Insulated Rail Joint defect recognition using frequency characteristics of the Axle Box Acceleration measurements







Master thesis

by

Paul Rouwenhorst

4302176

Date

13-08-2021

MSc committee

Chair: Dr. V.L. Markine – TU Delft (Railway Engineering)

Ir. J.M. Hendriks – TU Delft (Railway Engineering)

Dr. N. Bešinović – TU Delft (Transport & Planning)

Ir. D. van Dijk – Arcadis

Ing. M. Folkerts – Arcadis / ASSET Rail

Summary

In this thesis Axle Box Acceleration (ABA) data is used to determine the health condition of Insulated Rail Joints (IRJs) to investigate the potential of ABA data in predictive maintenance. Using frequency characteristics from the Continuous Wavelet Transformation (CWT), the main research question "How can Axle Box Acceleration data be used in automatic classification of specific defect types at Insulated Rail Joints?" is investigated and answered.

IRJs are crucial assets of the railway infrastructure's safety, as they are part of the signalling system. At the same time, IRJs are a weak link in the railway infrastructure because of a rail discontinuity which causes large impact forces and therefore a fast degradation of the asset. Moreover, degradation results in electrical failure and thus unavailability of the railway infrastructure which comes with high penalty costs. Whereas current maintenance strategies rely on reactive and preventive maintenance (time-consuming and cost-inefficient inspections), implementation of predictive maintenance concepts based on data would be beneficial in terms of cost-efficiency, safety and availability of the railway infrastructure. There is a need for a reliable source of data that ABA could provide. ABA is a measuring technique that provides data by means of (vertical) accelerations of the train wheel which are an indication of wheel-rail irregularities. ABA can therefore help in the detection of defects and identification of the condition of IRJs. In this thesis correlations between healthy IRJs and IRJs with battered head failure, lipping failure and loose fastening bolts are investigated as these defects are among the most frequent causes of IRJ failure and remain unexplored in former studies.

With the help of the CWT, the Wavelet Power Spectrum (WPS) is constructed, which is then divided into four main characteristic frequency bands. By investigating three features (total relative power, peak relative power and peak frequency) within each frequency band, frequency characteristics of each type of defect were determined. For each of these three main features, correlations in all frequency bands are investigated and discussed through theory. After that, a principal component analysis (PCA) and support vector machine (SVM) are applied to automatically classify the health condition of IRJs.

Battered head failure contained statistically significant increased powers for three out of four frequency bands, among which frequency band 200-600 Hz contained the largest relative increase. The results match with the theory in which the similarity of battered head failure with permanent deformation resulted in the expectation of similar frequency characteristics around 500-600 Hz. Lipping failure showed a statistically significant increase in frequency band 40-100 Hz which matches with the theoretical expectations. Since there is only a small increased impact force expected, the largest relative changes were expected in the low frequency range ($f<100$ Hz). In addition, there were no statistically significant differences between the battered head failure and lipping failure. It is recommended to compare two frequency bands to distinguish battered head failure from lipping failure. The last defect, loose fastening bolts, was not able to be studied because of limited data.

The found correlations were applied in machine learning algorithms. In this thesis, a PCA was successfully applied to reduce the dimensionality from 12 features to 3 principal components in which more than 85% of the variance within the data was kept which is beneficial for classification purposes. However, the found SVM classifiers for battered head failure and lipping failure performed badly on the given dataset because of the small dataset. Moreover, bad validation tests imply that the found classifiers were unstable and therefore unreliable. More data is needed to perform reliable classification without overfitting the algorithm.

As an answer to the main research question, it is not yet possible (with the results found in this thesis) to automatically classify specific defects at IRJs using frequency characteristics of ABA because this research used limited and inaccurate data. However, the applied methodology shows great potential since frequency bands 40-100 Hz for lipping failure and 200-600 Hz for battered head failure contained statistically significant increased powers. Since this study used limited data and a one year difference between ABA measurements and video footage for the health condition assessment was present, a future study with a large and accurate dataset is recommended to validate the results found in this research.

Content

1 Introduction 1
1.1 Background 1
1.1.1 Predictive maintenance 1
1.1.2 Axle Box Acceleration 1
1.1.3 Insulated Rail Joints 1
1.1.4 Literature review 2
1.2 Research objectives 3
1.2.1 Motivation 3
1.2.2 Research problem 3
1.2.3 Research questions 4
1.3 Research goal 4
1.4 Thesis outline 5
2 Theoretical background information 6
2.1 Insulated Rail Joint (IRJ) 6
2.1.1 IRJ function 6
2.1.2 IRJ degradation 7
2.2 IRJ failure modes 9
2.2.1 Battered head failure (BHF) 9
2.2.2 Lipping failure (LF) 10
2.2.3 Loose fastening bolts (LFB) 11
2.3 Axle Box Acceleration (ABA) 11
2.4 Continuous Wavelet Transform (CWT) 12
2.5 Machine learning 13
2.5.1 Principal Component Analysis (PCA) 13
2.5.2 Support Vector Machine (SVM) 13
3 Theoretical influence defects on ABA 14
4 Methodology 16
4.1 Data selection 17
4.2 IRJ localisation 17
4.3 Continuous Wavelet Transform 18
4.4 IRJ defect assessment 20
4.5 Assumptions and modifications 21
4.6 Correlation 22
4.7 Classification 23

5 Results 25
5.1 Correlation battered head failure 25
5.2 Correlation lipping failure 28
5.3 Correlation battered head failure vs lipping failure 30
5.4 Classification battered head failure 32
5.5 Classification lipping failure 34
6 Conclusion & discussion 37
6.1 Conclusion 37
6.2 Discussion 40
7 Recommendations and future prospects 43
7.1 Recommendations 43
7.2 Future prospects 44
References 45
Appendices 49
Appendix A | Types of IRJs 49
Appendix B | IRJ failure modes and video footage 51
Appendix C | IRJ maintenance 52
Appendix D | Axle Box Acceleration and Camino project 53
Appendix E | Time-frequency transformation 54
Appendix F | Continuous Wavelet Transformation 56
Appendix G | Machine learning 58
Appendix H | Principal Component Analysis 59
Appendix I | Support Vector Machine 60
Appendix J | Model to determine influence of frequency characteristics 61
Appendix K | Software and data selection 62
Appendix L | IRJ localisation 63
Appendix M | CWT and WPS 64
Appendix N | CWT results and eigenfrequencies 66
Appendix O | Vibration modes 69
Appendix P | Defect assessment scores 71
Appendix Q | Clarification assumptions 74
Appendix R | Results correlation BHF 76
Appendix S | Results correlation LF 79
Appendix T | Results correlation BHF vs LF 82
Appendix U| Results PCA BHF 83
Appendix V | Validation results SVM BHF 85

Appendix W | Results PCA LF 86
Appendix X| Test results SVM 88
Appendix Y | Validation results SVM BHF 89
Appendix Z | Peak frequency disadvantage 90

List of abbreviations

ABA	-	Axle Box Acceleration
BHF	-	Battered head failure
CWT	-	Continuous Wavelet Transform
DWT	-	Discrete Wavelet Transform
FT	-	Fourier Transform
IRJ	-	Insulated Rail Joint
LFB	-	Loose fastening bolts
LF	-	Lipping failure
PC	-	Principal Component
PCA	-	Principal Component Analysis
SAWP	-	Scale Average Wavelet Power
STFT	-	Short Time Fourier Transform
SVM	-	Support Vector Machine
WP	-	Wavelet Power
WPS	-	Wavelet Power Spectrum
WT	-	Wavelet Transform

1 Introduction

This report focuses on the automatic identification and classification of the health condition of Insulated Rail Joints. The introduction starts with background information, after which the research objective is described. The research objective includes the problem statement and research questions. Subsequently, the research goal and the relevant steps are described which are needed to answer the research questions. At last, the thesis outline summarizes the content of each chapter.

1.1 Background

This short introductory background chapter explains the concept of predictive maintenance in which the Axle Box Acceleration (ABA) determines the health condition of IRJs. Lastly, previous studies about ABA and the determination of the asset's health conditions are summarized in the literature review subchapter, which includes the remaining literature gap.

1.1.1 Predictive maintenance

From a cost-savings perspective, predictive maintenance has been utilized in the industrial world since the 1990s (Noria, 2020). There has been a change from a reactive mindset to a proactive mindset caused by the rise of new connected technologies. In reactive maintenance, an asset runs to failure and will then need to be replaced at high costs. Preventive maintenance tries to periodically maintain an asset, with the risk of maintaining the asset too early resulting in unnecessary costs. Predictive maintenance aims to optimize maintenance strategies by conducting maintenance only when completely necessary (Shickel, 2020). Data analytics are used to predict the demand and term for maintenance. This principle can also be applied in the rail industry.

In the field of railway infrastructure maintenance, contractors are responsible for the reliability, accessibility, and safety of railway infrastructure. By maintaining the railway infrastructure, the goal is to ensure a safe operation of the railway infrastructure against optimal costs. Predictive maintenance is interesting since it helps optimizing maintenance strategies by predicting failures, rather than correcting or preventing failures. There is a need to constantly inspect and monitor assets to determine their health which can be used for preventing failures.

1.1.2 Axle Box Acceleration

Axle Box Acceleration (ABA) is a potential measurement system to apply in predictive maintenance strategies (Nunez et al., 2014), but is yet not implemented because of unfamiliarity. ABA measures vertical, lateral and longitudinal accelerations by sensors mounted to the axle boxes of a train. Additionally, the speed of the train and its location are measured to link each output to a location. ABA measures vibrations of the wheel and can therefore indicate irregularities at the wheel-rail interface (Molodova et al., 2011). The biggest advantages of ABA (compared to the classical measurement systems such as Eddy Current and Ultrasonic measurements) are its fast measuring speed and low installation and operation costs. Moreover, ABA shows potential in early-stage (invisible) defect detection (Nunez et al., 2014). ABA can therefore be used for predictive maintenance, whereas Eddy Current and Ultrasonic measurements are effective for corrective maintenance because these techniques are most efficient at later stages of degradation (Molodova, Li, et al., 2014).

1.1.3 Insulated Rail Joints

In this master thesis, the application of ABA in the detection of defects located at Insulated Rail Joints (IRJs) is investigated. The IRJ is an asset that must be maintained robustly and adequately since it is a crucial component of the signalling system of the railway infrastructure. There are approximately 50.000 IRJs in the Netherlands (van den Heuvel et al., 2012) which are currently assessed by visual

inspections. This is very time consuming and therefore results in high costs. Besides this, the service life of an IRJ is 30-50% of the rail lifetime if no maintenance would be applied (Dangre, 2019), illustrating that IRJs are susceptible to maintenance.

1.1.4 Literature review

Data analytics in the field of railway engineering has a high potential for researchers all over the world. For example, using ABA many studies have been performed in China, Japan, South Korea, Italy, Poland, and the Netherlands to determine the condition of rails and wheels. Already in 1999 Polish researchers from the Technical University of Gdansk adopted unsprung-mass accelerations to investigate rail surface irregularities (Massel, 1999). Later on, Lee et al. (2012) used accelerometers mounted on the axle box and bogie to estimate lateral and vertical irregularities by a mixed filtering approach. More recently, An et al. (2017) investigated short-wave irregularities that appear due to rail welds in high-speed railways. They found that weld length and depth determine the vibration frequency and amplitude of ABA and described a quantitative relation between ABA and the wheel/rail contact force. Moreover, Song et al. (2020) were able to detect and categorize wheel polygonization by discovering a quantitative relation between the maximum ABA including their frequencies and the wheel polygon defects based on a parameter-lumped wheel-track model.

Adopting ABA to investigate rail surface irregularities has also been done by a group of researchers at the Delft University of Technology. For example, a study has been performed that categorized squats into three health conditions after applying the wavelet transform as the time-frequency tool. Three frequency bands were distinguished related to light, moderate and severe squats and it was possible to automatically detect squats with a hit rate of 78% for light squats and 100% for severe squats (Molodova, Li, et al., 2014). Follow-up studies by Li et al. (2015) improved the hit rate for small rail surface defects up to 85% by using longitudinal ABA and improved noise-reduction techniques.

Altogether, a lot of successful research has been performed in adopting ABA to discover squats, corrugation, weld- and wheel defects. However, only a few studies have been done to investigate defects at Insulated Rail Joints. Even though IRJs are a relatively frequent cause of delays and IRJs contain high maintenance costs, studies adopting ABA at IRJs lack. Molodova and Oregui et al. (2014) compared the ABA response of IRJs in good condition to three cases of poor condition with visible surface degradation, fastening damage and damage to the insulation layer. By investigating the maximum ABA response as well as its frequency response they were able to define certain characteristic frequencies of the investigated IRJ defects for the first time.

Later, in a feasibility study presented by Oregui et al. (2015), a potential to monitor and assess the condition of IRJs was shown by analysing the responses of impact excitations by a hammer or wheel impact. By comparing healthy IRJs and degraded IRJs they found two frequency bands related to the damaged IRJs, regardless of the type of damage. The first frequency range lies in the mid-frequency band between 420 and 600 Hz and shows the potential to be used for damage type assessment. The second frequency band, identified in the high-frequency range, is between 1000 and 1150 Hz. These frequencies were compatible with the pin-pin resonance of the track. A study of a larger set of IRJs is recommended as part of future work, as well as investigating the vibration modes at the two frequency ranges that were discovered. Such information would be useful for the monitoring systems of IRJs and their maintenance measures (Oregui et al., 2015). Additionally, further research resulted in an automatic detection algorithm that was able to correctly detect 84% of the IRJs with a false alarm rate of 19%. By applying the scale averaged wavelet power (SAWP), the local variations of the wavelet power spectrum can be determined. By calculating the SAWP in different frequency bands and introducing an empirical threshold, damaged IRJs could be automatically detected and distinguished from healthy IRJs (Molodova et al., 2016).

Whereas the method used by Molodova et al. (2016) proved to be successful in detecting damaged IRJs, it remains a challenge to classify the type of defects present. Being able to automatically classify IRJ defects is beneficial for subsequent maintenance decisions. In the above-mentioned studies, only three types of defects were investigated: visible permanent deformation, fastening damage and damage to the insulation layer. Furthermore, there is a challenge to distinguish even more types of defects that might occur at IRJs, for example, battered head or lipping IRJs. Especially in case of lipping failure in which the flow of steel might result in railhead contact, there is a risk of signal failure and therefore high penalty costs for a contractor. As the focus of a contactor lies on optimizing profit and therefore minimizing penalty costs, more research in defects resulting in electrical malfunction should be investigated. At last, automatic detection of defects at IRJs is not yet applicable in practice, because of a lack of sufficient validation evidence. Additional elaboration of the health condition assessment algorithm of IRJs is part of future work (Molodova et al., 2016).

1.2 Research objectives

In this subchapter, the research objectives are described by means of a motivation, problem context and associated research questions to solve the described problem statement and reach the goal.

1.2.1 Motivation

IRJs are important components of the railway signalling system but are also frequently cause of mechanical and electrical failure. Therefore, IRJ maintenance is crucial. However, IRJ maintenance is currently labour intensive and time-consuming because of yearly visual inspections. Moreover, inspections must be performed under dangerous circumstances closely to the operating railways. Lastly, current IRJ maintenance is a corrective and preventive measure rather than a predictive measure. Therefore, there is a potential to apply ABA for the determination of the health condition of IRJs.

1.2.2 Research problem

Although ABA shows great potential as a predictive maintenance tool, it is not feasible in practice yet. According to Molodova et al. (2016) elaboration of the health condition assessment algorithm of IRJs is part of future work. Moreover, Oregui et al. (2015) concluded that a larger set of damaged IRJs is needed as an extension of the study about the health condition of IRJs by impact excitations. Besides, battered head failure, lipping failure and loose fastening bolts are common defect types that are not yet being studied. Because time characteristics of ABA do not contain information about the health state of IRJs, a retrospective analysis should be performed to obtain more information (Molodova et al., 2016). Correlations between specific defects and their frequency characteristics need to be established to allow classification. There is a challenge to distinguish different types of defects that might occur at IRJs to create a fully automated system which can be used in comprehensive maintenance procedures. After all, gaining a more precise insight into the condition of an IRJ will be most cost-effective.

In this thesis battered head failure, lipping failure and loose fastening bolts are considered because these failure modes are common in practice and not yet studied. ***Figure 1*** gives an overview of the motivation to apply ABA in predictive maintenance and its entailing problems.

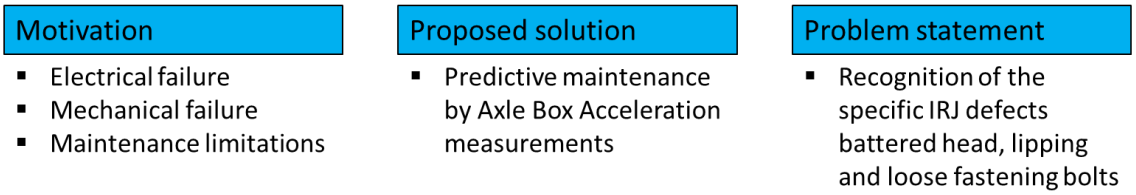


Figure 1 | *Schematic overview of the motivation of this study including the problem statement*

1.2.3 Research questions

Excessive research on the IRJ health condition assessment is needed to enlarge the feasibility of ABA in practice. The presence of correlations should be established, to allow for automatic classification. In this thesis battered head, lipping failure and loose fastening bolts are investigated because these defect types are common in practice and have not yet been studied. The following main research question is specified for this research:

"How can Axle Box Acceleration data be used in automatic classification of specific defect types at Insulated Rail Joints?"

Based on the main research objective, five sub-questions have been defined that need to be answered to reach the goal of this study.

1. **What are battered head failure, lipping failure and loose fastening bolts and what causes these defects?**
2. **What is the theoretical influence of battered head failure, lipping failure and loose fastening bolts on the dynamic behaviour of a passing train?**
3. **What processing techniques are required to study these theoretical effects from ABA data?**
4. **What characteristic frequency responses can be found in the data that are related to battered head failure, lipping failure and loose fastening bolts and are these in line with the theoretical influence?**
5. **How, and with what accuracy, can the obtained relations be used in classification of IRJs?**

1.3 Research goal

The goal of this thesis is to investigate the potential of ABA data for automatic classification of specific defect types at IRJs. If characteristic frequency responses are found, so called signature tunes, they can be applied in automatic classification algorithms to determine the health condition of IRJs. For a contractor, such information is useful to apply in predictive maintenance concepts. To reach this goal, the above-defined research questions need to be answered by means of literature and the available data.

The first two research questions will be answered from the literature. The third research question is related to the methodology, in which a time-frequency transform is applied to investigate the theoretical effects of the acceleration signal. As a result, correlations in the frequency response of ABA is investigated to answer the fourth research question. The identified frequency characteristic will then be used for classification by means of machine learning algorithms which answers the fifth research question. This thesis aims to find correlations in the frequency response of three types of IRJ defects to aid in automatic classification. ***Figure 2*** visualizes the goal of this master thesis.

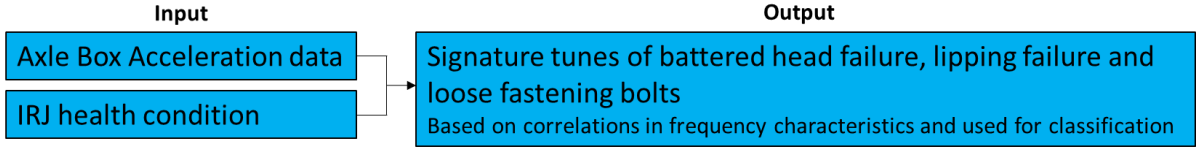


Figure 2 | *Goal of study by using ABA and IRJ health condition as an input to determine the signature tunes of battered head, lipping failure and loose fastening bolts based on correlations in frequency characteristics and which are used for classification*

1.4 Thesis outline

The structure of this report follows the steps described in the research goal which are in line with the specified research questions.

Chapter 2 describes the necessary theoretical background information about the IRJ and its behaviour which is needed to answer the first research question. Moreover, information on ABA is given and the Continuous Wavelet Transform (CWT) is explained which is a necessary step to answer the subsequent research questions. More extensive background information is written down in Appendices A-I.

Chapter 3 explain the influence of the defects on the dynamical behaviour of ABA and answers therefore the second research question. The theoretical effects are compared to the results and discussed in chapter 6.

Chapter 4 describes the methodology used to process the ABA signal into frequency characteristics which can be used for correlation and classification. It answers the third research question. Furthermore, results of the pre-processing steps are given which influences the final results of correlation and classification. Additionally, unavoidable assumptions and modifications were needed and explained in this chapter.

Chapter 5 gives the results of the correlation and classification assessment. First, the results of the correlation assessment are given for all three defect types separately. Thereafter, the results of the classification assessment are given for all three defect types. The results are extensively described and shortly concluded and discussed to answer the research questions 4 and 5. The correlations are described through theory. Classification has been applied by means of a Principal Component Analysis (PCA) and Support Vector Machine (SVM) to answer research question 5.

Chapter 6 summarizes the individual conclusion and discussion points of all research questions. Moreover, a final conclusion and discussion are given and an answer to the main research question of how ABA can be used to automatically classify certain defect types at IRJs is provided.

Chapter 7 explains the limitations and improvements of this study by means of recommendations. At last, a future prospect is described in which the subsequent steps towards a practical application is elaborated.

2 Theoretical background information

In this chapter theoretical background information is given about the IRJ and the three defect types which are investigated in this thesis. An answer to the first research question will be given. Moreover, information about ABA and its dynamical behaviour is given to answer the second research question. Furthermore, information about the Continuous Wavelet Transform (CWT) is provided which is part of important intermediate steps towards the frequency characteristics of the health condition of IRJs. At last, information about the applied machine learning techniques is given which will be used to answer the final research question about classification.

2.1 Insulated Rail Joint (IRJ)

Insulated Rail Joints (IRJs) are important assets of railway's infrastructure safety. In this subchapter, the function and degradation of IRJs are explained.

2.1.1 IRJ function

Insulated Rail Joints (IRJs) are essential components of the signalling system of the railway infrastructure. Railways are divided into block segments which are separated by IRJs. The insulation material of an IRJ separates an electrical current from other block segments. If there is no train on the segment, the electrical circuit is intact, indicating a free block segment resulting in a green light, see ***Figure 3a***. However, while a train is running over the segment, the current is disconnected turning the traffic light red, indicating an occupied block segment, see ***Figure 3b***.

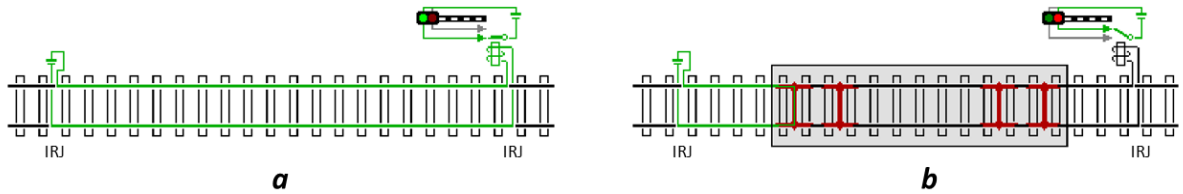


Figure 3 | *Depiction of the signalling system in a* ***(a)*** *clear track circuit and* ***(b)*** *occupied track circuit (Ruijters et al., 2016). The IRJs divide the track into a block segment and close an electrical circuit: the track is clear. The electrical circuit is disrupted by the wheelset of a train resulting in a red light: the block segment is occupied.*

IRJs are at the basis of the automatic train security (Dutch: Automatische Treinbeïnvloeding (ATB)) having two functions: keeping different track segments electrically separated while keeping the track mechanically together. ***Figure 4*** shows the elements of a glued IRJ in a side view and cross-section. It consists of two rails connected by a reinforced fishplate which is both glued and bolted together. The plastic glue is located in between the rails at the end-post as well as between the rail web and the fishplate. The rail foot is connected to the sleeper by a fastening system. The cables are connected to a battery or relay to create the current circuit. ***Appendix A*** explains two types of IRJs, glued and construction IRJs, as well as the two most common glued IRJs in practice, the ETS joint and NRG joint. Because of different joint components and therefore different joint stiffnesses, it is expected that the type of joint influences the frequency behaviour of ABA.

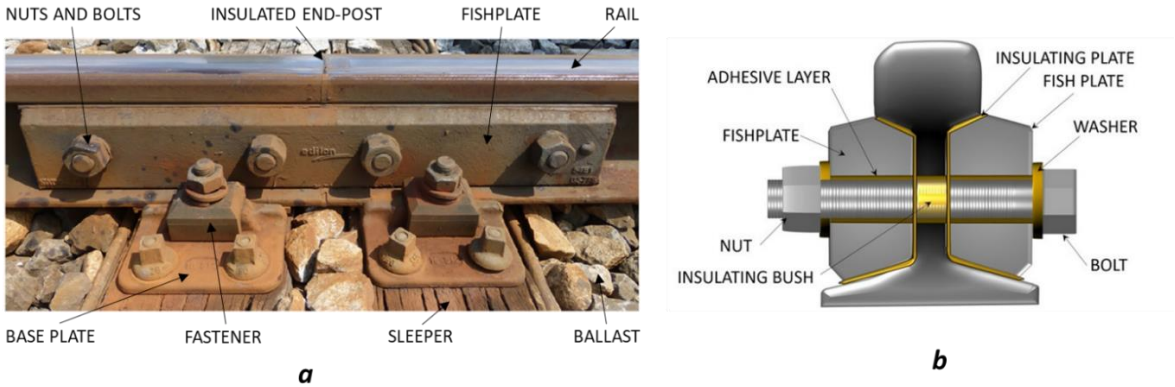


Figure 4 | (a) *Side view (Oregui et al., 2015) and* ***(b)*** *cross-section of the Insulated Rail Joint* (Indian Railways Institute Of Civil Engineering, n.d.-b).

2.1.2 IRJ degradation

IRJs are indispensable components of the railway's safety. However, at the same time, the joints are weak links of the railway infrastructure being a relatively frequent cause of failures as a result. At the joint, there is a discontinuity of track stiffness because of the insulating material. The elastic modulus of the insulating plastic (1590 MPa in case of the commonly used Nylon 66) could be over 100 times smaller than the elastic modulus of steel (210 GPa) (Mandal & Peach, 2010). The lower the elastic modulus, the lower the bending stiffness of the material. At the insulation gap, transient two-point contact occurs, indicating a discrete change in vertical momentum (Steenbergen, 2008). Such discontinuity results in very large impact forces which are detrimental to both the track and the wheels of the train.

Jenkins et al. (1974) investigated the wheel/rail vertical dynamics behaviour on rail joints. Characteristically for IRJs is the high short peak force P1 followed by a relatively longer P2 force as is shown in ***Figure 5***. The P1 force damages the local contact area of both the wheel and rail, whereas the P2 force damages mostly train and track components around the IRJ. An empirical relationship was described for the estimation of the impact forces P1 and P2 (Jenkins et al., 1974).

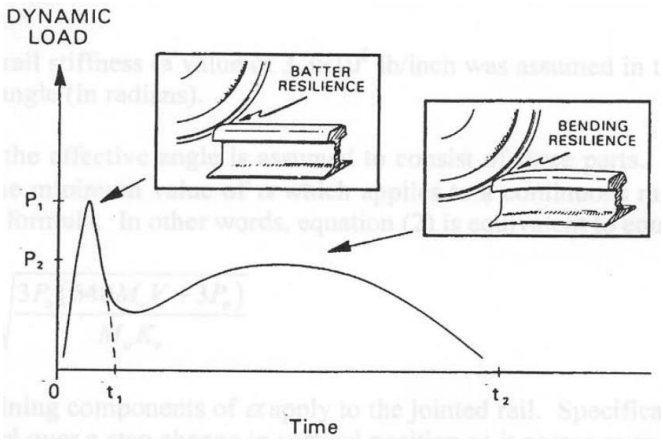


Figure 5 | *Vertical dynamic behaviour at the wheel-rail contact of an IRJ (Jenkins et al., 1974). P1 indicates the short local peak force, followed by a longer but lower P2 force.*

When analysing short-wave irregularities, which is the case at IRJs, two types of rolling contact need to be distinguished. A circular wheel rolling over a flat track has a continuous single point contact (given the fact that the wheel is perfectly circular and there are no irregularities of the track). Since there is no change in the direction of the wheel inertia, there is no dynamic amplification indicating there is no additional dynamic force, except for its static gravitational wheel-load.
However, at an IRJ there is a small gap between the rails filled up with a plastic material having a lower Young's Modulus. Such irregularity results in an increase in the dynamic amplification factor. ***Figure 6*** shows two possible concepts of two-point contact by considering the trajectory of the kinematic wheel gravity centre (red dashed line). Due to a track irregularity, the kinematic wheel gravity centre abruptly changes direction indicating transient two-point contact.

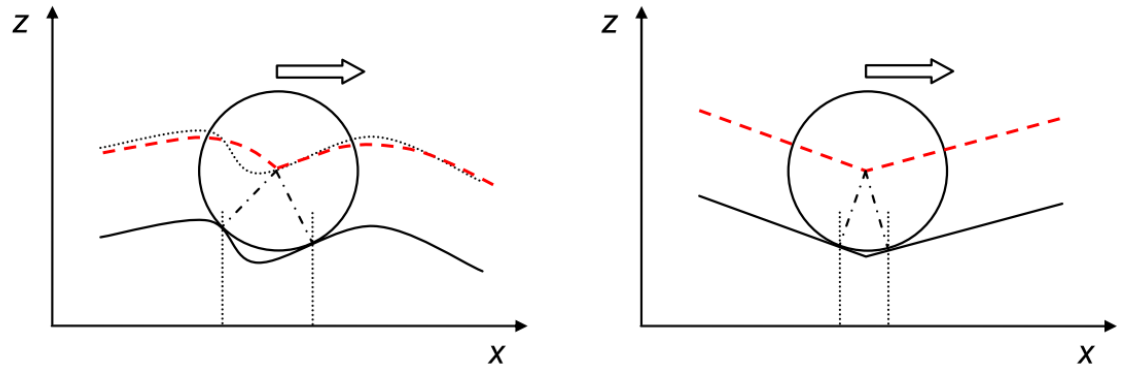


Figure 6 | *Two examples of transient two-point contact, in which an abrupt change in wheel trajectory (red dashed line) causes large impact forces (Steenbergen, 2008).*

Transient two-point contact needs to be avoided because of both large impact forces as well as high vibrations levels resulting in a high noise level. The large impact forces rise from the discrete change in vertical momentum p(t) which is related to the change of vertical velocity (v) times the constant wheel mass (m), see ***Figure 7***.

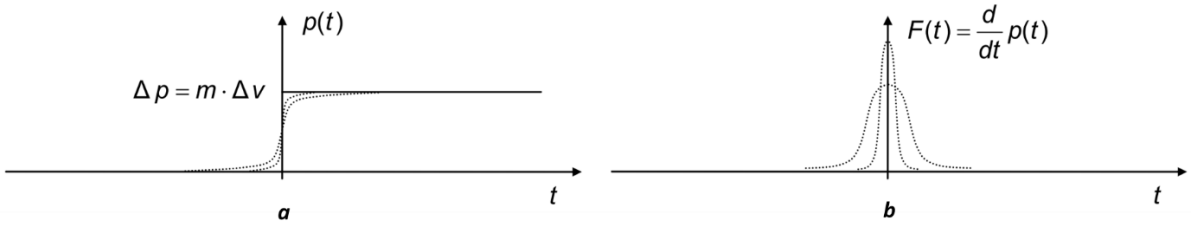


Figure 7 | (a) *Discrete change in vertical momentum p(t)* ***(b)*** *resulting in a high dynamic contact force F(t) (Steenbergen, 2008).*

2.2 IRJ failure modes

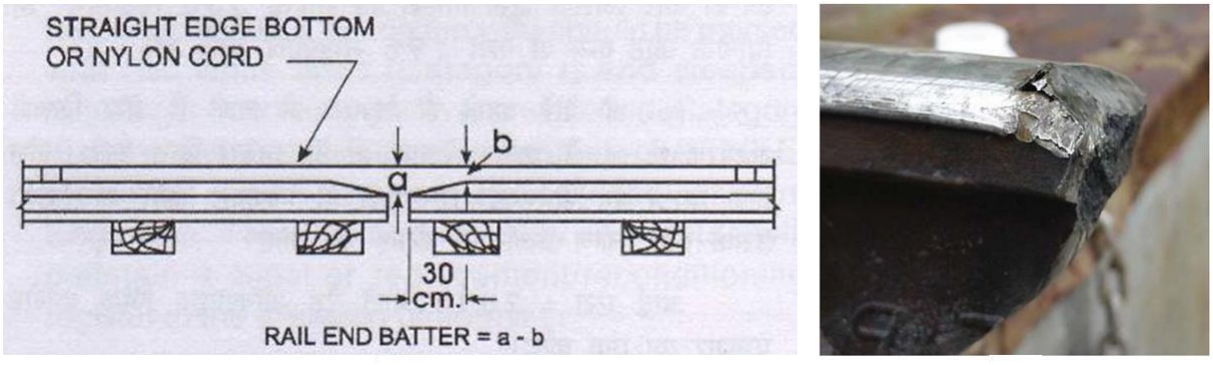
Due to a change in track stiffness, large impact forces occur at IRJs causing electrical and mechanical failure. As mechanical failure occurs, the joint loses its strength and will not be able to provide a physical connection anymore. In case of electrical failure, which is mostly a result of mechanical failure, the electrical track circuit can no longer be insulated causing signal failure. In terms of safety, mechanical failures have potentially catastrophic consequences as trains might derail but are rather uncommon. Electrical failure on the other hand is more common but has minor consequences in terms of safety (Ruijters et al., 2016). ***Table 1*** summarizes some important failure modes which are further clarified in ***Appendix B***. This thesis investigates battered head failure (BHF), lipping failure (LF) and loose fastening bolts (LFB) because these failure modes are common in practice and not yet studied. The subsequent subchapters 2.2.1, 2.2.2 and 2.2.3 describe the failure modes extensively. The current IRJ maintenance procedure is described in ***Appendix C*** which helps understand the reasoning of predictive IRJ maintenance.

