Assoc. Prof. Matti Rantatalo for the opportunity to conduct this exciting research and for his patient guidance and generous support. I would like also to thank my co-supervisors Prof. Jan Lundberg, Lecturer Johan Odelius and Researcher Taoufik Najeh for their valuable suggestions and comments on my research. I would also thank Prof. Uday Kumar for offering me the position to be part of the division. Many thanks to Researcher Praneeth Chandran, assoc. Lecturer Florian Thiery, Lecturer Stephen Mayowa Famurewa and Postdoc Iman Soleimanmeigouni for all the help I have received and the interesting discussions we have had. I am grateful to all the colleagues and friends from our division, out of our division for their support and accompany.

Finally, I would like to express my deepest gratitude to my family. My parents have supported and inspired me since I was a child. My deepest thanks to the wonderful kids I have, Harry and Terry. They have been understanding, patient and supportive during my whole study. Many thanks to my ex-wife Xiaolei Hu for helping take care of them when I was busy. Special thanks to my ex-girlfriend Yelyzaveta Kalianova for being more than supportive during my study but unfortunately won’t be able to witness this.

Yang Zuo

September 2022

Luleå, Sweden
Railway switches and crossings (S&Cs) are among the most important high-value components on a railway network and a single failure of such an asset could result in severe network disturbance, huge economical loss, and even severe accidents. Therefore, potential defects need to be detected at an early stage and the status of the S&C must be monitored to prevent such consequences. One type of defect that can occur is called a squat. A squat is a local defect like a dent or an open pit in the rail surface. In this thesis, a testbed including a full-scale S&C and a bogie was studied. Vibrations were measured for different squat sizes by an accelerometer mounted at the point machine, while the bogie was travelling along the S&C. A method of processing the vibration data and the speed data is proposed to investigate the feasibility of detecting and quantifying the severity of a squat. A group of features were extracted to apply isolation forest to generate anomaly scores to estimate the health status of the S&C. One key technology applied is wavelet denoising. The study shows that it is possible to monitor the development of the squat size introduced in the test bed by measuring point machine vibrations. The relationships between the normalised peak-to-peak amplitude of the vibration signal and the squat depth were estimated. The results also show that the proposed method is effective and can produce anomaly scores that indicate the general health status of an S&C regarding squat defects.
LIST OF APPENDED PAPERS

Paper 1

Paper 2
The main research work carried out for the appended papers was contributed by the thesis author as well as the co-authors. The contributions of the first author and the co-authors of the papers are listed in the table below:

1. Idea conception (Forming research questions and proposing methods)
2. Measurement setup and data collection
3. Data processing and analysis
4. Manuscript drafting
5. Revision and discussions
6. Final approval of the published version

<table>
<thead>
<tr>
<th></th>
<th>Paper1</th>
<th>Paper2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yang Zuo</td>
<td>3,4,5,6</td>
<td>1,3,4,5,6</td>
</tr>
<tr>
<td>Matti Rantatalo</td>
<td>3,5,6</td>
<td>3,5,6</td>
</tr>
<tr>
<td>Jan Lundberg</td>
<td>1,2,6</td>
<td>1,2</td>
</tr>
<tr>
<td>Taoufik Najeh</td>
<td>2,3,6</td>
<td>2</td>
</tr>
<tr>
<td>Johan Odelius</td>
<td>5,6</td>
<td>1,6</td>
</tr>
<tr>
<td>Florian Thiery</td>
<td></td>
<td>1,5,6</td>
</tr>
<tr>
<td>Praneeth Chandran</td>
<td></td>
<td>1,5,6</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>ABA</td>
<td>Axle Box Acceleration</td>
<td></td>
</tr>
<tr>
<td>AI</td>
<td>Artificial intelligence</td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>Bogie Acceleration</td>
<td></td>
</tr>
<tr>
<td>CBM</td>
<td>Condition-Based Maintenance</td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td>Corrective Maintenance</td>
<td></td>
</tr>
<tr>
<td>CNN</td>
<td>Convolutional Neural Network</td>
<td></td>
</tr>
<tr>
<td>CWT</td>
<td>Continuous Wavelet Transform</td>
<td></td>
</tr>
<tr>
<td>DCNN</td>
<td>Deep Convolutional Neural Network</td>
<td></td>
</tr>
<tr>
<td>DWT</td>
<td>Discrete Wavelet Transform</td>
<td></td>
</tr>
<tr>
<td>ECT</td>
<td>Eddy Current Testing</td>
<td></td>
</tr>
<tr>
<td>FE</td>
<td>Finite Element</td>
<td></td>
</tr>
<tr>
<td>IoT</td>
<td>Internet of Things</td>
<td></td>
</tr>
<tr>
<td>LCCs</td>
<td>Life-Cycle Costs</td>
<td></td>
</tr>
<tr>
<td>LOF</td>
<td>Local Outlier Factor</td>
<td></td>
</tr>
<tr>
<td>MPQY</td>
<td>Multi-Purpose Q and Y load detector</td>
<td></td>
</tr>
<tr>
<td>MSK</td>
<td>Million Swedish Kronor</td>
<td></td>
</tr>
<tr>
<td>NDT</td>
<td>Non-Destructive Testing</td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
<td></td>
</tr>
<tr>
<td>PdM</td>
<td>Predictive Maintenance</td>
<td></td>
</tr>
<tr>
<td>PM</td>
<td>Preventive Maintenance</td>
<td></td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
<td></td>
</tr>
<tr>
<td>QTA</td>
<td>Qualitative Trend Analysis</td>
<td></td>
</tr>
<tr>
<td>S&amp;C</td>
<td>Switch and Crossing</td>
<td></td>
</tr>
<tr>
<td>STFT</td>
<td>Short-time Fourier transform</td>
<td></td>
</tr>
<tr>
<td>TRV</td>
<td>Track Recording Vehicle</td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>United Kingdom</td>
<td></td>
</tr>
<tr>
<td>VGS</td>
<td>Video Gauge System</td>
<td></td>
</tr>
</tbody>
</table>
1.1 Background

The railway has always been an important part of the transportation infrastructure since the first day it was introduced. In recent years, many countries have adopted the philosophy of eco-friendly transportation and encouraged their citizens to use public transportation instead of driving personal cars. One approach to follow this philosophy is to enhance the existing public transportation network such as the railway. This approach makes the railway even more important.

The railway network has been subjected to improvement and expansion remarkably in the past decades due to both the evolvement of railway technologies and the increase in demands. The railway transportation system has been proven to be safe, economical, comfortable, and environment friendly, which makes it an attractive choice as public transportation. According to previous research, it can be considered one of the most efficient means of transportation, especially for travels with a range between 100 km to 1000 km (Esveld, 2001; Indraratna et al., 2011).

In Europe, both the railway freight and passenger traffic have increased during the past decade. Sweden has also experienced such a trend due to the growth of population, development of the economy and more frequent global trade (Trafikverket, 2019). Because of the challenges of severe congestion of both the roads and the sky, rising energy prices due to energy crises and the increasing emission restriction to save the environment, further shifting from air and road to rail can be anticipated (Stenström, 2014). The rise in traffic volume and load have reduced the dependability of the railway network and influenced the achieved operational performance and Quality of Service (QoS) in a negative way (Åhrén & Parida, 2009).

As connecting points in railway, the dependability of Railway Switches and Crossings (S&Cs) is facing challenges. Being safety-critical high-value assets, S&Cs in a railway network enable trains to switch between different tracks. To achieve such functionality, S&Cs include movable parts. This, together with discontinuities in the rail geometry and variability in the track support stiffness, cause S&C to have higher failure rates compared with a plain rail line (Kassa et al., 2006).

The current maintenance strategies and activities for S&Cs are cost intensive. Previous research has pointed out that in the United Kingdom (UK), S&Cs have consumed 24% of the maintenance and 23% of the renewal budget against only...
5% of the track length (Cornish et al., 2016). In Sweden, in 2018 alone, S&Cs have consumed over 530 Million Swedish Kronor (MSK), which is around 10% of the whole maintenance budget (Trafikverket, 2018).

Hashemian claimed that to be able to lower the maintenance cost, avoid unnecessary replacement and improve the safety, availability and efficiency of S&Cs, Condition-Based Maintenance (CBM) is needed (Hashemian, 2010). The immersion of modern techniques such as the Internet of Things (IoT), sensor technology and Artificial intelligence (AI) also contributes to push this trend of maintenance strategies from traditional Corrective Maintenance (CM) to Preventive Maintenance (PM) and eventually to CBM. Predictive Maintenance (PdM) is a subset of CBM. In CBM, the maintenance decisions are based on the previous and current state of the asset. PdM is a version of CBM where the prediction of future states is included in the formulation of the maintenance decisions.

The core of CBM is the condition monitoring process, where different types of signals are monitored using corresponding types of sensors or other appropriate condition indicators (Campos, 2009). Then the maintenance activities are performed only when necessary or just before failure occurs (Andersen & Rasmussen, 1999). The price of sensors is decreasing, which decreases the cost of condition monitoring and increases the potential gain.

1.2 Rail and S&C defects

Nowadays, the most common method of rail and S&C inspection is still manual on-site inspections, which means dedicated railway maintenance experts are needed to visit and inspect the rail regularly at fixed intervals. They base the inspection on measurement tools, their vision, experiences, and insights to detect any defects.

Another way of inspecting the rails is utilising a dedicated Track Recording Vehicle (TRV), which uses different optical sensors, accelerometers and gyro sensors for measuring the different irregularities.

Other non-destructive testing (NDT) techniques that have been adopted to evaluate the rail defects include utilising vision technology (e.g. laser/camera), ultrasound, Eddy Current Testing (ECT) system, strain gauges and accelerometers (Barke & Chiu, 2005).

Wear, fatigue and plastic deformations are among the most common types of defects observed in S&Cs. A squat is a common defect usually caused by Rolling Contact Fatigue (RCF) and appears on the running surface of the rails as a local depression and a dark spot containing cracks with either a circular arc or V-shape
(Grossoni et al., 2021). The root causes of squats are differential wear and plastic deformation according to previous study (Li et al., 2008).

1.3 Problem statement

Some existing methods of performing an inspection of rails and S&Cs are listed below. Their corresponding drawbacks and problems of each method are discussed.

The main problem of manual on-site inspection is that it could take a long time which would expose inspectors to hazardous situations and subject the achieved results to human errors. If an expert misses some abnormalities, they might lead to tragic and severe accidents (Gibert et al., 2016). This is especially true for S&Cs as they are one of the key components in the railway network.

The TRV system provides accurate measurements of the track condition. However, it is expensive to own and perform the inspection, thus the frequency of this type of inspection is usually low. Another issue with this method is that the optical sensors utilised are sensitive to the surrounding harsh environment. Due to the low frequency of inspections, severe track defects may not be detected in time and potential safety hazards may occur (Wei et al., 2016a).

The visual inspection approach is sensitive to light intensity and detection accuracy is poor in the morning and evening hours and certain weather conditions such as rain and snow will also influence the accuracy (Liu, S. et al., 2019). Lesiak et al. claimed that the laser scatterometry inspection is sensitive to dust and dirt on the lenses (Lesiak et al., 2015). Ultrasonic techniques suffer from some technical issues due to some environmental or operating conditions (Bombarda et al., 2021).

The main challenge of the ECT system is the lift-off effect that affects the ECT signal causing erroneous data interpretation (AbdAlla et al., 2019). The strain gauge is sensitive to electromagnetic interference, fragile, of excessive size, and high dependence on the temperature (Kouroussis et al., 2015).

One disadvantage of using Axle Box Acceleration (ABA) is that the maximal ABA excited by the wheel-track contact can be greater than 50 g; therefore, a larger operating range is required for the accelerometer. A trade-off in sensitivity is needed to meet the maximum measuring range. Besides, the vibration of the bearing itself and the vibration caused by the bearing defects would be mixed in the ABA measurements (Wei et al., 2016b). A roadside measurement system will be able to measure the status of the S&C more frequently, making it possible to detect defects at an earlier stage. Another disadvantage with ABA systems is the possibility of missing a specific squat defect due to the sinusoidal hunting movement of the measurement vehicle.
1.4 Purpose and objectives

The purpose of this study is to automate the process of squat detection and monitor the health status of S&Cs to get more frequent and robust updates of the status, reduce the cost of inspections, reduce system downtime and increase the safety of train operation.

The research objectives are:

I. Investigate the possibilities of using accelerometers mounted on the point machine to identify squats with different sizes and locations on S&Cs.

II. Propose an approach for automatic condition monitoring of S&Cs related to squats.

1.5 Research questions

Main RQ: How can vibration data collected at the point machine be used to detect defects and monitor railway S&Cs health?

The main research question was broken down into two research questions.

RQ1: How can squat defects be detected and quantified in S&Cs using vibration data collected at the point machine?

RQ2: How can squat anomalies of S&Cs be detected automatically to monitor their health status?

1.6 Linkage of research questions and appended papers

<table>
<thead>
<tr>
<th></th>
<th>Paper 1</th>
<th>Paper 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ 1</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>RQ 2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Paper 1 Proposed a method to analyse vibrations measured by accelerometers installed at the point machine to investigate the possibility to detect rail squats. The vibration signals used in this study were collected from a testbed. This study also investigated the possibility to quantify different squat sizes from the acquired signals.
Paper 2  Presented a method to analyse the vibration signals from the same test bed. Feature extraction and feature selection were applied to perform unsupervised machine learning algorithms for automatically monitoring the health status of an S&C.

1.7  Scope and limitations

The scope of this research is to study vibration data collected at the point machine to automatically monitor the health status of the S&C.