Table 1 | *Failure modes IRJ (Ruijters et al., 2016 & Dangre, 2019).*

Mechanical failure mode	Electrical failure mode
Broken fishplate or bolts	Lipping: flow of steel
Battered head and arc damage	Crushed end post
Permanent deformation	Sleeper shift
Insulation failure (cracks, peeling away)	Internal insulation failure

2.2.1 Battered head failure (BHF)

Battered head (in Dutch: “verslagen kop”) is the downward flow of steel as a result of impact forces (Dangre, 2019). As is explained by Jenkins et al. (1974), high impact forces (P1 and P2) occur at the wheel/rail contact of an IRJ. Especially the P1 force results in battered heads. As there is a change in joint stiffness and a lack of support at the insulating end post, the impact force causes the rail ends to flow downwards and become flattened (Dangre, 2019). Additionally, the fishplates are not able to keep the rails together because of the lack of support. Battered head failure is measured by comparing the difference in height at the location of the battered end post and the height of the rail 300 mm away as can be seen in ***Figure 8***. Rail end battering is acceptable up to 2 mm (Dangre, 2019).



a ***b***

Figure 8 | (a) *Schematic overview* (Indian Railways Institute Of Civil Engineering, n.d.-a) *and* ***(b)*** *depiction of rail end batter* (Nordco, n.d.).

2.2.2 Lipping failure (LF)

Lipping failure (in Dutch: "overkraging") is the flow of steel over the end post due to fatigue. The IRJ is repeatedly loaded by passing wheels, resulting in a localised elastic-plastic deformation, which is also called ratcheting failure. The wheel-rail contact pressure is normally defined by the Hertzian theory, in which a wheel load causes an elliptical contact pressure with its maximum stress located in the middle (Johnson, 1985). However, wheel-rail contact at the IRJ results in a contact singularity which means theoretically spoken an 'infinite pressure peak' (Dhanasekar, 2015). In this case, the contact pressure is not elliptical anymore and its maximum stress pressure is not in the middle, but at the edges, see ***Figure 9***.

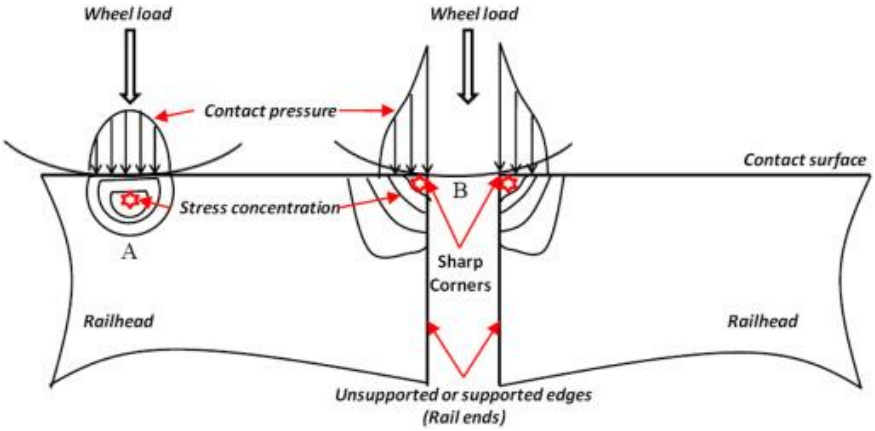
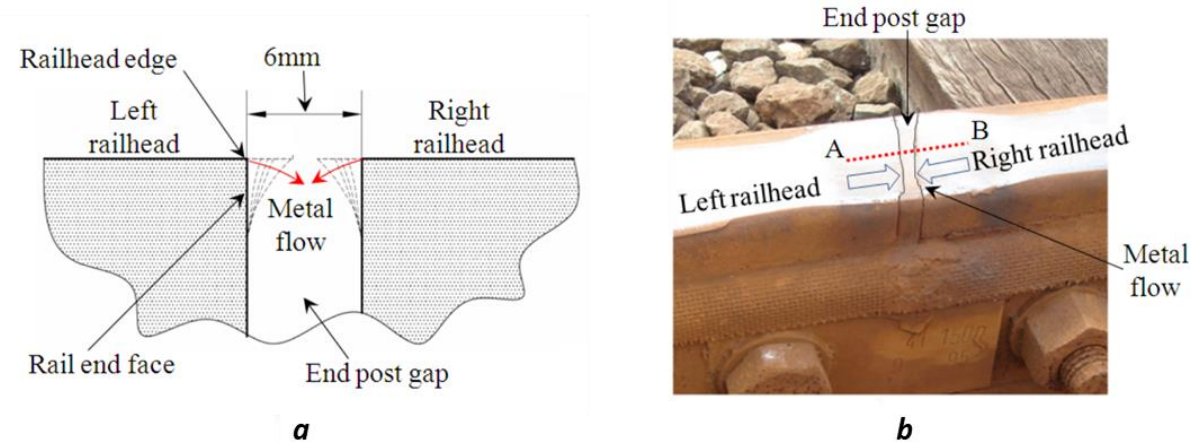


Figure 9 | *Contact pressure singularity causing lipping failure (Zong, 2013).*

Due to the high-stress concentration at the edges of the railhead combined with the repetitive wheel loads, the steel tends to flow inwards over the insulation end post (see ***Figure 10***). Especially during the summer months, additional thermal stresses cause the rails to expand thereby decreasing the gap between the railheads. Thermal expansion in combination with impact forces is detrimental to the IRJ by resulting in contact between the railheads leading to electrical failure (AMEC Rail et al., 2001).



a ***b***

Figure 10 | (a) *Schematic overview and* ***(b)*** *depiction of lipping failure, in which the steel flow might result in electrical failure (Dhanasekar, 2015).*

2.2.3 Loose fastening bolts (LFB)

Loose fastening bolts are at the basis of further IRJ degradation. The bolts keep the rail connected to the base plate and sleeper. A loss of connection drastically increases the IRJ degradation as additional high vibrations occur due to an increased dynamic force. Loose bolts initially start due to vibrations of the train. Subsequently, water can penetrate the empty spacing starting the corrosion process which accelerates the loosening of bolts. Besides, corrosion in combination with vibrations increases the development of fatigue cracks (Lachowicz & Lachowicz, 2021). ***Figure 11*** gives an example of a loose fastening bolt.



Figure 11 | *Loose fastening bolt* (Oregui et al., 2015).

2.3 Axle Box Acceleration (ABA)

Axle Box Acceleration (ABA) is a measurement system that measures vertical, lateral and longitudinal accelerations over time. The accelerometers are mounted to both ends of the two axles of a bogie as is visualized in ***Figure 12***. Next to the accelerometers, the position and the velocity of the train are recorded using GPS. In this thesis data from the Camino project has been used.

In the Camino project, accelerometers were mounted to the axle box of a SLT type (sprinter) and VIRM train, to measure the real-time balance of the train, as well as the switches and the quality of the rails themselves. The trains should operate on the normal daily schedule to ensure a continuous flow of data (ProRail, 2018). However, since the Camino project is a pilot project, only a small test dataset was shared and available. The accelerations are measured with a sampling frequency (f_s) of 6250 Hz, resulting in the possibility to investigate frequencies up to 3125 Hz according to the Nyquist frequency ($f_N = f_s/2$). The sampling frequency of the GPS is only 1 Hz, which could result in inaccurate results with deviations up to 25 meters (ProRail, 2020). Elaborated information of ABA and the Camino project is added to ***Appendix D***. In here, the shared data is described which was important for the pre-processing phase.

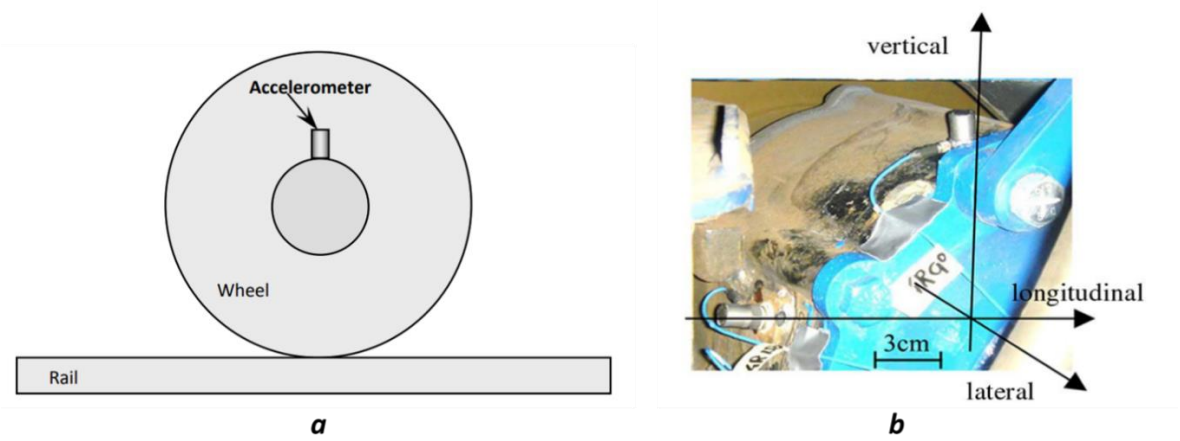


Figure 12 | (a) *Schematic side view and* ***(b)*** *real accelerometer mounted to the axle box (Molodova, Li, et al., 2014).*

2.4 Continuous Wavelet Transform (CWT)

To study the influence of battered head failure, lipping failure and loose fastening bolts on the ABA response, it is needed to perform a time-frequency transformation. According to Molodova et al. (2016), time characteristics of ABA do not contain information about the health state of IRJs and therefore a retrospective analysis should be performed to obtain more information. ***Appendix E*** explains the principle of time-frequency transformations and the advantages and disadvantages of multiple transforms.

In this thesis, a Continuous Wavelet Transform (CWT) is applied as the time-frequency transform because for IRJ localisation time information is just as important as frequency information. Since this method is not dependent on a fixed window size, high frequencies can be studied with a high time resolution and low frequencies with a high-frequency resolution. Characteristic for the CWT is the wavelet which is a "small wave-like oscillation that is localized in time" (Talebi, 2020). The localization of the wavelet in time allows the wavelet transform to obtain time information, whereas sines or cosines as used in FT are not localized in time and therefore result in loss of time resolution. There are several types of wavelets, each with its strength and weakness based on the trade-off between how compact and smooth a wavelet is and should be for its purpose. The CWT is described as follows (Nicoll, 2020):

$$W(\tau,s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} f(t)\psi^*\left(\frac{t-\tau}{s}\right) dt$$

In the CWT the integral is taken of the input signal f(t) which is multiplied by the complex conjugate (*) of function ψ(t), which is known as the mother wavelet. The wavelet consists of a family of wavelets deduced from the mother wavelet by a scaling (s) and shifting (τ) factor, hence $\frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right)$. These are also called child wavelets. A more detailed explanation of the CWT is given in ***Appendix F***.

2.5 Machine learning

One of the goals of this thesis is to classify rail defects based on their frequency characteristics and for that, a machine learning algorithm is used. Machine learning is a method within data analysis that automates analytical model building (SAS, 2021). Computers can learn from data, recognize patterns and make decisions automatically based on an algorithm, which is used for various purposes (i.e. facial recognition or decision making).

Below, the machine learning algorithms used in this thesis are explained. First, a Principal Component Analysis (PCA) is applied to reduce the dimensionality of the dataset which is beneficial for the selection of features used for classification (Pechenizkiy et al., 2004). Thereafter, the Support Vector Machine (SVM) is applied because of the benefit of applicability of higher dimensional spaces which is explained below. More information on machine learning algorithms is given in ***Appendix G***.

2.5.1 Principal Component Analysis (PCA)

PCA calculates how all variables relate to one another by determining two components: the direction and magnitude of the variation in the data. The direction with the most variation (highest magnitude) equals the first principal component. Then, a second principal component is constructed orthogonal to the first one, which includes the second most variation. Principle components are always perpendicular to one another to ensure they are statistically linearly independent of one another (Brems, 2017). By projecting data into a smaller space, the dimensionality of the features can be reduced. By transforming the data into different principal directions, a large part of the information in the data can be explained by means of principal components. The PCA is beneficial before application of the SVM because more information of multiple variables is considered which decreases the risk of overfitting the algorithm. An extensive explanation of the principle of PCA is given in ***Appendix H***.

2.5.2 Support Vector Machine (SVM)

Data points with differences in features can be classified by means of a machine learning algorithm. Classification is the process of predicting the class (or target) of data points (Asiri, 2018). An algorithm uses training data to understand the data and search for relations between input variables and their class. For this study, the health of the IRJ (class) is predicted using frequency information of the ABA signal at IRJs. More specifically, battered head failure and lipping failure are classified by means of a machine learning algorithm. For the sake and simplicity of this study, a linear classifier has been applied: the Support Vector Machine (SVM). Moreover, the SVM has the advantage to be effective in high dimensional spaces and it also performs well for a smaller dataset (Auria & Moro, 2008), compared to for example logistic regression.

The SVM algorithm searches for a hyperplane (decision line) in a N-dimensional space (N accounts for the number of features) that classifies the data points. Given a dataset with two classes (healthy and damaged) multiple hyperplanes (decision lines) can be drawn. The objective of SVM is to find a hyperplane that maximizes the distance between the data points of the two classes. By maximizing the distance, there is a built-in margin for future data points which provides more confident classification. An elaborated explanation including an example figure is given in ***Appendix I***.

3 Theoretical influence defects on ABA

ABA measures the vertical, lateral and longitudinal accelerations over time. The measured vibrations are an indication of wheel-rail irregularities. In this chapter, an explanation is given of the theoretical influence of the defects on the measured accelerations according to literature which answers the second research question. First, the general effects are summarized after which the expectations of battered head failure, lipping failure and loose fastening bolts are described.

General theoretical influences

The gap at an IRJ causes a discrete change in vertical momentum resulting in high vertical accelerations and forces. Within the time domain, such accelerations are clearly visible. According to Newton's second law ($F = m \cdot a$) force is linearly related to acceleration. So, the impact forces which occur at IRJs will be visible in ABA as high positive and negative accelerations. Moreover, high accelerations result in high vibration amplitudes. These amplitudes contain high powers in both time and frequency characteristics because power is proportional to the square of the vibration amplitude (Moebs et al., 2016).

A perfectly smooth rail without any interrupting assets or defects would show no or little power in each frequency (Steenbergen, 2008). Only train and track related eigenfrequencies will be expected to contain relatively low powers. However, due to the impact force at IRJs, vibrations with specific frequencies will occur which are typical for IRJs. However, if the IRJ is degraded, the track properties will change and so will the frequency response of ABA. These properties are investigated a lot by several researchers whose most important findings are summarized below.

Regarding the theory, frequency (f) is related to the velocity (v) and wavelength (L) by $f = v / L$ for low frequencies ($f < 100$ Hz), which is used by Molodova et al. (2016). The vibration of the wheel follows the rail with a relatively large wavelength and therefore the wave velocity is equal to the train velocity. A change in frequency could therefore be caused by a change in train velocity or a change in wavelength.

Moreover, Oregui et al. (2015) investigated the influence of the structure of the IRJ on the measured dynamics. Dynamic behaviour of the track at frequencies lower than 500 Hz are caused by the track components subgrade, ballast and sleeper, which all determine the track stiffness. They also discovered that sleeper distance, rail and railpad are track components that mostly determine the dynamic response of the track at high frequencies from 900 to 1200 Hz.

In fact, the frequencies are related to the track stiffness determined by its components. Track stiffness (k) was found to be dependent on the elastic modulus (E), the moment of inertia (I) and the length of the beam (l). Moreover, the mass of the system (both mass of the rail and train) influences the frequency behaviour of the track. ***Table 2*** shows the relative change in eigenfrequency due to a change of one of the properties which could be helpful to discover an increasing or decreasing trend in frequency. It can be seen that a decrease in mass and length increases the frequency. Moreover, the effect of decreasing the cross-sectional area and density does not affect the frequency. A decrease in inertia results in a lower frequency. Given these differences, the results from the theoretical equations can be compared and discussed to the results of the change in peak frequency due to battered head failure, lipping failure or loose fastening bolts. A detailed explanation of the formulas and relations is given in ***Appendix J***.

Table 2 | *Effect of halving the system properties on system frequency.*

Properties	Frequency
Baseline	--
m / 2	↑
/ 2	↓
A / 2	--
/ 2	↑
ρ / 2	--

Insulated Rail Joint related theoretical influences

Next to the general frequency behaviour of track components, research has been done for degraded IRJs. The most important findings are summarized next. Molodova and Oregui et al. (2014) found frequency characteristics related to permanent deformation at IRJs to be at 540 Hz. Besides that, Molodova et al. (2016) found that a change in frequency around 70 Hz was probably related to the change in stiffness of the IRJ as a result of a crack in the fastening system. Moreover, Oregui et al. (2015) found similar results of frequency characteristics related to plastic deformation. However, fastening damage was found at a completely different frequency band, around 566 Hz and 1040 Hz. In addition, damaged insulation layers were also typically found at high frequencies of 1040 Hz.

Theoretical influence battered head failure

Regarding the impact force, it is expected that battered head failure causes increased impact forces and therefore increased powers because the gap at the insulated end post enlarges due to battered head failure and increase the vertical momentum. Regarding the stiffness of the joint, it is expected that battered head failure influences both low and high-frequency ranges. Low frequencies are related to a changing track stiffness and high frequencies could be caused by the increased impact force and deviating rail properties. Since battered head failure coexists with joint settlement, it can be imagined that the ballast and sleeper properties change (e.g. decrease in stiffness). Besides that, changing rail properties (e.g. decrease in inertia) decreases the frequency according to ***Table 2***. At last, since battered head failure is somewhat similar to permanent deformation (but at a different location), it could be possible to find characteristic frequencies for battered head failure around 500-600 Hz.

Theoretical influence lipping failure

In case of lipping failure, there is hardly a change in vertical momentum because the steel flows over the insulated end post and therefore only a small increase in impact force. Thus, no or little increased powers are expected for lipping failure. Besides that, since there is only a small increased impact force expected, the largest relative changes are expected in the low-frequency range (f<100 Hz). However, the flow of steel might influence the properties of the insulation layer and thus high-frequency changes could also be expected at 1040 Hz (similarly to a damaged insulation layer).

Theoretical influence loose fastening bolts

For loose fastening bolts, there is no increased impact force expected because the fastening bolts do not influence the vertical momentum of the IRJ much. Therefore, no or little increases in powers are expected for LFB compared to healthy IRJs. Loose fastening bolts however decreases the joint stiffness. There is less fixation by the fastening system and therefore more room for IRJ movement. This behaviour is expected to see at low frequencies (f<100 Hz).

4 Methodology

The methodology is explained following the steps described in ***Figure 13,*** which also answers the third research question "What processing techniques are required to study these theoretical effects from ABA data?". First, the required data is extracted, after which the IRJs are localized within the data. Subsequently, the Wavelet Transform is applied, resulting in a Wavelet Power Spectrum (WPS). In parallel, the health condition of IRJs is assessed on battered head failure (BHF), lipping failure (LF), and loose fastening bolts (LFB). The outputs of the defect assessment and WPS are, after some modifications, combined to search for correlations and classification characteristics. Possible correlations in total power, peak power and peak frequency within each bandwidth will be investigated. Classification is obtained by first applying a principal component analysis, followed by a support vector machine as the classifier.

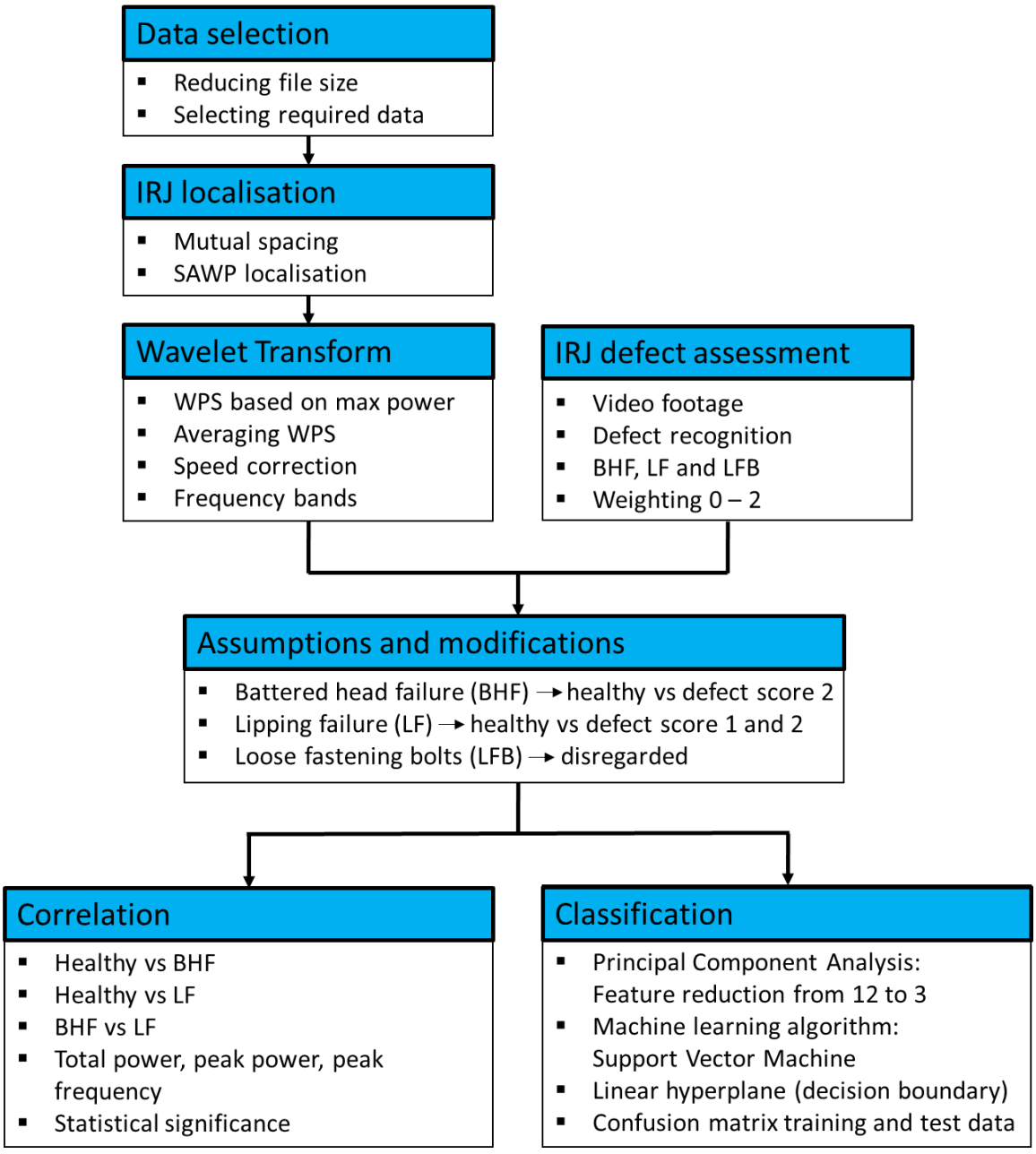


Figure 13 | *Schematic overview of methodology*

4.1 Data selection

The data retrieved from ProRail's Camino project consisted of one day of measurements monitored by two trains. As there were 4 accelerometers mounted on the axle boxes of each wheelset of a bogie, each measuring with a sampling frequency of 6250 Hz in the longitudinal, lateral, and vertical direction, terabytes of data were obtained. Since not all data is relevant, ProRail decided to share the data based on two conditions which is explained in ***Appendix D***. From the shared data only relevant information was needed to be selected which is explained in ***Appendix K***. For example, only the tracks that are within the contract region of the ASSET Rail company were obtained because the video footage of the track for the IRJ condition assessment was only available within this contract region. The output of data selection was a large data frame of 18 minutes (pre-processed) ABA data and information like GPS coordinates, velocity and "spoortak-id".

4.2 IRJ localisation

Within the 18-minute train ride, it was expected to locate many IRJs. The 18-minute train ride was divided into 4 parts, in which 9 batches of 2 minutes were present. Some parts contained multiple measurements which is beneficial for averaging the signal to increase reliability. By comparing a separate IRJ database with the registration (coordinates) of all IRJs within the contract region of ASSET Rail, the IRJs could be localized. However, due to a low GPS sampling frequency and inaccurate synchronisation to the ABA data inaccuracies up to 30 meters made IRJ localisation a challenge. ***Appendix L*** mentions the challenges of IRJ localisation and explains the performed steps to localise the IRJs. Using GPS coordinates, a Scale Average Wavelet Power (which is explained in chapter 4.3) and mutual spacing of IRJs it was possible to locate 60 IRJ pairs (left and right) and thus a total of 120 IRJs as can be seen in ***Table 3***.

Table 3 | *Overview obtained IRJs from batch measurements.*

Part	Batches	Multiple measurements	Number of IRJ pairs
1	1 & 2	2	10
2	3 & 4 & 5	3	20
3	6 & 7 & 8	3	18
4	9	1	12

Unfortunately, three issues were obstructing the localisation. Both front and rear measurements were needed to average the signal to obtain higher reliability. However, the front left accelerometer was broken and therefore it was not possible to average the signal, resulting in a 50% data reduction. Moreover, 6% of the IRJs were unable to be matched to the signal due to an unclear signal. Besides, for 3% of the data, a NaN value interrupted the usability of the data and had to be removed. It was not possible to apply a wavelet transform because of NaN values. Eventually, only 49 IRJs were left and used for the purpose of this study, as can be seen from ***Table 4***.

Table 4 | *Number of IRJs left due to inaccuracies of the data and IRJ localisation.*

	Number of IRJs	Percentage	IRJs left
Starting total	120	-	120
Front left error	60	50%	60
Unclear	7	6%	53
NaN	4	3%	49
		Total left	**49**

4.3 Continuous Wavelet Transform

As is explained in the theoretical background chapter the Continuous Wavelet Transform (CWT) is applied for the time-frequency transformation. This method is needed for the localisation of IRJs and the frequency extraction of the ABA signal. The CWT is described as follows (Nicoll, 2020):

$$W(\tau,s)=\frac{1}{\sqrt{|s|}}\int_{-\infty}^{+\infty} f(t)\psi^*\left(\frac{t-\tau}{s}\right)dt$$

The signal f(t) is multiplied by the complex conjugate (*) of the mother wavelet ψ(t). For this study, the Complex Morlet wave is used because, among other reasons, it is commonly used for the determination of the (electrical) signal's frequency content because of its feature of equal variance in both time and frequency domain (Devleker, 2016).

The Complex Morlet wavelet is a complex sine wave with a Gaussian envelope (Cohen, 2019) and is described as follows (Lee, 2018):

$$\psi(t)=\frac{1}{\sqrt{\pi B}}\exp^{-\frac{t^2}{B}}\exp^{\mathrm{j}2\pi Ct}$$

In the Complex Morlet wavelet, two parameters need to be chosen that define the trade-off between time precision and frequency precision. Those two parameters are the bandwidth (B) and center frequency (C), creating the Complex Morlet wavelet (cmorB-C). As the wavelet is already scaled and shifted, there is no interest in a predesigned particular frequency band. Therefore, the default center frequency (C) was set at 1 Hz (Tadel et al., 2019). Moreover, the center frequency (C or f_c) is coupled to the frequency and scale by $s=f_c/f$. More importantly, is the bandwidth (B) parameter which defines the width of the Gaussian. A wider Gaussian bandwidth decreases the time resolution but increases the frequency resolution (Cohen, 2019). Because of practical reasons, among them limited calculation power, the bandwidth is chosen to be 10 sec, resulting in the Complex Morlet wavelet cmor10-1. The scale array (s) is chosen between 0 and 3000 Hz with a logarithmic distribution composed of 300 steps.

Scale Average Wavelet Power

As is explained before, the Scale Average Wavelet Power (SAWP) is used for the localisation of IRJs. The SAWP captures the variation of the spectrum in a signal. It is used to determine fluctuations in the power of a predefined range of scales (Molodova, Li, et al., 2014) and is described as follows:

$$SAWP=\frac{\delta_j\delta_t}{C_\delta}\sum_{j=j_1}^{j_2}\frac{|W(s_j)|^2}{s_j}$$

Apart from the constant parameters (δ_j, δ_t and C_δ), SAWP is defined as the summation of the wavelet coefficients (W) over the scales s_{j1} to s_{j2}. As common frequency characteristics of IRJs are between 250 and 750 Hz (Molodova et al., 2016), the summation has been taken between the corresponding scales s_{j1} to s_{j2}. As a result, a time-SAWP graph can be constructed in which high values of SAWP correspond to the behaviour of IRJs. Other accelerations that occur at different frequencies, will not appear. Therefore, SAWP optimizes the localisation of IRJs.

Maximum wavelet power

After localizing the IRJs and applying the CWT, the Wavelet Power Spectrum (WPS) can be extracted. Wavelet Power (WP) is defined as the absolute value of the squared wavelet coefficients, as is described in the below formula. More information on the maximum wavelet power is given in ***Appendix M***.

$$WP = |W(\tau, s)^2|$$

Averaging and normalisation

Thereafter, as there are multiple measurements from the front and rear axle, the wavelet power is averaged. In some cases, it is possible to average double or triple measurements of the same IRJ, which enables to obtain more accurate and reliable results.

At last, the WP is normalised for the train speed. Molodova (2013) investigated the influence of speed on the acceleration response and concluded that there is a linear relation between train speed and ABA power between 60 and 140 km/h. Therefore, all wavelet powers are divided by the train speed in m/s.

Frequency bands

For the sake and simplicity of this thesis, the WPS is divided into 4 different frequency bands. The frequency bands are chosen based on the frequency characteristics described in the literature and based on the results of the WPS. ***Table 5*** gives an overview of the proposed frequency bands and an explanation is given in ***Appendix M***.

Table 5 | *Applied frequency bands.*

	Frequency bands
Low-frequency band	25 – 40 Hz 40 – 100 Hz
Middle-frequency band	200 – 600 Hz
High-frequency band	650 – 1250 Hz

Of each frequency band, three features are obtained: the total power, peak power and peak frequency. The total power is being calculated by integrating the area within the bandwidth using the trapezoidal rule. As the widths of the frequency bands differ from each other, the total power is divided over the width of the bandwidth resulting in total relative power. The peak power is the maximum value of relative power within each frequency band. The peak frequency is defined as the frequency at which the peak power occurs. ***Appendix N*** gives examples and an explanation of a scalogram and WPS as a result of the CWT. Besides that, train and track related eigenfrequencies are determined.

Eigenfrequencies

Eigenfrequencies are important in the investigation of a signal in its frequency domain. As with all objects, the rail track vibrates based on its natural frequencies, also known as eigenfrequencies. In fact, the track can be seen as a mass-spring system. If the system is forced from its equilibrium position due to an external force, the system will vibrate with the natural frequency that is specifically related to the system. The vibration modes are dependent on the track parameters. As the signature tunes of defects might be dependent on the natural frequencies of the track, the natural frequencies need to be considered (Molodova, 2013). From the literature, the vibrations modes and natural frequencies are summarized in ***Appendix O***.

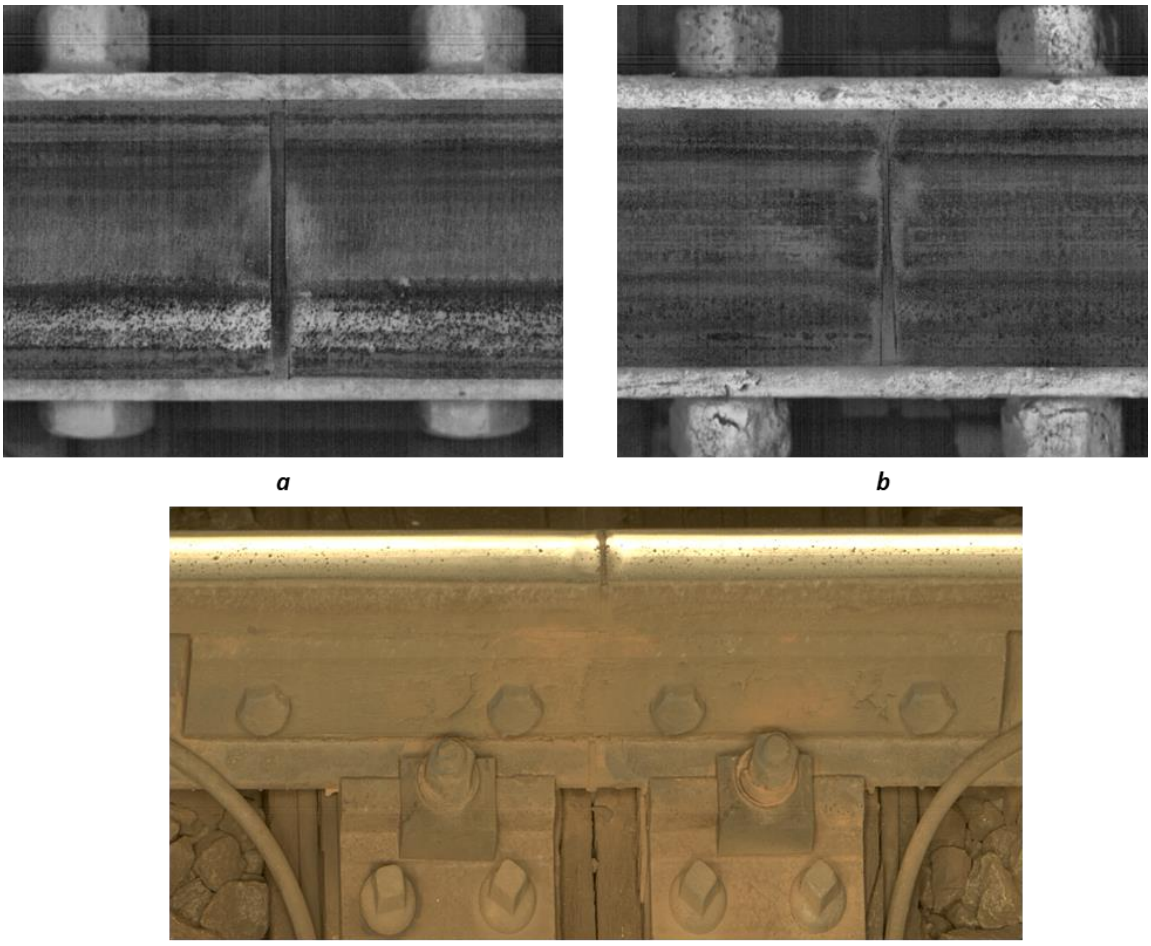
4.4 IRJ defect assessment

To determine the condition of the IRJs, video images in the possession of ASSET Rail have been investigated. A video train has recorded the top, inner and outer sides of the rails with a high frame rate and resolution. The top view images are suitable to observe rail surface defects, like permanent deformation, battered head failure and lipping failure. The side views are used to observe the condition of the fastening bolts and fishplate. For this thesis, the IRJ condition was assessed to failure modes battered head and lipping failure from the top view images. In addition, loose bolts from the side views were assessed.

Battered head failure is the downward flow of steel as a result of impact forces (Dangre, 2019). It is mostly coupled with arc damage, which is the white rash visible in images due to arcs. Moreover, circular deformation of the railhead at the insulating end post is an indication of battered head failure, see ***Figure 14a***.

Lipping failure is the flow of steel over the insulated end post as a result of fatigue (Dangre, 2019). It is visible from the top view images as the width of the insulated end post is smaller than originally, see ***Figure 14b***.

Loose fastening bolts can be recognized as corrosion around the bolt is visible. The right fastener bolt in ***Figure 14c*** is an example of the side view of an IRJ with corroded and therefore (most likely) loose fastener bolts.



a *b* *c*

Figure 14 | *Video images of* ***(a)*** *battered head failure,* ***(b)*** *lipping failure and* ***(c)*** *loose (corroded) fastening bolts.*

The IRJs were assessed on the presence of battered head failure, lipping failure, and loose fastening bolts. A weighting between 0 and 2 was given to the severity of the defect, in which a 0 indicates no presence of the defect, 1 indicates a light defect and 2 corresponds to moderate or severe defects, see ***Table 6***. ***Appendix B*** gives a clear overview of the damage types including its defects scores

Table 6 | *Weighting of defect assessment.*

Defect	Not present	Light	Severe
Weighting	0	1	2

4.5 Assumptions and modifications

During pre-processing of the data, difficulties led to the need for some assumptions and modifications. First of all, there is a one year difference between the instance of acceleration measurements and video footage. The video footage is used to determine the health condition of the IRJs, which is matched to the acceleration response. The data from the Camino project was only available for a single day in May 2019 whereas the video footage dated from May and August 2020. Therefore, it was not possible to perfectly match the health condition of IRJs to the acceleration response of the signal. Records from the maintenance log did mention that there was a maintenance operation within the intervening period, but without any details about the specific maintenance actions at each IRJ. Therefore, knowing the year difference between measurements and video footage, and the fact that there were maintenance operations within the intervening period, resulted in a lot of uncertainty and unreliability in the further analyses.

There are two scenarios possible. First, a defect might occur and grow within the intervening period. In this scenario, the health condition is overestimated to the time instance of measuring. The second scenario consists of the possibility that maintenance has been applied to the IRJ in the period between acceleration recording and video assessment. In this case, the IRJ would be assessed as healthy based on the video footage, whereas at the time instance of measuring the health condition was different. In this second scenario, the health condition would be underestimated to the time instance of measuring.

Based on these remarks, it was decided that the IRJs scored with a defect score of 1 for battered head failure were excluded from the analysis. IRJs with battered head failure defect score 1 do contain too much uncertainty and unreliability. Simultaneously, it is assumed that IRJs with defect score 0 for battered head failure are also healthy from battered head failure at the time of acceleration recording. Similarly, battered head failure scored with a defect score of 2 is assumed to be also present to a certain extent at the time of measurement.

Essentially, the same assumption applies to IRJs with lipping failure. However, due to limited data IRJs with lipping failure defect score 1 were included in the analysis. The clarification of the second assumption of including defect score 1 for lipping failure is reasonable because there is little difference between an IRJ with defect score 1 and 2 in the case of lipping failure, see ***Appendix P***. In both cases there is lipping failure visible and therefore the assumption can be made that the defects were also to a certain extent present at the time of recording.

In case of loose bolts, there was not enough data available to perform an analysis. Moreover, IRJs which are assessed with loose bolts, often are paired with other defects such as battered head failure or lipping failure. It was therefore decided to ignore the investigation of the results of IRJs with loose bolts.

The last assumption is related to healthy IRJs. It is possible that IRJs do not contain battered head failure and are therefore given a defect score of 0, but are degraded by lipping failure (with a defect score of 2). The IRJ cannot be assessed as healthy. Therefore, only IRJs which score 0's for all the defects are defined as healthy IRJs. Moreover, from the maintenance log, it was possible to acquire information about (completely) renewed IRJs. A total of 5 IRJs are replaced by a new one 3 weeks in advance of the measurements. These IRJs are completely certain to be healthy at the moment of measurement and are therefore reliable data points to distinguish between healthy and degraded IRJs. More details on these assumptions and their implications on the results are described in ***Appendix Q***.

As a result of the assumptions and modifications only 9 IRJs were assessed as healthy, see ***Table 7***, of which 5 IRJs are three weeks old (relative to the measuring date). The other four IRJs were assessed to be healthy from the video images as they did score zeros for all failure types. So, it was assumed that one year before the IRJs were also healthy. It was not possible to clarify if there was any maintenance during the intermediate period as there were no extensive maintenance logs. Most of the IRJs were assessed to have battered head failure with a defect score of 2 (12 out of 49 IRJs). Additionally, 11 IRJs were found to have lipping failure during the video footage assessment (defect score 1 and 2 together). Only a limited number of joints appeared with loose bolts (4 out of 49). Due to the limited number of data points for this defect, classification of loose bolts is not performed. Lastly, a few of IRJs were assessed to contain a different defect outside the scope of this study and were therefore also excluded from the analysis. However, these defects were interesting for the discussion.

Table 7 | *Overview of the number of healthy and damaged IRJs according to the defect assessment.*

Total IRJS = 49	**Score**		
	0	**1**	**2**
Healthy (all 0's)	**9**	-	-
Battered head	15	22	**12**
Lipping	38	**5**	**6**
Loose bolts	36	9	4
Other	29	18	2

4.6 Correlation

Healthy IRJs and degraded IRJs were investigated for possible correlations in each of the 3 features total power, peak power and peak frequency. A scatter plot of the IRJs (both healthy and degraded IRJs indicated with colour codes) was showing a feature, total power for example, for two frequency bands on the axes. The characteristics within all frequency bands were plotted in two figures for all the features.

Besides that, the increases between the mean values of healthy IRJs and degraded IRJs were calculated and tested to be statistically significant. For this, Welch's T-test is applied which is a modification of the independent samples T-test (Glen, 2021b). Whereas the independent samples T-test is quite unreliable for unequal sample sizes and variances, Welch's T-test tends to perform better (Glen, 2021b). Besides that, the assumption of normality is applicable to the data used in this thesis. The Welch's T-test determines whether two sample means are significantly different by stating in the null hypothesis for the test that the means of both groups are equal. In fact, the test answers the question "could these differences have occurred by random chance?" (Glen, 2021b). In this thesis, the two groups are healthy IRJs and IRJs with either battered head failure or lipping failure.

The formula for Welch's T-test is described below,

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

in which X_n is the mean value of the dataset (n=2), s_n the standard deviation of the dataset and N_n the length of the dataset. The t-value is a measure for the "difference between two groups and the difference within the groups" (Glen, 2021a). The smaller the t-value, the more similarity there is between groups. The t-value can be translated into a p-value, which is "the probability that the results from the sample data occurred by chance" (Glen, 2021a). P-values vary from 0% 9 (p=0) to 100% (p=1). The lower the p-value, the higher the chance the data did not occur by chance. The upper threshold (alpha) whether the datasets are equal or different from each other is commonly set at 5% (Glen, 2021a). However, the threshold value depends on the type of test. If a one-sided test is performed which investigates the chance that X is larger than Y then the threshold value is 5%. However, if a two-sided test is performed which investigates the chance that X is lower or larger than Y then the threshold value should be taken as 2.5%. Below, the threshold values are explained which are used in this thesis.

Battered head failure

In case of the features total relative power and peak relative power, an increase in power is expected for IRJs with battered head failure. Therefore, a one-tailed T-test is applied with a corresponding alpha of 0.05. In case of peak frequency, its behaviour cannot be predicted and therefore a two-tailed T-test is applied with a corresponding alpha of 0.025.

Lipping failure

In case of lipping failure, it is more difficult to expect increased powers in all frequency bands. The defect does not specifically increase the vertical momentum as is explained in chapter 3. For this reason, the statistical significance of all three features is tested two-sided with a corresponding alpha of 0.025.

4.7 Classification

After the correlation assessment, IRJs were classified by means of the Support Vector Machine (SVM). First, to allow for validation of the constructed classifier, the dataset was randomly separated into a training set (60%) which trains the algorithm and a test set (40%) which validates the algorithm. ***Table 8*** gives the number of healthy IRJs and IRJs with battered head failure and lipping failure for both training and test datasets. Unfortunately, by dividing the dataset into two parts, there was only a limited number of IRJs left. Since there were only a limited number of times IRJs present with loose fastening bolts (n=5), this defect is excluded from classification.

Table 8 | *Overview number of healthy IRJs and IRJs with battered head failure and lipping failure used for classification.*

	Training dataset	**Test dataset**
Healthy	5	4
Battered head failure	7	5
Lipping failure	5	5

The 12 features (total relative power, peak relative power and peak frequency) are independent of each other and characteristic for the IRJ. Instead of investigating each of the 12 features, the Principal Component Analysis (PCA) can reduce the dimensionality of the dataset into a limited number of components that explain the variety seen in the data. This way the interpretability of all features at once is increased without losing much information (Jollife & Cadima, 2016). The PCA was constructed on the training dataset and fitted to the test dataset.

The three features were applied in an SVM. The objective of SVM is to find a hyperplane that maximizes the distance between the data points of the two classes. The hyperplane acts as a decision boundary: IRJs on one side of the hyperplane are predicted healthy and IRJs on the other side of the hyperplane are predicted to be damaged. The results of the classification were denoted in a confusion matrix which shows the true and false positives, as well as true and false negatives. For example, the IRJ can be predicted to be healthy and being actually healthy (true negative). But, an IRJ can also be predicted to be damaged, but being actually healthy (false positive). The performance score of the SVM algorithm is calculated by the number of correctly predicted IRJs divided over the total number of IRJs.

5 Results

After pre-processing and selection of the data based on the described assumptions, the frequency behaviour of degraded IRJs was analysed in comparison to healthy IRJs. The results section is divided into two parts. Firstly, the correlations between healthy and degraded IRJs are described and in the second part, the results of the classification of healthy and degraded IRJs are presented.

The correlation results are subdivided into three subchapters: correlations between healthy IRJs and IRJs with battered head failure, correlations between healthy IRJs and IRJs with lipping failure and differences in correlations between IRJs with battered head failure and IRJs with lipping failure. IRJs with loose fastening bolts are present in the dataset only a limited number of times (n=5) and were therefore excluded from the analysis. First, the results are described after which a conclusion is given which answers the fourth research question. Thereafter, the results and conclusions are discussed and a reflection to the theory is made. Healthy IRJs are distinguished in new and healthy in the graphs, but the term healthy is used in the text.

The classification results are subdivided into two subchapters in which IRJs with battered head failure and lipping failure are classified by means of a PCA and SVM.

5.1 Correlation battered head failure

To answer the fourth research question regarding characteristic frequency responses, correlations within the three features total relative power, peak relative power and peak frequency were investigated. The relative difference between the mean values of healthy IRJs and degraded IRJs was determined as well as its statistical significance by means of a Welch's T-test. Since total relative power and peak relative power behave similarly, the correlation results are only extensively described for total relative power. The graphs of peak relative power and peak frequency are added to ***Appendix R***.

5.1.1 Correlation BHF

Total relative power

In ***Figure 15*** the total relative power of healthy IRJs is compared to the total relative power of IRJs with battered head failure. A difference between total relative power for healthy IRJs and total relative power for battered head IRJs is present within all frequency bands. Most individual IRJs with battered head failure show increased total relative power in all frequency bands as compared to healthy IRJs. A cluster of healthy IRJs is visible with relatively low power in the lower-left corner of both total relative power graphs. The IRJs with battered head failure are more distributed over relatively higher powers. There is one outlier visible (denoted with superscript 1) which contains increased power in frequency band 200-600 Hz and 650-1250 Hz. Moreover, two healthy IRJs (denoted with superscripts 2 and 3) contain relatively large power in the same frequency bands. In ***Appendix R*** the outlier IRJs are elaborated on and discussed.

Apart from the individual data points, the mean relative power is indicated by a large open circle and the size and direction of the increase are visualized by the black arrow. There is a clear difference in the total relative power of healthy IRJs and IRJs with battered head failure. In ***Table 9*** the mean power for every frequency band for both healthy and damaged IRJs including the standard deviation are given as well as the relative power increase. The results from statistical analysis are also presented in the table.

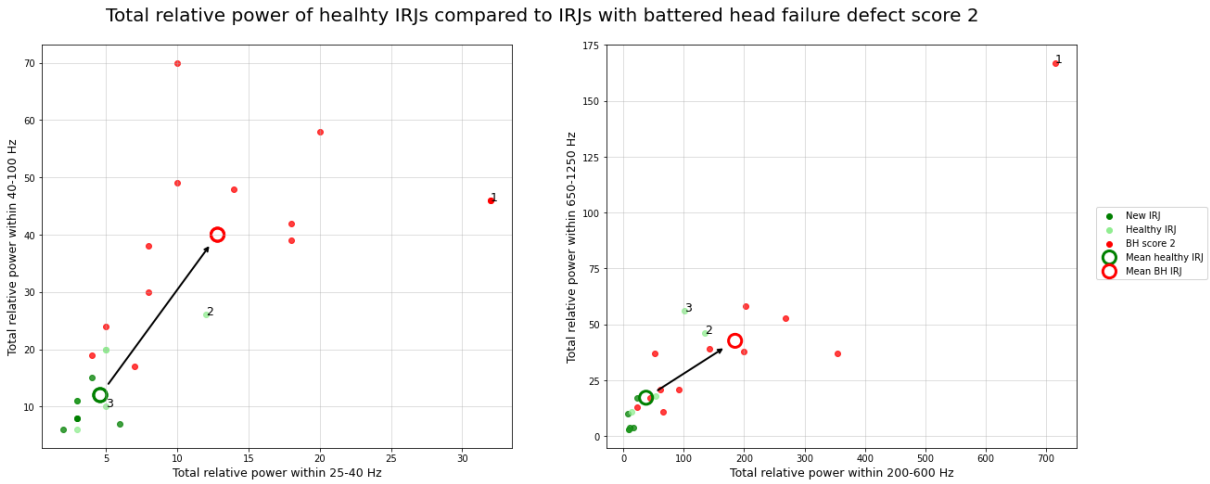


Figure 15 | *Total relative power of healthy IRJs compared to IRJs with battered head failure defect score 2.*

From ***Table 9*** it can be seen that the largest increase in total relative power occurs in frequency band 200-600 Hz, with a factor of 4.54. A change from a total relative power of 37.6 to 185.0 is visible in ***Figure 15***. Additionally, this increase is found to be statistically significant ($p_{200\text{-}600Hz}$ = 0.029). Moreover, an increase in power within frequency band 25-40 Hz and 40-100 Hz are statistically significant with even lower p-values ($p_{25\text{-}40Hz}$ = 0.0062 & $p_{40\text{-}100Hz}$ = 0.000053). Only the increase in frequency band 650-1250 Hz is not statistically significant ($p_{650\text{-}1250Hz}$ = 0.10).

Peak relative power

The peak relative power shows a similar behaviour as the total relative power. Since a higher total power is often a result of a higher peak power because total power is the area under the PSD, it is not surprising to see a similar response. The peak power in all frequency bands increases in case of battered head failure as can be seen from ***Table 9***. The largest increase in peak relative power is found in the same frequency band as for total relative power having frequency values between 200-600 Hz and a factor increase of 4.23. Furthermore, the same outliers as for total relative power are present and result in high standard deviations.

Peak frequency

The peak frequency slightly increases within frequency bands 40-100 Hz and 200-600 Hz but decreases in frequency band 650-1250 Hz. A cluster of healthy IRJs can be distinguished around approximately 60 Hz, which is close to the eigenfrequency of the train (see ***Appendix N***). The lowest frequency band does not change between healthy and battered head IRJs. The largest increase is found in frequency band 200-600 Hz with a factor of 1.12 and the largest decrease is found in frequency band 650-1250 Hz with a factor of 0.88. However, all deviations are found to be statistically insignificant (p-value > 0.025) which implies that it is uncertain to relate the change in peak frequency to battered head failure.

Table 9 | *Overview of the mean and standard deviation of the three features total relative power, peak relative power and peak frequency of healthy and battered head IRJs. The last two columns describe the increasing factor from the mean value of healthy IRJs to battered head IRJs and its statistical significance in terms of a P-value derived from Welch's T-test. Total relative power and peak relative power are tested one-sided (alpha=0.05) whereas peak frequency is tested two-sided (alpha=0.025).*

	Frequency band	Healthy IRJs		Battered head IRJs		Factor increase	P-value
		Mean	Std	Mean	Std		
Total relative power	25 – 40 Hz	4.6	2.9	12.8	8.0	2.69	**0.0062**
	40 – 100 Hz	12.0	6.6	40.0	15.8	3.30	**0.000053**
	200 – 600 Hz	37.6	45.2	185.0	195.6	**4.54**	**0.029**
	650 – 1250 Hz	17.2	18.8	42.7	41.8	2.27	**0.10**
Peak relative power	25 – 40 Hz	8.6	6.6	22.4	12.4	2.49	**0.0055**
	40 – 100 Hz	29.8	20.2	104.3	43.3	3.40	**0.000082**
	200 – 600 Hz	86.0	99.0	393.0	429.0	**4.23**	**0.037**
	650 – 1250 Hz	47.5	59.7	115.6	134.5	2.21	**0.17**
Peak frequency	25 – 40 Hz	30.0	4.6	30.1	5.1	1.00	**0.97**
	40 – 100 Hz	63.9	7.3	71.0	8.5	1.10	**0.078**
	200 – 600 Hz	359.8	50.4	396.3	90.2	**1.12**	**0.20**
	650 – 1250 Hz	991.8	96.9	870.8	149.2	**0.88**	**0.050**

5.1.2 Conclusion BHF

A statistically significant increase is found in frequency bands 25-40 Hz, 40-100 Hz and 200-600 Hz for the total relative power and peak relative power, among which frequency band 200-600 Hz shows the largest absolute and relative increase. Furthermore, although an increase and decrease in peak frequencies are observed, no statistically significant influence of battered head failure on the peak frequency could be established in this study.

According to the theory (see chapter 3), it was expected that battered head failure results in increased impact forces at IRJs and therefore an increase in vertical momentum would be evident. Furthermore, a larger vertical momentum involves larger accelerations and therefore higher power since power is proportional to the square of the amplitude of the acceleration signal. Besides that, from the results it is concluded that battered head failure does not significantly influence peak frequency and joint stiffness.

5.1.3 Discussion BHF

Although the largest total relative power increase is found in frequency band 200-600 Hz, the lowest p-value is found in frequency band 40-100 Hz. Since Welch's T-test is dependent on the standard deviation, a relatively large standard deviation involves lower t-values and therefore larger p-values. From ***Table 9*** it can be seen that the standard deviation for power within frequency band 200-600 Hz is relatively large for both healthy IRJs and IRJs with battered head failure thereby partly explaining the higher p-value. The relatively large standard deviations are caused by outliers like the beforementioned outlier denoted with superscript 1.

From the theory it is known (see chapter 3) that sleeper spacing (and thus type of sleeper) influences the peak frequency response. Since the IRJs are a random mix of ETS joints and NRG joints, the statistically insignificant deviations in peak frequency could be related to a different type of sleeper and thus not related to battered head failure.

5.2 Correlation lipping failure

The next chapter investigates the frequency behaviour of IRJs with lipping failure to answer the second part of the fourth research question. A relative increase between the mean values of healthy IRJs and degraded IRJs was determined as well as its statistical significance. Because total relative power and peak relative power behave similarly, the correlation results are only extensively described for total relative power. The graphs of peak relative power and peak frequency are added to ***Appendix S***. The appendix also shows an additional analysis in which an outlier was removed from the data before correlation analysis.

5.2.1 Correlation LF

Total relative power

IRJs with lipping failure contain increased powers in every frequency band as can be seen from the mean values in ***Figure 16***. However, the increase is much lower than in case of IRJs with battered head failure. The same two healthy IRJs as in the results of battered head failure (superscripts 2 and 3) are within the range of IRJs with lipping failure. Moreover, three IRJs with lipping failure (30%) are situated in the cluster of healthy IRJs (denoted with superscripts 4, 5 and 6) with low powers in all frequency bands.

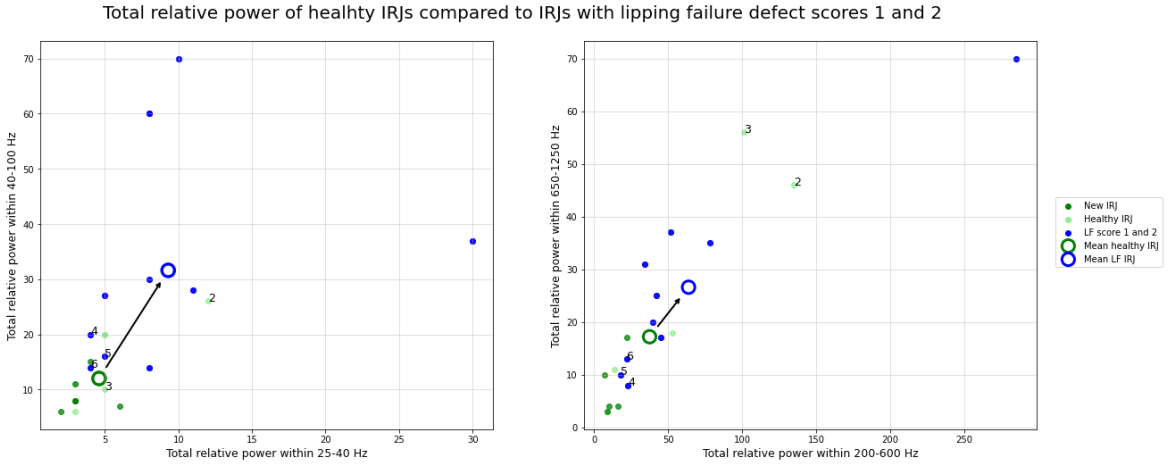


Figure 16 | *Total relative power of healthy IRJs compared to IRJs with lipping failure defect scores 1 and 2.*

The largest increase in total relative power between healthy IRJs and IRJs with lipping failure is found in frequency band 40-100 Hz with an increase factor of 2.61, as can be seen from ***Table 10***. This power increase is statistically significant ($p_{40\text{-}100Hz}$ = 0.012). All other frequency bands show small increases in total relative power. However, since there is no statistical significance (p-values > 0.025) in these frequency bands, it implies that there is no demonstratable impact of lipping failure on the amount of energy within these frequency bands.

Peak relative power

From the peak relative power, a similar trend is visible as depicted in ***Table 10***. Especially, the peak relative power for frequency band 40-100 Hz shows a large increase between healthy IRJs and IRJs with lipping failure having an increase factor of 2.68. Again, this is the only frequency band that shows statistical significance between healthy IRJs and IRJs with lipping failure ($p_{40\text{-}100Hz}$ = 0.0059).

Peak frequency

There is a lot of variation in peak frequency both for healthy IRJs as well as for lipping failed IRJs as can be seen from the high standard deviations presented in ***Table 10***. No large differences are visible in peak frequency, except for frequency band 650-1250 Hz. A statistically significant decrease ($p_{650\text{-}1250Hz}$ = 0.0053) is obtained with a relative factor of 0.86 and an absolute decrease of 90 Hz.

Table 10 | *Overview of the mean and standard deviation of the three features total relative power, peak relative power and peak frequency of healthy IRJs and IRJs with lipping failure. The last two columns describe the increasing factor from the mean value of healthy IRJs to IRJs with lipping failure and its statistical significance in terms of a P-value derived from Welch's T-test. All three features are tested two-sided (alpha=0.025).*

	Frequency band	Healthy IRJs		Lipping failure IRJs		Factor increase	P-value
		Mean	Std	Mean	Std		
Total relative power	25 – 40 Hz	4.6	2.9	9.3	7.7	1.95	**0.11**
	40 – 100 Hz	12.0	6.6	31.6	19.3	**2.61**	**0.012**
	200 – 600 Hz	37.6	45.2	63.9	79.6	1.57	**0.45**
	650 – 1250 Hz	17.2	18.8	26.6	18.3	1.42	**0.39**
Peak relative power	25 – 40 Hz	8.6	6.6	15.4	14.2	1.71	**0.23**
	40 – 100 Hz	29.8	20.2	82.2	44.4	**2.68**	**0.0059**
	200 – 600 Hz	86.0	99.0	134.2	175.3	1.44	**0.54**
	650 – 1250 Hz	47.5	59.7	71.0	56.3	1.36	**0.50**
Peak frequency	25 – 40 Hz	30	4.6	32.1	5.8	1.07	**0.40**
	40 – 100 Hz	63.9	7.3	68.8	10.3	1.07	**0.30**
	200 – 600 Hz	359.8	50.4	395.2	85.1	1.11	**0.23**
	650 – 1250 Hz	991.8	96.9	850.9	75.7	**0.86**	**0.0053**

5.2.2 Conclusion LF

There is a statistically significant increase in frequency band 40-100 Hz for the total relative power and peak relative power. Power increases in the other frequency bands are present but are found to be statistically insignificant. From the theory, it was expected that only a small power increase would be obtained which is the case according to the data. In the peak frequency analysis, a statistically significant decrease in peak frequency is observed in frequency band 650-1250 Hz.

5.2.3 Discussion LF

The three IRJs with lipping failure that are situated in the cluster of healthy IRJs (denoted with superscript 4, 5 and 6) contain low powers in all frequency bands, which is remarkable for degraded IRJs. One possible explanation is the beforementioned mismatch between the video footage used for defect assessment and the time of recording. Besides that, since there is no statistically significant difference in three of the four frequency bands, it can also be concluded that IRJs with lipping failure do not differ much from healthy IRJs concerning to the three investigated features total relative power, peak relative power and peak frequency.

The change in peak frequency within frequency band 650-1250 Hz implies a change in pin-pin resonance. However, as the peak frequency decreases from approximately 990 Hz to 850 Hz, a change in pin-pin resonance is unlikely as pin-pin resonance occurs at a higher frequency of approximately 1160 Hz (Molodova, 2013). The peak frequency might be related to a change in joint stiffness as is explained in chapter 3. It is difficult to discuss the precise effect of lipping failure on a change in peak frequency due to the dependency on many factors involved. However, a decrease in inertia could decrease the peak frequency. Besides that, from the theory it is known that the sleeper

distance influences the peak frequency. Since there are two types of IRJs that are supported differently (see ***Appendix A***), the deviation in peak frequency might be caused by the sleeper distance and therefore type of sleeper.

5.3 Correlation battered head failure vs lipping failure

Besides comparing the separate defects to healthy IRJs it is also interesting to investigate if there exist any differences between the considered defects. Finding such differences might open up possibilities for defect specific classification. Therefore, IRJs with battered head failure and lipping failure are compared to each other in this section concerning the same features (total relative power, peak relative power and peak frequency).

5.3.1 Correlation BHF vs LF

The total relative power of healthy IRJs, IRJs with battered head failure and IRJs with lipping failure are plotted in ***Figure 17***. It can be seen that IRJs with battered head failure have more power in all frequency bands than IRJs with lipping failure. From the right graph in ***Figure 17***, it can be observed that IRJs with lipping failure are more related to healthy IRJs than to IRJs with battered head failure concerning the total relative power. A similar trend is visible for the feature peak relative power for which graphs are included in ***Appendix T***. For the peak frequency, no significant differences between battered head and lipping failure are present (see ***Appendix T***).

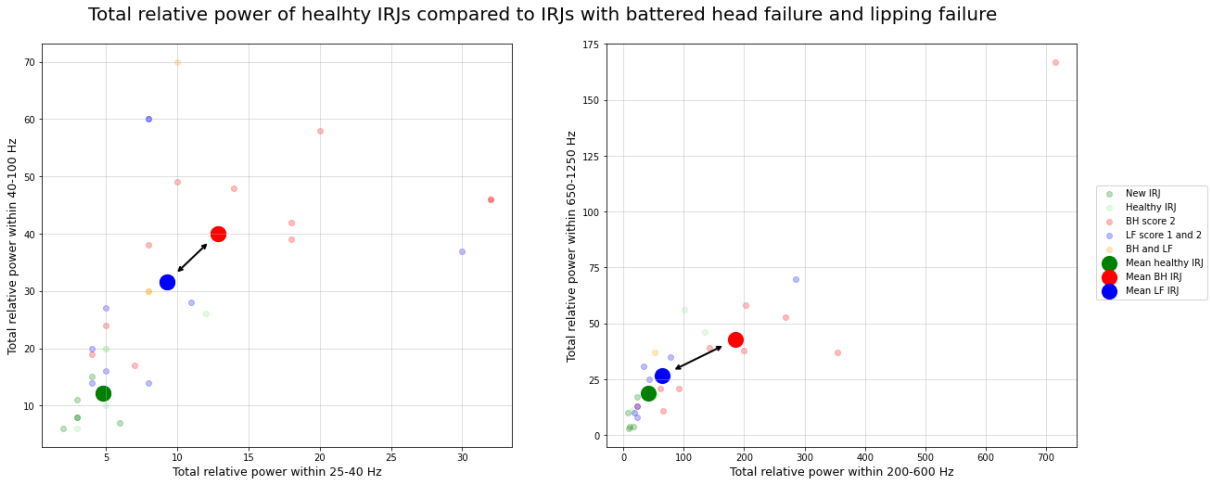


Figure 17 | *Total relative power of healthy IRJs, IRJs with battered head failure and IRJs with lipping failure.*

From ***Table 11*** it can be seen that the largest difference of total relative power between battered head failure and lipping failure is in frequency band 200-600 Hz. Battered head failure contains 2.90 times more power than lipping failure. Besides, the peak relative power is 2.93 times larger in case of battered head failure compared to lipping failure. Although there is a power difference visible, these differences are found to be statistically insignificant.

Table 11 | *Overview of the mean and standard deviation of the three features total relative power, peak relative power and peak frequency of IRJs with lipping failure and IRJs with battered head failure. The last two columns describe the increasing factor from the mean value of IRJs with lipping failure to IRJs with battered head failure and its statistical significance in terms of a P-value derived from Welch's T-test. All three features are tested two-sided (alpha=0.025).*

	Frequency band	Lipping failure IRJs		Battered head IRJs		Factor increase	P-value
		Mean	Std	Mean	Std		
Total relative power	25 – 40 Hz	9.3	7.7	12.8	8.0	1.38	0.31
	40 – 100 Hz	31.6	19.3	40.0	15.8	1.27	0.28
	200 – 600 Hz	63.9	79.6	185.0	195.6	2.90	0.069
	650 – 1250 Hz	26.6	18.3	42.7	41.8	1.60	0.25
Peak relative power	25 – 40 Hz	15.4	14.2	22.4	12.4	1.46	0.24
	40 – 100 Hz	82.2	44.4	104.3	43.3	1.27	0.25
	200 – 600 Hz	134.2	175.3	393.0	429.0	2.93	0.076
	650 – 1250 Hz	71.0	56.3	115.6	134.5	1.63	0.31
Peak frequency	25 – 40 Hz	32.1	5.8	30.1	5.1	0.94	0.40
	40 – 100 Hz	68.8	10.3	71.0	8.5	1.03	0.60
	200 – 600 Hz	395.2	85.1	396.3	90.2	1.00	0.98
	650 – 1250 Hz	850.9	75.7	870.8	149.2	1.02	0.69

It is difficult to distinguish IRJs with battered head failure from IRJs with lipping failure based on total relative power, peak relative power or peak frequency. However, there is a small difference visible from the clusters of IRJs. For total relative power and peak relative power, lipping failure has a higher relative power in frequency band 40-100 Hz given a low relative power in frequency band 25-40 Hz. For example, a cluster of IRJs with lipping failure is visible from the right total relative power graph of ***Figure 17***. The peak relative power in bandwidth 650-1250 Hz for lipping failure is higher than for battered head failure given a low power within frequency band 200-600 Hz. Considering a combination of two frequency bands could therefore be beneficial for distinguishing lipping failure from battered head failure. However, no statistical test was performed in this study investigating multivariate relationships and therefore this remains a matter for future research.