The research presented in this thesis has the following main limitations:

- Only one type of rail defect, rail squats, was investigated.
- The bogie speed in the test bed was limited to under 2 m/s.
- The study was limited to experiments from a testbed with one type of point machine and one type of bogie wagon.
- The measurement data-set size was limited. Only three repetitions were performed per test case.
2.1 Squat detection

A simplified method to detect the squat defects using laser scatterometry was presented by Lesiak et al. and both the simulation and experiment results achieved from artificial and real defects verified that it is effective in squat detection (Lesiak et al., 2015). The authors analysed the behaviour of laser beam passing through the artificial patterns of defects carved on the surface of rail head. However, this method encountered the challenges of dust and dirt on the laser transmitter or receiver. Ye et al. utilised 3D reconstruction techniques to improve the performance of laser-based track inspection and the results demonstrated the feasibility of using such techniques in railway systems (Ye et al., 2019). The authors built a novel 3D perceptual system based on a low-cost 2D laser sensor. The study verified that the proposed method for rail and crossing-nose inspection was feasible. However, the issues of dust and dirt on the laser transmitter or receiver in operation environment remained as a key concern. Faghih-Roohi et al. proposed a Deep Convolutional Neural Network (DCNN) approach to analyse image data for the detection of rail surface defects and the experiments showed promising results and demonstrated its capability (Faghih-Roohi et al., 2016). The results of different DCNN architectures with different sizes and activation functions were tested to explore the efficiency of the proposed DCNN for detection and classification. However, the lens of the camera remained sensitive to dusts and dirt from the operation environment. Another approach of semi-supervised squat detection for imbalanced image data were presented by Kassa et al. and it was concluded that the approach was a reasonable alternative for improving the performance of rail track squat detection (Kassa et al., 2006). Two models were proposed within the study. One model was generated using a commercial MultiBody System (MBS) software and the other was based on a multibody dynamics formulation. The results from both models showed that the maximum lateral displacement occurred at similar position and agreed with one another. However, these were simulated results and field experiments to verify them were not performed. A monitoring method using ultrasonic techniques was proposed by Kaewunruen & Ishida to measure the crack propagation of squats. The results showed that the propagation can be roughly estimated to be linear to accumulated passing tonnages up to a certain degree (Kaewunruen & Ishida, 2016). This finding could help railway authorities to plan more effective and efficient maintenance. However, the actual detection and measurement of the squat size were not discussed. An embedded system
based on ECT for online detecting and locating rails defects was presented by Alvarenga et al. and it claimed a classification accuracy of 98% (Alvarenga et al., 2021). The proposed method aimed to interpret Eddy Current signals by using wavelet transforms. The processed data were further classified using a Convolutional Neural Network (CNN). However, well-labelled data in railway field are very rare and can be labour and cost intensive. A method to evaluate the track geometry and the load of fixed railway crossings using processed train gauge signals was proposed by Oßberger et al. and the results could be used as input for the condition monitoring process of the S&C (Oßberger et al., 2017). The superposition of the corresponding 2D profiles and the 3D reconstruction allowed a quantitative measurement of the geometry changes during actual train service. However, this study only focused on the crossing nose and not the whole S&C.

Another way of detecting squats is using the ABA measurements. Bocciolone et al. investigated the feasibility of ABA-measurement-based rail status diagnosis (Bocciolone et al., 2007). The signal processing techniques of produced some RMS band values and associated the growing of one band level to the short-pitch corrugation wavelength off-line. But the algorithm was possible to be implemented on-line to make real-time decisions. However, relationship between the level of the vibration and the depth of the short-pitch corrugation was not established in this study. Molodova et al. used the Finite Element (FE) model to perform a parametric study analysing the relationship between maximum ABA signal amplitude and the train speed for squats (Molodova, Maria et al., 2015). A practical method was proposed to represent the relationship between the train speed and corresponding characteristics of ABA at each squat. The parameter study also indicates that the major frequency characteristics of ABA at squats were related to the natural frequencies of the track. However, the squat detection part was not included in the study. An automatic detecting method for track surface squats using ABA measurements was also presented and validated (Molodova, Maria et al., 2014). The automatic squat detection algorithm used wavelet spectrum analysis and estimated the squat locations. However, this approach aimed to perform squat detection for general rails, and it was not dedicated to performing condition monitoring for S&Cs. Wei et al. used the Bogie Acceleration (BA) collected by the accelerometers installed on the bogie frame for detecting both squats and corrugations (Wei et al., 2019). The study pointed out two main disadvantages of using ABA. One is the large values of the ABA that can exceed 50 g, therefore the acceleration sensor must have a trade-off between the maximum measuring range and the sensitivity to detect the track defects. Another problem of the ABA is the vibration caused by the bearing defects are mixed in the measurements. Therefore, the study chose BA signal instead. The results showed the proposed method could detect squats defects from the in-service train successfully. However, this approach was not designed to perform dedicated condition monitoring for S&Cs.
2.2 Condition monitoring for S&C

S&Cs are critical asserts in railway network and the condition monitoring for S&C is an important topic. Many studies have been conducted previously in this area. One study presented an algorithm using Qualitative Trend Analysis (QTA) to detect and diagnose faults in switches (Silmon & Roberts, 2010). The authors claimed that the increased fault diagnosis capability had the potential to contribute significantly towards the achievement of the 30% reduction in track Life-Cycle Costs (LCCs). However, the proposed method was designed for the switch system only and did not consider defects such as squats. Taştimur et al. proposed a vision-based condition monitoring approach for S&Cs using hierarchical SVM (Taştimur et al., 2016). The authors claimed that the performance of the proposed approach was superior to previous methods found in the literatures. However, the proposed method only could identify the S&C from provided images but not the defects. Liu et al. experimented with two different systems, one equipped with a 3-D accelerometer and a speed detection sensor to describe crossing degradation and the other using a Video Gauge System (VGS) to detect and quantify ballast conditions (Liu, X. et al., 2018). However, the measurements were sensitive to the speed and the type of the trains and only data from the same train type and with a similar speed could be used as input. Boogaard et al. presented a method of utilising both accelerometers and a strain gauge mounted 50 mm below the crossing frog (Boogaard et al., 2018). Only the vibration data from the furthest measuring point from the tip of the nose were processed in the study. The results showed the advantages of combining two different measuring methods for monitoring the crossing nose. However, the study did not cover monitoring the whole S&C. Barkhordari et al. proposed a method of employing a track-side system measuring the track acceleration to monitor ballast degradation (Barkhordari et al., 2020). However, this method could not provide continuous condition monitoring. Another study presented a condition monitoring system for railway crossing geometry via both measured and simulated track responses. A calibration method was proposed to improve both the agreement between measured and simulated sleeper displacements for lower frequency track response and for the dynamic track response at high frequencies. However, the system had not been proven to be able to detect the change in crossing geometry over time and to quantify the geometry change.

So far, there has no study been found to combine both the advantage of vibration related squat detection techniques to the condition monitoring health status of S&Cs. This study proposes a new method for squat detection and S&C health status monitoring related to squats. To achieve this, the vibration data measured by the accelerometer mounted at the point machine are used as input data.
3.1 Vibration measurements

Traditionally, some methods have been used to detect rail defects such as ultrasonic test vehicle (Kondo et al., 1996), ECT system (Mohan et al., 2011), strain gauge instrumented wheelsets (Magel et al., 2008), image-based visual system (Bojarczak, 2013) and Multi-Purpose Q and Y load detector (MPQY) (Delprete & Rosso, 2009) etc. Analysing vibration measurements is another method that have been widely used. It can be used to detect and diagnose defects for both the wheels and the rails. The advantages of this method are the cheap price, convenient installation and the integrity of the contained information (Wei et al., 2019; Molodova, Marija et al., 2011; Najeh et al., 2021).

Detecting defects using vibration signals is generally conducted based on features extracted from the signals in either time, frequency, or time-frequency-domain. The extracted features are then applied as the input for further analysis. An example of a time-domain vibration signal measured from an S&C subjected to an excitation of a moving bogie is shown in Figure 3.1.

An accelerometer is a device that measures the vibration in terms of acceleration of an object. The force generated by vibration (acceleration) causes the mass to compress the piezoelectric material, which produces an electrical charge that is proportional to the force applied on it. The charge is proportional to the force, and the mass is a constant. By applying Newton’s second law, the charge is also proportional to the acceleration.

3.2 Wavelet denoising

The concept of wavelet transform can be traced back to 1909. The concept of wavelet transform is closely related to the Fourier transform. The Fourier transform is a useful tool to analyze the constant frequency components of a signal. However, it is impossible to tell at what instant a particular frequency
component occurs. Short-time Fourier transform (STFT) uses a sliding window to find spectrogram and provides the information in both time and frequency domains. However, the length of window limits the resolution in frequency. Wavelet transform solves the problem by using small wavelets functions with limited duration (Chun-Lin, 2010). Wavelet transform can be divided into two categories, namely, Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT) (Torrence & Compo, 1998). CWT is a powerful tool for time-frequency analysis and can be viewed as replacing the short-time Fourier transform’s time-frequency window $g_{t,\xi}$ with a “time-scale window” $\psi_{a,b}$. The CWT can be defined as follows according to the study conducted by Johanson (Johansson, 2005):

A function $\psi$ with

$$\int_{\mathbb{R}} \psi(x) \, dx = 0 \quad (3.1)$$

is called a wavelet. For a given function $f(x)$, and a selected wavelet $\psi$, its CWT is defined as

$$W_{\psi} f (a, b) = \int_{\mathbb{R}} f(x) \psi\left(\frac{x-b}{a}\right) \, dx \text{ for all } a, b \in \mathbb{R}_+ \times \mathbb{R} \quad (3.2)$$

where

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \quad (3.3)$$

The function $\psi$ is called the mother wavelet. It is chosen to be localised at $x = 0$ and at some $\omega = \omega_0 > 0$ (and/or $\omega = -\omega_0$). $a$ is a scaling/dilation factor that controls the width of the wavelet and $b$ is a translation parameter that controls its location.

DWT decomposes a signal into a set of mutually orthogonal wavelet basis functions and is defined as follows (Torrence & Compo, 1998):

$$DWT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) \psi\left(\frac{x-b}{a}\right) \, dx \quad (3.4)$$

where

$$a = 2^j, \quad b = k2^j, \quad (k, j \in \mathbb{Z}) \quad (3.5)$$

Wavelet denoising utilises DWT to decompose the original signal to obtain the wavelet coefficients, thresholding the coefficients and reconstructing the signal with reverse DWT (Peng & Chu, 2004). Here $a$ is called the scale factor and
represents the scaling of the function, and \( b \) is called the shift factor and represents the temporal offset of the function.

### 3.3 Scale averaged wavelet power (SAWP)

To evaluate fluctuation in power over a range of scales, the SAWP time series over scales \( s_1 \) to \( s_2 \) is defined as follows (Torrence & Compo, 1998):

\[
\overline{W}_n^2 = \frac{\delta_f}{C_\delta} \sum_{j=1}^{j_2} \left| \frac{W_n(s_j)}{s_j} \right|^2
\]  

(3.6)

where

\[
s_j = s_0^{2^{j-1}}, j = 0, 1, \ldots, J
\]  

(3.7)

\[
J = \delta_j^{-1} \log_2 \left( N\delta_f / s_0 \right)
\]  

(3.8)

\( C_\delta \) is scale independent and a constant for the selected wavelet function, \( \delta_j \) is a factor for scale averaging, \( \delta_f \) is the sampling period and \( j_1, \ldots, j_2 \) represent scales over which the SAWP is computed. \( s_0 \) is the smallest resolvable scale and \( J \) determines the largest scale. \( W_n(s) \) is the CWT of a discrete sequence. \( N \) is the number of points in the time series (Kaiser & Hudgins, 1994). SAWP is utilised to detect the power burst in the vibration signal when a wheel hits a squat or a joint gap. This power time series will be used later to extract some features.

### 3.4 Machine learning

Machine learning addresses the question of how to build computer programs that improve automatically through learning experience (Jordan & Mitchell, 2015). Machine learning algorithms can mainly be divided into three categories, namely, supervised learning, unsupervised learning, and reinforcement learning (Jordan & Mitchell, 2015). Another way of categorising machine learning is according to the learning scenarios. Supervised learning, unsupervised learning, semi-supervised learning, transductive inference, on-line learning, reinforcement learning, and active learning are the main categories (Mohri et al., 2018).

Supervised learning is a machine learning approach that is defined by its use of labelled datasets to predict the output. Depending on the predicted output type, it could be either a regression problem when the outputs are quantitative, or a classification problem when the outputs are qualitative (Hastie et al., 2009). In
other words, in classification problems, the labels are discrete and in regression problems, the label are continuous (Nasteski, 2017). These two tasks are similar, and both can be viewed as a function approximation task. A few widely applied supervised machine learning algorithms are support-vector machines, linear regression, logistic regression, Naive Bayes, linear discriminant analysis, decision trees, k-nearest neighbour algorithm, neural networks, and similarity learning.