5.3.2 Conclusion BHF vs LF

The largest difference is obtained in frequency band 200-600 Hz which was expected from the theory. Moreover, the smallest increase is found in frequency band 40-100 Hz which implies that lipping failure increases with a similar rate in this frequency band compared to battered head failure. However, using the features total relative power, peak relative power and peak frequency separately, it is not possible to statistically distinguish IRJs with battered head failure and IRJs with lipping failure. But, considering a combination of two frequency bands could be beneficial for distinguishing the two defects. The latter observation is however not statistically tested in this study.

5.3.3 Discussion BHF vs LF

No statistical difference is found between IRJs with battered head failure and IRJs with lipping failure. A statistically significant difference is difficult to expect since the exact behaviour is simply not known. An interesting observation is the visible cluster within the high-frequency ranges. It looks like IRJs with lipping failure contain relatively higher powers in frequency band 650-1250 Hz than IRJs with battered head failure, given a low power in frequency band 200-600 Hz. This observation should be tested to be significant. However, a Welch's T-test is not able to test multiple variables. A possible solution could be the application of a Multivariate ANOVA test, which compares multiple variables and determine their statistical difference.

5.4 Classification battered head failure

After analysing the correlations, the three features are used for classification. The results of the classification are divided into two parts. The first part gives the results from the Analysis (PCA) and in the second part, the confusion matrix is shown as a result of the decision boundary which is constructed by the Support Vector Machine algorithm (SVM). The results of PCA and SVM are shown first for IRJs with battered head failure, followed by the results of IRJs with lipping failure.

To allow for validation of the constructed classifier, the dataset was randomly separated into a training set (which trains the algorithm) and a test set (which validates the algorithm). Moreover, the results are controlled by means of a switch in train and test data. If a similar decision line is constructed on the test dataset it would prove that the decision line is reliable. Also, if the results of the confusion matrix are better on the test dataset, the SVM could be improved. A worse result implies an unreliable SVM.

5.4.1 Principal Component Analysis BHF

The PCA captures a large part of the information of all 12 variables (3 features x 4 bandwidths) in only 3 principal components. From ***Table 12*** it can be seen that at least 87% of the variation present in the data is contained in the first three principal components. An explained variance of 80% is acceptable for descriptive purposes (Minitab, 2019) and therefore the first three components are selected to use for classification. A 3D plot of the three principal components including a front, side and top view, is added to ***Appendix U***. Based on visual inspection, it appears that PC1 and PC3 are important in the classification of healthy and battered head IRJs.

Table 12 | *Percentage of data information that is captured in 3 components*

Principal Component	Percentage of variance explained
PC 1	60 %
PC 2	14 %
PC 3	13 %
Total	**87 %**

5.4.2 Support Vector Machine BHF

Training dataset

The machine learning algorithm SVM is applied to the three principal components of the training set, which results in a three-dimensional decision plane as depicted in ***Figure 18***. All healthy IRJs are separated from IRJs with battered head failure resulting in a performance score of 1 (see ***Figure 18***). Moreover, it can be seen that the decision plane has maximized the distance between the data points of the two classes, which is expected based on the cost function of the SVM algorithm.

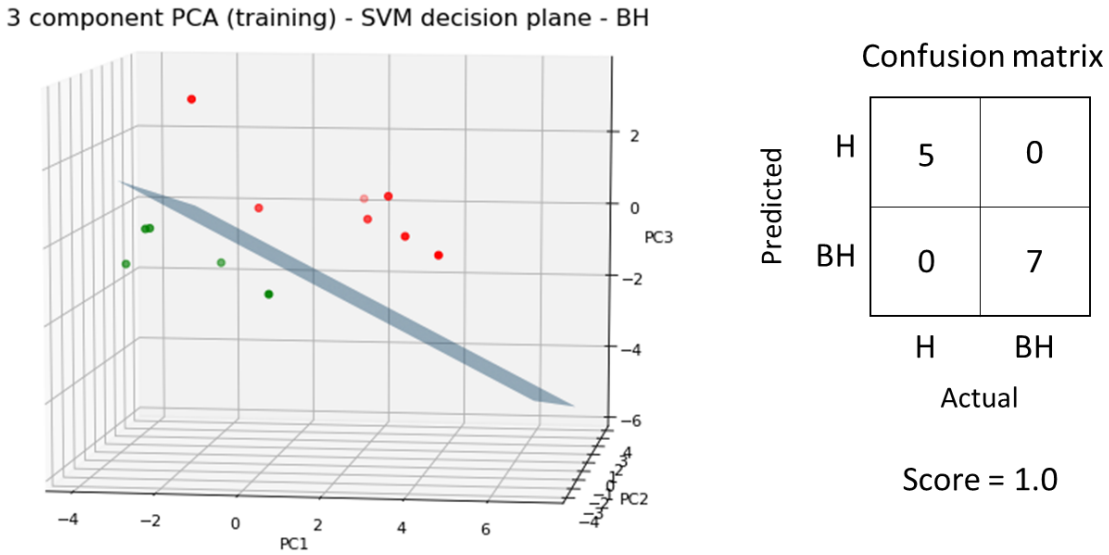


Figure 18 | *3D plot PCA and SVM decision plane for battered head failure including the confusion matrix and training score. The confusion matrix presents the results in terms of correctly predicted healthy (H) or battered head failed (BH) IRJs.*

Test dataset

The obtained classifier is tested and the results are added in ***Appendix X***. Applying the previously found decision plane to the test data, a performance score of 0.56 is achieved. Out of the four healthy IRJs, only one IRJ was predicted to be healthy correctly. Out of the five battered head IRJs on the other hand, only one IRJ was falsely classified as healthy. Four IRJs were correctly predicted to have battered head failure.

Since the algorithm does not perform well on the test dataset, it implies that the classifier found by applying the SVM algorithm on the training set is not stable and thus unreliable. To test whether this is due to an unrepresentative training dataset, the SVM algorithm was also applied to the test dataset and validated on the training dataset. Therefore, a validation test has been performed in which the SVM is constructed based on the test dataset and tested on the training dataset. If a similar decision plane existed, it would mean that the SVM performed well, producing a stable classifier. The results are stated in ***Appendix V***. The overall performance score increased (SVM training score of 1.0 and test score of 0.83) which implies a better classification might have been possible. However, it can also be concluded that a different plane was constructed which implies that the found classifier to this dataset is unstable and unreliable.

5.4.3 Conclusion BHF

The PCA was able to capture 87% of the variation within the data in only 3 components, which is an acceptable percentage for descriptive purposes (Minitab, 2019). Having a small number of components explaining a large part (>80%) of the variance in the data is beneficial for classification purposes. Using more variables induces the risk of overfitting during classification.

The found classifier is unstable and unreliable because of a low test performance score. Moreover, a validation test confirmed that the application of SVM on this dataset is not suitable because a different decision plane was constructed indicating an unstable and unreliable performance.

5.4.4 Discussion BHF

The application of a PCA was useful to reduce the dimensionality of a dataset. This way a lot of information present in the data was included in only a few principal components. However, a disadvantage of PCA is that the physical meaning of the components is gone which makes it difficult to interpret the results through theory. The PCA is therefore a useful tool before application of a machine learning algorithm, but a separate correlation assessment is needed to fully understand the responses.

The SVM was found unstable for the given dataset and therefore unreliable due to a low performance score and a different validation plane. The SVM is therefore not applicable for the dataset used in this study. Due to the data splitting in a training and test dataset, the number of data points in both sets was limited resulting in unrepresentative datasets for the whole set of IRJs. Moreover, the one year difference in ABA data and video footage for the determination of IRJ health condition results in a lot of uncertainty. One IRJ which is assessed wrongly could influence the classifier a lot. Therefore, more data and a better defect assessment are needed to obtain higher reliability.

5.5 Classification lipping failure

Similarly, the classification of lipping failure is studied. Firstly, the results from the PCA are described and in the second part, the confusion matrix is shown as a result of the decision boundary which is constructed by the SVM.

5.5.1 Principal Component Analysis LF

From ***Table 13***, it can be obtained that 86% of the variance in the data is explained by the first three principal components. Since more than 80% of the variance is captured by these components, continuing with the first three components for classification is acceptable (Minitab, 2019). A 3D plot of the three principal components including a front, side and top view, is added to ***Appendix W***. Based on visual inspection, it appears that PC1 and PC2 are most important in the classification of healthy and IRJs with lipping failure.

Table 13 | *Percentage of data information that is captured in 3 components.*

Principal Component	Percentage of variance explained
PC 1	54 %
PC 2	19 %
PC 3	13 %
Total	**86 %**

5.5.2 Support Vector Machine LF

Training dataset

The machine learning algorithm SVM is applied to the three principal components of the training dataset, which results in a decision plane as depicted in ***Figure 19***. A confusion matrix is constructed, including the performance score of the SVM algorithm. Four out of five healthy IRJs are separated correctly from IRJs with lipping failure and all five IRJs with lipping failure are correctly assessed as IRJs with lipping failure resulting in a performance score of 0.9.

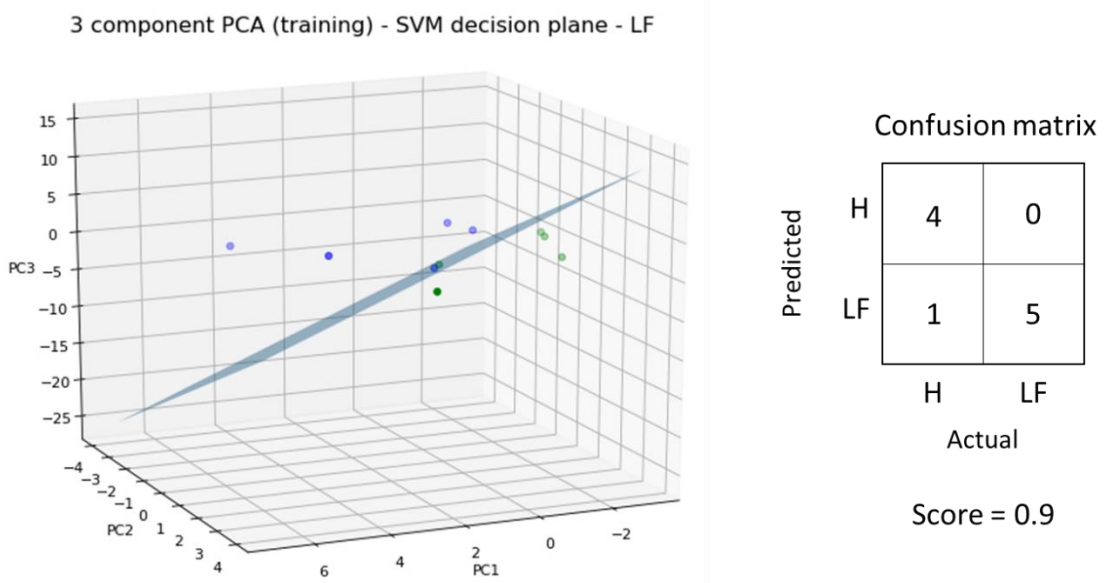


Figure 19 | *3D plot PCA and SVM decision plane for lipping failure including the confusion matrix and training score. The confusion matrix presents the results in terms of correctly predicted healthy (H) or lipping failed (LF) IRJs.*

Test dataset

The found decision plane is applied to the test dataset and the results are added in ***Appendix X***. Applying the same decision plane to the test data, a performance score of 0.67 is achieved. Out of the four healthy IRJs, only one IRJ was predicted to be healthy correctly. On the other hand, all five IRJs with lipping failure were correctly predicted to have lipping failure. According to the test, it looks like the decision plane is strict compared to healthy IRJs because only 25% of the healthy IRJs were predicted correctly, whereas there is a margin left for IRJs with lipping failure which are all predicted correctly. Besides that, the training dataset could have included a lot of healthy IRJs compared to less healthy IRJs in the test dataset. Therefore, more data should be used to overcome the problem of unequal distributions.

Since the found classifier is based on a small dataset and not able to predict the condition of IRJs with high accuracies, it implies that the classifier is again not stable on this dataset and thus unreliable. Therefore, a validation test has been performed in which the SVM is constructed based on the test dataset and tested on the training dataset. The results are summarized in ***Appendix Y***. The overall performance score increased (SVM training score of 0.89 and test score of 0.78) which implies a better decision plane was possible. However, it can also be obtained that a different plane (perpendicular to the previous decision plane) was constructed which implies that the classifier is unstable and unreliable.

5.5.3 Conclusion LF

The PCA was able to capture 86% of the variation within the data in only 3 components, and therefore performing classification with only 3 components is suitable (Minitab, 2019).

The Support Vector Machine algorithm is unstable and unreliable because of a mere test performance score. Moreover, a validation test confirmed that the application of SVM on this dataset is not suitable because a different decision plane was constructed indicating an unstable and unreliable performance.

5.5.4 Discussion LF

The SVM was found unstable for the applied dataset and therefore unreliable due to a low performance score and a different validation plane. The SVM is therefore not applicable for the dataset used in this study. Due to the data splitting in a training and test dataset, the number of data points decreased too much. Moreover, the one year difference in ABA data and video footage resulted in a lot of inaccuracy and uncertainty. Therefore, more data and a better defect assessment are needed to obtain higher reliability.

6 Conclusion & discussion

In this thesis, ABA data is used to investigate the health condition of IRJs which can be used for predictive maintenance. Using frequency characteristics from the Continuous Wavelet Transformation the main research question “How can Axle Box Acceleration data be used in automatic classification of specific defect types at Insulated Rail Joints?” was investigated and answered. Moreover, five sub-questions were answered to reach the goal of using ABA data for automatic classification of specific defect types at IRJs.

6.1 Conclusion

In the previous chapter, the results of correlation and classification were described. In the next chapter, conclusions are given following the structure based on the 5 research questions and the main research question.

6.1.1 Conclusion cause defects

The first research question is answered which explains the phenomenon of battered head failure, lipping failure and loose fastening bolts and what causes these defects. All three defects are caused by large impact forces at IRJs. These large impact forces rise from the discrete change in vertical momentum due to a discontinuity of track stiffness because of the insulating material. In case of battered head failure the impact forces, in combination with a lack of support due to unfixed fishplates, causes the rail ends to flow downwards and become flattened. In case of lipping failure, on the other hand, a high-stress concentration at the edges of the railhead combined with the repetitive impact loads, causes the steel to flow inwards over the insulated end post. Especially during the summer months, additional thermal stresses cause the rails to expand thereby decreasing the gap between the railheads resulting in electrical failure. So impact forces combined with a lack of support cause battered head failure and impact forces combined with thermal stresses cause lipping failure. The biggest difference is related to the fixation of the fishplate and therefore joint stiffness. Besides that, both defects may occur at the same time. At last, loose fastening bolts initially start due to vibrations of the train. Then, a loss of connection drastically increases the IRJ degradation as additional high vibrations occur due to an increased dynamic force.

6.1.2 Conclusion theoretical influence

The second research question is answered which explains the theoretical influence of the defects on the dynamic response of ABA. The power within the frequency domain is proportional to the square of the amplitude of the acceleration signal. Since large impact forces result in large accelerations according to Newton’s second law and large accelerations result in vibrations with higher amplitudes, higher powers within the frequency domain were expected. Therefore, total relative power and peak relative power are interesting features to investigate. Besides that, peak frequency gives information in which frequency range the defects are active.

In case of battered head failure, in which the track discontinuity widens, increased impact forces were expected and thus increased powers in all frequency bands, especially around the high-frequency range of 500-600 Hz which is similar to permanent deformation. Lipping failure, on the other hand, was expected to contain less increased power in all frequency bands because only small extra vertical momentum is obtained. Besides that, it was expected that especially low-frequency ranges would gain increased powers because the wheel will follow the rail instead of subjecting impact forces as a result of the short track discontinuity. Loose fastening bolts were expected to influence the lower frequency ranges (f< 500 Hz) as the stiffness of the IRJ decreases.

Peak frequencies will change as a results of changing track stiffness. The elastic modulus, the moment of inertia and the beam length for example determine the track stiffness which can be explained through a simplified beam model. However, it is difficult to understand and predict the changing frequency behaviour due to battered head failure, lipping failure or loose fastening bolts because multiple parameters are influenced and changed and the proportion of each parameter is unknown.

6.1.3 Conclusion methods

The third research question is answered which explains the applicability of the Continuous Wavelet Transform. The CWT proved to be useful in the transformation from ABA in the time domain to the frequency domain. There was a need for time resolution to locate the IRJs within the ABA signal, but at the same time, frequency resolution was needed to investigate the power and frequencies. For this thesis, a medium frequency resolution was sufficient since large frequency bands were used, but in order to extract more precise information of the frequency response (with smaller frequency bandwidths), it is recommended to apply a higher frequency resolution. Furthermore, the health condition of IRJs was determined based on video footage. These high-resolution video images were useful in the determination of battered head failure, lipping failure and loose fastening bolts, however, some expertise is recommended. The determination of the health condition is difficult and subjective. Besides, the one year difference between ABA measurements and video footage did not contribute to accurate and reliable results.

6.1.4 Conclusion correlation

From the results, the following research question “What characteristic frequency responses can be found in the data that are related to battered head failure, lipping failure and loose fastening bolts and are these in line with the theoretical influence?” is answered and summarized below.

IRJs with battered head failure contained statistically significant higher power in frequency bands 25-40 Hz, 40-100 Hz and 200-600 Hz, among which frequency band 200-600 Hz contained the largest absolute and relative increase in total relative power and peak relative power. Besides that, no statistical significance was revealed for a change in peak frequency due to battered head failure. The increased powers are in line with the theoretical influence as the largest increase was expected in frequency band 200-600 Hz.

IRJs with lipping failure only contained statistically significant higher power in frequency band 40-100 Hz. This frequency band also showed the largest relative increase in total relative power and peak relative power. This is in accordance with the theory from which it was expected to find differences in the lower frequencies ($f<100$ Hz) as lipping failure does hardly affect the impact force. Therefore ABA captures relatively high powers in low frequencies compared to the powers in high frequencies. Moreover, a statistically significant decrease in peak frequency was found in frequency band 650-1250 Hz which could be related to a change in stiffness due to e.g. a change in width of the insulation layer.

In addition, it was not possible to statistically distinguish IRJs with battered head failure from IRJs with lipping failure. However, considering a combination of two frequency bands could be beneficial for distinguishing the two defects since clusters of IRJs with battered head failure and lipping failure were visible within two combined frequency bands. The latter observation is however not statistically tested in this study.

Loose fastening bolts could not be investigated in this thesis using the available data. A future study must be performed in which a larger dataset is investigated.

6.1.5 Conclusion classification

From the results the following research question “How, and with what accuracy, can the obtained relations be used in classification of IRJs?” is answered and summarized below.

The PCA was useful as a dimension reduction tool before the application of a machine learning algorithm because more than 85% of the variance within the data was kept which is beneficial for classification purposes. However, the found classifiers (as a result of the SVM) for both IRJs with battered head failure and IRJs with lipping failure were unstable using the available dataset and therefore unreliable. Low performance scores were obtained as well as bad validation test results. More data is needed to perform reliable classification without overfitting the algorithm.

6.1.6 Conclusion main research

Overall, this thesis found characteristic frequency bands related to battered head failure and lipping failure as well as insights into the applied methodology and its possibilities for predictive maintenance. The main research question “How can Axle Box Acceleration data be used in automatic classification of specific defect types at Insulated Rail Joints?” is answered and concluded below.

Application of ABA data in the health condition classification of IRJs can be achieved by means of frequency characteristics of the ABA response. Statistically significant increased powers in frequency band 200-600 Hz are correlated to IRJs with battered head failure. Moreover, statistically significant increased powers in frequency band 40-100 Hz are correlated to IRJs with lipping failure. The features total relative power, peak relative power and peak frequency can therefore be useful in the automatic classification of defect types which answers the first part of the main research question.

The applicability of PCA and SVM as a classifier could be useful in the automatic classification of the health condition of IRJs which answers the second part of the main research question. However, due to a limited dataset, the results shown in this thesis are uncertain and overestimated. Therefore, the results should be interpreted carefully.

A large remark should be mentioned that the data used in this thesis is not accurate because of two reasons. Firstly, only little data was obtained and secondly, a one year difference existed between the ABA measurements and video footage used for the determination of the health condition. For this reason, a validation study is recommended.

6.2 Discussion

In the discussion, the results and conclusions that are important and influence the results of this research are discussed. The same structure as for the conclusions is followed in the discussion.

6.2.1 Discussion cause defects

It is trivial that the discontinuity at an IRJ causes large impact forces which are detrimental to both the track and train wheels. However, both defects (battered head failure and lipping failure) are different failure modes. More factors are influencing the failure modes. Battered head failure is related to weak fishplates, but it can be discussed which factors influence the fishplate. Possibly, factors occur that are similar for multiple defects. It could be possible that the two failure modes are related to each other like e.g. battered head failure and permanent deformation could be related to each other because the increased gap and impact force result in a jump of the wheel, causing a second impact force slightly further causing permanent deformation. However, since battered head failure and lipping failure are completely different failure modes, such similarity (apart from the impact force) is illogical. Besides, the design of a gap in between the rails is not favourable from a maintenance perspective but is functional for the signalling system. However, other detection systems like the axle counter fulfil a similar function, but without causing impact forces.

6.2.2 Discussion theoretical influence

It is difficult to understand the theoretical influence of specific defects on the dynamic response of ABA. Whereas it can be imagined that IRJs with battered head failure contain higher powers in the whole frequency range, it is more difficult to imagine for IRJs with lipping failure. Moreover, the changing peak frequencies are even more difficult to understand. A separate study could be done alongside a practical ABA investigation that focuses on the theory of one defect type. However, the biggest challenge is that defects do not occur on their own, but are a combination of multiple defects. It can be discussed whether ABA measures the defect or the cause of the defect.

In case of battered head failure and loose fastening bolts, large impact forces might damage the subgrade and ballast for example, resulting in a change in stiffness and therefore frequency, in frequencies lower than 500 Hz. However, this implies that ABA measures the cause (degraded subgrade or ballast) rather than the effect (defect). It can be discussed whether the cause and effect are always related to each other. It could be possible that one cause has multiple effects. Therefore, it could be possible that ABA has measured other (unknown) defects as well. However, from a maintenance perspective, gaining information on the cause of defects could also be valuable so that the cause can be solved rather than solving the effect.

6.2.3 Discussion methods

Given the processing steps from the methodology, in each step improvements could be implemented to carry out a more accurate and reliable research. In the data selection, incomplete data was found which resulted in a limited data set. A low GPS frequency rate and inaccurate synchronization resulted in a large error within the IRJ localisation. The CWT was executed for a moderate frequency resolution because of limited computing power. A higher frequency resolution could be beneficial for the determination of the peak frequency (no disadvantages as explained in ***Appendix Z***) and to decrease the frequency bandwidths (to increase accuracy). Also, frequency bands could be chosen differently which could match better to the frequency characteristics. To obtain reliable results, a lot of averaging is needed. In the case of this study, averaging led to a further reduction of the amount of data.

The health condition of IRJs was assessed based on one year later video footage (compared to the time of measurement), which entailed a lot of uncertainty. Regarding the defect assessment, in case of battered head failure for example, it was difficult to assess an IRJ with a defect score of 1. Due to the one year difference, the IRJ could be either healthy (score 0) or severely damaged (score 2). Moreover, all other assumptions which are described in ***Appendix Q*** can be discussed whether the assumptions are valid or not. The last remark on the defect assessment is that it was a rather subjective process. One person would assess the IRJ differently from another person. Having one person doing all the assessments results in the advantage of all IRJs being assessed in a similar stepwise approach. However, the disadvantage is that this assessment is subjective and most likely non-repeatable.

6.2.4 Discussion correlation

The increased powers for battered head failure and lipping failure were in accordance with the expectations from the theory. But, in case of healthy IRJs, the IRJ could be either new or healthy. New IRJs are often subjected to low powers assuming the replacement has been executed perfectly fine. Healthy IRJs on the other hand are slightly degraded and will therefore contain increased powers compared to new IRJs. For a contractor, it is important to determine which IRJs are still healthy and which IRJs are degraded too much and must be maintained. To find this balance, the healthy IRJs which are assessed with all 0's are therefore more representative in practice than new IRJs.

Regarding the peak frequency, it could be influenced by many other components. For example, the predefined bandwidths determine the boundaries of the peak frequency. However, if a certain bandwidth was smaller or divided into two parts, other peak frequencies could be found. Moreover, other peak frequencies could be found if a higher frequency resolution would have been applied. The deviations in peak frequency might be related to track components rather than the defect itself as is discussed in 6.2.1 and 6.2.2. The sleepers influence the low-frequency range ($f<500$ Hz) and sleeper distance influence the high-frequency range ($f>500$ Hz). It could be possible that the increased powers and changing peak frequencies are related to the different types of IRJs (ETS joint and NRG joint) with different sleeper materials (wood and concrete) and different sleeper spacing rather than the defect itself. Besides, the speed possibly influences the lower frequencies because frequency is related to speed and wavelength. However, Molodova (2013) concluded that velocity does not influence frequency nor wavelength, which contradicts to beforementioned relation. It can be discussed whether the train speed influences the frequency.

At last, due to the hunting motion of the wheelset and therefore a changing contact area, one IRJ could have been measured differently from the other. This oscillation might have influenced the amount of power that should be taken into consideration. Therefore, averaging signals is important to disregard the changing hunting behaviour.

6.2.5 Discussion classification

Unfortunately, the one year difference in ABA data and video footage for the determination of IRJ health condition resulted in a lot of uncertainty. One IRJ which is assessed wrongly could influence the classifier a lot. Besides, approximately 50% of the healthy IRJs were new and therefore very healthy which is not beneficial for classification. In practice, IRJs will be degraded more, but still be healthy. Therefore, more data and a better defect assessment are needed to obtain higher reliability.

Regarding the PCA, it could have been better to fit the PCA on the whole dataset, rather than fitting the PCA to only the training dataset. More information would have been included in the principal

components increasing the reliability. However, too much information of the test data is given which decreases the reliability of validation. It is recommended to have at least 75-100 samples containing an equal distribution of healthy and degraded IRJs (Beleites et al., 2013). Important is that the defect assessment is performed well to obtain accurate results.

6.2.6 Discussion main research

Overall, it can be discussed whether the results are accurate because of limited data and a one year difference between ABA measurements and video footage for the health condition assessment. The chosen frequency bands could be selected differently which may or may not be more beneficial. Besides, the feature peak frequency might not have added value to the classification because of the drawback described in ***Appendix Z***. Moreover, one might discuss whether the selected features (total relative power, peak relative power and peak frequency) are well enough for classification purposes. Since the methodology shows potentiality, a future study with a large and accurate dataset is recommended to validate the results found in this research.

Although this study specifically does not add value to contractors, the possibilities of ABA in predictive maintenance concepts are still potential. From literature, it is known that automatic detection of squats is already possible and therefore automatic detection of IRJ defects should also be feasible.

7 Recommendations and future prospects

There were a lot of limitations within this study which all came down to limited data and inaccurate validation. Therefore, in the recommendations, it is described what should be improved to obtain reliable results from the applied methodology. Moreover, a future prospect is given in which it is described what steps are needed before ABA can be used in predictive maintenance.

7.1 Recommendations

As explained above, most limitations of this study came down to limited data and inaccurate validation. Therefore, a large dataset is necessary to increase reliability. This means the trains with accelerometers should measure the whole track multiple times. Because accelerations measured on different trains might deviate from each other it is recommended to use only one train which measures repetitively, approximately 5 to 10 times. The repetitive measurements can be used to average the signal to erase abnormal measurements and therefore to increase reliability. For a reasonable study, it is recommended to have at least 75-100 samples per class (Beleites et al., 2013). A sample size with 5-25 independent samples achieve acceptable performances but could be completely random with high uncertainty and overestimation.

Furthermore, it remains a challenge to perfectly detect each IRJ within the data. It is recommended to use a better GPS (DGPS for example) with a measuring accuracy of approximately 1 m. However, it is still important that the synchronisation works well which ensures the exact location of each data point. In the ideal scenario, a match of coordinates of IRJs and ABA should be sufficient to detect IRJs. This way, no intermediate localisation steps are needed such as SAWP response and mutual spacing.

Once the ABA signal is obtained at the right position of the IRJ, it is important to apply a wavelet with a high frequency resolution. A higher frequency resolution would describe the peak frequency in a more precise way which could be an interesting characteristic of the health condition of IRJs. A higher frequency resolution would also improve the accuracy of the measured powers of the frequencies. Besides, smaller frequency bands could be chosen to capture more and more accurate information. As is discussed before, the frequencies might shift due to a change in stiffness for example, which should be kept within the chosen frequency bands.

Another important improvement concerns the defect assessment. There should not be a big difference between the time of measuring and the time of assessing the health condition of IRJs. The degradation of IRJs is a rather slow process, but to increase reliability it is important to assess the health condition no later than 1 after measuring. It should be prevented that intermediate maintenance actions change the health condition of IRJs. It is also advised to assess the health condition of IRJs based on inspections by multiple experts. As the assessment is rather subjective, one must discuss the health condition with each other.

A validation study is recommended to verify the results found in this thesis. Furthermore, future studies should be performed to investigate whether the found correlation translates to a qualitative or even quantitative relation.

7.2 Future prospects

In this research, correlations were studied and key frequency bands are described. These frequency bands are of importance for the automatic classification of the health condition of IRJs. Currently, IRJ maintenance is either corrective or preventive maintenance. Yearly visual inspections determine the need for maintenance. These visual inspections might be replaced by data, which predicts the health condition of IRJs. It is expected that, generally, data will help in or even replace the decision making of maintenance. Besides, ABA is a cheap way of gaining a continuous flow of data. Moreover, relatively simple algorithms should be able to help to assess the health condition of IRJs.

Before practical application, future studies must be performed which verify the results found in this thesis. It is recommended to have at least 75-100 healthy samples and 75-100 degraded samples per class. Besides, other defects should be considered and all results (of different defects) should be combined. Additionally, the features total power, peak power and peak frequency should be determined (among extra new features) which can be applied in classification. It should be investigated that the found frequency characteristics are specific for the defect rather than multiple other defects. Apart from the theory, a machine learning algorithm might distinguish healthy IRJs from degraded IRJs. However, to find quantitative relations, the correlations should be clarified through theory.

In the short term, ABA might assist in visual inspections. It is expected that using frequency characteristics of IRJs (obtained from ABA) enable an automatic distinction in healthy IRJs and degraded IRJs. The healthy IRJs would therefore not be visually inspected anymore. Only the degraded IRJs need a further inspection in order to determine the precise health condition and the subsequent maintenance actions. Such application of ABA is already favourable in terms of cost efficiency, but also in terms of safety.

In the long term, degraded IRJs might be automatically classified by their defect type. It is interesting to know from the data what the health condition of IRJs is and which specific defects are present. Using the information of the present defect, maintenance procedures might be anticipated. In case of light battered head failure, grinding maintenance could remove the defect. In case of severe battered head failure, the IRJ needs to be completely replaced with high maintenance costs as a result. Therefore, it would be beneficial to detect battered head failure in the early stage. In case of lipping failure, a risk of signal failure occurs which could be simply prevented by immediate maintenance actions in which the insulating end post is grinded. It is therefore important to detect lipping failure before occurrences of signal failure.

A change to predictive maintenance would benefit the current maintenance procedure. Corrective maintenance could be prevented which reduces the high additional penalty costs. Moreover, from a cost-savings perspective, time-consuming and periodically scheduled maintenance could be optimized as visually inspections will be taken over by automatic defect detection algorithms.

References

AMEC Rail, AMEY Rail, Balfour Beatty Rail Maintenance Ltd, Balfour Beatty Rail Renewalds Ltd, Centrac, First Engineering, Grant Rail, GT Railway Maintenance Ltd, Jarvis Rail, & Serco Rail Maintenance Ltd. (2001). *Insulated Rail Joints (IRJs) Civil Engineering*. Civil Engineering Conference.

Asiri, S. (2018). *Machine Learning Classifiers*. Towards Data Science. https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623

Auria, L., & Moro, R. A. (2008). *Support Vector Machines (SVM) as a technique for solvency analysis*. Berlin: German University for Economic Research (DIW).

Beleites, C., Neugebauer, U., Bocklitz, T., Krafft, C., & Popp, J. (2013). Sample size planning for classification models. *Analytica Chimica Acta*, *760*, 25–33. https://doi.org/10.1016/j.aca.2012.11.007

Brugman, A. (n.d.). *Hoe bouw je een beeld op met Fourier-transformatie?* Retrieved April 1, 2021, from https://www.natuurkunde.nl/artikelen/3511/hoe-bouw-je-een-beeld-op-met-fouriertransformatie

Cohen, M. X. (2019). A better way to define and describe Morlet wavelets for time-frequency analysis. *Neurolmage*, *199*, 81–86. https://doi.org/10.1101/397182

Dangre, H. (2019). A Review On Insulated Rail Joints (IRJ) Failure. *International Journal of Advanced Research and Publications*, *3*(1), 5–9.

Degen, J. B. H., Rietberg, R. R., Jolink, R. G. J., & Van Nimwegen, C. D. (2019). *Werkmethodiek: Inspectie & Onderhoud ES-las*. Unpublished internal company document.

Devleker, K. (2016). *Understanding Wavelets - MATLAB*. Matlab. https://nl.mathworks.com/videos/series/understanding-wavelets-121287.html

Dhanasekar, M. (2015). *Best Practice Manual for Insulated Rail Joints*. *October 2013*. https://doi.org/10.13140/RG.2.1.3770.0569

Esveld, C. (2001). Modern Railway Track, second edition. In *Modern railway track* (pp. 210–212). MRT productions.

ETS spoor. (2021). *Elektrische Scheidingslassen (Lijmlassen) - ETS Spoor*. https://www.etsspoor.nl/producten/bovenbouwmaterialen/lijmlassen/

Gandhi, R. (2018). *Support Vector Machine — Introduction to Machine Learning Algorithms*. Towards Data Science. https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

Glen, S. (2021a). *T Test (Student's T-Test): Definition and Examples*. StatisticsHowTo.com. https://www.statisticshowto.com/probability-and-statistics/t-test/

Glen, S. (2021b). *Welch's Test for Unequal Variances*. StatisticsHowTo.com. https://www.statisticshowto.com/welchs-test-for-unequal-variances/

Grassie, S. L. (1996a). Measurement of railhead longitudinal profiles: A comparison of different techniques. *Wear*, *191*(1–2), 245–251. https://doi.org/10.1016/0043-1648(95)06732-9

Grassie, S. L. (1996b). Short wavelength rail corrugation: Field trials and measuring technology. *Wear*, *191*(1–2), 149–160. https://doi.org/10.1016/0043-1648(95)06755-8

Indian Railways Institute Of Civil Engineering. (n.d.-a). *Chapter II The Maintenance Of Permanent Way*. Retrieved May 3, 2021, from https://indianrailways.gov.in/railwayboard/uploads/codesmanual/IRPWM/PermanentWayManualCh2_data.htm

Indian Railways Institute Of Civil Engineering. (n.d.-b). *Glued Insulated Joint*. Retrieved March 25, 2021, from http://www.iricen.gov.in/ModelRoom/G1_Glued joint.html

Jenkins, H. H., Stephenson, J. E., Clayton, G. A., Morland, G. W., & Lyon, D. (1974). The effect of track and vehicle parameters on wheel/rail vertical dynamic forces. *Railway Engineering Journal*, *3*(1).

Johnson, K. L. (1985). *Contact mechanics* (Ninth edit). Cambridge University Press.

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. In *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* (Vol. 374, Issue 2065). Royal Society of London. https://doi.org/10.1098/rsta.2015.0202

Keshari, K. (2018). *PySpark Programming - Integrating Speed With Simplicity*. Edureka. https://medium.com/edureka/pyspark-programming-e007e68fbccb

Lachowicz, M. B., & Lachowicz, M. M. (2021). Influence of corrosion on fatigue of the fastening bolts. *Materials*, *14*(6), 1–17. https://doi.org/10.3390/ma14061485

Lee, G. R. (2018). *PyWavelets Documentation Release 1.0.1*.

Mandal, N. K. N., & Peach, B. (2010). An Engineering Analysis of Insulated Rail Joints: A General Perspective. *International Journal of Engineering Science and Technology*, *2*(8), 3964–3988.

Massel, A. (1999). Power spectrum analysis - Modern tool in the study of rail surface corrugations. *NDT and E International*, *32*(8), 429–436. https://doi.org/10.1016/S0963-8695(99)00020-1

Minitab. (2019). *Interpret the key results for Principal Components Analysis*. https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/principal-components/interpret-the-results/key-results/

Moebs, W., Ling, S. J., & Sanny, J. (2016). *University Physics Volume 1*. OpenStax.

Molodova, M. (2013). *Detection of early squats by axle box acceleration*. Doctoral thesis, Delft University of Technology.

Molodova, Maria, Li, Z., & Dollevoet, R. (2011). Axle box acceleration: Measurement and simulation for detection of short track defects. *Wear*, *271*(1–2), 349–356. https://doi.org/10.1016/j.wear.2010.10.003

Molodova, Maria, Li, Z., Nunez, A., & Dollevoet, R. (2014). Automatic detection of squats in railway infrastructure. *IEEE Transactions on Intelligent Transportation Systems*, *15*(5), 1980–1990. https://doi.org/10.1109/TITS.2014.2307955

Molodova, Maria, Oregui, M., Núñez, A., Li, Z., & Dollevoet, R. (2016). Health condition monitoring of insulated joints based on axle box acceleration measurements. *Engineering Structures*, *123*, 225–235. https://doi.org/10.1016/j.engstruct.2016.05.018

Molodova, Maria, Oregui, M., Nunez, A., Li, Z., Moraal, J., & Dollevoet, R. (2014). Axle box acceleration for health monitoring of insulated joints: A case study in the Netherlands. *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, 822–827. https://doi.org/10.1109/ITSC.2014.6957791

Nicoll, A. (2020). *The Wavelet Transform for Beginners - YouTube*. https://www.youtube.com/watch?v=kuuUaqAjeoA&ab_channel=AndrewNicoll

Nordco. (n.d.). *Nordco Rail Flaw Defects Identification Handbook*. Retrieved May 3, 2021, from https://www.nordco.com/Media/Assets/General-Files/NordcoRailFlawDefectsIdentificationHandbook.pdf

Noria. (2020). Predictive Maintenance Explained. *Reliable Plant*.

Nunez, A., Hendriks, J., Li, Z., De Schutter, B., & Dollevoet, R. (2014). Facilitating maintenance decisions on the Dutch railways using big data: The ABA case study. *Proceedings - 2014 IEEE International Conference on Big Data, IEEE Big Data 2014*, *19*, 48–53. https://doi.org/10.1109/BigData.2014.7004431

Oregui, M., Molodova, M., Núñez, A., Dollevoet, R., & Li, Z. (2015). Experimental Investigation Into the Condition of Insulated Rail Joints by Impact Excitation. *Experimental Mechanics*, *55*(9), 1597–1612. https://doi.org/10.1007/s11340-015-0048-7

Pechenizkiy, M., Tsymbal, A., & Puuronen, S. (2004). *PCA-based Feature Transformation for Classification: Issues in Medical Diagnostics*.

Powell, V. (n.d.). *Principal Component Analysis explained visually*. Setosa.Io. Retrieved June 9, 2021, from https://setosa.io/ev/principal-component-analysis/

ProRail. (2018). *Sensoren op reizigerstreinen | ProRail*. ProRail. https://www.prorail.nl/nieuws/sensoren-op-reizigerstreinen

ProRail. (2020). *Camino data guide*. Unpublished internal company document.

Rout, A. R. (2020). *Advantages and Disadvantages of Logistic Regression*. Geeks for Geeks. https://www.geeksforgeeks.org/advantages-and-disadvantages-of-logistic-regression/

Ruijters, E., Guck, D., Noort, M. Van, & Stoelinga, M. (2016). Reliability-centered maintenance of the electrically insulated railway joint via fault tree analysis: A practical experience report. *Proceedings - 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN 2016*, 662–669. https://doi.org/10.1109/DSN.2016.67

SAS. (2021). *Machine Learning: What it is and why it matters*. SAS. https://www.sas.com/en_us/insights/analytics/machine-learning.html

Shickel, D. (2020). *Reactive Vs. Preventive Vs. Predictive Maintenance | Prometheus Group*. https://www.prometheusgroup.com/posts/reactive-vs-preventive-vs-predictive-maintenance

Spijkers, J. M. J., Vrouwenvelder, A. W. C. M., & Klaver, E. C. (2005). *Structural Dynamics CT 4140*.

Steenbergen, M. J. M. M. (2008). *Wheel-rail interaction at short-wave irregularities*. Doctoral thesis, Delft University of Technology.

Tadel, F., Pantazis, D., Bock, E., & Baillet, S. (2019). *Time Frequency - Brainstorm*. Brainstorm. https://neuroimage.usc.edu/brainstorm/Tutorials/TimeFrequency

Talebi, S. (2020). *The Wavelet Transform. An Introduction and Example*. Towards Data Science. https://towardsdatascience.com/the-wavelet-transform-e9cfa85d7b34

Taspinar, A. (2018). *A guide for using the Wavelet Transform in Machine Learning*. https://ataspinar.com/

van den Heuvel, J., Lamper, A., Smulders, J., Fischer, R., & van Asten, A. (2012). *Handboek SAM - Handleiding SysteemAssetManagement*. ProRail.

Voestalpine Railpro BV. (2020). *Plaatsings-, onderhouds- en sloop-voorschrift energy-las (nrg-las)*.

Voestalpine Railpro BV. (2021). *Next generation high performance insulated rail joint*. https://www.voestalpine.com/track-solutions-netherlands/en/products/Permanent-way/Insulated-Rail-Joint/

Appendices

Appendix A | Types of IRJs

There are two types of IRJs: glued and construction IRJs. Both joints are built up the same way, except for the type of insulation material. For construction IRJs this is a plastic plate whereas for glued IRJs a plastic glue is used. Both materials fulfil the electrical insulation function. However, considering the mechanical function, the construction IRJ lacks the absorption of high compression and tension forces. The glued IRJ on the other hand creates a better homogeneous joint as a whole, resulting in the ability of absorption of those forces as well as bending and fatigue (Esveld, 2001). Therefore, the glued IRJs are the new standard joints imposed by the Dutch railway infrastructure manager ProRail.

There are two common glued IRJs in practice, the ETS joint and the eNeRGy (NRG) joint.

ETS joint

The ETS joint was the standard joint imposed by ProRail. This type of joint is pre-isolated and glued with fast hardening two-component glue (ETS spoor, 2021). As can be seen in the left image of ***Figure A1,*** the joint is supported by double wooden sleepers for additional support. The fastening system consists of a plastic fitting cap mounted by a bolt and washer.

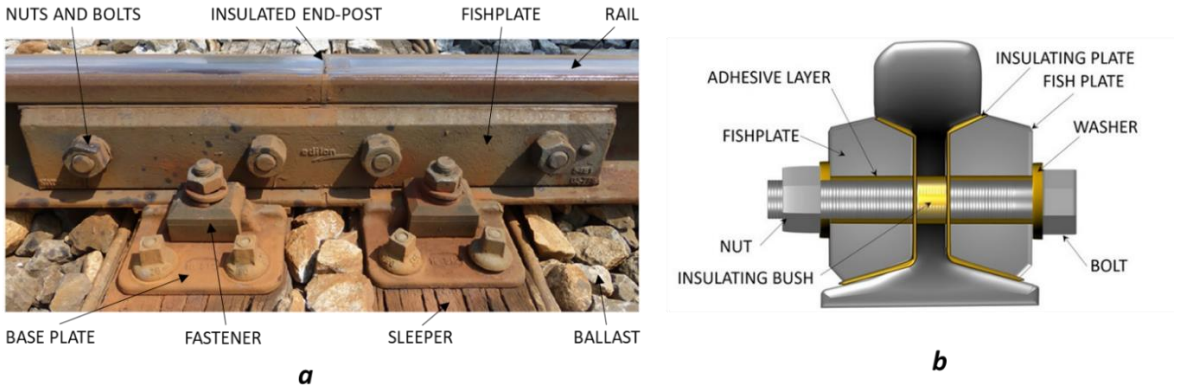


Figure A1 |(a) *Side view (Oregui et al., 2015) and* ***(b)*** *cross-section of the Insulated Rail Joint* (Indian Railways Institute Of Civil Engineering, n.d.-b)

NRG joint

Since 2016 is the eNeRGy-joint is the new standard imposed by ProRail (Voestalpine Railpro BV, 2021). Since IRJs were a frequent cause of disruption, ProRail searched for a better and more durable joint that was less sensitive to disruptions. Voestalpine Railpro came with the solution with the NRG joint, which moves along with the rails to reduce the dynamic impact energies. Also, the NRG joint is more resistant to different climate circumstances (Voestalpine Railpro BV, 2020). ***Figure A2*** shows an NRG joint in which the floating IRJ is clearly visible as well as the tension spring clamps. Moreover, the joint is mainly supported by concrete sleepers compared to the ETS joint which is supported by wooden sleepers.



Figure A2 | *eNeRGy-joint* (Voestalpine Railpro BV, 2021) .

The differences between the two types of joints described are summarized in ***Table A1***.

Table A1 | *Overview differences ETS and NRG joint.*

ETS joint	**NRG joint**
4 bolts	6 bolts
Bolt clamp fastener	Tension (spring) clamp fastener
Supported by double wooden sleepers	Floating joint supported by concrete sleepers

Appendix B | IRJ failure modes and video footage

Due to the large impact forces, an IRJ can fail in several ways. Failure of an IRJ can be categorized as mechanical failure and electrical failure. All failure types are to a certain extent caused by the wheel-rail impact load. ***Table B1*** summarizes some important failure modes of the IRJ which will be explained shortly. In case of mechanical failure, the fishplate or bolts could crack and break apart, resulting in a deviation in joint stiffness. Even loose bolts could cause changes in joint stiffness. Also, insulation failure might occur resulting in cracked or peeled away end posts from the joints bars and rail, which again weakens the IRJ. Battered head failure is a result of high impact forces and a small joint settlement that causes the rail end to flow downwards and become flattened (Dangre, 2019). A battered head mostly comes with arc damage, as the large impact force causes arcs that cover the railhead. Plastic deformation is caused due to the impact force after the gap, at a wavelength of approximately 40-50 mm. The wavelength is typical for permanent deformation as the vibrations at 540 Hz is common (Molodova, Oregui, et al., 2014). A speed of 100 km/h (27 m/s), for example, results in a theoretical wave pattern of 50 mm (27[m/s] / 540 [Hz] = 0.050 [m]).

Electrical failure is always related to a lack of insulation, which results in a current flow from one rail to the other. A common electrical failure mode is lipping, which is the flow of steel due to fatigue. A repetitive impact force weakens the steel which results in a flow of steel in the direction of the gap with a physical connection as a result. Furthermore, such failure might occur when the railhead is crushed and the crushed steel particles form a connection at the end post. Moreover, crushed end posts may result in empty spaces which can be filled up with water. In that case, the IRJ is not able to insulate the rails and a flow of current is possible. In addition, a shift of sleepers or internal insulation failure are typical causes of electrical failure as the steel parts of the bolts and fishplate might connect to the rail (Dangre, 2019).

Table B1 | *Failure modes IRJ (Ruijters et al., 2016 & Dangre, 2019).*

Mechanical failure mode	**Electrical failure mode**
Broken fishplate or bolts	Lipping: flow of steel
Battered head and arc damage	Crushed end post
Permanent deformation	Sleeper shift
Insulation failure (cracks, peeling away)	Internal insulation failure

Appendix C | IRJ maintenance

Current IRJ maintenance consists of a combination of corrective and preventive maintenance, as is summarized in ***Figure C1***. In case of signal failure, there is a high priority to solve the problem immediately to minimize the penalty costs. This is an example of corrective maintenance. Periodical inspections are performed to decrease the risk of immediate signal failures, which is a form of preventive maintenance. To ensure the functionality of an IRJ, a contractor is obligated to follow the rules prescribed by ProRail. Every section of the railway must be minimally inspected once a year. Depending on the importance, the age and the condition of the track, the number of inspections must be increased. During the inspection, a minimum of two people is needed for safety reasons. The mechanic is in charge of the maintenance whereas the security guard oversees the safety (Degen et al., 2019).

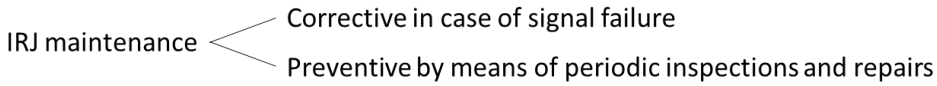


Figure C1 | *IRJ maintenance.*

During an inspection, the construction and geometry are analysed by checking the fastening system, defect types and contamination among others. Based on the observations, a decision is made to perform maintenance right away or to plan it for later. Rail maintenance is dangerous as it is not only performed during the night but also during the day while trains are operational.

Contractors aim to avoid corrective maintenance as it involves high penalty costs by applying preventive inspections. However, from a cost-savings perspective, time-consuming and periodically scheduled maintenance is not optimal. A continuous stream of data that monitors the IRJs might be beneficial in reducing the number of required inspections. There is no need to check healthy IRJs that are constantly monitored, resulting in time savings. A shift to predictive maintenance would benefit the current maintenance procedure. The ABA measurements might be used to achieve this.

Appendix D | Axle Box Acceleration and Camino project

The Axle Box Acceleration (ABA) is a measurement system that measures vertical, lateral and longitudinal accelerations over time. The accelerometers are mounted to both ends of the two axles of a bogie. Next to the accelerometers, the position and the velocity of the train are recorded using GPS. Experimenting with ABA to measure the rail profile has been done from the early 70s in which Grassie (1996) mounted an accelerometer to a trolley. Later studies by Grassie (1996b) verified the advantage of measuring at higher speeds. More recently, researchers at the Delft University of Technology studied the applicability of ABA to detect surface defects such as squats (Molodova et al., 2011). ProRail closely followed the development of ABA at the TU Delft and engaged an alliance in 2018 with the TU Delft, TNO and college infrastructure manager Next Generation Infrastructures in a project called the Camino project (ProRail, 2018). ProRail's goal is to apply innovative solutions to ensure the railway's infrastructure safety, reliability, sustainability and to be cost-efficient with the purpose to prevent malfunctions and delays (ProRail, 2018). Accelerometers were mounted to the axle box of an SLT type (sprinter) and VIRM train, to measure the real-time balance of the train, as well as the switches and the quality of the rails themselves. The trains operate on the normal daily schedule to ensure a continuous flow of data. It is expected that problems like track settlement, damage to switches and rail failures can be measured premature (ProRail, 2018).

The accelerations are measured with a sampling frequency (f_s) of 6250 Hz, resulting in the possibility to investigate frequencies up to 3125 Hz according to the Nyquist frequency ($f_N = f_s/2$). Frequencies above the Nyquist frequency (f > 3125 Hz) will be aliased back to frequencies between 0 and 3000 Hz resulting in improper frequencies that are not present. The sampling frequency of the GPS is only 1 Hz, which could result in inaccurate results with deviations up to 25 meters (ProRail, 2020). Due to the large amount of data ProRail was not able to share all the data collected by the trains. Therefore, the decision was made to share parts of the data based on two threshold values, a lower and upper threshold. The train trip was divided into batches having a duration of 2 minutes. If any measurement exceeded the upper threshold within a batch, the complete 2-minute batch was stored and shared (ProRail, 2020). These 2-minute batches are very suitable for the investigation of defects in its frequency domain. Moreover, during the 2-minute train ride a distance of approximately 4 km is travelled, including about 10 healthy and/or damaged IRJs (120 [km/h] = 33 [m/s] = 4000 [m/120 sec]). Additionally, if accelerations exceeded the lower threshold, the individual measurement was saved. However, for this study's purposes (investigating frequency responses of time signals) these individual measurements are not useful and therefore only the batch measurements were analysed. ***Figure D1*** gives a schematic example of the shared data, divided into batches and shared if exceeding the thresholds.

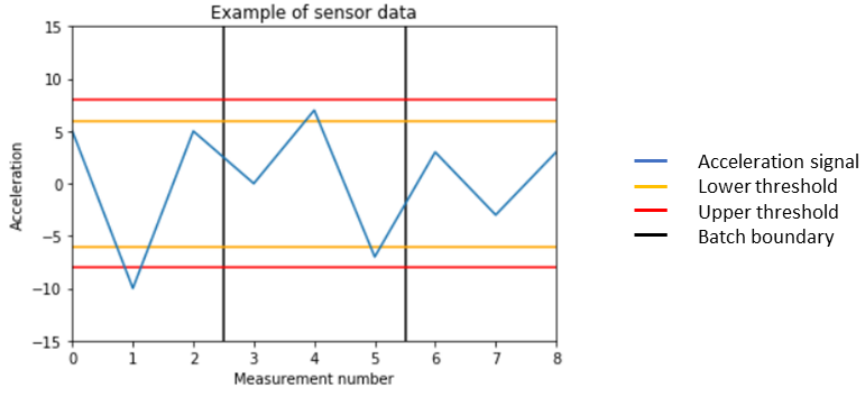


Figure D1 | *Schematic example of shared data from Camino project* (ProRail, 2020). *The measurements are divided into batches of two minutes. For the Camino project it is measured with a sampling frequency of 6250 Hz resulting in 750.000 measurements per batch (6250[Hz]*120[s]).*

Appendix E | Time-frequency transformation

The objective of this thesis is to relate IRJ defects to characteristics in the frequency response. To achieve this, a time-frequency transformation should be applied. The most common time-frequency transform is the Fourier Transform (FT). Applying the FT to a time-domain signal transforms the signal to the frequency domain. A peak in the time domain indicates the amplitude of a signal and the width of the peak the period and frequency of the signal. Peaks in the frequency domain indicate the presence of waves with specific frequencies in the signal.

A (single) signal mostly consists of multiple frequencies which can be decomposed into multiple sine waves with each having a single frequency. By applying the time-frequency transform a signal can be investigated from a different perspective by getting information about the amplitude of the frequencies. ***Figure E1*** describes the decomposition of a signal into sinusoids by applying a time-frequency transformation method.

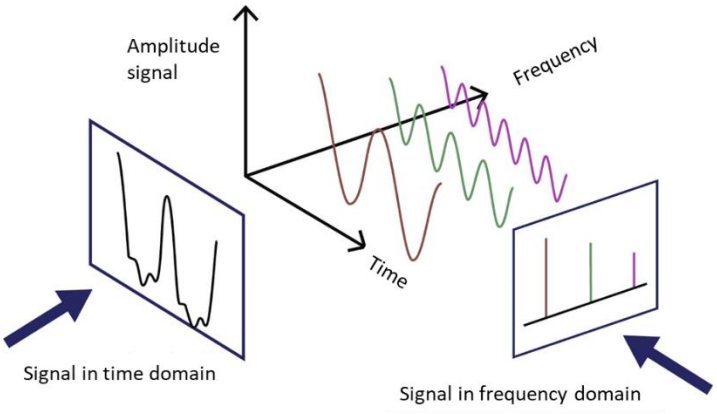


Figure E1 | *Signal decomposition by a time-frequency transform* (Brugman, n.d.).

There are many time-frequency techniques among which the FT is most popular. This method is perfectly suited for stationary signals, in which the signal's frequency does not change in time but remains constant over time. Since the FT is applied over the whole signal, time information will be lost. As ***Figure E1*** shows, the FT changes a time signal into the frequency domain, but at the expense of the time information. Therefore, a time-series signal has non-zero time resolution, but zero resolution on the frequency axis (see ***Figure E2a***). On the other hand, time-series signals transformed to the frequency domain using the FT have non-zero resolution in the frequency domain, but zero resolution in the time domain (see ***Figure E2b***). In conclusion, applying the FT over the whole signal fails to provide frequency information for a localized signal region in time (Nicoll, 2020).

Since ABA gives a non-stationary signal, there is a need for time information. Again, there are several ways to overcome this problem of zero time resolution among which the Short Time Fourier Transform (STFT) is most common. The STFT splits the signal into multiple parts with an equal length, assuming that some portion of the non-stationary signal is stationary (Nicoll, 2020). Only a small section of the signal at a time is analysed by a technique called windowing. By applying the FT over a short time window, both time and frequency resolution are obtained, but both at medium resolution. However, the disadvantage of an STFT is that the size of the window is unchangeable resulting in a fixed time and frequency resolution (see ***Figure E2c***). So, with the STFT it is not possible to know what frequencies exist at what time instance, but it is possible to know what frequency bands exist at what time intervals (Nicoll, 2020). There is a dilemma between choosing a narrow window resulting in a high time resolution but a poor frequency resolution, or the other way around, a wide window that results in a poor time resolution, but a good frequency resolution. This trade-off is known as the uncertainty principle (Taspinar, 2018). The output from time/frequency techniques is

a 3D Fourier spectrum plot with on the vertical axis the frequency, on the horizontal axis the time and on the z-axis the magnitude (amplitude).

As an alternative, the Wavelet Transform (WT) is introduced. The WT is able to divide the spectrum into differently shaped (but equally sized) windows, also known as multiresolution analysis (Nicoll, 2020). In case of low frequency values, a high frequency resolution can be obtained at the expense of time resolution. However, a lower time resolution is sufficient because there is little variation in time for low frequencies. Meanwhile, for high frequencies there is a lot of variation in time, resulting in the necessity of a high time resolution. In other words, and as Taspinar stated (2018), "the WT makes a trade-off; at windows in which time-dependent features are interesting it has a high resolution in the time domain and at windows in which frequency-dependent features are interesting it has a high resolution in the frequency domain" (Taspinar, 2018). ***Figure E2d*** shows the principle of WT in which the shape of the window is adjustable per application.

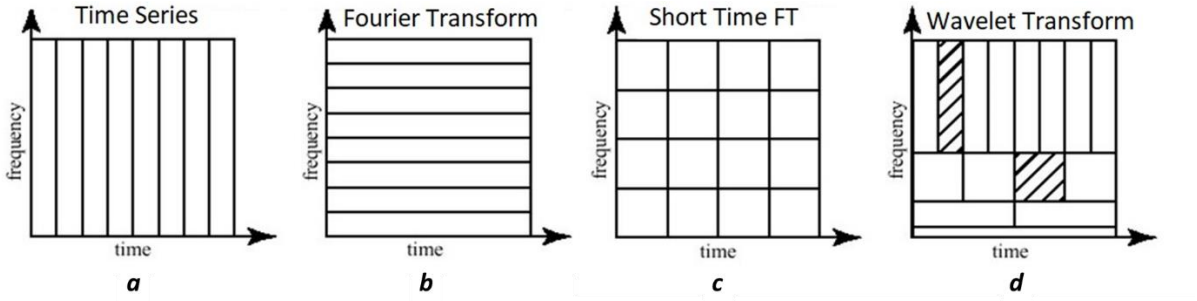


Figure E2d | *A schematic overview of the time and frequency resolutions of the different time-frequency transforms* (Taspinar, 2018).

A study of a time signal in the frequency domain is interesting to study its frequency characteristics. Every object on earth will vibrate due to an external force. According to the structural dynamics each object will vibrate with a specific frequency, also known as its eigenfrequency. By extracting frequency values from a time signal, information about objects might be obtained.

Appendix F | Continuous Wavelet Transformation

This study uses the Continuous Wavelet Transform (CWT) as the time-frequency tool to investigate the frequency characteristics because of the need for both time and frequency resolution. Since this method is not dependent on a fixed window size, high frequencies can be studied with a high time resolution and low frequencies with a high frequency resolution.

A wavelet is a "small wave-like oscillation that is localized in time" which is the basis function for the wavelet transform and acts as a window function (Talebi, 2020). Characteristic for wavelets is the short (mostly symmetrical) excitation with a zero amplitude left and right of the excitation. The localization of the wavelet in time allows the WT to obtain time information, whereas sines or cosines as used in FT are not localized in time and therefore result in loss of time resolution. In ***Figure F1*** the difference between a sine-wave and a wavelet over time is clearly visible. Whereas the sine-wave is infinitely long in time, the wavelet is localized in time (Taspinar, 2018). There are several types of wavelets, each with its own strength and weakness based on the trade-off between how compact and smooth a wavelet is and should be for its purpose. As long as the wavelet satisfies two mathematical conditions, the wavelet can be used for time-frequency analysis. First, a wavelet must contain finite energy. This means that "the wavelet is localized in time and frequency, integrable and the inner product between the wavelet and the signal always exists" (Taspinar, 2018). The second requirement for a wavelet is that the mean value of a wavelet is always zero to ensure that the wavelet is integrable and the inverse of the wavelet transform exists (Taspinar, 2018).

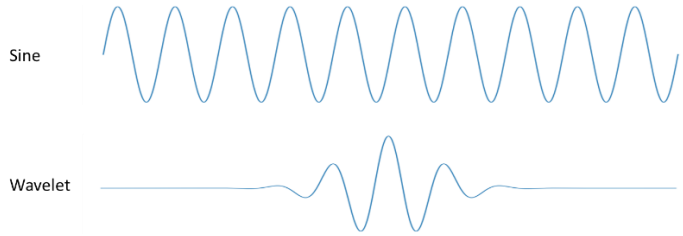


Figure F1 | *Difference between sine-wave used for an FT and wavelet used for a WT (Taspinar, 2018).*

The wavelet transform can be applied in two ways, via a discrete or continuous way. The Discrete Wavelet Transform (DWT) is mostly used for denoising and compression of signals. However, the CWT is better suited for time-frequency analysis of oscillatory signals as the way of discretizing the scale and translation parameters is done in a more precise way. The higher the number of scales, the finer the scale discretization and the higher the analysis accuracy (Devleker, 2016).

The CWT is described as follows (Nicoll, 2020):

$$W(\tau,s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} f(t)\psi^*\left(\frac{t-\tau}{s}\right) dt$$

In the CWT the integral is taken of the input signal f(t) which is multiplied by the complex conjugate (*) of function ψ(t), which is known as the mother wavelet. The wavelet consists of a family of wavelets deduced from the mother wavelet by a scaling (s) and shifting (τ) factor, hence $\frac{1}{\sqrt{|s|}}\psi\left(\frac{t-\tau}{s}\right)$. These are also called the child wavelets. As a result, the CWT calculates a set of wavelet coefficients (W(τ,s)) which depend on the parameters shift (τ) and scale (s). The shift is nothing more than the time at a different time instance. The wavelet scale is related to the inverse of frequency ($s=f_c/f$), in which f_c is a constant representing the center frequency of the wavelet. The CWT is similar to the FT,

but instead of multiplying the signal by an exponential function (sines or cosines), it is now multiplied with the family of wavelets (M. Molodova, 2013 & Nicoll, 2020).

The procedure of multiplying the signal with the shifted wavelet at different locations in time is called convolution (Taspinar, 2018). When the frequency of the wavelet ψ(t) matches part of the input signal f(t) with a very similar frequency, the output of wavelet coefficients W(τ,s) will be high. By repeating the convolution process for different scales, the wavelet coefficients matrix W(τ,s) is constructed and visualized in a time-scale graph, also called a scalogram.

The term scale (s) determines the accuracy of the scalogram. A scale vector is defined to obtain the desired width of the wavelet for each frequency. A stretched wavelet (large value of s) is able to resolve low-frequency components of the signal, but with a bad time resolution. This makes sense since the scale is the inverse of frequency. Also, a smaller (squished) wavelet (small value of s) is better able to resolve high frequency components of the signal at a good time resolution. By changing the scale of the window function, the WT is able to resolve high and low frequency components with both the desired frequency as time resolution (Nicoll, 2020). ***Figure F2*** shows the effect of decreasing (a) and increasing (b) the scale of a wavelet.

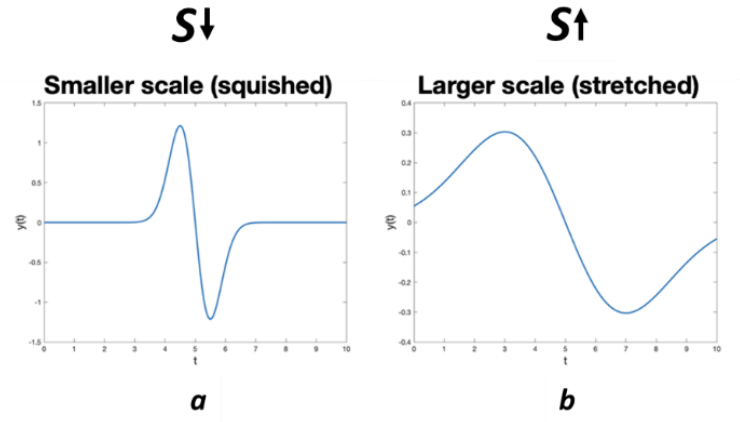


Figure F2 | *Effect of a* ***(a)*** *smaller scale and* ***(b)*** *larger scale to a wavelet* (Talebi, 2020).

Appendix G | Machine learning

One of the goals of this thesis is to classify rail defects based on their signature tunes and for that, a machine learning algorithm is used. As explained before, machine learning is a method within data analysis that automates analytical model building (SAS, 2021). Computers are able to learn from data, recognize patterns and make decisions automatically based on an algorithm.

The learning process of machines can be either supervised or unsupervised, as can be seen in ***Figure G1***. In case of supervised learning, a set of data is provided as well as the target (class) to which the data belongs. Therefore, a model is trained by knowing the outcome of the model in supervised learning. In the case of unsupervised learning, the outcome is unknown.

Since the defect labels and health conditions of the IRJs are known in this study, a type of supervised learning will be used. Within supervised learning, there are two categories, classification and regression (see ***Figure G1***). Regression is used to describe a numerical trend within the data and to predict numerical values based on the trendline. Classification, however, predicts the category to which the data belongs. In case of this study, the aim is to predict whether an IRJ is healthy or damaged and therefore a classification technique will be applied. There are many classification algorithms, among which logistic regression is most commonly used. However, for this study, the Support Vector Machine (SVM) algorithm is being used because of the benefit of the applicability of higher dimensional spaces.

Within the unsupervised learning techniques there are two subsets which are clustering and dimensionality reduction (see ***Figure G1***). In case of clustering, patterns are formed by means of clustering. Data points close to each other most probably belong to the same category and are labelled the same within clustering. The second subset is dimensionality reduction which is a powerful method to decompose a larger set of data into a smaller, better interpretable, dataset. Dimensionality reduction techniques can be used before classification to improve the (selection of) features used for classification (Pechenizkiy et al., 2004). For this study, the Principal Component Analysis (PCA) will be used which reduces the number of features from 12 to only 3 to increase interpretability and improve classification results.