Unsupervised learning is a type of machine learning algorithm that learns patterns from unlabelled data (Mohri et al., 2018). The machine is forced to build a compact internal representation of its world and then generate imaginative content from it (Hinton & Sejnowski, 1999). Another way to conceptualise this process is that the algorithms are left on their own to discover and present the interesting structure in the training data (Mahesh, 2020). The unsupervised learning algorithms learn patterns from the training data. When new data are introduced, it uses the previously learned pattern to recognize to which cluster the new data belong. Clustering, feature reduction and anomaly detection are the three main applications. Principal Component Analysis (PCA), manifold learning and autoencoders are a few examples of unsupervised learning algorithms for dimensionality reduction. Isolation forest, Local Outlier Factor (LOF) and minimum covariance determinant are typical unsupervised machine learning algorithms used in anomaly detection. K-means, hierarchical clustering, DBSCAN, affinity propagation, mean shift and Gaussian mixture models are some popular unsupervised machine learning algorithms used for clustering. Since the fact that a large amount of data is unlabelled from railway maintenance and the healthy data are dominant compared with the data with defects (Hajizadeh et al., 2016), it is suitable to apply unsupervised machine learning algorithms to perform anomaly detection.

Reinforcement learning directly takes inspiration from how humans learn from experience. It features an algorithm that improves by itself and learns from new situations using a trial-and-error method (Kaelbling et al., 1996). Another more recent understanding of reinforcement learning is that it studies how to use past data to enhance the future manipulation of a dynamical system (Recht, 2019). In another word, the goal is to find a sequence of inputs that drives a dynamical system to maximize a given object starting with minimal knowledge of the system. There are two main strategies for solving a reinforcement learning problem. The first is to search in the entire space of behaviors to find one that performs well in the given environment. This approach has been applied in genetic algorithms and genetic programming, and some other novel search techniques (Schmidhuber, 1996). The second is to utilise statistical techniques and dynamic programming methods to estimate the reward of taking actions in the provided environment (Kaelbling et al., 1996).
3.5 Anomaly detection

Anomaly detection aims to detect anomalous or abnormal data points from a provided dataset and discover enthralling and rare patterns in the dataset (Ahmed et al., 2016). Anomalies refer to the patterns in data that do not conform to the defined notion of normal behaviour (Chandola et al., 2009a). For example, Figure 3.2 illustrates anomalies in a simple two-dimensional data set. The data has two normal regions, N1 and N2. Most observations lie in these two regions. Points that are sufficiently far away from these regions, for example, points o1 and o2, and all points labelled o3, are considered as point anomalies. Figure 3.3 shows an example of a contextual anomaly. The black point is considered as normal compared with its neighbouring points as it follows the general trend. The red point depicted in the same figure has the same value and cannot be decided as an anomaly by checking the value itself. However, when it is compared with the neighbouring points, there is a sudden change in the pattern and therefore can be considered as an anomaly in terms of the given context. Such deviations are known as contextual anomalies. Figure 3.4 displays an example of a collective anomaly scenario. The points in the red circle are normal when viewed as individuals. However, when these points are compared to the entire data as a collection, they are anomalies.

![Figure 3.2 An example of point anomalies in a two-dimensional data set. (Adapted from Chandola et al. 2009)](image)
Figure 3.3 An example of contextual anomaly in a time series. (Adapted from Chandola et al. 2009)

Figure 3.4 An example of collective anomaly in a time series. (Adapted from Chandola et al. 2009)
4.1 Research approach and process

Research can be defined in many ways. The common meaning of research is to search for new knowledge. Research is defined as a piece of original contribution to the existing stock of knowledge helping for its advancement (Kothari, 2004). According to Kumar, research can be classified depending on three criteria: the application, the objectives and the enquiry mode (Kumar, 2018). The current study belongs to applied research, with exploratory, descriptive and explanatory objectives. A quantitative approach (Creswell & Creswell, 2017) was utilised to investigate how the squat defects on an S&C can be detected and quantified and how it can be automated to generate an anomaly score to indicate the health status of the S&C. Figure 4.1 illustrates the main research process carried out to answer the research questions.

Figure 4.1 Research design process
The field experiment and vibration measurements were performed in a previous study. The focus of this study was to design methods to process the available data by applying both statistical model and machine learning model to achieve new knowledge.

4.2 Research methodology

4.2.1 Data generation and collection

The experiment was performed in a testbed including a full-scale S&C and a moving bogie. The accelerometer was mounted on the point machine. Both the vibration signal and the corresponding speed information were measured and saved while the bogie was travelling along the S&C with artificially introduced squats. The S&C used in the experiment has a dimension of 1:16 and a length of 38.14 m. A simplified illustration of the testbed is shown in Figure 4.2. The rails are labelled in the figure from rail 1 to rail 4. The squats are labelled from A to K.

The sensor used for this study was a miniature accelerometer of type KS91C. It has a measuring range of 0.3–37,000 Hz, with a sensitivity of $10 \pm 20\%$ mV/g and the resonant frequency greater than 60 kHz (+25 dB). The position of the accelerometer is visualised in Figure 4.2 and Figure 4.3. The vibration in the z-direction was measured. The accelerometer was glued to the aluminium holder, which was mounted on one rod of the point machine.

![Figure 4.2 Schematic diagram of the test setup and accelerometer placement](image)

![Figure 4.3 Sensor mounted on one rod of the point machine for extra protection](image)
To simulate two different squat levels, the squats were manually introduced stepwise with 1 mm and 4 mm maximum depth. The positions and the measured dimensions of the squats are listed in Table 4.1. The squat with 1 mm depth is around 42 mm in diameter and the squat with 4 mm depth is about 63 mm in diameter. $S_0$ and $S_1$ are two stop blocks of the S&C on each end in the through direction. Wheels 2 and 4 travel first on rail 4 and then switch to rail 3 while wheels 1 and 3 always travel on rail 1. The point machine is located 5.86 m away from the stop block $S_0$. Figure 4.4 shows what a real squat and an artificially introduced squat in the testbed look like.

![Figure 4.4 An example of Squat defect (a) A real squat (b) A manually induced artificial squat](image)

### Table 4.1 Measurements of squats' dimensions

<table>
<thead>
<tr>
<th>Rail Nr.</th>
<th>Squat name</th>
<th>From $S_0$ (m)</th>
<th>Squat diameter (mm)</th>
<th>Max depth (mm)</th>
<th>Squat diameter (mm)</th>
<th>Max depth (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>A</td>
<td>5.70</td>
<td>43</td>
<td>1.2</td>
<td>62</td>
<td>3.7</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>6.70</td>
<td>41</td>
<td>1.0</td>
<td>61</td>
<td>3.9</td>
</tr>
<tr>
<td>1</td>
<td>C</td>
<td>7.27</td>
<td>42</td>
<td>1.0</td>
<td>63</td>
<td>3.7</td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>10.68</td>
<td>42</td>
<td>1.0</td>
<td>66</td>
<td>4.4</td>
</tr>
<tr>
<td>1</td>
<td>E</td>
<td>12.47</td>
<td>0</td>
<td>0</td>
<td>65</td>
<td>3.7</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>18.04</td>
<td>42</td>
<td>1.1</td>
<td>65</td>
<td>4.2</td>
</tr>
<tr>
<td>1</td>
<td>G</td>
<td>19.32</td>
<td>42</td>
<td>1.0</td>
<td>64</td>
<td>3.7</td>
</tr>
<tr>
<td>1</td>
<td>H</td>
<td>28.02</td>
<td>42</td>
<td>1.5</td>
<td>62</td>
<td>4.7</td>
</tr>
<tr>
<td>3</td>
<td>I</td>
<td>29.23</td>
<td>42</td>
<td>1.4</td>
<td>62</td>
<td>4.3</td>
</tr>
<tr>
<td>3</td>
<td>J</td>
<td>32.14</td>
<td>42</td>
<td>1.2</td>
<td>63</td>
<td>4.4</td>
</tr>
<tr>
<td>1</td>
<td>K</td>
<td>34.00</td>
<td>42</td>
<td>1.1</td>
<td>61</td>
<td>4.1</td>
</tr>
</tbody>
</table>

### 4.2.2 Data processing

The method to solve RQ 1 is described in detail in Figure 4.5. The vibration signal was first processed to remove the constant signal level, then high-pass filtered, wavelet denoised and converted from time domain to spatial domain. The
signals were resampled and then synchronised with the smoothed and re-sampled speed data. The positions of squats were used to match the final processed vibration signal and identify different events.

The mean value of the vibration signal was first deducted from the original measured signal. This step ensures that the constant component of the signal was removed. Since the focus of this study was to detect impact events, a third order high-pass filter with cut-off frequency at 100 Hz was applied to the original signal to remove some part of the low frequency components.

To apply wavelet denoising, some parameters and the thresholding method should be decided. The denoising was set at a level nine decomposition empirically. The wavelet function should reflect the features presented in the signal in the time domain. However, since the primary interest in this study was the SAWP time series, different types of wavelet functions would yield the same qualitative results (Torrence & Compo, 1998). Symlet 4 (sym4) was chosen. There were some methods that could be used to determine the denoising thresholds. According to previous study, the influence of different methods on the SAWP is insignificant (Torrence & Compo, 1998). Empirical Bayesian with median thresholding was chosen.

The original vibration data was sampled in the time domain. However, the interesting aspect in the study was where the squats are located. With the help of the logged speed information, the data in time domain was converted into the spatial domain. Since the speed was not constant, the converted spatial domain signal had a non-constant distance interval.

To remove sudden changes in the speed measurements, convolution technique was used to smooth the speed data. Resampling was applied to both the speed signal and the vibration signal. Since the speed signal was sampled at 1 Hz and the vibration signal at 51.2 kHz, upsampling was applied to the speed signal. The converted vibration signal in the spatial domain had non-constant sampling frequency due to varying speed condition. Therefore, resampling to 51.2 kHz was applied using interpolation technique.

To perform the spatial domain conversion, synchronisation of the vibration data and the speed estimation data was performed. The actual start time of the vibration data and the speed data were not the same. The data were aligned by assuming both signals stop at the same time. Furthermore, parts of the signals where the measured speed was constantly zero were removed due to that part of the signal were recorded before the bogie started to move. The technique used to help synchronising the vibration signals in the spatial domain with the expected events was utilising signatory common impulse response as reference points. The event chosen was when the front wheel hit the first rail joint, which happened for all test cases. This point was used to align the vibration signal in
the spatial domain to the expected events such as rail hitting squats, rail gaps or stop blocks.

*Figure 4.5 Data processing for RQ 1*
The method to solve RQ 2 is described in detail in Figure 4.6. The vibration signals were initially filtered with a third-order Butterworth band-pass filter with 50 Hz and 2.5 kHz cutoff frequencies. The band-pass filter was applied to filter away the frequencies with noise and preserve the frequencies with useful information. A wavelet magnitude scalogram was utilised as a tool to decide the cutoff frequency of the band-pass filter as shown in Figure 4.7. It showed that the main...
energy of the response for the squat defect was around 200 Hz to 400 Hz. There was also a second frequency band around 500 Hz to 2000 Hz. The filtered signals were aligned and truncated to equal length. This step made it possible to compare the results from different runs in the results. It could also be implemented in future studies to accurately extract the position information.

![Magnitude scalogram of squat G in a 4 mm case](image)

Furthermore, the signal was down-sampled to one-tenth of the original frequency. As the band-pass filter has a cutoff frequency as high as 2.5 kHz, the original signal with sampling frequency at 51.2 kHz contains redundant information. A sampling frequency at 5 kHz was enough to preserve all the useful information. To make the calculation easier, 5.12 kHz was applied. The output signals were processed in two separate ways after that.

In one way, the signals were segmented into 400 equal-sized segments and nine corresponding time-domain features were extracted. The features used in this study were RMS, standard deviation, shape factor, kurtosis, skewness, peak-to-peak amplitude, impulse factor, crest factor and clearance factor. In the other way, wavelet denoising was applied. The denoising was set at a level nine decomposition, with Symlet 4 wavelet. The empirical Bayesian denoise method was utilised with median thresholding and level-dependent noise estimator. The SAWP was calculated from the output signal. Two features, the number of peaks and the total peak power, were extracted from the SAWP time series and assigned to each segment. Totally 11 features were generated for the entire study.
4.2.3 Data modelling

Linear regression:

Linear regression analysis predicts the value of one variable based on the value of another variable. The variable to be predicted is usually named the dependent variable. The variable used to predict the other is called the independent variable.

The aim of this analysis is to estimate the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. A more straightforward way to understand linear regression is that it tries to fit a straight line or surface that minimizes the error between predicted and actual values. One popular method to measure the error is the “least squares” method.

To answer RQ 1, the influence of non-constant speed need to be analysed. Since the measured speed was in a small interval of 0 m/s to 1.5 m/s, it was reasonable to assume the relationship between the speed normalised peak-to-peak amplitude and the squat depth is approximately linear. Therefore, two linear models were proposed to estimate the relationship between the speed normalised peak-to-peak amplitude and the squat depth for squat F and G as shown in Figure 4.2.

Isolation Forest:

The Isolation Forest algorithm was initially proposed in 2008 (Liu et al., 2008). The authors used two quantitative properties of anomalous data points in each sample:

- Few – the anomalous data points are the minority and consist of fewer instances
- Different – the anomalous data have values that are very different from those of normal ones

Based on the properties of anomalies to be few and different, they are easier to be isolated compared to normal points. The algorithm is described in detail as follows. Given a set of observations, the isolation forest algorithm selects a random sub-sample of the observations and assigns them to a binary tree. The branching of the tree starts by selecting a random feature from d-dimensional features. Then branching is done on a random threshold in the range of the selected feature. If the value of one observation is less than the selected threshold, it goes to the left branch; otherwise, it goes to the right. With such an approach, a node is split into left and right branches. This process should continue recursively until all data points are completely isolated or when the
maximum depth is reached. The above steps are repeated to construct random binary trees until all observations are isolated. Those points that are easier to isolate and with smaller path lengths will thus have higher anomaly scores.