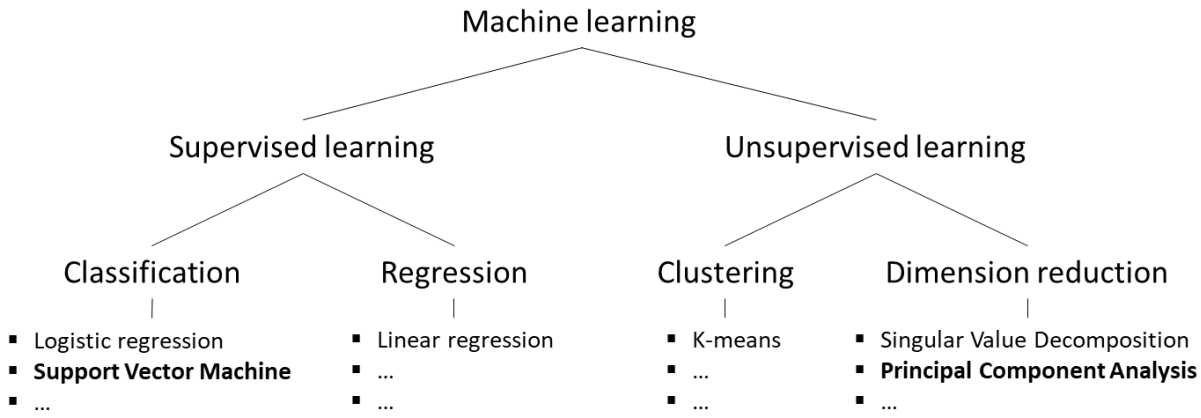


Figure G1 | *Simplified overview of the properties of machine learning. The models used in this study are emphasised in bold.*

Appendix H | Principal Component Analysis

Figure H1 explains the principle of the PCA in a 2D example. A dataset consists of two dimensions, x and y, which can be plotted as points in a plane. The variation of the data is projected to the horizontal and vertical axis below the left graph, still containing a lot of variation. Both features remain important to obtain all information of the data. Now, a PCA is applied in which a line with a certain direction and magnitude captures the most variation of the data. This is the first principal component (red line). A second line, perpendicular to the first principal component, captures the second most variation in the second principal component (green line). There is only limited variation within this direction. By rotating the two principal components, the right plot can be created. The projection on the horizontal and vertical line now shows a lot of variation in the first principal component, but almost no variation in the second component. By only considering the first principal component, most of the information of the data is still present. Dropping the second principal component does not result in a loss of information. The 2D data set is now captured with a high accuracy in a 1D data array.

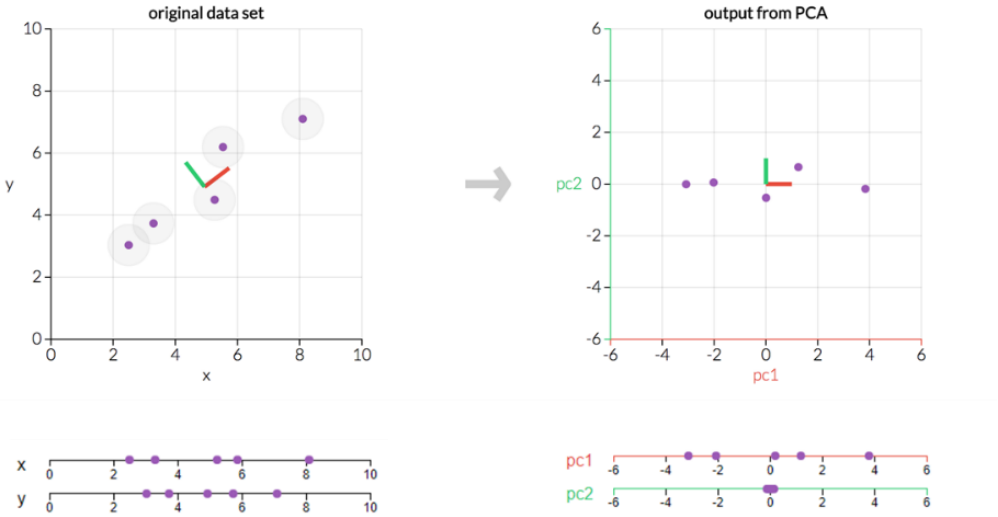


Figure H1 | *Example of principle of PCA* (Powell, n.d.).

Appendix I | Support Vector Machine

The goal of this thesis is to automatically classify the health condition of IRJs. For that, a machine learning algorithm is applicable which classifies data points with differences in features. The most common linear classifier is logistic regression, which determines the probability that a given data point is equal to either 0 or 1, event success or event failure, healthy or damaged, by means of the sigmoid function. Logistic regression is easy to implement and interpret. However, logistic regression has the disadvantage of overfitting in case of limited data and higher dimensional datasets (Rout, 2020). Therefore, an alternative technique, the Support Vector Machine (SVM) algorithm is applied in this study. It has the advantage to be effective in high dimensional spaces. It also performs well for a smaller dataset (Auria & Moro, 2008).

The SVM algorithm searches for a hyperplane in an N-dimensional space (N accounts for the number of features) that classifies the data points. Given a dataset with two classes (healthy and damaged) multiple hyperplanes (decision lines) can be drawn. The objective of SVM is to find a hyperplane that maximizes the distance between the data points of the two classes. By maximizing the distance, there is a built-in margin for future data points which provides more confident classification. The support vectors, the data point closest to the hyperplane, determine the position and orientation of the hyperplane (Gandhi, 2018). ***Figure I1*** describes the principle of SVM in which there is only one optimal hyperplane (***Figure I1b***) among all other possible hyperplanes (***Figure I1a***). The hyperplane in a two-dimensional space is a line, whereas for a three-dimensional space it is a plane. In SVM, classification is based on the output of the linear function representing the hyperplane.

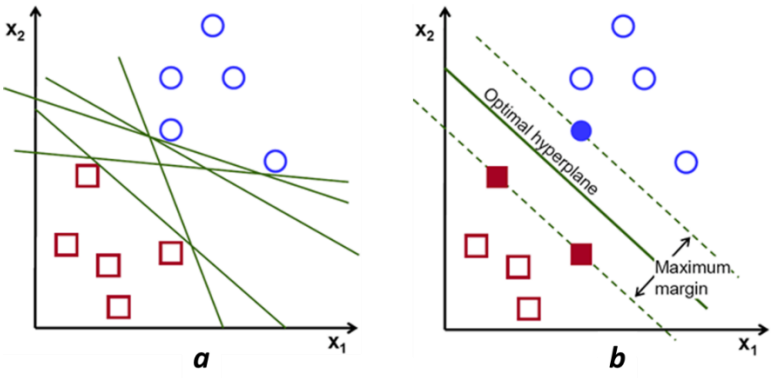


Figure I1 | *Principle of Support Vector Machine as a machine learning classifier* (Gandhi, 2018).

Appendix J | Model to determine influence of frequency characteristics

The relation between track stiffness and (natural) frequency might be explained by a simplified model as is described in ***Figure J1***. Let's simplify a short rail with an IRJ as a beam on two supports with a length of one sleeper spacing (l = 60 cm). An impact force (F) is applied at the IRJ in the middle of the beam. Due to the force, the beam will deflect (u). The magnitude of deflection is dependent on the track stiffness (k) of the simplified beam. Moreover, track stiffness itself is dependent on the track parameters elastic modulus (E), the moment of inertia (I) and the length of the beam (l). The formula for track stiffness as is written in equation (1) is characteristic for a beam on two rotational supports. The model of the beam is visualized in ***Figure J1a***.



Figure H1 | (a) *Simplified representation of rail and IRJ between two sleepers* ***(b)*** *which is modelled by a single-mass-spring system* (Spijkers et al., 2005).

The model given in ***Figure J1a*** can be simplified in a single-mass-spring system as is described in ***Figure J1b***. In here, a mass is fixed to a spring and will vibrate due to the external (impact) force. The mass consists of the mass of half of a rail and the mass of the train as is described in equation (3). The (natural) frequency of the system's vibration can be determined using equation (2) which is dependent on the track stiffness and mass.

$$k = \frac{F}{u} = \frac{48EI}{l^3} \quad \textbf{(1)} \qquad f = \frac{1}{2\pi}\sqrt{\frac{k}{m}} \quad \textbf{(2)} \qquad m = \frac{\rho A l}{2} + m_{train} \quad \textbf{(3)}$$

Now, all equations are known to determine the causes of the change in frequency. The parameters are related to the properties of a rail track and given in ***Table J1*** and the results are given in ***Table J2***.

Table J1 | *Properties of the rail track with a rail profile UIC54.*

Parameter	Unit	Value
m (train)	[kg]	10000
E	[N/m2]	210E9
I	[m4]	2350
L	[m]	0.6
A	[m2]	70E-4
ρ	[kg/m3]	7800

Table J2 | *Effect of halving the system properties on system frequency.*

	Frequency [Hz]	
Baseline	52.66	--
m / 2	74.48	↑
/ 2	37.24	↓
A / 2	52.68	--
/ 2	149.01	↑
ρ / 2	52.68	--

Appendix K | Software and data selection

Software

As the size of the data was too big for conventional programming software like Python, the data was first loaded and filtered in PySpark, a collaboration of Python and Apache Spark. Spark is an "open-source cluster-computing framework" with its advantages of speed and ease of use (Keshari, 2018). In fact, PySpark is a Python Application Programming Interface (API) for Spark that has the benefits of Python's simplicity and the power of Apache Spark to deal with Big Data. In this thesis, the DataFrame content of PySpark is used which is the collection of structured data similar to the rows and columns of an Excel sheet. Later on, after the selection of the useful data, Python is used to perform the analyses.

Data selection

The large DataFrame consisted of 29 columns and 60 million rows. Besides the accelerations, there were useful columns like GPS coordinates as well as velocity. Moreover, the "spoortak-id" was necessary for determining on which specific track the train has driven. In order to reduce the size of the DataFrame for the purpose of this master thesis, the data was selected by applying commands like sort, drop and filter. An example of the filtering process is to obtain only the tracks that are within the contract region of the ASSET Rail company. The data was then filtered based on the coordinates that lie within the contract region. This way, it was possible to obtain video footage of the track for the IRJ health condition assessment. Besides that, the sprinter train data was filtered out and only the intercity train data was used in this thesis.

As a result, there were 9 batches obtained each describing a two-minute train ride. With a sampling frequency of 6250 Hz, each batch consisted of 750.000 measurements. These files were able to be saved and opened in Python for the remaining analysis.

Appendix L | IRJ localisation

From the raw batch data, the acceleration data from the IRJs needed to be localized. A second database has the registration of all IRJs within the contract regions. The coordinates of the IRJs were used to localise the joints in the batch files. However, as the coordinates of the IRJ database were RD-coordinates ("Rijksdriehoekscoördinaten") indicating the lateral coordinate as X-coordinates and longitudinal coordinates as Y-coordinates a transformation was needed to change the X-Y-coordinates into GPS coordinates. Using the locations of the IRJs as well as the location of the ABA data, the accelerations around the IRJs could be obtained. However, due to the IRJ coordinate transformation and low GPS sample frequency inaccuracies up to 30 meters made IRJ localisation a challenge.

By comparing the coordinates of the acceleration signal to the IRJ coordinates, most of the time it was not possible to match the signal to the IRJ with 100% certainty. Surrounding acceleration peaks might be matched to the IRJ, but also to welds and crossings.
The mutual distance between the IRJs is known. By assuming the fixed spacings, it was possible to obtain the correct IRJ acceleration more accurately. However, it was still not possible to automatically detect the IRJs with a high certainty due to the alternating behaviour of the acceleration signal. As a solution, rather than searching for the locations in the time domain of the acceleration signal, this was done in the frequency domain. By applying a wavelet transform in the first place of the whole signal and average the output scales to the expected frequencies, it was possible to obtain a better and more accurate result. This method is known as the Scale Average Wavelet Power, which is further explained in chapter 4.3. In fact, the frequency response is filtered to only pass frequencies between 250 and 750 Hz. These frequencies are obtained from literature to be characteristic for IRJs (Molodova, 2013 & Molodova et al., 2016).

Although, the automatic detection of IRJs using the SAWP resulted in a more accurate localisation, the challenge remained to distinguish IRJs from crossings because they behave similarly, both in time characteristics as well as in frequency characteristics. Eventually, controlling the output by hand and using theoretical expectations, it was possible to obtain the correct signal to the IRJ. If it was still not possible to match the signal to the IRJ, the joint was removed from the analysis. ***Figure L1*** below gives an example of ABA data in the time domain (***Figure L1a***) and the associated SAWP response (***Figure L1b***).

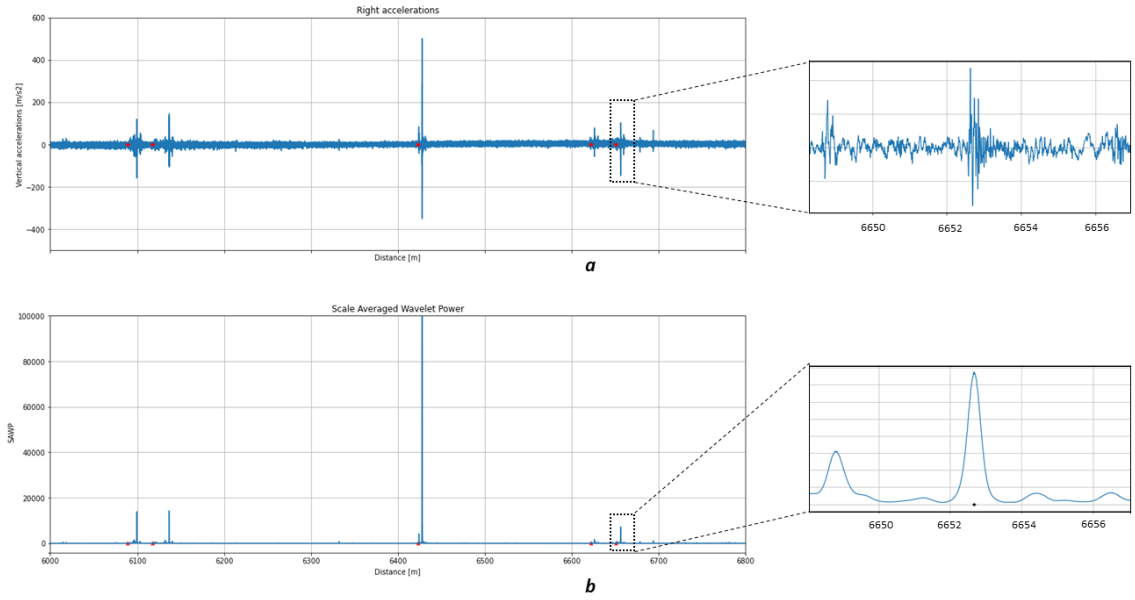


Figure L1 | *IRJ localisation by means of* ***(a)*** *acceleration signal and* ***(b)*** *SAWP*

Appendix M | CWT and WPS

Continuous Wavelet Transform

In a CWT the mother wavelet needs to be determined which influences the time and frequency resolutions. This study used the Complex Morlet wave as the mother wavelet as we expect that the shape of a Morlet wave matches the shape of the signal best. Moreover, there are no sharp edges in a Complex Morlet wavelet which minimizes the ripple effects that might be misinterpreted as oscillations (Cohen, 2019). From previous ABA studies, the Morlet wave has been used with proper results (Molodova, 2013). Although these studies used the normal Morlet wavelet, the Complex Morlet wavelet is suited better for this study as the complex part (which implies the phase lag) is important to create smooth transitions with a clearer scalogram as a result. Lastly, the Complex Morlet wavelet is commonly used for the determination of the (electrical) signal's frequency content because of its feature of equal variance in both time and frequency domain (Devleker, 2016).

Maximum wavelet power

Wavelet power (WP) is defined as the absolute value of the squared wavelet coefficients. An area of 1.5 meters around the IRJ is considered, 75 cm at each side. For each frequency, the maximum power within the area is taken to construct the final WPS. High powers for high frequencies are likely to occur faster and therefore closer to the end post of the IRJ than for low frequencies, which mostly occur at some distance from the IRJ. The high individual powers benefit the results of the study.

Frequency bands

In this thesis, the WPS is divided into 4 different frequency bands. The frequency bands are chosen based on the frequency characteristics described in the literature and based on the results of the WPS. As is mentioned in chapter 3, 500 to 600 Hz is interesting to study for battered head failure. It is assumed to obtain visible changes in this frequency region as well for battered head failure as its behaviour might be similar to plastic deformed IRJs. For both damage types, there is a permanent deformation at the railhead but at a different spot. Fastening damage is expected to occur at a lower frequency band between 70 and 100 Hz and damage to the insulation layer (which is interesting in case of lipping failure) is expected to occur at high frequencies of approximately 1040 Hz. Besides, the frequency bands are chosen based on clear visible peaks in the WPS of several IRJs. ***Figure M1*** shows two examples of WPS in which those peaks are visible.

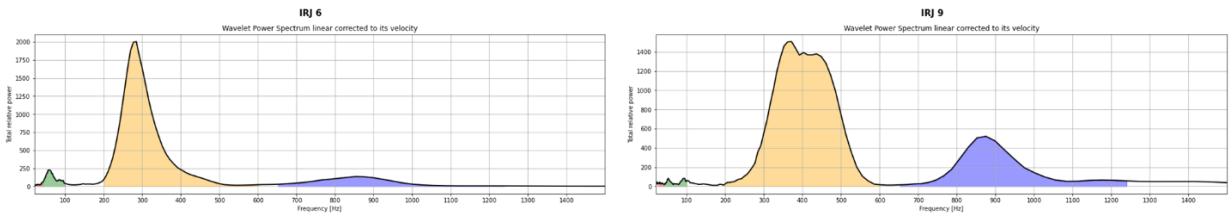


Figure M1 | *An example of the WPS of two IRJs which are divided into the four frequency bands*

Two bands are in the lower frequencies between 25 and 40 Hz and between 40 and 100 Hz. Since the number of scales at a low frequency is very high, indicating a high frequency resolution, a small frequency band is sufficient. For the middle frequencies, a band is chosen between 200 and 600 Hz. This frequency band is expected to contain a lot of power because these frequencies are typical for IRJs (Molodova et al., 2016). The last frequency band contains the higher frequencies from 650 to 1250 Hz. ***Table M1*** gives an overview of the proposed frequency bands.

Of each frequency band, three features are obtained: the total power, peak power and peak frequency. The total power is being calculated by integrating the area within the bandwidth using the trapezoidal rule. As the width of the bandwidths differs from each other, the total power is divided over the width of the bandwidth resulting in a total relative power. The peak power is the maximum value of relative power within each frequency band. The peak frequency is defined as the frequency at which the peak power occurs. The three features within each frequency band are stored in a Data Frame resulting in a total of 12 features per IRJ.

Table M1 | *Proposed frequency bands*

	Frequency bands
Low frequency band	25 – 40 Hz 40 – 100 Hz
Middle frequency band	200 – 600 Hz
High frequency band	650 – 1250 Hz

Appendix N | CWT results and eigenfrequencies

After matching the acceleration signal to the selected IRJs, the continuous wavelet transform (CWT) was applied. Using the CWT, accelerations in the time domain are transformed into powers in the time-frequency domain, in which power is defined as the squared wavelet coefficients and is depicted in a scalogram. ***Figures N1a and N1b*** show examples of a scalogram of a healthy IRJ and damaged IRJ, in which the IRJ (middle peak) was in between the two thermite welds left and right of the IRJ. Characteristic for thermite welds is the power in the low-frequency range (between 50-60 Hz) and most of all the power between approximately 200 and 250 Hz (Molodova, 2013). The IRJ was distinguished from thermite welds by the power in high frequencies between 250 and 750 Hz and even higher frequencies above 750 Hz. The damaged IRJ resulted in increased powers compared to the healthy IRJ, especially for frequencies around 400 Hz. ***Figures N1c and N1d*** show the cross-section at the location of the same healthy and damaged IRJ. Where a scalogram gives frequency, power and time information, the WPS only gives frequency and power information.

The WPS was extracted and averaged for both front and rear measurements, as well as for multiple repeated measurements as can be seen in the figure. The front and rear axle and multiple measurements showed similar behaviour. However, it can be seen that the front axle measured higher accelerations and corresponding powers than the rear axle. This could be because the rear axle is affected by the front axle. Possibly, the front axle in free air suffers higher frequencies and absorbs some energy and the rear axle follows with a lower magnitude. Moreover, the front axle seemed to have power in higher frequencies than the rear axle which can be seen at frequencies around 1000 Hz in ***Figure N1d***. Again, this might be caused by surrounding axles.

Figure N2 shows the averaged and normalized WPS (power divided by speed) which is the input for the relation and classification assessment. The WPS is divided into four frequency bands in which two frequencies bands describe the low frequencies (25-40 Hz and 40-100 Hz), one frequency band the middle frequencies (200-600 Hz) and the last one high frequencies (650-1250 Hz). Of all four frequency bands, the total power, peak power and peak frequency are collected and stored in a DataFrame. Therefore, for each IRJ, 12 features are describing the IRJ. The unit of total power remains relative since the power is divided by the train speed. The same goes for peak power, which is a relative number. Therefore, the total power is described as total relative power and the peak power is described as the peak relative power. The peak frequency keeps the unit of Hertz.

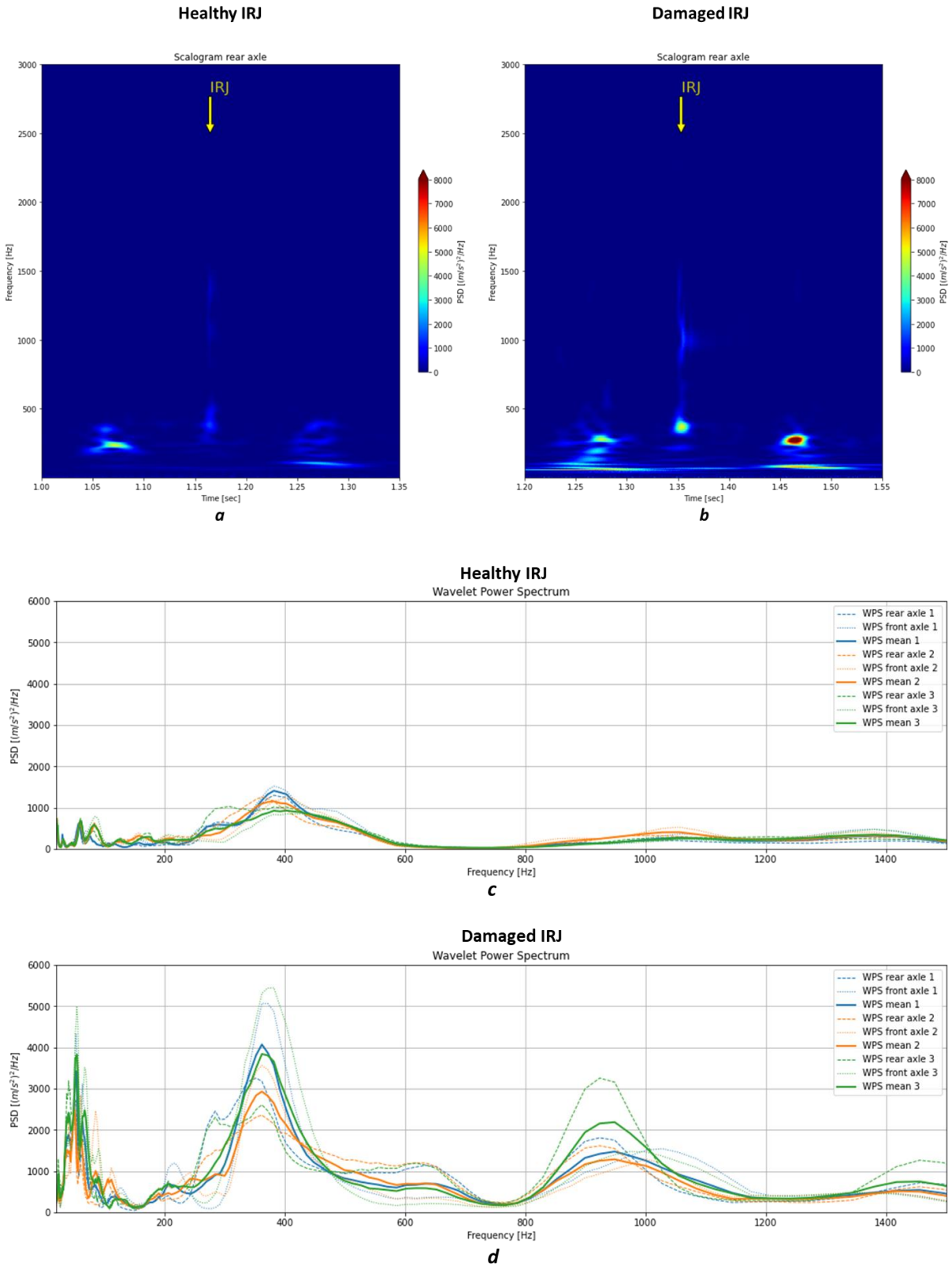


Figure N1 | *Example of frequency characteristics of a healthy and damaged IRJ as an output of the CWT:* ***(a)*** *scalogram of healthy IRJ,* ***(b)*** *scalogram of damaged IRJ,* ***(c)*** *WPS of the same healthy IRJ and* ***(d)*** *WPS of the same damaged IRJ. The scalograms only show the response for the rear axle of one measurement, whereas the WPS shows front and rear responses, including the average response, as well as the responses for three repetitive measurements.*

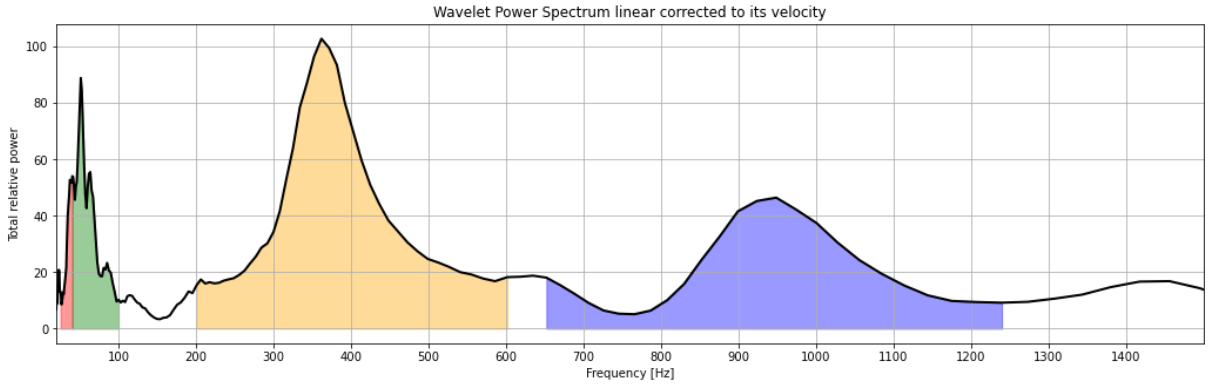


Figure N2 | *Example of the WPS of the same damaged IRJ as above which is constructed by averaging multiple measurements and linearly corrected for its velocity. The WPS is divided into four frequency bands, in which two frequency bands describe low frequencies (red and green), one frequency band describes the middle frequencies (yellow) and the last one describes high frequencies (purple).*

Additionally, it was noticed that on a normal straight track without any interrupting assets, there was a constant frequency present which is depicted in ***Figure N3***. Especially, frequencies at approximately 63 Hz were visible which are probably related to the eigenfrequency of the train. The exact eigenfrequency is specific for each train and dependents on e.g. mass and speed. Moreover, at frequencies between 300 and 400 Hz and approximately 650 Hz, semi-constant powers were present. The powers in these frequency ranges could also be related to eigenfrequencies of the train or track or it could be considered as noise. From literature (see theoretical information subchapter vibration modes), it is known that sleeper anti-resonance occurs at approximately 300 Hz and rail resonance at approximately 1000 Hz. Therefore, it is most likely to conclude that the frequencies around 300 and 400 Hz coincide with sleeper anti-resonance. The frequencies at approximately 650 Hz might be related to rail resonance, but could either be related to one of the eigenfrequencies of the train or noise.

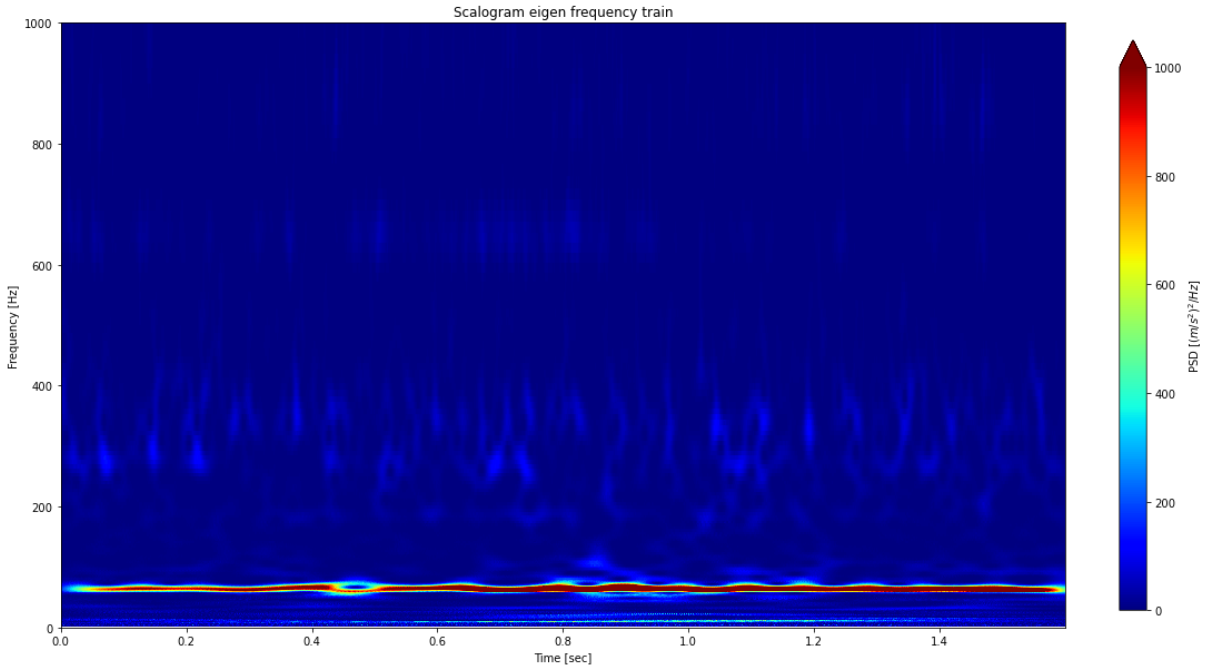


Figure N3 | *Scalogram in which the eigen frequencies of the train and/or track or noise are visible*

Appendix O | Vibration modes

As explained in chapter 3 high frequencies are related to the eigenfrequencies of the track which are dependent on track characteristics such as track stiffness. Eigen frequencies are important in the investigation of a signal in its frequency domain. As with all objects, the rail track vibrates based on its natural frequencies, also known as eigenfrequencies. The track can be seen as a mass-spring system. If the system is forced from its equilibrium position due to an external force, the system will vibrate with the natural frequency that is specifically related to the system. The vibration modes are dependent on the track parameters. As the signature tunes of defects might be dependent on the natural frequencies of the track, the natural frequencies need to be considered (Molodova, 2013).

In a study by Molodova (2013) the vibration modes of a ballasted track were investigated. Four track vibration modes were found, which are summarized in ***Table O1*** and visualized in ***Figure O1***. The first natural frequency was found at a low frequency of approximately 40 Hz which matches the vertical vibration of the rail which moves together with the sleepers. The vibration mode is called full track resonance, which is also known as in-phase vibration. The frequency of 40 Hz was found for a loaded track, whereas for an unloaded track the full track resonance frequency was found to be around 90 Hz. In ***Figure O1a*** it is visible that the rail and sleepers move together. The second vibration mode was found at 300 Hz representing sleeper anti-resonance. In this vibration mode, the sleepers experience resonance even though the rails hardly move at all, see ***Figure O1b***. The third vibration mode around 1000 Hz is called rail resonance or out-of-phase resonance, in which the rail vibrates in the opposite direction compared to the sleepers, see ***Figure O1c.*** The light green coloured sleepers move up whereas the dark green coloured sleepers move down. Depending on the rail pad parameters the frequency deviates in the mid/high frequency range. The last vibration mode is found at high frequencies of approximately 1160 Hz and is called pin-pin resonance. It is the movement of the rail matching a vibration with a wavelength of twice the sleeper spacing. In here, the rails are pinned to the un-displaced sleepers, as can be seen in ***Figure O1d***.

Table O1 | *Ballasted tack vibration modes (Molodova, 2013).*

Resonance	Definition	Frequency
Full track resonance	Vertical in-phase vibration of the rail and sleepers together	90 Hz (unloaded track) 38 Hz (pre-loaded track)
Sleeper anti-resonance	Only sleeper resonance; no (or small) rail movement	300 Hz
Rail resonance	Vertical out-of-phase vibration of the rail relative to the sleeper	1000 Hz
Pin-pin resonance	Rail vibration where the rails are pinned to the sleepers resulting in a wavelength of twice the sleeper spacing	1160 Hz

As the rails at IRJs are expected to behave differently due to their discontinuity in track stiffness, the track eigenfrequencies might differ. The track vibration modes could be related to the formation of defects, as is the case of corrugation. Therefore, it is important to distinguish the frequency characteristics of the defects from the natural track frequencies (Molodova, 2013).