Chandola et al. claimed a few advantages of Isolation Forest (Chandola et al., 2009b). First, it had a low linear time complexity and a small memory requirement. Next, it was suitable to deal with high dimensional data with irrelevant attributes. Lastly, it could be trained with or without anomalies in the training set, plus it could provide detection results with different levels of granularity without re-training.

To answer RQ 2, Isolation Forest anomaly detection was used to detect the anomaly segment of the vibration data. The detailed process of choosing different parameters such as the proper segment size, scaling technique, and feature selection was described and discussed in chapter 5.2.
5.1 Results and discussion related to RQ1

RQ1: How can squat defects be detected and quantified in S&Cs using vibration data collected at the point machine?

RQ1 was mainly answered in paper one. The first step performed in this research was the literature review. The research related to the use of vibration signals to detect the rail defects and the research about condition monitoring of S&Cs were investigated. The results of current studies were analysed. The results showed that analysing the vibration data is a promising approach for detecting the squats of an S&C. The initial results from the testbed also indicated that the vibration data changes when the squats defect level changes. This is shown in Figure 5.1. This was the first proof from data analysis that showed the possibility of using the peak-to-peak amplitude to detect the squats defects. The difference was clearer when a squat was nearer the measuring point and when a squat was located further away, the difference became less significant.
Figure 5.1 Comparison for different squat levels: (a) No squat case (b) 1 mm depth’s squats (c) 4 mm depth’s squats.
The maximum peak-to-peak amplitude values were extracted from the processed vibration signal. The maximum peak-to-peak amplitude of squat F and G are visualised in Figure 5.2 and Figure 5.3, respectively. As the speed of the bogie was non-constant, the peak-to-peak amplitude values could not be directly compared with each other. From all the squats introduced, squat F and squat G were chosen to be further analysed because they were not far away from the accelerometer and there were no joint gaps nearby. From those two figures, we could observe that for squat F and squat G, the speed was around 1m/s. It was reasonable to assume a proportional relationship between the amplitude and the speed within such a small interval. Therefore, all measured amplitude for squat F and G were divided by the corresponding speed to get an estimation of the amplitude at a fixed speed of 1m/s. This assumption might not be valid when the speed variation became larger so that more complex method might be needed to normalize the peak-to-peak amplitude.

Figure 5.2 Peak-to-peak amplitude data for squat F on rail 3 (a) no squat case (b) 1 mm squat case (c) 4 mm squat case
The mean and the standard deviation of the estimated acceleration together with the fitted linear regression model for squat F are visualised in Figure 5.4. It showed that the mean amplitude value increases from 0.7605 g to 1.2355 g when the squat depth increases from 1 mm to 4 mm. The standard deviations for 1 mm and 4 mm cases were 0.3434 g and 0.088 g, respectively. The fitted linear model was:

\[ y = 0.255x + 0.2942 \]  \hspace{1cm} (5.1)

This curve fitting was reasonable because the data points were limited. When more data could be available from field measurements, more advanced curve fitting would be possible to apply. Different approaches could be compared by calculating the error of the fittings on new data.
The mean and the standard deviation of the estimated acceleration together with the fitted linear regression model for squat G are presented in Figure 5.5. The mean amplitude value increased from 0.3499 g to 1.6286 g when the squat depth increased from 1 mm to 4 mm. The standard deviations for 1 mm and 4 mm cases were 0.1437 g and 0.7054 g, respectively. The fitted linear model was:

\[ y = 0.388x + 0.0452 \]  

(5.2)
By analysing these two squats, it could be observed that the fitted linear model were different from one another. Therefore, more analysis should be performed when larger dataset from field tests became available.

5.2 Results and discussion related to RQ2

RQ2: How can squat anomalies of S&Cs be detected automatically to monitor their health status?

This research question was mainly answered in paper two. The first step to answer the research question was to choose how the recorded signal should be cut into segments. A few different segment sizes were tested, and it was observed that dividing the signal into 400 segments was a reasonable choice. Higher number of segments would cut the signal from one squat into a few segments and low number of segments would make one segment containing multiple squats. The results of the influence of segment size are shown in Figure 5.6.

![Figure 5.6 Example of anomaly scores for a case with 4mm artificial squats and with different segment size](image)

The time-domain features used in this study were RMS, standard deviation, shape factor, kurtosis, skewness, peak-to-peak amplitude, impulse factor, crest factor and clearance factor. The features extracted from SAWP were the number of peaks and the total peak power. In total, 11 features were generated. The features extracted for the study are shown in Fel Ogiltig självreferens i bokmärke.
Two different approaches for feature selection were utilised. The first approach made use of PCA. The accumulated PCA feature importance score is presented in Figure 5.7 (a). The first five features in PCA space captured 96.55% of all the useful information. The second approach employed the Laplacian score and correlations between the features. First, the Laplacian feature importance score was calculated and ranked. The results are shown in Figure 5.7 (b). The y axle was not the Laplacian score but a score representing feature importance as defined in the Matlab implementation. A lower Laplacian score means a more relevant feature; thus a higher feature importance score. Then correlations between the most significant feature and the others were calculated. All other features that had a correlation value higher than 0.9 to the most significant one was removed. The above steps were repeated until there were no two features with a correlation value higher than 0.9. As a result, the remaining features were feature numbers 10, 11, 9, 8, 5 and 6.
The results of utilising different features groups were compared. Figure 5.8 shows one example of the anomaly scores for a test with 4mm squat case with different features. By comparing the anomaly scores for those three cases, however, it is evident that the difference is small and from a performance point of view, either of them could be used. However, a lower dimension of features
was preferred. The PCA space features were linear combinations of the original features and were difficult to interpret. Based on the above two arguments, the Laplacian score selected features were used for further study.

A threshold was needed to decide what should be considered as an anomaly. The knee point method was applied on the sorted anomaly score as is shown in Figure 5.9. The results indicated that the anomalies should be around 12% of the total segments. This was verified by plotting the anomalies with 88 percentiles together with the vibration signal. One example of the results for a 4 mm case is presented in Figure 5.10. By using the 88 percentiles, most of the anomalies were found without introducing unexpected false alarms.
Figure 5.10 Example anomaly detection for 4 mm squat case run 1 with 88 percentiles as a threshold, the anomalies match the pattern of the events (squats and joint gaps)

All the test cases and the corresponding anomaly scores above the 88-percentile threshold are presented in Figure 5.11. This showed that with increased squat depth, more anomalies were found indicating the health status of the S&C was degraded. It can also be observed that the different test runs were well aligned at the beginning, but with the different speed profile for each run, the spotted defects also encountered different drift, and thus were not well aligned any more. This, however, would not influence the results of utilising the anomaly scores as an indicator to monitor the health status of the S&C.

Figure 5.11 Anomaly scores above the 88-percentile threshold for all test cases showing the larger squats cause both anomalies spotted and lower anomaly scores

A few indicators could be calculated to represent the health status of the S&C. One such indicator could be to calculate the sum of anomaly score for each test and then calculating the mean value of the sum for each test scenario with three repetitions. From the test data, we observed the scores of 11.65, 20.31 and 29.59 for the S&C with healthy, 1 mm and 4 mm deep squat cases, respectively. Another indicator could be the mean value of the number of anomalies for each test scenario. From the test data, the average number of anomalies are 18.67, 32.67 and 45.00 for the S&C with healthy, 1mm and 4 mm deep squat cases, respectively.
CHAPTER 6. CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

6.1.1 Conclusions related to RQ1
RQ1: How can squat defects be detected and quantified in S&Cs using vibration data collected at the point machine?

The following conclusions can be drawn from the investigations performed to answer RQ 1:

- The study shows that accelerometers placed within the protective environment within a point machine could be utilised for monitoring defects such as squats along the S&Cs of the railway infrastructure.

- The proposed methodology was able to detect six out of seven squats of 4 mm depth within a range of approximately 13 m from the accelerometer mounted on the point machine.

- The proposed methodology was able to detect four out of six squats of 1 mm depth within a range of approximately 13 m from the accelerometer.

- It was challenging to extract and locate accurately both 1 mm and 4 mm depth squats that are further than approximately 22 m away from the accelerometer.

- The mean normalised amplitude value for squat F increased around 62% when the squat depth increases from 1 mm to 4 mm.

- The mean normalised amplitude value for squat G increased 365% when the squat depth increases from 1 mm to 4 mm.

- A linear model could be used to fit the normalised amplitude versus squat depth for squats F and G within the collected data. Larger data set could however require a different model.
6.1.2 Conclusions related to RQ2

RQ2: How can squat anomalies of S&Cs be detected automatically to monitor their health status?

The following conclusions can be drawn from the investigations performed to answer RQ 2:

- The extracted features from both the time domain vibration signal and the SAWP with the proposed signal-processing procedure was able to be used as input to anomaly detection algorithms to detect squat defects.
- Skewness, peak to peak amplitude, crest factor, clearance factor, number of peaks and total peak power were ranked to be the top features for anomaly detection.
- The selected five PCA space features explained more than 96% of all the variance in the features.
- Anomaly-detection algorithms could be utilised to generate anomaly scores to indicate the health state of S&Cs regarding squat defects. Using knee point technique, 12% of the total segments of all nine instances were determined to be anomalies.
- The mean value of the total anomaly scores for each test scenario increased from 11.65 to 20.31 and 29.59 for S&Cs with healthy, 1 mm deep, and 4 mm deep squat cases, respectively. The values for 1 mm and 4 mm cases were almost 1.7 and 2.5 times greater compared to the healthy case, respectively.
- The mean value of the number of anomalies for each test scenario increased from 18.67, 32.67 and 45.00 for S&Cs with healthy, 1 mm deep, and 4 mm deep squat cases, respectively. The values for 1 mm and 4 mm cases were almost 1.7 and 2.5 times greater compared to the healthy case, respectively.
- An isolation forest algorithm could be utilised to generate anomaly scores to identify squat defects.

6.1.3 Conclusions related to the main research question

Main RQ: How can vibration data collected at the point machine be utilised to detect defects and monitor railway S&Cs health?

In this study, a method was proposed to estimate the relationship between different squat levels and the corresponding normalised signal amplitude. An approach was proposed to process the vibration data to extract different features to detect defects such as squats and estimate the health status of the rails at the S&C. It is possible to use the signal processing approach described in the current study and combine it with anomaly detection techniques such as isolation forest to estimate the health status of S&Cs. Skewness, peak-to-peak
amplitude, crest factor, clearance factor, number of peaks and total peak power are some promising features.

6.2 Future Research

Some possible future research topics and approaches are listed below:

- Collect more data from the test bed to quantify the relationships among the parameters such as the speed, depths and distance to the sensor to the maximum peak-to-peak amplitude in a controlled environment. With controlled variables, influence of each factor could be studied separately.
- Evaluate the proposed method for real case to detect and quantify squats in an S&C.
- Propose and implement a machine learning algorithm to learn from the patterns of the healthy S&Cs and perform continuous anomaly detection on them.
- Evaluate the influence of other parameters such as train type, type of S&C, load and speed on the indicators and extend the presented method.
- Combine the method to answer RQ 2 and the concept of federated learning to propose a nation-wide condition monitoring system for S&Cs.


Johansson, E. (2005). Wavelet theory and some of its applications


Appended Papers
Point Machine Vibration Analysis for Squats Detection of Railway Switches and Crossings
Point Machine Vibration Analysis for Squats Detection of Railway Switches and Crossings

Yang Zuo,1 Jan Lundberg,1 Taoufik Najeh,1 Johan Odellius1 and Matti Rantatalo1

Abstract
Railway switches and crossings (S&C) are among the most important high-value components in a railway network and a failure of such an asset could result in severe network disturbance. Therefore, potential defects need to be detected at an early stage to prevent traffic disturbing downtime or even severe accidents. A squat is a common defect of S&C that has to be monitored and repaired to reduce such risks. In this study, a testbed including a full-scale S&C and a bogie wagon was developed. Vibrations were measured for different squat sizes by an accelerometer mounted at the point machine. A method of processing the vibration data and the speed data is proposed to investigate the possibility of detecting and quantifying the severity of a squat. One key technology used is wavelet denoising. The study shows that it is possible to monitor the development of the squat size on the rail up to 13.46 metres from the point machine. The relationships between the normalised peak-to-peak amplitude of the vibration signal and the squat depth were also estimated.

Keywords
Railway switches & crossings, vibration, squats analysis, condition monitoring, wavelet denoising, fault detection

Introduction
Railway switches and crossings (S&C) play an important role in a railway system by enabling trains to switch between different tracks. To achieve such functionality, S&C include movable components. This, together with the discontinuities in the rail geometry and variability in the track support stiffness, cause higher failure rates compared with plain line tracks.1

Therefore, the maintenance cost of S&C often comprises a considerable part of the total maintenance budget of a railway system. In Sweden, approximately 8 per cent of the budget was dedicated to S&C maintenance in 2009.2 In 2018, S&C maintenance consumed 530 MSEK, which is almost 10 per cent of the total maintenance cost.3 It is reported that S&C in the United Kingdom could consume up to one-third of the whole maintenance budget.4 The cost varies from country to country, depending on S&C type and deployed maintenance strategy. However, the maintenance cost is not the only cost that should be taken into account. A more complete calculation of Life Cycle Cost (LCC) of S&C is analysed by Nissen A.5

To avoid failures of these important junction points and reduce the cost, condition monitoring of S&C and shifting from corrective maintenance to preventive maintenance are needed. Many researchers have proposed different condition monitoring technologies to detect defects that could lead to failures of S&C. One study monitors the motor current and the force in the drive bar to detect switch defects using the Kalman filter.6 Another study utilises qualitative trend analysis of current, force and displacement to detect the fault.7 Previous studies focused on measuring different parameters of S&C such as the voltage, current, force, displacement, etc. of the point machine, during the switching procedure. Those data are later utilised to detect anomalies associated with possible failures or degraded statuses. A review of the existing Fault detection diagnostics (FDD) techniques for railway S&C is performed by Moussa A.8 Different anomaly detection techniques are used to process the data and detect possible defects, such as using a rule-based decision process with the help of the Kalman filter9 and net energy analysis (NEA).10 Such measurement data can also be applied to implement a closed-loop feedback control of a lift-and-drop railway track switch actuator.11 These approaches all focus on defects of the functionality of a point machine. However, other parts of S&C also need to be monitored and assessed, such as different rail parts including the stock rail, the switch rail and the crossing nose/frog.

Rail surface defects can be detected by using manual on-site inspections, vehicle-based measurements using visual systems (camera/laser)12,13 or eddy current systems.14,15 A manual inspection could be subjected to the effect of human errors and can expose inspectors to hazardous situations. The method of using a dedicated track inspection vehicle/eddy current system would acquire accurate measurements, but the frequency of measurements would be limited due to the high cost and long track possession time. In the past decade, with the price of sensors decreasing, there are more and more studies that use inertial sensors mounted on the train to detect faults on the tracks. Lederman G. introduced an implicit

1Luleå University of Technology, Sweden

Corresponding author:
Yang Zuo, Department of Operation and Maintenance, Luleå University of Technology, Laboratorievägen 14, 971 87 Luleå, Sweden.
Email: yang.zuo@ltu.se

Prepared using sage-jdb [Version: 2017/02/017 v.2.9]
Table 1. Squats’ dimension measurements

<table>
<thead>
<tr>
<th>Rail No.</th>
<th>Squat name</th>
<th>From S₀ (m)</th>
<th>Squat diameter (mm)</th>
<th>Max depth (mm)</th>
<th>Squat diameter (mm)</th>
<th>Max depth (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>A</td>
<td>5.7</td>
<td>43</td>
<td>1.2</td>
<td>62</td>
<td>3.7</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>6.7</td>
<td>41</td>
<td>1.0</td>
<td>63</td>
<td>3.9</td>
</tr>
<tr>
<td>1</td>
<td>C</td>
<td>7.27</td>
<td>42</td>
<td>1.0</td>
<td>63</td>
<td>3.7</td>
</tr>
<tr>
<td>3</td>
<td>D</td>
<td>10.68</td>
<td>42</td>
<td>1.0</td>
<td>66</td>
<td>4.4</td>
</tr>
<tr>
<td>1</td>
<td>E</td>
<td>12.47</td>
<td>0</td>
<td>0</td>
<td>65</td>
<td>3.7</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>18.04</td>
<td>42</td>
<td>1.1</td>
<td>65</td>
<td>4.2</td>
</tr>
<tr>
<td>1</td>
<td>G</td>
<td>19.32</td>
<td>42</td>
<td>1.0</td>
<td>64</td>
<td>3.7</td>
</tr>
<tr>
<td>1</td>
<td>H</td>
<td>28.02</td>
<td>42</td>
<td>1.5</td>
<td>62</td>
<td>4.7</td>
</tr>
<tr>
<td>3</td>
<td>I</td>
<td>29.23</td>
<td>42</td>
<td>1.4</td>
<td>62</td>
<td>4.3</td>
</tr>
<tr>
<td>3</td>
<td>J</td>
<td>32.14</td>
<td>42</td>
<td>1.2</td>
<td>63</td>
<td>4.4</td>
</tr>
<tr>
<td>1</td>
<td>K</td>
<td>34</td>
<td>42</td>
<td>1.1</td>
<td>61</td>
<td>4.1</td>
</tr>
</tbody>
</table>

There are many defects and failure modes of S&C. A squat is one common defect of the S&C. A study conducted by Ilaria G. et al. stated that the most common cause of failure at the crossing panel is a squat on a casting, which comprises approximately a third of all S&C failures. The development and cause of squats are discussed and explained in detail in a study presented by Grassie SL. Squats at an early stage could be difficult to identify. There has been some previous research dedicated to modelling the process of squats’ development. Daniel WJ pointed out that the rate of squats growth measured corresponds to a power-law with a low exponent. Freimanis A. highlighted the approach of peridynamic modelling of squat distribution and the growth. However, these types of models are developed to predict the growth of squats but not to detect them. Another study uses a parametric approach to analyse the axle box acceleration (ABA) signal and visualise the relationship between the signature tone and the train speed. It also states that high-frequency components (e.g. 1000 and 2000 Hz) could occur for light squats. Another study presented a method for squat detection using wavelet analysis of vibration data extracted from ABA. It introduced a new system for automatic squat detection. However, with this approach, a large amount of repetition data is needed to cancel out the random noise. Also, this method is applied generally to all tracks and is not specifically focused on identifying squats at S&C. Another research focused on explaining the relationship between white etching layers (WELs) and the forming of squats and therefore proposes using more regular and relatively shallow grinding to control the development of the squats. However, such an approach could potentially increase the maintenance frequencies and it has not been proven that the total costs would be lower.

This study investigates the possibility of detecting and quantifying squats of different severities by processing the vibration signal measured at the point machine. The relationship between the vibration amplitude and the severity is estimated. By detecting the squats and following the evolution of the squat severity, maintenance actions could be planned and scheduled more efficiently.

Materials and methods

**Track layout and the testbed**

In this study, a method is presented to investigate the possibility of detecting and quantifying squat defects on an S&C. The experiment was performed in a testbed including a full-scale S&C and a bogie wagon. The accelerometer was mounted on the point machine. This setup provides a robust and protective environment for the accelerometer with easy access to electrical power. The vibration signal and the corresponding speed information were measured and saved while the bogie was travelling along the S&C with artificially introduced squats. The main layout of the testbed is shown in Figure 1. Another related study was performed with the same S&C to estimate the wear size using deep learning.
The S&C used in the experiment has a dimension of 1:16 and a length of 38.14 metres located at Luleå University campus. A bogie was used to travel through the S&C to generate the vibration. A simplified illustration of the testbed is shown in Figure 2. The rails are labelled in the figure from rail 1 to rail 4. The squats are labelled from A to K. To simulate two different squat levels, the squats were manually introduced step-wise with 1 mm and 4 mm maximum depth. The actual positions and the measured dimensions of the squats are listed in Table 1. The squat with 1 mm depth is around 42 mm in diameter and the squat with 4 mm depth is about 63 mm in diameter. S0 and S4 are two stop blocks of the S&C on each end in the through direction. Wheel 2 and 4 travel first on rail 2 and then switch to rail 3 while wheel 1 and 3 always travel on rail 1. The point machine is located 5.86 metres away from stop block S0. Figure 3 shows what a real squat and an artificially introduced squat in the testbed look like.

Measurement setup

Accelerometer

A previous study shows that the expected movements of the trains hardly generate frequencies higher than 20 kHz.\textsuperscript{24} Molodova M. et al.\textsuperscript{24} further stated that squats-related frequencies in their system have high energy peaks around 300 Hz and 1060-1160 Hz. Besides, the high-frequency response component could reach up to 2 kHz. Some pre-tests were performed with the testbed, and it showed that most of the impact responses of interests could be captured within the 10 kHz range. However, to be able to catch possible higher frequency, an accelerometer that can measure from 0.3 to 37,000 Hz was installed. Some key technical parameters of the accelerometer used are listed in Table 2.

The accelerometer used in this study was mounted inside the point machine on the rods that were connected to the switch blade. The detailed installation location of the accelerometer can be seen in Figure 2. Here x, y and z coordinates are defined. The mounted accelerometer was used to measure the vibration in z-direction.

Data acquisition

The point machine used in the experiment is electrical mechanical. During the experiment, the vibration signal and the corresponding speed information were measured and logged when the bogie was moving along the rail of the S&C from one end to the other.

A system was built to measure and save the vibration and speed data. The accelerometer measured the acceleration and was then acquired by a conventional data acquisition system (DAQ). An optical sensor was utilised to estimate the wheel speed due to the bogie being manually pushed and the moving speed being non-constant. Both data were later acquired by Arduino Uno Wi-Fi V2. The system was implemented in VI code running in LabVIEW 2019 and saved onto an external hard drive. The sampling frequency of the system is 51.2 kHz.

Test runs

The experiment was performed as follows. Squats were introduced step-wise with two squat levels at 1 mm and 4 mm depth. Tests were also performed before the squats were
created as base line references. The test runs and the recorded data were organised as shown in Table 3.

**Post-processing**

The post-processing procedure of the measured signals is described in Figure 5. The vibration signal was high-pass filtered, wavelet denoised and then synchronised with the smoothed and re-sampled speed data. After synchronising the vibration and speed data, the vibration data was converted from time domain to spatial domain. Then the data is resampled to the same sampling frequency as the original time domain signal. The expected events chart is used to match the final processed vibration signal and identify impact events. The steps are described in detail in the following paragraphs.

**High-pass filtering**

Since the focus of this study was to detect impact events, a third order high-pass filter with cut-off frequency at 100 Hz was applied to the original signal to remove low frequency components such as the variant mean value.

**Wavelet denoising**

To emphasise the impact events in the acceleration signal and to reduce noise, wavelet denoising was performed. Wavelets have been applied as a denoising technology for vibration data ever since it was first introduced by Donoho and Johnstone. The wavelet transform is a very powerful tool for time-frequency analysis and the cornerstone of wavelet denoising. It can be viewed as replacing the short-time Fourier transform’s “time-frequency window” $g_{t,c}$ with a “time-scale window” $\Psi_{a,b}$. The continuous wavelet transform can be defined as follows.

A function $\Psi$ with

$$\int_{\mathbb{R}} \Psi(x) dx = 0$$

is called a wavelet. For every $f$, $\Psi$ defines the continuous wavelet transform

$$W_{\Psi} f(a,b) = \int_{\mathbb{R}} f(x) \Psi(\frac{x - b}{a}) dx$$

for all $a, b \in \mathbb{R} \times \mathbb{R}$

where

$$\Psi_{a,b} = \frac{1}{\sqrt{a}} \Psi\left(\frac{x - b}{a}\right)$$

The function $\Psi$ is called the mother wavelet. It is chosen to be localised at $x = 0$ and at some $\omega = \omega_0 > 0$ (and/or $\omega = -\omega_0$).

The wavelet denoising technology utilises both wavelet transform and reverse wavelet transform. First, a wavelet transform was conducted and the values of the coefficients underneath a certain threshold were considered as noise and was thereby scaled down or set to zero. Then a reverse wavelet transform was applied to reconstruct a denoised version of the original signal. More details of the wavelet denoising technology implementation and application are presented in Wavelet Denoising by Luo and Zhang.

The effect of wavelet denoising for the experiment data can be seen in Figure 4. It shows the acceleration signal in the time domain of one run from $S_1$ to $S_2$ with no manually introduced squats. Figure 4a shows the acceleration signal after applying high-pass filter before wavelet denoising. Figure 4b shows the acceleration signal after wavelet denoising. The dashed line represent the speed data after smoothing. In this case, the peaks seen in the acceleration signal are generated by different impact events during the experiments such as when the wheels interact with the joints, for example. For the original signal, only a few impulse responses with a distinct amplitude were visible. However, the noise level was suppressed in the denoised signal, which enhances the signal to noise ratio and more impact events become visible.

![Figure 4. Comparison of the signal before and after wavelet denoising in no squat case (a) The signal after applying high-pass filter before denoising (b) The signal after denoising Time to spatial domain conversion](image)

The original vibration data was sampled in the time domain. However, what is interesting in the study is where the squats are located. With the help of the logged speed information, it is possible to convert the time domain data into the spatial domain. Since the speed is not constant, the
converted spatial domain signal will have a non-constant distance interval.

**Smoothing speed signal**

Since in this experiment the bogie was pushed manually, the speed variation should be limited and there shouldn’t be any sudden change of speed. However, certain parts of the measured speed data had sudden changes due to measurement error. To remove such effect, convolution technology was used for smoothing the speed data before applying upsampling.

**Resampling**

Resampling was applied to both the speed signal and the vibration signal. Since the speed signal was sampled at 1 Hz and the vibration signal at 51.2 kHz, upsampling was applied to the speed signal. The converted signal in the spatial domain had non-constant sampling frequency due to the speed is none constant. Therefore, resampling to 51.2 kHz was applied using interpolation technology.

**Synchronising**

To be able to perform the spatial domain conversion, the synchronising of the vibration signal and the speed estimation data is necessary. The actual starting time of the vibration signal and the speed signal was not exactly the same due to them being controlled by two different path with the acceleration data transmitted via a cable and the speed data transmitted via WiFi. The signal was aligned by assuming both signals stop at the same time. Further, the signals where the measuring speed is constant zero were removed due to that part of the signal being recorded before the bogie started to move and do not contribute to the analysis.

The technique used to help synchronising the vibration signals in the spatial domain with the expected events is utilising signatory common impulse response as reference points. The event chosen is when the front wheel hits the first rail joint, which happens in all test cases. This point is used to align the vibration signal in the spatial domain to the expected events. With such an approach, the acceleration data is better synchronised with the expected events. Figure 6 shows an example of the acceleration signal after post-processing of pushing the bogie from S_0 to S_1 with 11 squats with 4 mm depth. The corresponding speed curve is also presented with dashed line. With the help of the generated expected events chart, it is apparent that after the matching, the expected events match the acceleration signal impulse responses. This is especially true for the first 7 squats. It is worth pointing out that because of using such a technique, the first impulse response will always be the event of front wheels hitting the first joint gap, which is shown at the distance 0 metres.