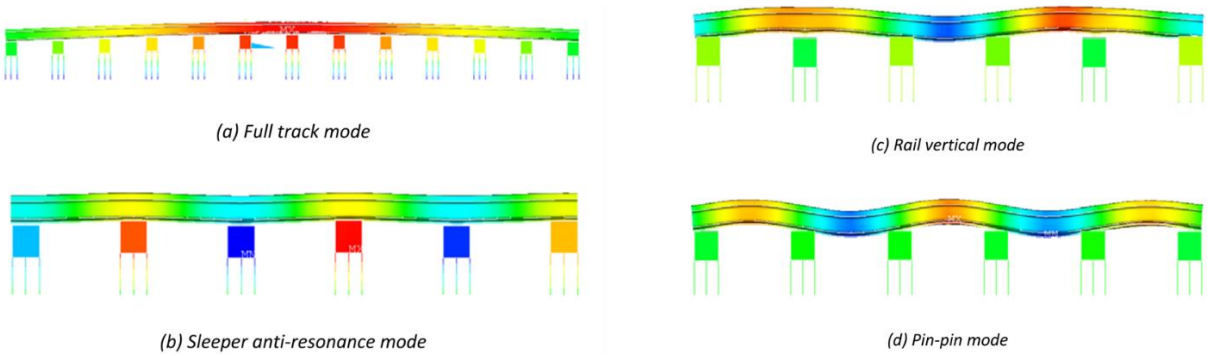


Figure O1 | *Track vibration modes:* ***(a)*** *Full track resonance;* ***(b)*** *Sleeper anti-resonance;* ***(c)*** *Rail resonance;* ***(d)*** *Pin-pin resonance. The red colour indicates the magnitude of the upward deflection and the blue colour indicates the magnitude of the downward deflection. The green colour indicates a zero deflection (Molodova, 2013).*

Appendix P | Defect assessment scores

The health condition of IRJs is visually assessed. In the below ***Figure P1*** video images of battered head failure, lipping failure and loose bolts are shown. For battered head failure and lipping failure, top view images are used because both defects occur at the top surface of the rail and insulating end post. For loose bolts, side views (inner and outer) are used to assess the health condition of the IRJs. Moreover, it is looked at other visible defects, like permanent deformation and corrugation, which can be seen in ***Figure P2***. These failures are excluded from this study but are interesting to clarify deviating frequency behaviour if needed. In order to assess the IRJs similarly, it is looked at a few attention points for all defects.

Battered head failure

Battered head failure can be recognized from visible arc damage. Due to the large impact force at an IRJ, sparks might be developed. In case of battered head failure, the railhead permanently deforms resulting in a larger gap at the insulating end post. The degradation of the railhead results in an even larger impact force, and therefore larger arc damage. Moreover, due to the sparks, the insulating end post weakens and will deform more easily. Additionally, the fishplates are not able to keep the rails together because of the lack of support. The deformation of the IRJ causes an even larger impact force. Arc damage is visible in the top view images as a white area with small black dots on one side of the rail. In ***Figure P1,*** arc damage is visible on the right side of the insulting end post, which means the train travels from left to right. The IRJ is assessed with a defect score of 1 for battered head failure if arc damage is visible. Defect score 1 differs from defect score 2 if there is circular damage at the end post. The black horizontal line in the below figure deviates at the end post, which indicates a circular deformation at the end post. The large impact force causes a combination of arc damage and circular (deformation) damage of the railhead at the insulating end post. Both are an indication of battered head failure and are visible from the top view images.

Lipping failure

Lipping failure is easily visible from the top view images as the width of the end post decreases. Lipping failure is after all the flow of steel over the insulating end post. Due to fatigue, the repetitive forces weaken the railhead which results in an inward flow. The smaller width of the insulating end post is therefore an indication of lipping failure. The width difference between defect scores 1 and 2 are arbitrarily chosen. There is a clear difference in the example below, but in some cases, the difference was less visible. Therefore, for the purpose of this study, it has been chosen to take both defect scores 1 and 2 into consideration.

Loose bolts

Loose bolts are assessed based on side-view images. Only the bolts which fix the rail to the rail plate is being assessed. It is looked at the formation of corrosion which is an indication of looseness. There is after all a possibility for water to get stuck between the loose space of bolts. Water and air combined with steel are subjective to corrosion. As can be seen in the first image, the bolt looks healthy as there is no formation of rust. Moreover, the length of the bolt itself looks normal. The second image is assessed with a defect score of 1 because of the formation of rust. The light brown colour indicates the presence of rust. Besides, the length of the bolt itself looks normal. It can be distinguished from defect score 3 if the length of the bolt is a little bit shorter. This is a little more visible if compared to other bolts with a normal length. At last, a little spring is visible which indicates the bolt is loose. After all, the spring should not be visible due to the fixation of the bolt. All three features result in the assessment of defect score 2.

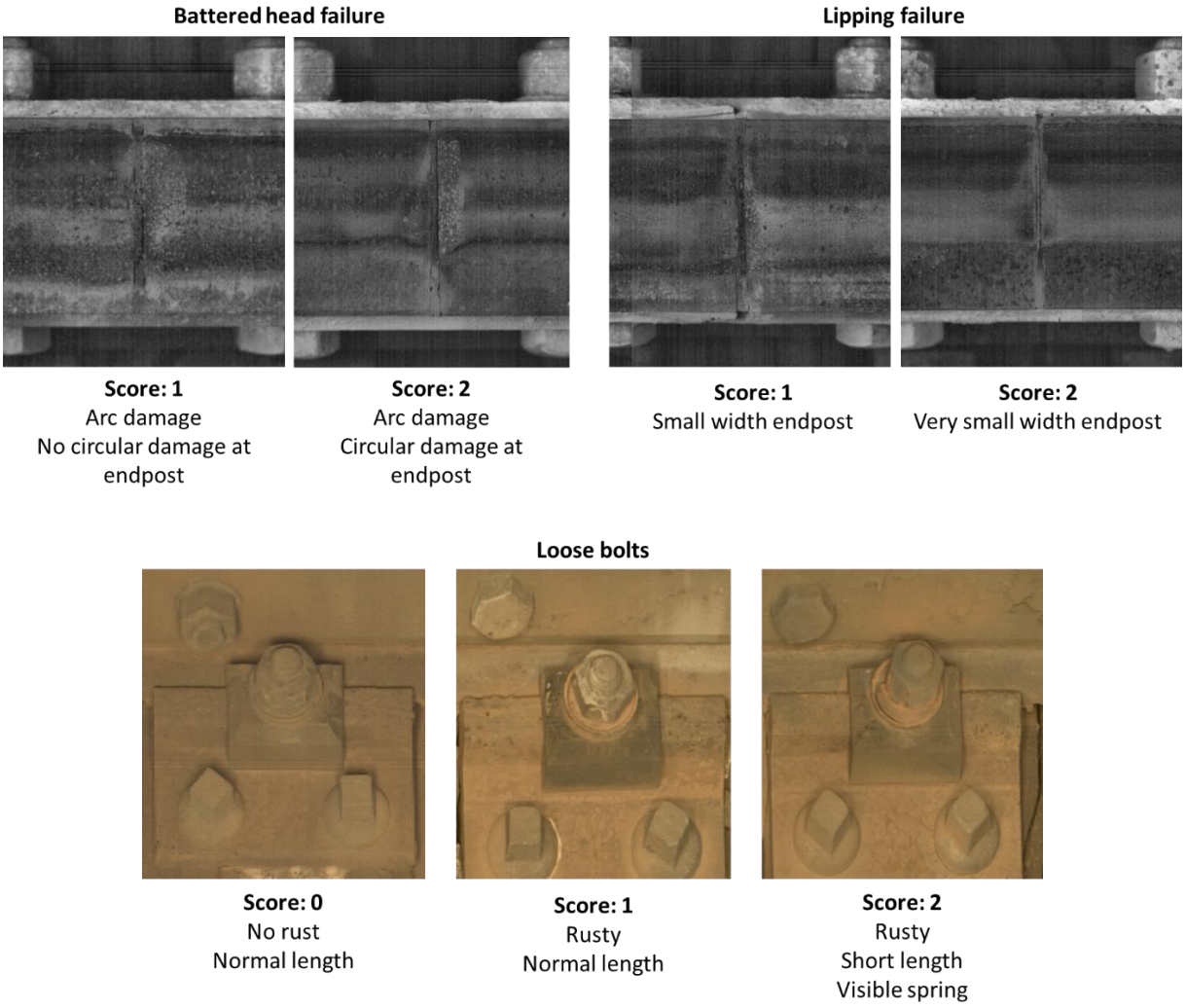


Figure P1 | *Video images of the studied defects and its defect scores, which are determined based on the characteristics written below the scores.*

There is also looked at other defects, which might interrupt the frequency behaviour of IRJs. It is therefore important to assess the complete condition of the IRJ. Permanent deformation and corrugation are two examples of defects that are visible from the video footage. Hanging sleepers for example is a common defect of IRJs, but is not clearly visible from the video images and could therefore not be assessed.

<u>Permanent deformation</u>

Permanent deformation is caused by the impact force at a distance of approximately 50 mm from the insulating end post and is visible from top view images as a deviation in the black horizontal line, see ***Figure P2***. The black horizontal line indicates the normal running line of the wheel of the train, which is similar to battered head failure. However, due to deformation, the running line deviates sideways. Permanent deformation differs from battered head failure in the location of deviation. In case of permanent deformation, the deviation is at a short distance from the insulating end post (about 50 mm), whereas for battered head failure it is right at the end post.

Corrugation

Corrugation is wear of the rail surface in a wave-form and can be classified as either short pitch rail corrugation or long pitch corrugation. From the video images, some short pitch rail corrugation was visible, with a wavelength of 25-80 mm. Is it caused by the pin-pin rail resonance, which is the movement of the rail matching to a vibration with a wavelength of twice the sleeper spacing. In here, the rails are pinned to the un-displaced sleepers (Steenbergen, 2008). Due to vibrations with a frequency of approximately 1160 Hz, repetitive forces as a result of the vibration's amplitude, degrade the rail top surface. The wavelength (L) is after all related to the speed of the train (v) and frequency (f) of the vibration (L = v/f). Corrugation causes an increase in noise emission and gives rise to the formation of other failure types such as RCF. It has therefore a large impact on maintenance. Short pitch rail corrugation can be seen as shiny crests and darker troughs which are visible from the video footage. ***Figure P2*** gives an example of corrugation.

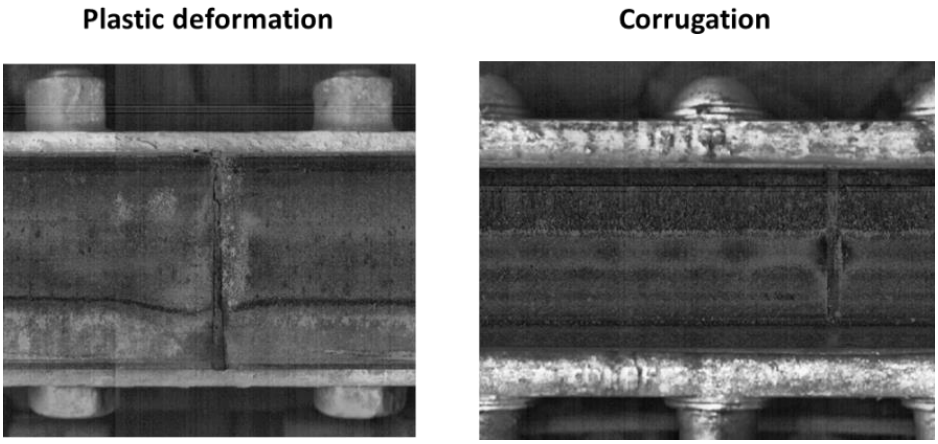


Figure P2 | *Examples of plastic deformation and corrugation*

Appendix Q | Clarification assumptions

Four assumptions had to be made to overcome the problem of inaccurate data due to a year difference in measurements and video footage. First of all, it is assumed that only healthy IRJs are used as comparison to degraded IRJs. Healthy IRJs are IRJs with zero's for all defects or are new IRJs that are recently replaced (compared to the measuring date). Secondly, it is assumed that for battered head failure only healthy IRJs are compared to IRJs with battered head failure with defect score 2. IRJs with defect score 1 are uncertain and could be overestimated or underestimated. The IRJ can be either healthy from measurements but degraded if it comes down to the video image assessment. Also, the IRJ could be damaged but the IRJ looks healthy from the video images. In between maintenance could have been executed. Secondly, in case of lipping failure, only healthy IRJs are compared to IRJs with lipping failure with both defect scores 1 and 2. There was little data available and the difference between defects scores 1 and 2 were difficult to distinguish. And lastly, loose bolts had to be neglected as there was too little data available.

The graph in ***Figure Q1*** shows a 3D graph in which the total power of two frequency bands are plotted on the x and y-axis and the defect score of battered head failure is visible as a colour, in which green equals defect score 0, orange equals defect score 1 and red equals defect score 2. For the low frequencies (left figure) two green outliers are visible. Moreover, for the higher frequencies (right graph) there are also green values in between orange and red values. It can be concluded that battered head with a defect score of 0 does not imply a healthy IRJ. This makes sense since other defects could be present as well. Therefore, it is assumed to use only healthy IRJs as a comparison. Healthy IRJs are obtained from the defect assessment if all defects are given a 0 and IRJs that are recently replaced for a new one. Using this assumption it is possible to conclude that the green values belong indeed to a completely healthy IRJ. Furthermore, the second assumption is also based on the results of ***Figure Q1***. Especially for high frequencies, orange outliers are visible that show large powers, but are assessed with only intermediate defects (defect score 1). IRJs with battered head failure defect score 1 are unreliable and should therefore be neglected. Once removing these values, there is a clear difference between healthy and damaged IRJs.

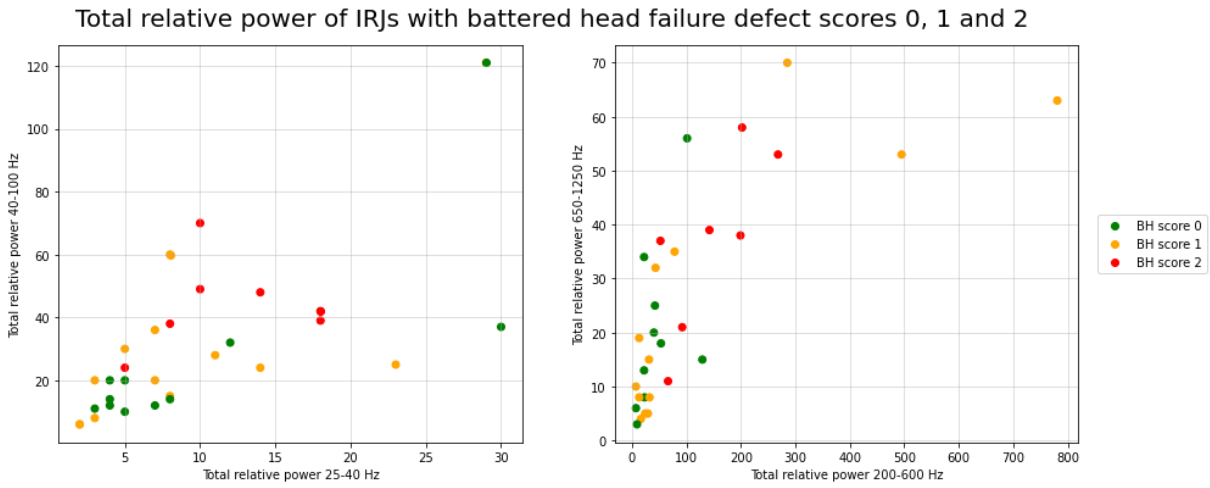


Figure Q1 | *Example of the total relative power of battered head failure which verifies the assumption to use only healthy IRJs and IRJs with battered head defect score 2.*

Figure Q2 shows the assumption to use both defect scores 1 and 2 for lipping failure as they behave similarly. Moreover, the new and healthy IRJs are shown, instead of lipping failure with defect scores of 0.

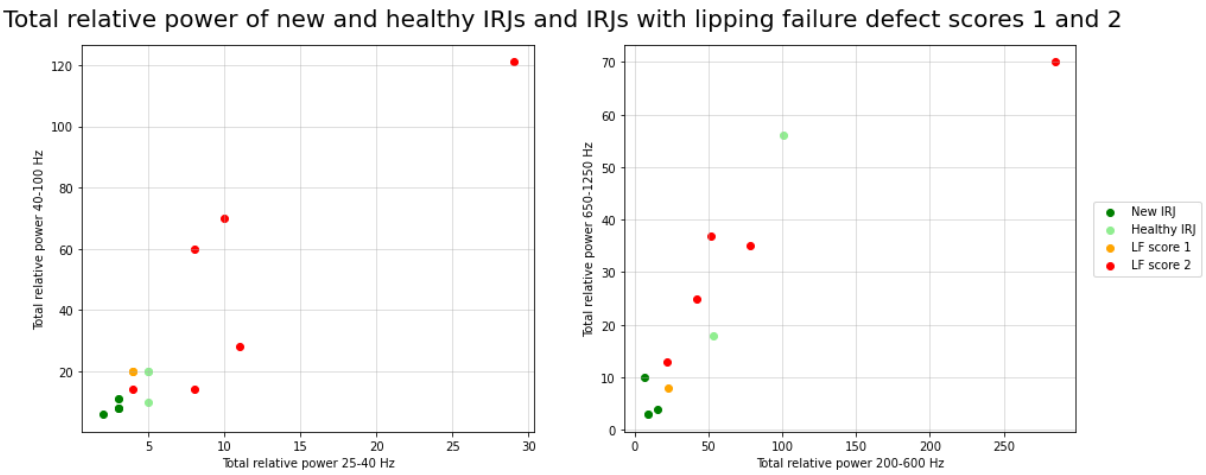


Figure Q2 | *Example of total relative power of lipping failure which verifies the assumption to use only new and healthy IRJs and IRJs with lipping failure defect score 1 and 2.*

Appendix R | Results correlation BHF

In this appendix, the results of peak relative power and peak frequency are presented for IRJs with battered head failure as well as an elaborated explanation of the outliers.

Peak relative power

The peak relative power shows a similar behaviour as the total relative power as can be seen in ***Figure R1***. After all, a higher total power is often a result of a higher peak power because total power is the area of the PSD. The peak power in all frequency bands increases in case of battered head failure as can be clearly seen from the open circles which indicate the mean peak relative power for healthy (green) and battered head (red) IRJs. Frequency band 200-600 Hz shows the greatest increase in peak relative power with a factor of 4.23. However, the same outliers as for total relative power result in high standard deviations. IRJ no. 1 contains high peak relative powers for all frequency bands, except for 40-100 Hz. Moreover, IRJ no. 2 contains a high peak relative power within frequency band 25-40 Hz, whereas both IRJs no. 2 and 3 contain excessive peak powers in frequency bands 200-600 Hz and 650-1250 Hz.

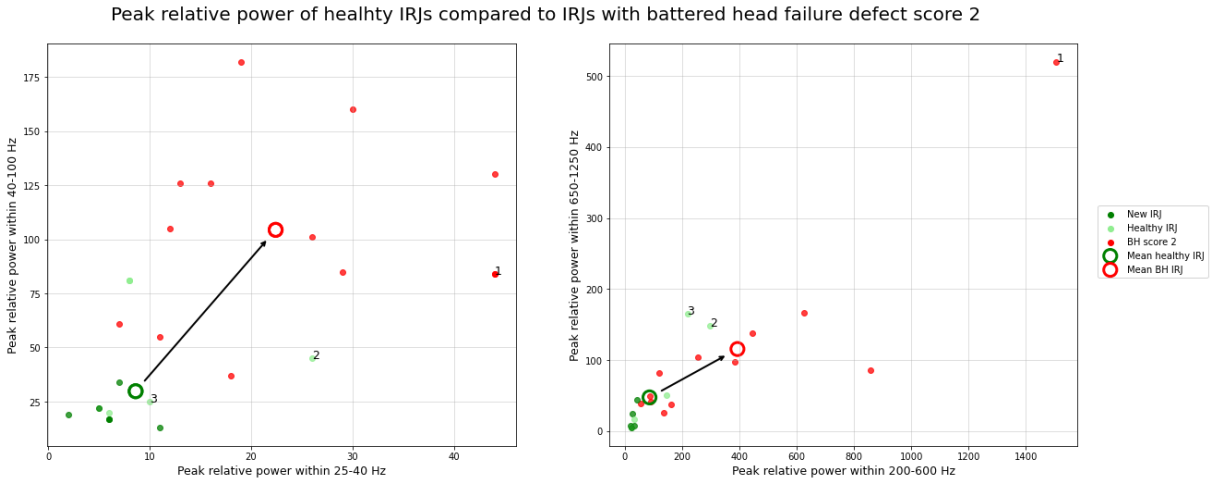


Figure R1 | *Peak relative power of healthy IRJs compared to IRJs with battered head failure defect score 2*

Peak frequency

The peak frequency is defined as the frequency at which the peak power occurs within the boundaries of the given frequency band and is visualized for the analysed IRJs in ***Figure R2***. Rather than the positive increase in case of total relative power and peak relative power, it can be seen that for some frequency bands the peak frequency increases, but for other frequency bands decreases or remains the same. Overall, there is some variation in peak frequency visible. There is no change in peak frequency from the mean peak frequency in bandwidth 25-40 Hz. This can be explained by the fact that this frequency band does not show a clear peak, but a rather rectangular total power area. There is no remarkable frequency that exists within this frequency band. There is a small increase of peak frequency within frequency band 40-100 Hz. Regardless of the individual outliers, a cluster around 70 Hz is visible for IRJs with battered head failure, whereas the healthy IRJs show a lower peak frequency around 60 Hz. It is known from ***Appendix O*** that frequencies at approximately 63 Hz are related to the eigenfrequency of the train. In case of healthy IRJs, this eigenfrequency remains most significant, whereas for IRJs with battered head failure, the peak frequency is most likely to increase a little bit to approximately 70 Hz.

Within frequency band 200-600 Hz there is an increase in peak frequency from healthy IRJs to IRJs with battered head failure. The mean peak frequency of healthy IRJs is at 360 Hz and there is also a cluster of healthy IRJs visible at approximately 375 Hz. The mean peak frequency of battered head IRJs is also at approximately 390 Hz, but a cluster is visible at 440 Hz. So there is also an increase visible between the two clusters. Lastly, a decrease is visible for the peak frequency within bandwidth 650-1250 Hz. The mean value of healthy IRJs changes from 990 Hz to 870 Hz in case of IRJs with battered head failure. Moreover, clusters of healthy and battered head IRJs are visible. Especially the cluster of battered head IRJs is consistent at approximately 870 Hz. However, all differences are not statistically significant and therefore it is concluded that the mean peak frequencies are equal for healthy IRJs and battered head IRJs. The three outliers in total relative power and peak relative power do not act differently in the peak frequency plots.

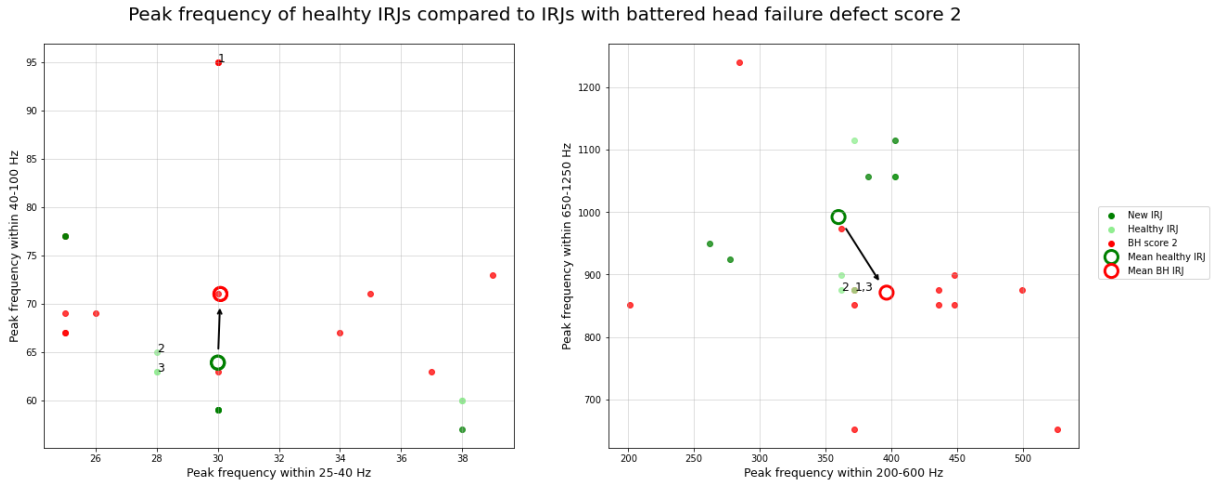


Figure R2 | *Peak frequency of healthy IRJs compared to IRJs with battered head failure defect score 2*

Outliers

The large increase in frequency band 200-600 Hz is mainly caused by one individual IRJ (denoted with superscript 1 in ***Figure R1***). Moreover, there is a relatively large standard deviation for powers in frequency band 200-600 Hz and 650-1250 Hz, which also indicates the possibility of an outlier. It is found that the IRJ is assessed with severe battered head failure, but also loose bolts, permanent deformation and a crack in the insulation layer. A combination of these failure modes might explain the high total relative powers in 3 out of 4 frequency bands.

Moreover, two healthy IRJs show a relatively large power within frequency bands 200-600 Hz and 650-1250 Hz (denoted with superscripts 2 and 3 in ***Figure R1***) of which one IRJ (superscript 2) also shows a larger power within frequency bands 25-40 Hz and 40-100 Hz. In case of IRJ no. 2, it could be seen from the video images that maintenance had been performed in terms of grinding. One might hypothesize that the grinding action has taken place after the measurements. Therefore, the IRJ could have been more damaged than is assessed. In case of IRJ no. 3, there was no evidence of any maintenance actions. However, it could be seen that the adjacent rail suffers from severe plastic deformation and battered head failure. It might be possible that the adjacent left rail influences the accelerations of the right accelerometer with higher powers of the left accelerations as a result. Video images of the outlier IRJs are given below.

IRJ superscript 1

IRJ superscript 1 with battered head failure, permanent deformation, loose bolts and a crack in the insulation layer, as can be seen in ***Figure R3***.

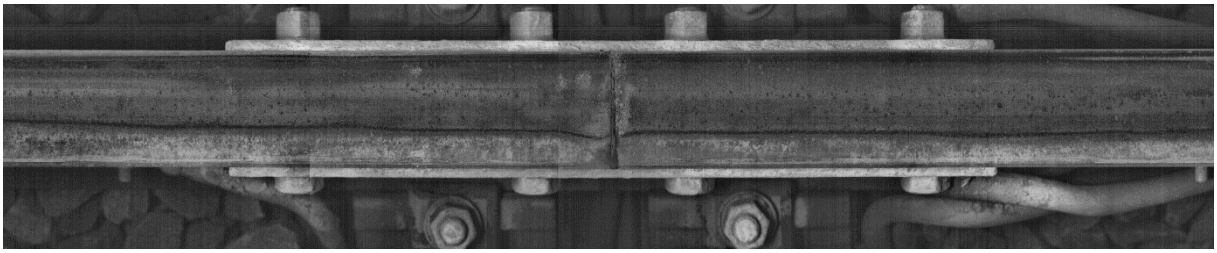


Figure R3 | *Degraded IRJ superscript 1*

IRJ superscript 2

In ***Figure R4*** it is clearly visible that the IRJ has been grinded during maintenance, which might be the reason for a high amount of power while being a healthy IRJ.



Figure R4 | *Grinded IRJ superscript 2*

IRJ superscript 3

The top image of ***Figure R5*** shows the (right) IRJ of superscript 3 which is assessed as a healthy IRJ. However, the high powers could be explained by the degraded adjacent left rail. The adjacent left IRJ contains severe permanent deformation and battered head failure, as can be seen in the bottom image of ***Figure R5***.

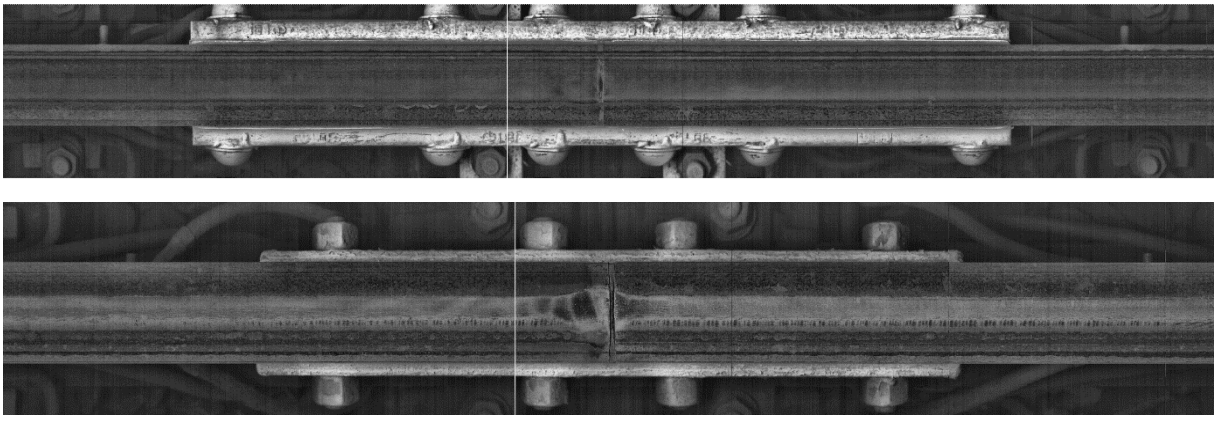


Figure R5 | *Top image: right IRJ superscript 3. Bottom image: adjacent left IRJ superscript 3 with severe permanent deformation and battered head failure*

Appendix S | Results correlation LF

Outlier

Lipping failure was investigated by comparing the healthy IRJs to the IRJs assessed with defect scores 1 and 2 for lipping. From ***Figure S1***, it is apparent that there is an outlier in the lower frequency bands, denoted with a superscript 1. This IRJ presents with 5 times more power within the lowest frequency band as compared to the average of all degraded IRJs and even 7 times more power within frequency band 40-100 Hz. However, no clear outliers appear when looking at the higher frequency bands (right graph in ***Figure S1***). Interestingly, the power of the outlier seems to increase linearly in the lower frequency bands. Taken together with the fact that for the higher frequency bands there is no IRJ with high deviating powers, it might not concern an outlier, but rather an extreme case of lipping failure possibly combined with another defect. From the video assessment results, it can be seen that the beforementioned IRJ contains a combination of lipping failure, loose bolts and a loss of insulation material as can be seen in ***Figure S2***. The mean relative powers are influenced a lot by the one IRJ having a large deviation in power in the lower frequency bands and therefore this IRJ will not be considered in the subsequent steps. The removal of the IRJ decreases the difference in healthy IRJs and IRJs with lipping failure and therefore improves the investigated relationships.

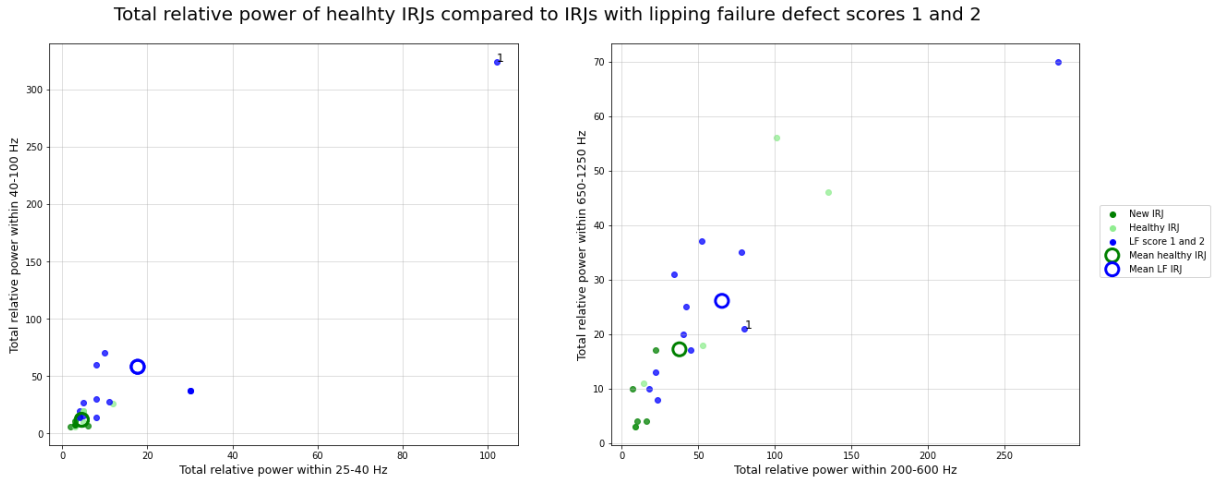


Figure S1 | *Total relative power of healthy IRJs compared to IRJs with lipping failure defect scores 1 and 2. The IRJ with superscript 1 is removed from further analysis because of its high deviation in power in the lower frequency bands.*

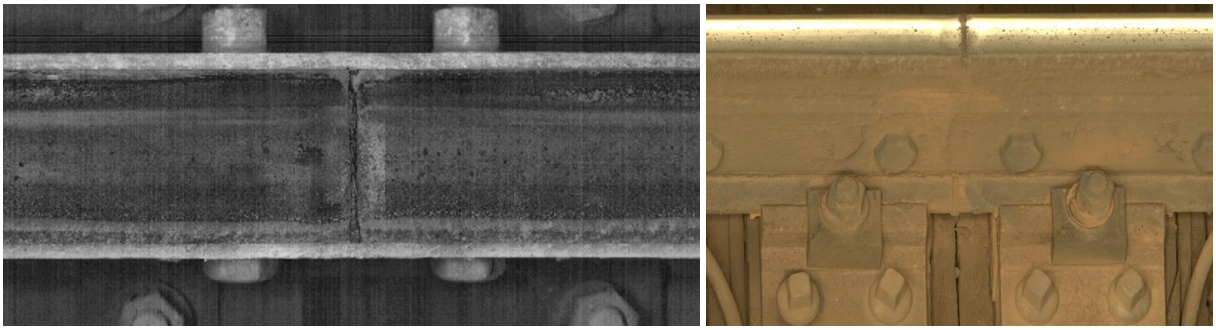


Figure S2 | *Video footage of the above-described outlier IRJ which contains lipping failure, but also loose fastening bolts and a loss of insulation material*

Peak relative power

The peak relative power graphs are shown in ***Figure S3***. A clear trend is visible, especially the peak relative power in frequency band 40-100 Hz. Moreover, there are two healthy IRJs (superscript 2 and 3) that show relatively high peak relative powers, especially in the two high-frequency bands. Three lipping failure outliers (superscript 4, 5 and 6) are within the cluster of healthy IRJs for all frequency bands.