**Expected events chart**

In a railway system, the vibrations are generated by the wheel and rail interaction. The condition of the wheel and the rail will directly influence how these vibration signals look. Defects of the wheel or rail will generate larger impulse responses compared to a new system due to the increased dynamic force between the wheel and the rail surface. In the testbed, when a wheel hits squats, joint gaps or stop blocks, the amplitude of the vibration signal should increase. These increased impulse responses are called impact events in this paper. A chart of expected impact events was generated by utilising the physical location information of the squats, the joint gaps or the stop blocks. The expected events when the bogie travels from S_0 to S_1 are illustrated in Figure 6.

Prepared using sagolcite
Markers show different expected events. The star symbol represents wheel 2 hits a squat, the plus symbol represents wheel 4 hits a squat, the circle represents wheel 1 hits a squat, the square represents wheel 3 hits a squat, triangle represents other events.

Amplitude value extracting

The signal was further sliced into segments to analyse the vibration signal near different events such as hitting a gap or encountering a squat. There are many different ways to define the amplitude of a signal. In this study, a maximum peak-to-peak amplitude is used.

Results and discussions

Squat detection

A 0.5 metres’ window on both sides of the expected events chart was utilised to extract the corresponding impulse responses for different squats. Because of the limitation in speed measurement accuracy, not all squats could be detected. The results are listed in Table 4 for the 1 mm case and 3 for the 4 mm case, respectively. The results show that it is possible to extract and locate 4 out of 6 squats with a 1 mm depth that is within a 13.46 metres’ range from the accelerometers. It is also possible to detect and locate 6 out of 7 squats with 4 mm depth within the same range. Squats located more than 22.16 metres away can not be extracted and located correctly with the current implementation for both the 1 mm and 4 mm case. One possible reason is the error of speed measurement accumulates with time, thus the further the squat is located from the starting point, the more uncertainty it has. Another reason is that the amplitude is smaller when the squat is further away from the accelerometer. The results above show that it is possible to implement an automatic squat detection system with the processed vibration signal in the spatial domain within a certain distance.

Table 4. Squat detection for 1 mm case

<table>
<thead>
<tr>
<th>Squat name</th>
<th>Detected</th>
<th>Location error &lt; 0.5m</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>B</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>C</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>D</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>E</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>F</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>G</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>H</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>I</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>J</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>K</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 5. Squat detection for 4 mm case

<table>
<thead>
<tr>
<th>Squat name</th>
<th>Detected</th>
<th>Location error &lt; 0.5m</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>B</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>C</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>D</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>E</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>F</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>G</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>H</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>I</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>J</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>K</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

General trends of the impact events

Figure 7 visualises the changing of impulse responses when the squats size increases. The first impulse response at 0 meter is the event of the front wheel hitting the first joint gap. The first plot shows the reference case with no squats. The second plot is of 1 mm squat depth level case and the third one is 4 mm. Both the amplitude of the impulse response and the number of identifiable impulse responses increase with a larger squat depth. The amplitude decreases and the synchronisation becomes less accurate with increased travelling distance due to the accumulated speed measurement error. The square represents the location of the point machine. The triangles represent events of rails hitting squats, gaps, crossing nose and stops. The subscript f and b represent if it is the front wheel or the back wheel that is involved in the event. In general, the selected events match the acceleration signal impulse responses. This is true especially for the first 7 squats. As a trend, the events further away from the measuring point have smaller impulse response amplitude.

Notably, although the switch is in the ‘through’ status, and rail 1 is not directly attached to the point machine, the amplitudes of these events were still large. Some of them are even larger than the impulse response on rail 2 or 3, as can be seen in Figure 7. One example is squat G and F. The interpretation is that the vibration is transmitted through the bogie. How the wheels are positioned on the track when hitting the squat could be an important factor. However, to be able to draw a concrete conclusion, more data must be collected and analysed.

Normalising peak-to-peak amplitude to 1 m/s case

The maximum peak-to-peak amplitude data were extracted from the processed vibration signal with the help of the expected events chart. The original data of squat F and G are visualised in Figure 8 and Figure 9 respectively. As the speed of the bogie is not constant, the peak-to-peak amplitude values should not be directly compared. To estimate the relationship between the amplitude and the speed for each squat requires a large amount of repetition data at a different speed. In this study, such data are not available. From all the squats introduced, squat F and squat G were chosen to be further analysed. They have a few advantages that are listed below:

- Located in the middle of the S&C, speed of bogie was high.
- Not close to S1 where many events were happening at the same time.
- Not too far away from S1b, so the amplitude value can be detected.
- Close to each other and the speed of the bogie was within 1 ± 0.2 m/s.
Figure 7. Amplitude comparison for different squat levels: (a) No squat case (b) 1 mm depth of squats (c) 4 mm depth of squats

For squat F and squat G, it is reasonable to assume a proportional relationship between the amplitude and the speed within such a small interval. All measured amplitude for squat F and G are, therefore, divided by the corresponding speed to get an estimation of the amplitude at the speed of 1 m/s.

Figure 8. Peak-to-peak amplitude data for squat F on rail 3 (a) no squat case (b) 1 mm squat case (c) 4 mm squat case

Statistics of the estimated amplitude for 1 m/s case

The mean and the standard deviation of the estimated acceleration is plotted against the corresponding speed for no squat, 1 mm squat and 4 mm squat for squat F. The results are visualised in Figure 10. It shows that the mean amplitude value increases from 0.7605 g to 1.2355 g when the squat depth increases from 1 mm to 4 mm. The standard deviations for 1 mm and 4 mm case are 0.3434 g and 0.088 g, respectively. Similar results can be seen in Figure 11. The mean amplitude value increase from 0.3499 g to 1.6286 g when the squat depth increase from 1 mm to 4 mm. The standard deviations for 1 mm and 4 mm case are 0.1437 g and 0.7054 g, respectively. Squat F is located at 12.18 metres away from the point machine and squat G at 13.46 metres.

Figure 9. Peak-to-peak amplitude data for squat G on rail 1 (a) no squat case (b) 1 mm squat case (c) 4 mm squat case

Linear Regression for estimated amplitude versus squat depth

One degree polynomial curve fitting is utilised to model the relationship between the peak-to-peak amplitude to the depth of squat G and F. This fitting use the least square error method. The dashed line in Figure 11 shows the fitted linear model for normalised peak-to-peak amplitude to the depth of squat F:

\[ y = 0.255x + 0.2942 \]  

where y is the normalised peak-to-peak amplitude values and x is the squat depth. The root mean squared error of the fit is 0.2578 g, the coefficient of determination \( R^2 \) is 0.7086 which means that the model predicts 70.86% of the variance in the variable y.
The dashed line in Figure 10 shows the fitted linear model for normalised peak-to-peak amplitude to the depth of squat G:

\[ y = 0.388x + 0.0452 \]  \hspace{1cm} (5)

where \( y \) is the normalised peak-to-peak amplitude value and \( x \) is the squat depth. The root mean squared error of the fit is 0.3231 g, the coefficient of determination \( R^2 \) is 0.7818 which means that the model predicts 78.18% of the variance in the variable \( y \).

Conclusions and future works

This study demonstrates it is possible to use the presented method to detect and quantify squats of different severities by processing the vibration signal measured at the point machine. In detail, the following conclusions are drawn:

- It is possible to extract and locate 4 out of 6 squats with 1 mm depth that is within 13.46 metres range from the accelerometer.
- It is challenging to extract and locate accurately both 1 mm and 4 mm depth squat that is further than 22.16 metres away from the accelerometer.
- The mean normalised amplitude value for squat F increases from 0.7605 g to 1.2355 g when the squat depth increases from 1 mm to 4 mm with standard deviations 0.3434 g and 0.088 g, respectively.
- The mean normalised amplitude value for squat G increases from 0.3499 g to 1.6286 g when the squat depth increases from 1 mm to 4 mm with standard deviations 0.1437 g and 0.7054 g, respectively.
- It is possible to fit a linear model to the normalised amplitude versus squat depth for squats F and G with the data collected.

The conclusions above apply to the low-speed experiment under 2m/s. The findings indicate that it is possible to compose an automatic squats detection algorithm with high accuracy if the squats are as deep as 4 mm and the squats are not more than 13.46 metres away. Since the experiment involves the bogie having a very low speed, the result might vary when the speed is much higher. However, it is reasonable that in a higher speed case, it would be easier to detect the squats on the rail with this approach, due to the increased energy of the impact force. The research method presented in this paper is still valid.

One possible future work is to collect more data from a real railway system to verify the method presented in this study. Another future work could be implementing a machine learning algorithm to learn from the patterns of the healthy S&G and perform continuous anomaly detection on it.

References


Squat Detection of Railway Switches & Crossings Using Wavelets and Isolation Forest

Squat Detection of Railway Switches and Crossings Using Wavelets and Isolation Forest

Yang Zuo *, Florian Thiery, Praneeth Chandran, Johan Odelius and Matti Rantatalo

Division of Operation and Maintenance Engineering, Luleå University of Technology, 97187 Luleå, Sweden
* Correspondence: yang.zuo@ltu.se; Tel.: +46-76-126-3904

Abstract: Railway switches and crossings (S&Cs) are critical, high-value assets in railway networks. A single failure of such an asset could result in severe network disturbance and considerable economical losses. Squats are common rail surface defects of S&Cs and need to be detected and estimated at an early stage to minimise maintenance costs and increase the reliability of S&Cs. For practicality, installation of wired or wireless sensors along the S&C may not be reliable due to the risk of damages of power and signal cables or sensors. To cope with these issues, this study presents a method for collecting and processing vibration data from an accelerometer installed at the point machine to extract features related to the squat defects of the S&C. An unsupervised anomaly-detection method using the isolation forest algorithm is applied to generate anomaly scores from the features. Important features are ranked and selected. This paper describes the procedure of parameter tuning and presents the achieved anomaly scores. The results show that the proposed method is effective and that the generated anomaly scores indicate the health status of an S&C regarding squat defects.

Keywords: railway switch and crossing; vibration; squat; anomaly detection; unsupervised machine learning; anomaly score; point machine

1. Introduction

In recent years, rail transportation has gained significant attention due to its potential to relieve road and air congestion and environmental problems. Railway traffic in Europe has experienced a significant rise in both transporting passengers and freight in Europe [1,2]. In EU-15 countries, the passenger-kilometres and the rail freight ton-kilometres increased 28% and 15%, respectively, between 1990 and 2007 [3]. The increased volumes of freight and passenger traffic are challenges that need to be addressed because they set higher requirements on the maintenance and renewal process. To keep the railway transportation efficient, comfortable and safe under such circumstances, innovative maintenance techniques of the critical components are vital.

Railway switches and crossings (S&Cs) are important components of railway transportation infrastructure. A failure in the S&C could lead to delays globally in the system and considerable economical loss. Since S&Cs include movable parts, and are the discontinuous points of the rail geometry, they encounter high failure rates [4]. Maintaining and renewing the S&Cs across the rail network is expensive [5]. According to Cornish et al. [6], S&Cs have consumed 24% of the maintenance and 23% of the renewal budget against only 5% of the track miles in the U.K. In 2018 alone, S&C cost 530 MSEK, which is around 10% of the entire maintenance budget in Sweden [7]. In the worst case, such a failure could even result in catastrophic accidents due to derailments.

Due to safety concerns and their high maintenance costs, monitoring the status of S&Cs and performing preventive maintenance is important. Many studies have been performed to monitor the status of S&Cs. Most of the studies use wayside mounted systems. Liu et al. [8] experimented with two different systems. One was equipped with a 3D accelerometer and a speed detection sensor to describe crossing degradation, and the
other was a video gauge system (VGS) to detect and quantify ballast conditions. However, the measurements were sensitive to the speed and the type of the train. Data from the same train type and with similar speeds were needed. Boogaard et al. [9] presented a method of utilising both accelerometers and a strain gauge mounted 50 mm below the crossing frog. Only the vibration data of the furthest measuring point from the tip of the nose were presented in the study. The results showed the advantages of combining two different measuring methods for monitoring the crossing nose. However, the proposed approach was focused on measuring the dynamics of the frog in the S&C. Barkhordari et al. [10] proposed a method of employing a wayside system to measure the track acceleration to monitor ballast degradation. However, this method does not provide continuous condition monitoring. Milosevic et al. [11] developed a condition-monitoring approach of railway crossing geometry by using measured and simulated track responses. Kerrouche A. et al. [12] proposed an experimental strain-measurement approach for monitoring the crossing nose of railway S&Cs. However, both of these studies focused on the crossing nose instead of the whole S&C. The Axle Box Acceleration (ABA) system can also be used to monitor the status of the S&C. Wei et al. evaluated the degradation at a railway crossing using ABA measurements [13]. However, the study focused mainly on the uneven deformation between the wing rail and crossing nose and local irregularity in the longitudinal slope of the crossing nose.

Squats are one type of rail defect. According to Grosomini et al. [14], one-third of the recorded failures at the crossing panel are squat-related. Molodova et al. presented a series of studies on utilising ABA to explore the influence of different parameters and to implement an automatic squat-detection method [15,16]. However, these studies are aiming for normal tracks, and the situation for an S&C is more complicated. In addition, the ABA signal is dependent on the property of the axle box, the condition of the wheel axle bearings and the wheel profiles. Cho [17] proposed a similar method for detecting squat defects using the ABA measurement with signal processing and wavelet spectrum analysis. This study has the same drawback as the other ABA-based methods.

As critical components in railway infrastructure, S&Cs are required to be reliable in order to prevent delays and avoid fatal accidents [22]. Nowadays, manual inspection at fixed intervals is still the most commonly used way to assess the status of S&Cs [18]. These manual inspections encounter human errors and can lead to severe accidents. Manual inspection can also place inspectors in danger as regular physical access to the railway is inevitable. A plausible solution to this conundrum can be to automate the process of squat detection and monitor the health status of S&C to obtain more frequent updates of the status information, reduce the cost of inspections, reduce system down-time and increase safety. Anomaly-detection techniques are suitable for finding the segments of S&C that contain squats among the healthy data.

In most of the prior studies using wayside monitoring techniques, sensors were either installed on the side, underneath the rails or mounted on the sleeper to collect the data. These approaches of installing wired or wireless sensors are not practical due to an increased risk of damaged power and signal cables or sensors themselves under normal operation or during maintenance activities. A possible solution to overcome this issue is to make use of the protective environment within the point machines to host the sensors. This study proposed an approach of positioning the accelerometer inside the point machine to estimate the overall health condition of the S&C. The accelerometer was installed on one rod of the point machine with customised aluminum holder. This positioning provides good protection for the accelerometer against harsh weather conditions. An electrical power supply is also easily accessible from the point machine. Previously, the sensors were either installed on the axle box of the train [13,15–17], on the bogie of the train [19–21], directly on the rail [22,23] or underneath the rail [24,25]. This study proposes a new processing procedure which combines classical time-domain features with features derived from scale-averaged wavelet power (SAWP) with the help of wavelet techniques and utilises an unsupervised anomaly-detection algorithm called isolation forest to predict the
anomaly score. This combination has not been yet utilised to process vibration data from the railway application. The previous studies used only time domain features [13,16,26] and supervised machine-learning algorithms [1,26]. The objective of this study is to enable continuous monitoring of the S&C to estimate its general health condition and to reduce the human interventions on track for the inspection purpose. Previous studies focused only on individual defects on normal rails [16,27,28].

The rest of the paper is organised as follows. Section 2 presents the materials and methods. Section 3 presents the results and discussions. Section 4 presents the conclusions and the future works.

2. Materials and Methods

The basic idea behind the current study is that the vibration at the point machine is affected by the squats on the rail head of the S&C. Squats may lead to defects, which can result in system failure during normal railway operations. The vibration is the result of a dynamic response to the wheel–rail interaction. If the rail has squats, then the vibration signal will also change its property. Therefore, analysing the vibrations can be effective in estimating the health status of the S&C.

The experiment for this study was carried out along a testbed including a full-scale S&C and a 6-tonne bogie wagon. Two levels of squats were introduced manually with 1 mm and 4 mm maximum depth. The vibration sensor is mounted at the point machine. Several signal-processing steps were applied to the original signal and 11 features were extracted for each segment of the signal. The features were the root mean square (RMS), standard deviation, shape factor, kurtosis, skewness, peak-to-peak amplitude, impulse factor, crest factor and clearance factor from time domain and the number of peaks and the total peak power from the SAWP. These features were used as input to an unsupervised anomaly-detection algorithm named isolation forest to predict if a section contained squat defects or not. By combining the results of each individual segment, the health condition of the whole S&C could be assessed. A detailed description of the methods used for this study are described in the sub-sections below.

2.1. Track Layout and the Testbed

In this study, an approach was presented to investigate how to detect and evaluate the health status of an S&C regarding squat defects by using unsupervised machine learning. The experiment was performed with a testbed located at Luleå University of Technology including a full-scale S&C and a 6-tonne bogie wagon. This bogie wagon has two axles, and the distance between them is 2.5 m. The S&C used has a dimension of 1.16 and a length of 38.14 m. The accelerometer was mounted on the point machine to provide a protective environment for the accelerometer and easy access to electricity. The vibration signal and the corresponding speed information were measured. The test site is shown in Figure 1 and an illustration of the testbed is shown in Figure 2. The squats were labelled from A to K. S0 and S1 were two stop blocks mounted on the two ends of the rails in the through direction. The point machine is 5.86 m from the stop block S0.

![Figure 1. View of the test site.](image-url)
Figure 2. Test bed schematic diagram and accelerometer placement. The squats are labelled from A to K. S0 and S1 are stop blocks in the through direction.

To simulate two different squat levels, the squats were manually introduced stepwise with 1 mm and 4 mm maximum depths. The dimensions of the squats with two different levels were measured and are presented in Table 1. The sensor used was KS91C. It has a measuring range of 0.3–37,000 Hz, sensitivity was $10 \pm 20\%$ mV/g and the resonant frequency was greater than 60 kHz (+25 dB). The position of the accelerometer is visualised in Figures 2 and 3. The vibration in the $z$-direction was measured. The accelerometer was glued to the aluminum holder which was mounted on one rod of the point machine.

Table 1. Dimension measurements of the two squat levels.

<table>
<thead>
<tr>
<th>Squat Name</th>
<th>Squat Diameter 1 (mm)</th>
<th>Max Depth 1 (mm)</th>
<th>Squat Diameter 2 (mm)</th>
<th>Max Depth 2 (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>43</td>
<td>1.2</td>
<td>62</td>
<td>3.7</td>
</tr>
<tr>
<td>B</td>
<td>41</td>
<td>1.0</td>
<td>61</td>
<td>3.9</td>
</tr>
<tr>
<td>C</td>
<td>42</td>
<td>1.0</td>
<td>63</td>
<td>3.7</td>
</tr>
<tr>
<td>D</td>
<td>42</td>
<td>1.0</td>
<td>66</td>
<td>4.4</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
<td>65</td>
<td>3.7</td>
</tr>
<tr>
<td>F</td>
<td>42</td>
<td>1.1</td>
<td>65</td>
<td>4.2</td>
</tr>
<tr>
<td>G</td>
<td>42</td>
<td>1.0</td>
<td>64</td>
<td>3.7</td>
</tr>
<tr>
<td>H</td>
<td>42</td>
<td>1.5</td>
<td>62</td>
<td>4.7</td>
</tr>
<tr>
<td>I</td>
<td>42</td>
<td>1.4</td>
<td>62</td>
<td>4.3</td>
</tr>
<tr>
<td>J</td>
<td>42</td>
<td>1.2</td>
<td>63</td>
<td>4.4</td>
</tr>
<tr>
<td>K</td>
<td>42</td>
<td>1.1</td>
<td>61</td>
<td>4.1</td>
</tr>
</tbody>
</table>

Figure 3. Sensor mounted on one rod of the point machine for extra protection.
2.2. Test Procedure and Data Acquisition

The experiment was performed as follows. Three different test cases were performed. The bogie wagon travelled from $S_0$ to $S_1$ without squats, with squats of 1 mm depth and with squats of 4 mm depth. Each test case was repeated 3 times. In total, 9 instances were recorded. The vibration data were measured with the accelerometer installed at the point machine. A data acquisition platform DAQ9174 was utilised to capture the vibration data and feed them to the computer directly. The sampling frequency of the platform was 51.2 kHz. The speed was measured with a customised tachometer with Hall effect sensor A3144 and neodymium magnets. An Arduino Uno unit was utilised to send the revolution per minute (RPM) data of the left back wheel via WiFi to the computer. The controlling system was implemented in VI code running in LabVIEW 2019.

2.3. Signal-Processing Procedure

The signal-processing procedure for this study is described in Figure 4. The vibration signals were initially filtered with a third-order Butterworth band-pass filter with 50 Hz and 2.5 kHz cutoff frequencies. The band-pass filter was used to filter away the frequencies with noise and preserve the useful information. A wavelet magnitude scalogram was utilised as a tool to help decide the cutoff frequency of the band-pass filter. The process is explained by using the following example. A piece of vibration data with a squat defect was extracted and evaluated with wavelet transform. Figure 5 presents the wavelet magnitude scalogram of squat G in a 4 mm case. It showed that the main energy of the response for the squat defect was around 200 Hz to 400 Hz. There was also a second frequency band around 500 Hz to 2000 Hz. This implied how the band-pass filter should be designed. The filtered signals were aligned and truncated to equal length. This step makes it possible to compare the results from different runs in the results. It could also be utilised in future studies to accurately extract the position information. Further, the signal was down-sampled to one-tenth of the original frequency. As the band-pass filter has a cutoff frequency as high as 2.5 kHz, the original signal with sampling frequency at 51.2 kHz contains redundant information. A sampling frequency at 5 kHz was enough to preserve all the useful information. To make the calculation easier, 5.12 kHz was applied. The output signals were processed in two separate paths after that. On one path, the signals were directly segmented into 400 equal-sized segments and 9 corresponding time-domain features were extracted. The features used in this study were RMS, standard deviation, shape factor, kurtosis, skewness, peak-to-peak amplitude, impulse factor, crest factor and clearance factor. On the other path, wavelet denoising was applied. The denoising was set at a level 9 decomposition, with Symlet 4 wavelet, Empirical Bayesian denoise method with median thresholding and level-dependent noise estimator. The SAWP was calculated from the output signal. Two features, the number of peaks and the total peak power, were extracted from the SAWP time series and assigned to each segment. In total, 11 features were generated. The extracted features are also described in Table 2.

![Figure 4. Signal-processing diagram.](image-url)
Figure 5. Magnitude scalogram of squat G in a 4 mm case.

Table 2. Extracted features.

<table>
<thead>
<tr>
<th>Feature Number</th>
<th>Feature Type</th>
<th>Feature Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>time domain</td>
<td>RMS</td>
</tr>
<tr>
<td>2</td>
<td>time domain</td>
<td>standard deviation</td>
</tr>
<tr>
<td>3</td>
<td>time domain</td>
<td>shape factor</td>
</tr>
<tr>
<td>4</td>
<td>time domain</td>
<td>kurtosis</td>
</tr>
<tr>
<td>5</td>
<td>time domain</td>
<td>skewness</td>
</tr>
<tr>
<td>6</td>
<td>time domain</td>
<td>peak to peak amplitude</td>
</tr>
<tr>
<td>7</td>
<td>time domain</td>
<td>impulse factor</td>
</tr>
<tr>
<td>8</td>
<td>time domain</td>
<td>crest factor</td>
</tr>
<tr>
<td>9</td>
<td>time domain</td>
<td>clearance factor</td>
</tr>
<tr>
<td>10</td>
<td>SAWP</td>
<td>number of peaks</td>
</tr>
<tr>
<td>11</td>
<td>SAWP</td>
<td>total peak power</td>
</tr>
</tbody>
</table>

2.4. Wavelets

The concept of wavelet transform can be traced back to 1909 when Harr introduced the first wavelet. Wavelet transform can be divided into two categories, namely, continuous wavelet transform (CWT) and discrete wavelet transform (DWT).

CWT is a very powerful tool for time-frequency analysis and can be viewed as replacing the short-time Fourier transform’s “time-frequency window” $g_{t,\omega}$ with a “time-scale window” $\Psi_{a,b}$. However, calculating all wavelet coefficients at all scales is computationally expensive, and it contains a high amount of redundant information. DWT is a good alternative in some cases. DWT works similarly to a band-pass filter and it can be performed for a signal on several levels. Each level decomposes the original signal into approximations (the low-frequency part) and details (the high-frequency part). The next level of DWT is carried out on the approximations of the previous level. Mathematically, the DWT of a function $f(x)$ is defined as the integral transform of $f(x)$ with wavelet functions $\Psi_{a,b}(x)$, when scales and positions are based on powers of two. It is defined as follows:

$$DWT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) * \Psi \left( \frac{x - b}{a} \right) dx$$ (1)
where

\[ a = 2^j, b = k2^j, \left( k, j \in \mathbb{Z}^2 \right) \]  \hspace{1cm} (2)

Here \( a \) is called the scale factor and represents the scaling of the function, and \( b \) is called the shift factor and represents the temporal offset of the function. Wavelet denoising utilises DWT to decompose the original signal to obtain the wavelet coefficients, thresholding the coefficients and reconstructing the signal with reverse DWT [29]. Wavelet denoising has been widely utilised to denoise different vibration signals. Chen et al. proposed a wavelet denoising method for the vibration signals collected from wind turbines [30]. Chegini et al. [31] proposed a new application using imperial wavelet transform denoising in bearing fault diagnosis. He et al. [32] constructed a distributed acoustic sensor technology using multi-level wavelet decomposition denoising for condition monitoring of the heavy-haul railway. More details of the wavelet denoising technology implementation and application can be found in “Wavelet Denoising” by Luo and Zhang [33].

When applying wavelet denoising, a few parameters and the thresholding method needed to be decided. The maximum level of decomposition depended on signal length (N). The data acquired yielded a maximum number of decomposition levels of 21. Levels of coefficients influenced the kurtosis of the signal [34]. Increasing the number of levels of decomposition would lead to more aggressive denoising but also distort the output signal more. Empirical testing with different levels yielded 9. The wavelet function should reflect the features presented in the signal in the time domain. However, since the primary interest in this study was the SAWP time series, the different types of wavelet functions would yield the same qualitative results [35]. Symlet 4 (sym4) was chosen. There were a few methods that could be used to determine the denoising thresholds. Empirical Bayesian, block James-Stein, false discovery rate, minimax estimation, Stein’s unbiased risk estimation and universal threshold were tested. The influence on the SAWP is insignificant [35]. Since the signal without noise was not available, a quantitative comparison could not be performed in this case. Empirical Bayesian with median thresholding was chosen.

2.5. SAWP

The SAWP time series over scales \( s_1 \) to \( s_2 \) is defined as follows [35]:

\[ W_n^2 = \frac{\delta_j \delta_l}{C_j} \sum_{s_j} \left| W_n(s_j) \right|^2 \]  \hspace{1cm} (3)

where

\[ s_j = s_0 2^{j1}, j = 0, 1, \ldots, f \]

\[ f = \delta_j \left( \frac{1}{2} \log_2 \left( \frac{N s_2}{s_1} \right) \right) \]  \hspace{1cm} (4)

\[ C_j \] is scale independent and a constant for the selected wavelet function, \( \delta_j \) is a factor for scale averaging, \( \delta_l \) is the sampling period and \( j_1, \ldots, j_2 \) represent scales over which the SAWP is computed. \( s_0 \) is the smallest resolvable scale and \( f \) determines the largest scale. \( W_n(s) \) is the continuous wavelet transform of a discrete sequence. \( N \) is the number of points in the time series [36].

This can be utilised to examine fluctuations in power over a range of scales, which is exactly what was needed to detect the power burst in the vibration signal when a wheel hits a squat or a gap. This power time series will be utilised later to extract two peak-related features. The threshold for the detection of peaks was set to \( 2.5 \times 10^{-3} \text{ g}^2 \). The threshold was chosen empirically. Figure 6 shows an example of the identified peaks in a 4 mm squat depth case. The corresponding number of peaks in each segment and their total power were calculated.
Figure 6. Scale-averaged wavelet power (SAWP) peaks found for a 4 mm squat case.

Isolation Forest

An isolation forest is an unsupervised anomaly-detection technology based on the idea of isolating anomalies instead of profiling the normal points. Given a set of observations, the isolation forest algorithm selects a random sub-sample of the observations and assigns them to a binary tree. The algorithm starts by selecting a random feature from d-dimensional features. A split is then done on a random threshold in the range of the selected feature. If the value of one observation is less than the selected threshold, it goes to the left branch; otherwise, it goes to the right. With such an approach, a node is split into left and right branches. This process continues recursively until all data points are completely isolated or when the max depth is reached. The above steps are repeated to construct random binary trees until all observations are isolated. Those points that are easier to isolate and with smaller path lengths will thus have higher anomaly scores. A comprehensive description of the isolation forest algorithm is given by Liu F.T. et al. [37].

The 31 features extracted were first scaled using normalisation. The scaled features were evaluated and selected by using both PCA and Laplacian score. After applying the isolation forest algorithm, each segment received an anomaly score. The threshold of anomaly score to separate the healthy data and the anomalies were decided by finding the knee point. After the hyperparameters were decided, two possible indicators were proposed.

3. Results and Discussions
3.1. Segmentation

All signals were aligned to have the same starting points and truncated to 350,000 samples for each signal. The signal was segmented into 20 segments and the anomaly scores achieved cannot pinpoint the precise defect location. To be able to obtain more accurate positioning of anomalies the number of segments was increased to 200 and 400, respectively. In the 400 segment case, since the speed of the bogie never exceeds 2 m/s each segment corresponds to around 0.17 s and will not be more than 0.34 m. This achieves a resolution that can be used in identifying the individual defects. The results of the influence of segment size are shown in Figure 7.
3.2. Feature Extraction

A total of 11 features were extracted from both the processed time-domain signal and the SAWP time series. The features can be grouped into two categories. The RMS, standard deviation, shape factor, kurtosis, skewness, peak to peak amplitude, impulse factor, crest factor and clearance factor are time-domain statistical features. The number of peaks and total peak power are extracted from the SAWP. All the extracted features are summarised in Table 2.

3.3. Feature Scaling

The two most used types of feature-scaling techniques are normalisation and standardisation. Normalisation is also referred to as max-min scaling, and standardisation is also referred to as Z-score normalisation. The normalisation scales the input feature values to the range of [0, 1], while standardisation converts the input feature values to obtain zero mean and a unit standard deviation. Since the PCA algorithm requires input features to have zero mean and a unit standard deviation, the features were standardised.

3.4. Feature Selection

Two different approaches for feature selection were utilised. The first method utilised PCA. The accumulated PCA feature importance score is presented in Figure 8. The first five features in the PCA space captured 96.85% of all the useful information. The second method employed the Laplacian score for feature selection. The redundant features were removed using the cross-correlation values between the features. Usually, the Laplacian score is defined as \( L_f = 1 - s_f \) where \( s_f \) is a score for each feature [38]. However, MATLAB only uses the second term \( s_f \) which represents the feature importance. Therefore, a lower Laplacian score is equivalent to a higher feature importance score, which indicates the corresponding feature is more important. The Laplacian feature importance was calculated and ranked using MATLAB and the results are shown in Figure 9. The correlations between the most significant feature and the others are calculated. The procedures for removing correlated features are as follows. The most important feature was selected and the cross correlation between it and the rest of the features was calculated. The features that had a higher correlation value than 0.9 with the most important feature were removed. This procedure was repeated for the second most important feature in the remaining feature set. This process stopped when there were no two features left that had a cross correlation value higher than 0.9. As a result, the remaining features are features 10, 11, 9, 8, 5 and 6.
Figure 8. Accumulated PCA feature ranking.

Figure 9. Laplacian score-based feature ranking

The results of utilising different groups of features were compared. Figure 10 shows an example of the anomaly scores for a test case with no squat with different feature groups. The anomaly scores using PCA space features and the Laplacian score-selected features are shifted down with 0.5 and 1 correspondingly for better visualisation. The numerical comparison using mean root squared error (MRSE) is presented in Table 3. The anomaly scores generated by using all features and the PCA space features are very similar. This can be explained because the selected 5 PCA space features explain 96.55% of the variance of all features combined. The anomaly scores generated by using Laplacian score-selected features and the PCA space features are also very similar. This shows both feature selection approaches generate similar anomaly scores and are acceptable. However, the anomaly scores generated by using all features and the Laplacian score-selected features are slightly more different with around double MRSE values. The Laplacian score approach only removed five redundant features that are highly correlated with the selected most important features and the anomaly score difference is still small. Plotting and comparing the anomaly scores for those two cases verifies that the difference is so small that it is reasonable to assume similar performance.
Figure 10. Example anomaly score for a no-squat case with different features.

<table>
<thead>
<tr>
<th>Test Run</th>
<th>All Features vs. PCA MRSE</th>
<th>All Features vs. Laplacian MRSE</th>
<th>PCA vs. Laplacian MRSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 mm_1</td>
<td>$8.23 \times 10^{-4}$</td>
<td>$1.90 \times 10^{-3}$</td>
<td>$6.12 \times 10^{-4}$</td>
</tr>
<tr>
<td>0 mm_2</td>
<td>$8.26 \times 10^{-4}$</td>
<td>$1.80 \times 10^{-3}$</td>
<td>$5.13 \times 10^{-4}$</td>
</tr>
<tr>
<td>0 mm_3</td>
<td>$1.00 \times 10^{-3}$</td>
<td>$2.10 \times 10^{-3}$</td>
<td>$5.24 \times 10^{-4}$</td>
</tr>
<tr>
<td>1 mm_1</td>
<td>$7.91 \times 10^{-4}$</td>
<td>$1.80 \times 10^{-3}$</td>
<td>$6.26 \times 10^{-4}$</td>
</tr>
<tr>
<td>1 mm_2</td>
<td>$8.55 \times 10^{-4}$</td>
<td>$2.10 \times 10^{-3}$</td>
<td>$8.40 \times 10^{-4}$</td>
</tr>
<tr>
<td>1 mm_3</td>
<td>$8.27 \times 10^{-4}$</td>
<td>$1.90 \times 10^{-3}$</td>
<td>$7.28 \times 10^{-4}$</td>
</tr>
<tr>
<td>4 mm_1</td>
<td>$7.46 \times 10^{-4}$</td>
<td>$2.00 \times 10^{-3}$</td>
<td>$7.59 \times 10^{-4}$</td>
</tr>
<tr>
<td>4 mm_2</td>
<td>$8.45 \times 10^{-4}$</td>
<td>$2.10 \times 10^{-3}$</td>
<td>$7.91 \times 10^{-4}$</td>
</tr>
<tr>
<td>4 mm_3</td>
<td>$8.23 \times 10^{-4}$</td>
<td>$2.10 \times 10^{-3}$</td>
<td>$7.63 \times 10^{-4}$</td>
</tr>
</tbody>
</table>

From a performance point of view, either group of features could be utilised for further study. However, because of the curse of dimensionality, a higher dimension of features leads to exponentially increased computational efforts [39]. Therefore, it is reasonable to choose either the PCA space features or the Laplacian score-selected features. A drawback with PCA space features is that the generated features are linear combinations of the original features and they become less interpretable and lose their physical meanings [40]. Those two reasons combined justify that it is reasonable to choose Laplacian score-selected features for further study.

3.5. Threshold for Anomaly Score

A threshold should be provided to decide what an anomaly is. The descend-sorted anomaly scores against the index of all 9 instances are plotted in Figure 11. Each instance contains 400 segments and these 9 instances contain 3600 segments in total. The knee point method was applied and it was found that the point with index 436 was the knee. This corresponded to around 12% of the total segments. Therefore, the 88th percentile should be used as the threshold. This is verified by plotting the anomalies using the 88th percentile together with the vibration signal. An example of the results for a 4 mm case is presented in Figure 12. By using the 88th percentile, most of the anomalies were found without introducing unexpected false alarms.
3.6. Anomaly Indicator for the Whole Switch

All the test cases and the corresponding anomaly score above the threshold are presented in Figure 13. This shows clearly that with increased squat depth more anomalies are found, which indicates the health status of the S&C is degraded. It can also be observed that the different test runs were well aligned at the beginning; however, with a different speed profile for each run, the spotted defects also encounter a different drift. They are no longer well aligned after a while. This, however, would not influence the result as utilizing the anomaly score as an indicator of the health status of the S&C.

Figure 13. Anomaly scores above the 88th percentile threshold for all test cases.

One indicator could be calculating the sum of all anomaly scores and using the mean value for each test scenario. From the test data, the observed scores are 11.65, 20.31 and 29.59 for the S&C with healthy, 1 mm deep and 4 mm deep squat cases.

Another indicator could be the mean value of the number of anomalies for each test scenario. From the test data, the average number of anomalies was 18.67, 32.67 and 45.00 for the S&C with healthy, 1 mm deep and 4 mm deep squat cases.
4. Conclusions and Future Works

The present study demonstrates that it is possible to use the proposed method to extract features and utilise unsupervised anomaly-detection techniques, such as the isolation forest to detect the squat defects. The following conclusions can be drawn:

- The study shows that accelerometers placed within the protective environment within a point machine can be utilised for monitoring defects such as squats along the S&Cs of the railway infrastructure.
- The signal-processing procedure of extracting features from both the time domain vibration signal and the SAWP is effective and promising.
- Skewness, peak to peak amplitude, crest factor, clearance factor, Nr. of peaks and total peak power are ranked to be the top features for anomaly detection.
- The selected five PCA space features explain 96.55% of all the variance in the features.
- Anomaly-detection algorithms can be utilised to generate anomaly scores to indicate the health state of the S&C regarding squat defects. Using knee point technique, 12% of the total segments of all nine instances were determined to be anomalies.
- The mean value of the total anomaly scores for each test scenario increase from 11.65 to 20.31 and 29.59 for the S&C with healthy, 1 mm deep, and 4 mm deep squat cases. The values for 1 mm and 4 mm cases are almost 1.7 and 2.5 times greater compared to the healthy case, respectively.
- The mean value of the number of anomalies for each test scenario increases from 18.67, 32.67 and 45.00 for the S&C with healthy, 1 mm deep and 4 mm deep squat cases. The values for 1 mm and 4 mm cases are almost 1.7 and 2.5 times greater compared to the healthy case, respectively.
- An isolation forest algorithm is suitable for anomaly detection related to the squat defects.

Since isolation forest is an unsupervised machine-learning technique, no labelled data are needed to train the model. By learning from the unlabelled data, a model is built and can be utilised to perform anomaly detection on the new data. It is promising to utilise such an approach to enhance the safety and reliability of S&C. One future study would be to verify the approach with data from S&C in a real railway network. Another interesting future study could be to take into consideration such parameters as train type, load and speed among the indicators and extend the method. The future study will also include enhancing the current data set and carrying out a comparative study where the results of the proposed unsupervised anomaly-detection model will be compared to other anomaly-detection methods such as neural networks. In the future, a nationwide condition-monitoring system for S&Cs could be developed by combining such an approach and the concept of federated learning.

**Author Contributions:** Conceptualisation, Y.Z. and M.R.; methodology, Y.Z., M.R., F.T. and P.C.; software, Y.Z. and F.T.; formal analysis, Y.Z.; data acquisition, Y.Z. and M.R.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z., P.C., M.R., F.T. and J.O.; visualisation, Y.Z.; supervision, M.R. and J.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** Great thanks to Jan Lundberg and Taouﬁk Najeh for the design and building of the complete test rig and performing all experimental tests and collecting all data.

**Conflicts of Interest:** The authors declare that there is no conflict of interest.

**Sample Availability:** Samples of the compounds are available from the authors.
References

20. Dumiriu, M. Fault detection of damper in railway vehicle suspension based on the cross-correlation analysis of bogie accelerations. Mechatronics 2019, 20, 102. [CrossRef]


33. Luo, G.; Zhang, D. *Wavelet Denoising*; IntechOpen: London, UK, 2012; Chapter 4. doi: 10.5772/37424. [CrossRef]


Squat Detection in Railway Switches & Crossings Using Point Machine Vibration

Yang Zuo

Railway maintenance with AI