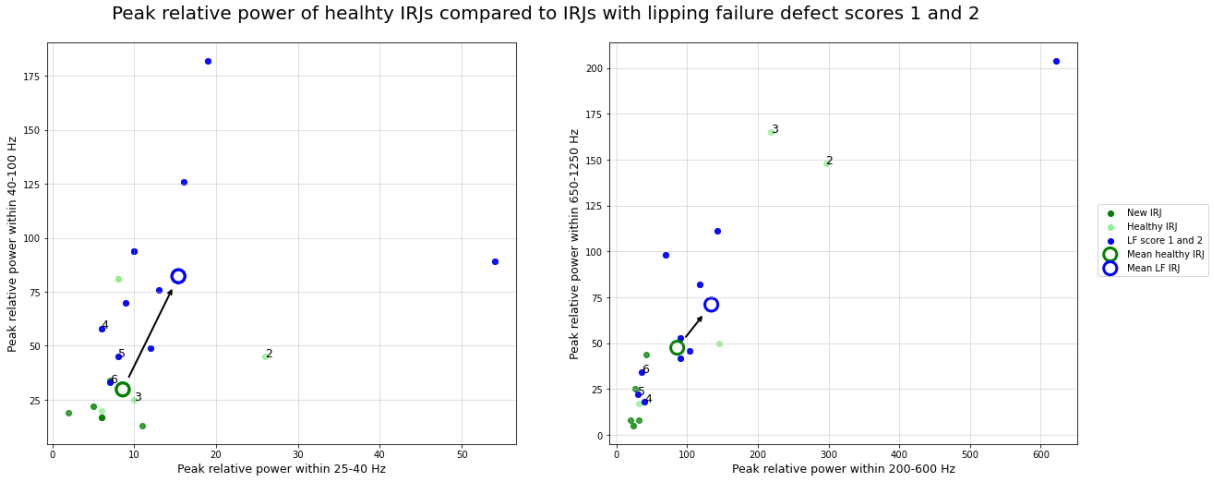


Figure S3 | *Peak relative power of healthy IRJs compared to IRJs with lipping failure defect scores 1 and 2*

Peak frequency

From ***Figure S4*** it can be seen that there is a lot of variation in peak frequency within the two low-frequency bands 25-40 Hz and 40-100 Hz. The frequencies are distributed over the whole width of the frequency band for both healthy IRJs and IRJs with lipping failure. In case of the peak frequency within frequency band 40-100 Hz, there is a slight upward shift indicating an increase of peak frequency. However, due to two data points that contain a peak frequency lower than 55 Hz the mean increase is slightly reduced. The same applies to one healthy datapoint with a relatively large peak frequency of 77 Hz. A healthy cluster can be distinguished around approximately 60 Hz, which is close to the eigenfrequency of the train.

Similarly, a larger difference in peak frequency is visible in the higher frequency bands, especially bandwidth 650-1250 Hz. The mean peak frequency of healthy IRJs is higher than IRJs with lipping failure within bandwidth 650-1250 Hz. A cluster of lipping failure data points is located at a peak frequency of 860 Hz, which is also the mean peak frequency within this frequency band. The peak frequency in bandwidth 200-600 Hz is more distributed, but for healthy IRJs the value varies around 360 Hz. Except for one lipping failure data point, the peak frequency of lipping failure is generally higher than the healthy peak frequency and a cluster around 440 Hz seems to be present.

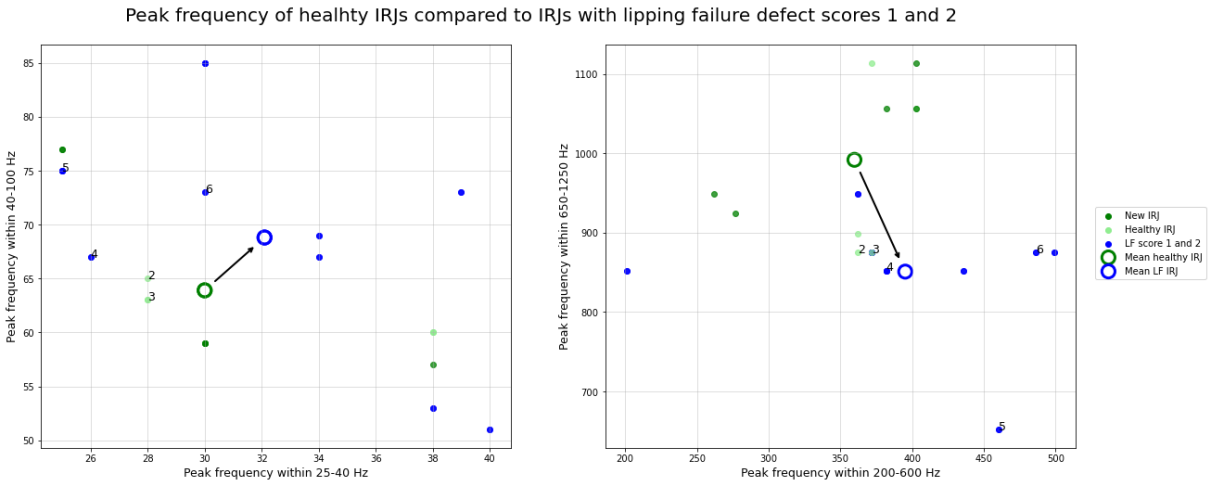


Figure S4 | *Peak frequency of healthy IRJs compared to IRJs with lipping failure defect score 1 and 2*

Appendix T | Results correlation BHF vs LF

Peak relative power

From ***Figure T1*** it can be seen that the peak relative power of battered head failure and lipping failure both increase similar to the total relative power. No statistically significant differences can be obtained. Therefore, this might suggest that lipping failure and battered head failure are related considering the frequency characteristics.

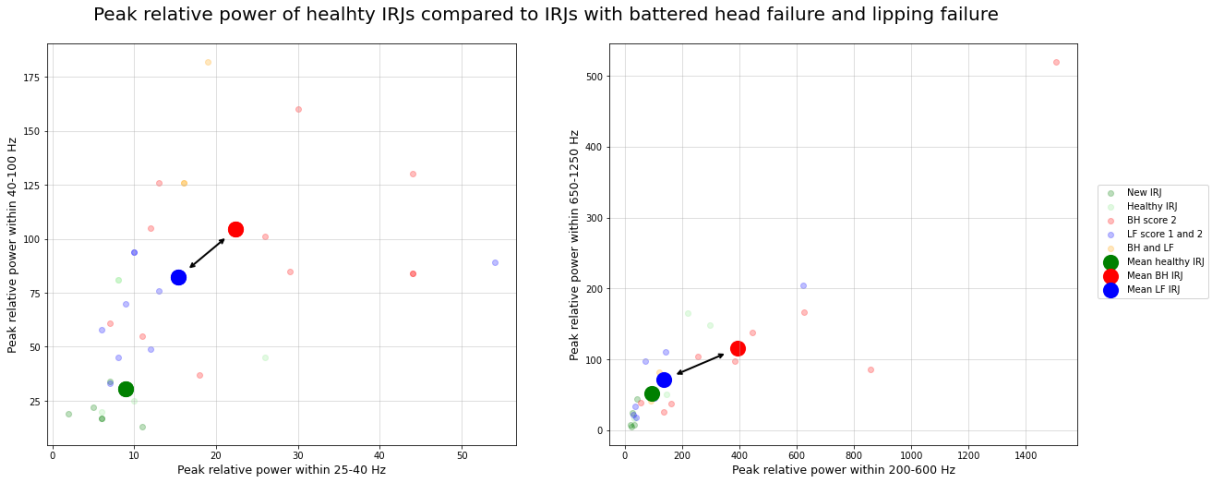


Figure T1 | *Peak relative power of healthy IRJs, IRJs with battered head failure and IRJs with lipping failure.*

Peak frequency

From ***Figure T2*** it can be seen that the peak frequencies of battered head failure and lipping failure behave similarly, compared to healthy IRJs. Although there is a small difference visible compared to healthy IRJs, there is hardly any change visible between battered head failure and lipping failure and thus no statistically significant difference.

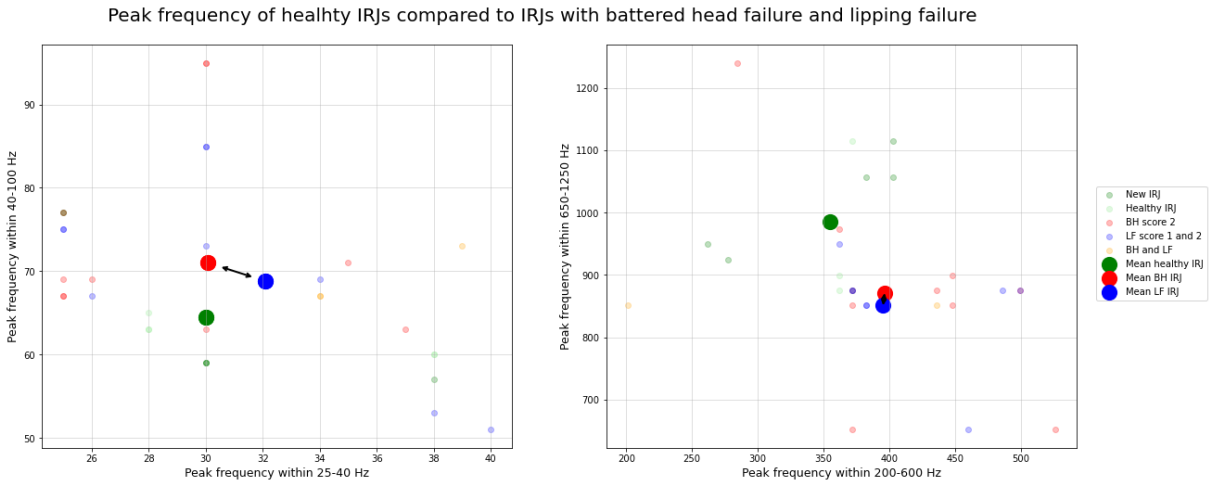


Figure T2 | *Peak frequency of healthy IRJs, IRJs with battered head failure and IRJs with lipping failure.*

Appendix U| Results PCA BHF

Training dataset

In ***Figure U1*** below a 3D plot of the PCA of IRJs with battered head failure is shown, including a front, side and top view to clarify the 3D plot. A cluster of 5 battered head data points can be distinguished from the healthy data points within the PC1-PC2 plot. Two battered head points are within the cluster of healthy points. The best distinction can be seen in the PC1-PC3 plot. All battered head data points are separated from healthy data points. In the PC2-PC3 view also a cluster of healthy IRJ separated from a cluster of battered head IRJs can be observed. However, two battered head data points are within the cluster of healthy IRJs. Based on these observations it appears that PC1 and PC3 are important in the classification of healthy and battered head IRJs.

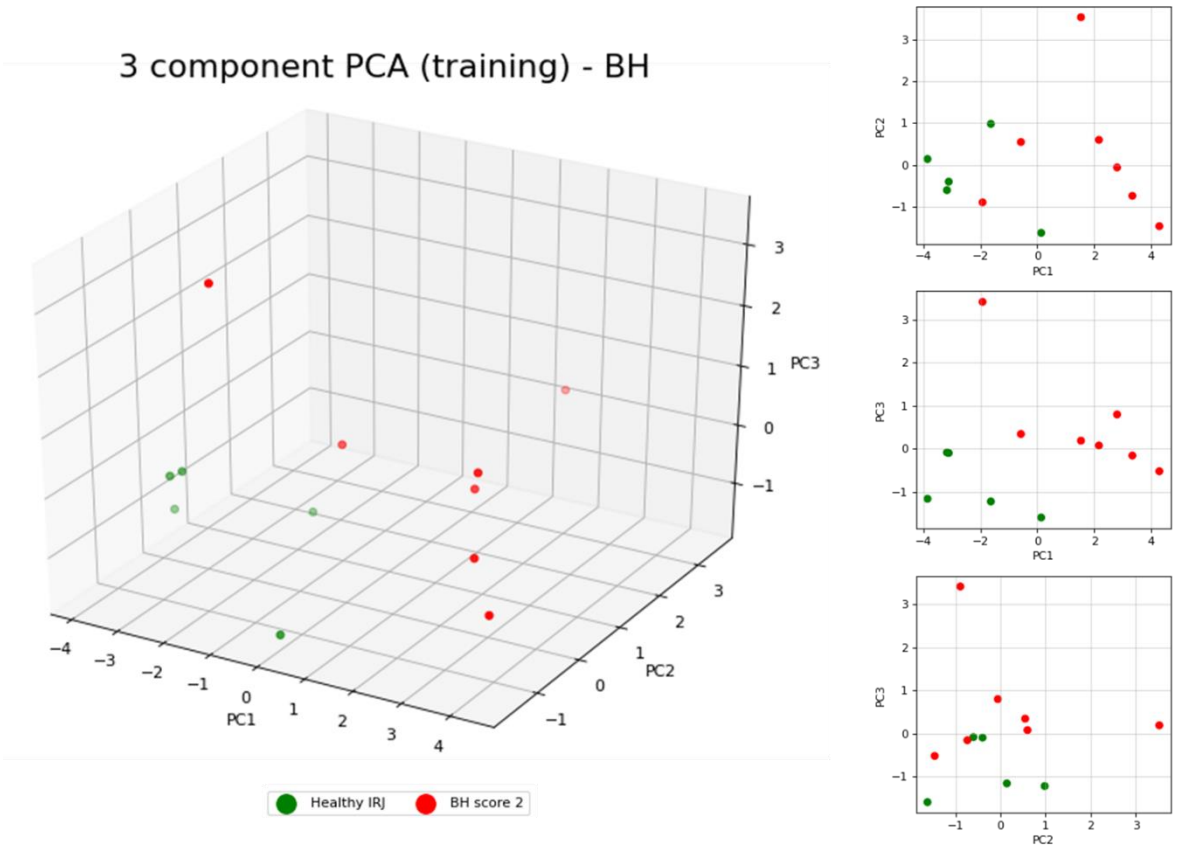


Figure U1 | *3D plot including 2D views with the results of the PCA of the training dataset of IRJs with battered head failure*

Test dataset

In ***Figure U2*** the 3D plot and 2D views of the PCA are given for the test dataset. There is a difference visible in the PC1-PC2 plot in which the red dots (IRJs with battered head failure) are separated from the green dots (healthy IRJs). However, both other views (PC1-PC3 and PC2-PC3) do show a little more overlap.

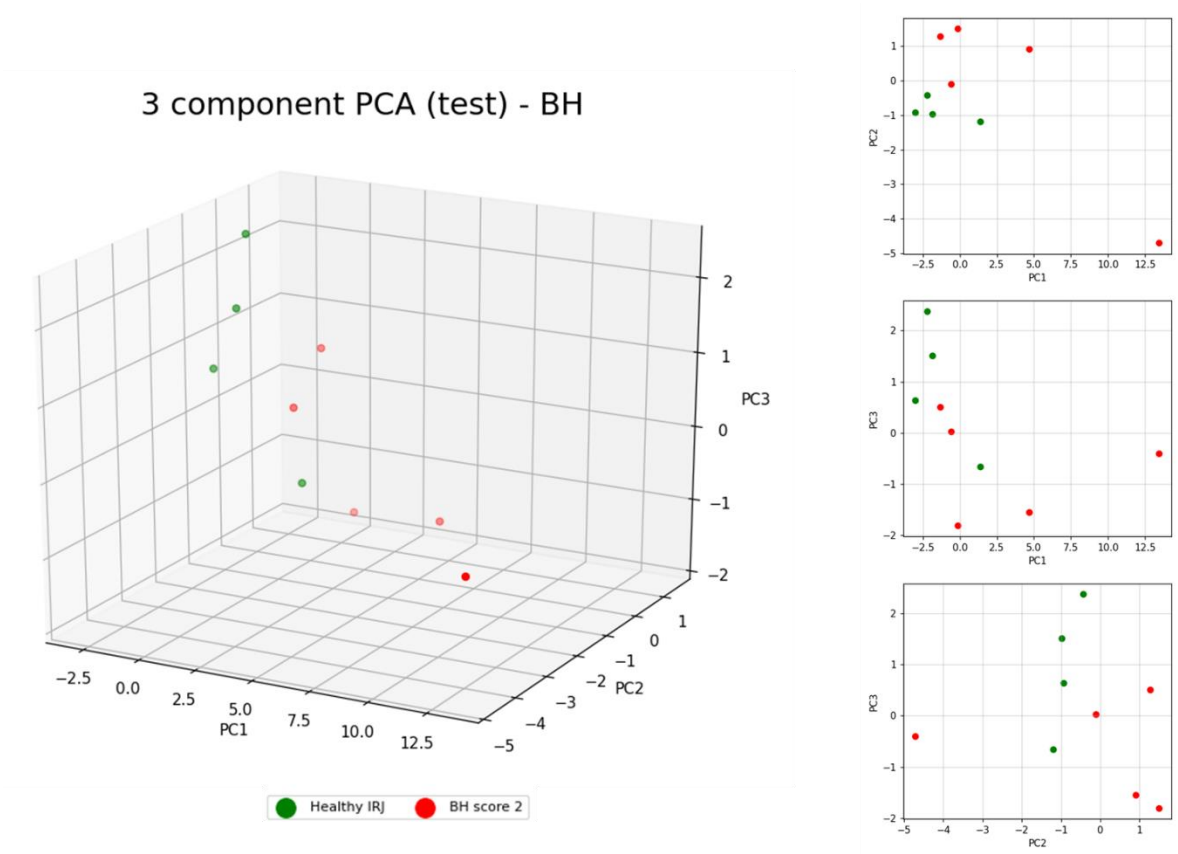


Figure U2 | *3D plot including 2D views with the results of the PCA of the test dataset of IRJs with battered head failure*

Appendix V | Validation results SVM BHF

For validation purposes, the training dataset and test dataset were switched. So, the SVM is constructed based on the test dataset rather than the training dataset. ***Figure V1*** shows the decision plane trained on the test dataset. A performance score of 1.0 is obtained which is equal to the first classifier. However, ***Figure V2*** shows the decision plane tested on the training dataset. It can be seen that the performance score increased from a mere 0.56 to 0.83. Moreover, the decision plane is totally different from the first decision plane. Therefore, it must be concluded that the SVM is not stable and unreliable.

Training SVM on the test dataset

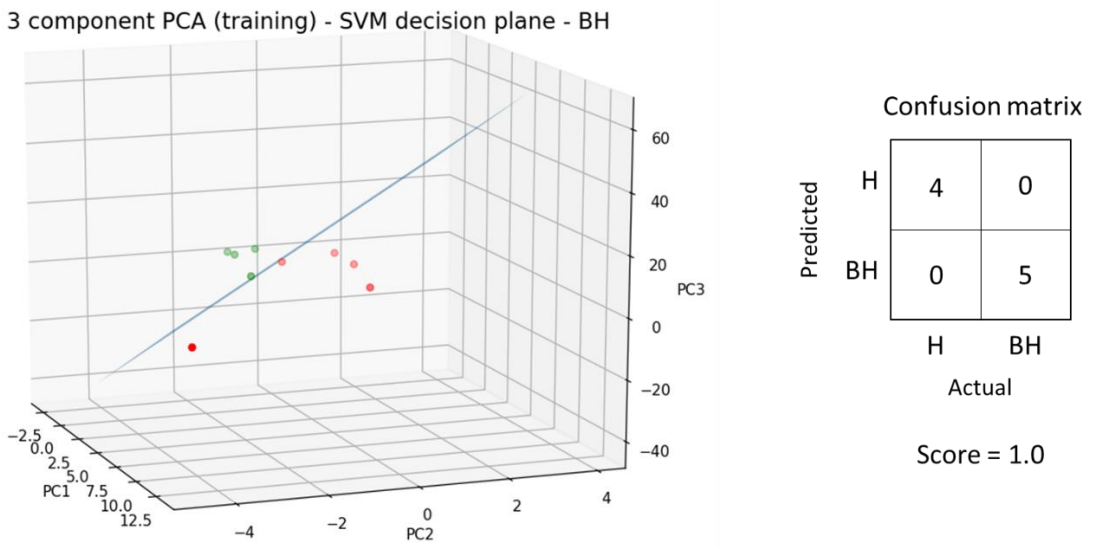


Figure V1 | *3D plot PCA and SVM decision plane including the confusion matrix and training score. The confusion matrix presents the results in terms of correctly predicted healthy (H) or battered head failed (BH) IRJs.*

Testing SVM on the training dataset

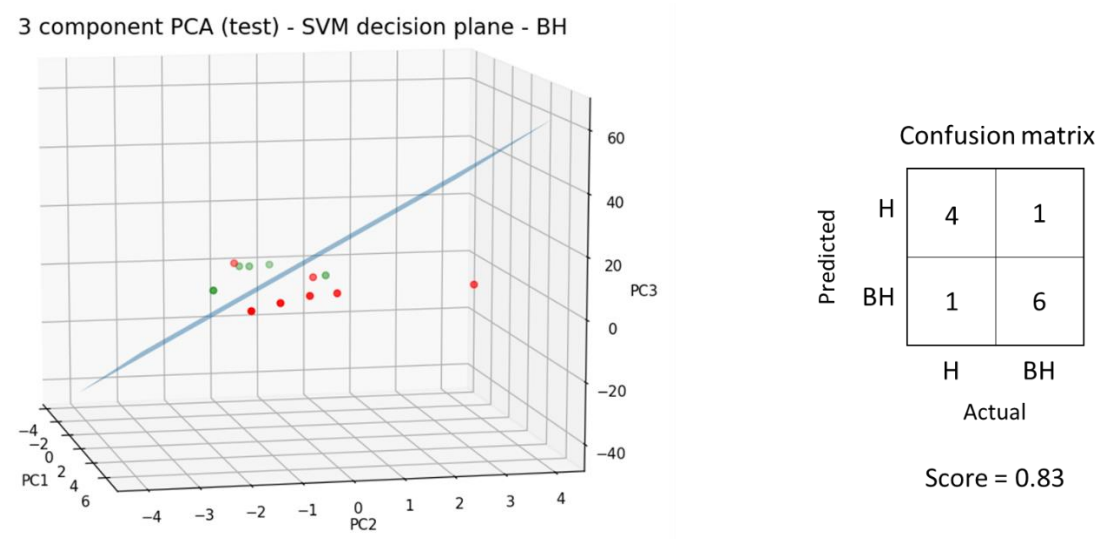


Figure V2 | *3D plot PCA and SVM decision plane including the confusion matrix and training score. The confusion matrix presents the results in terms of correctly predicted healthy (H) or battered head failed (BH) IRJs.*

Appendix W | Results PCA LF

Training dataset

In ***Figure W1*** the 3D plot of the principal components is given, in which the green colour indicates the healthy IRJs and the blue dots represent the IRJs with lipping failure. From the first two principal components, a clear difference is visible between the green and blue data points. Moreover, 4 out of 5 IRJs with lipping failure within PC1 and PC3 are clustered in the upper right corner and separable from the healthy IRJ located in the lower-left corner. However, PC2 compared to PC3 gives a large overlap. There is no distinction visible between these two principal components.

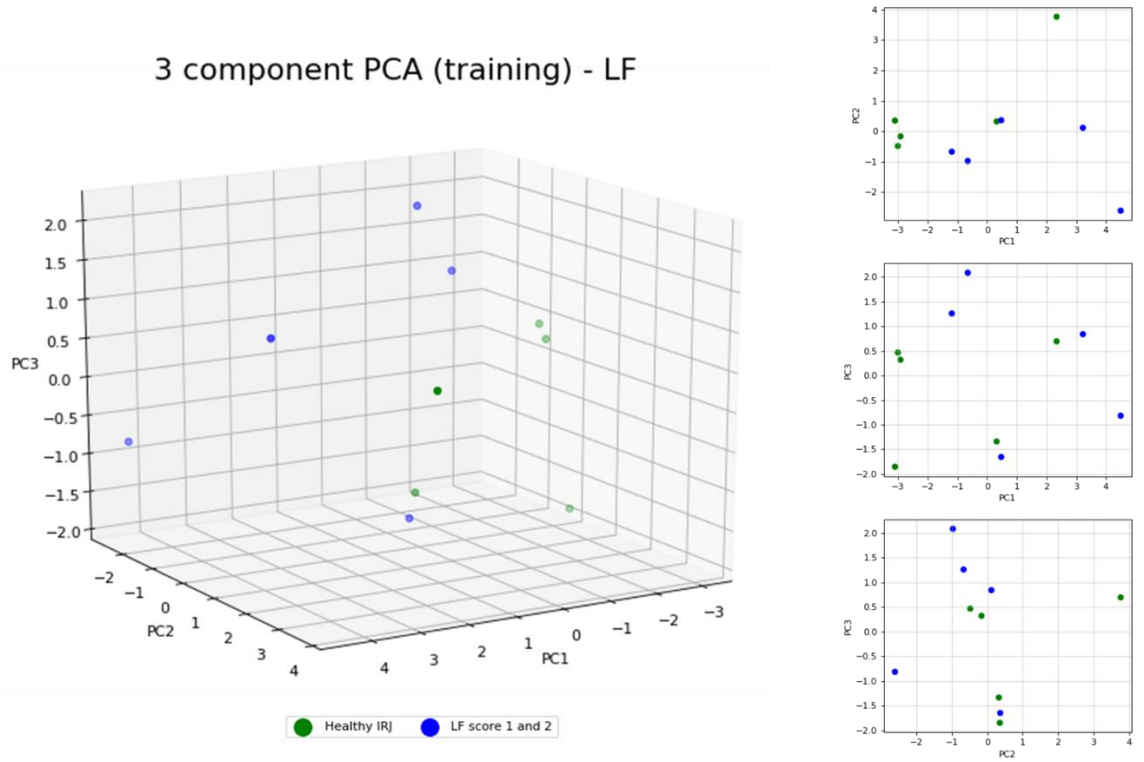


Figure W1 | *3D plot including 2D views with the results of the PCA of the training dataset of IRJs with lipping failure*

Test dataset

In ***Figure W2*** the 3D plot of the principal components of the test dataset is given. The healthy IRJs are again clustered in the upper left corner of the PC1-PC2 plot and the IRJs with lipping failure are clustered in the lower right corner. There is more overlap visible in the two other principal component views.

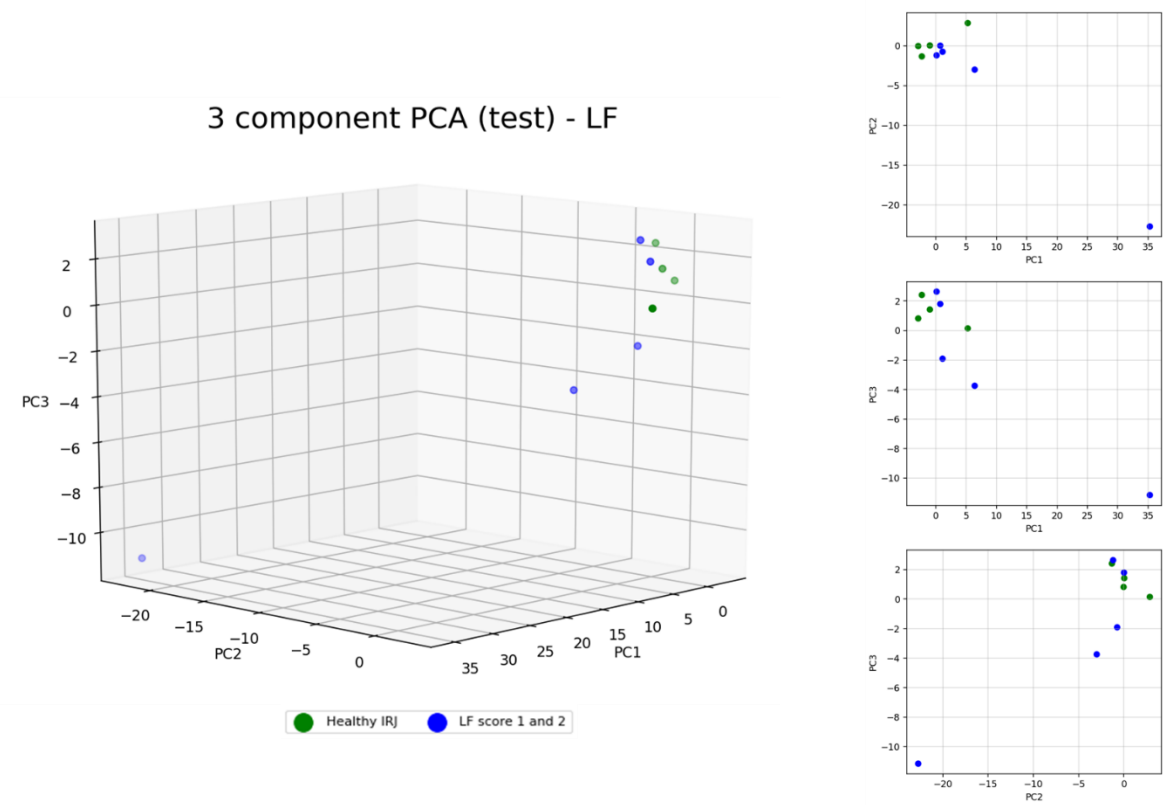


Figure W2 | *3D plot including 2D views with the results of the PCA of the test dataset of IRJs with lipping failure*

Appendix X| Test results SVM

Results SVM Battered head failure

A bad performance score of 0.56 is obtained of which it can be concluded that the classification is inaccurate and unreliable. ***Figure X1*** shows the 3D graph of the test results including the confusion matrix and performance score. It can be seen that 3 of the 4 healthy IRJs are predicted to contain battered head failure. Therefore, the classifier looks strict for healthy IRJs.

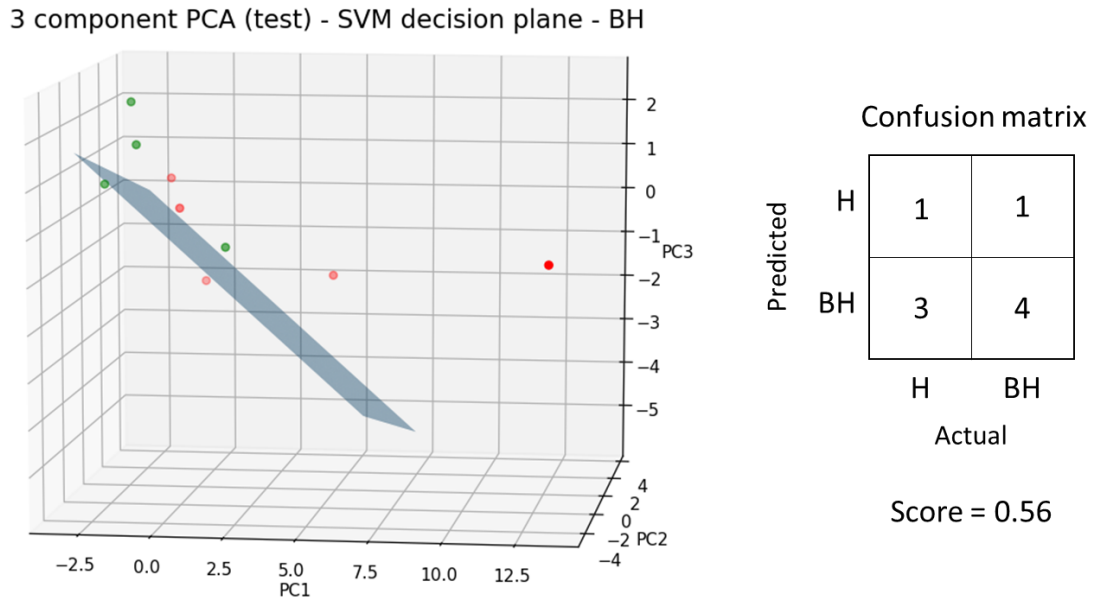


Figure X1 | *Test results of SVM battered head failure*

Results SVM Lipping failure

A performance score of 0.67 is obtained of which it can be concluded that the classification is not very accurate. ***Figure X2*** shows the 3D graph of the test results including the confusion matrix and performance score. It can be seen that 3 of the 4 healthy IRJs are predicted to contain lipping failure. Again, the classifier looks strict for the healthy IRJs.

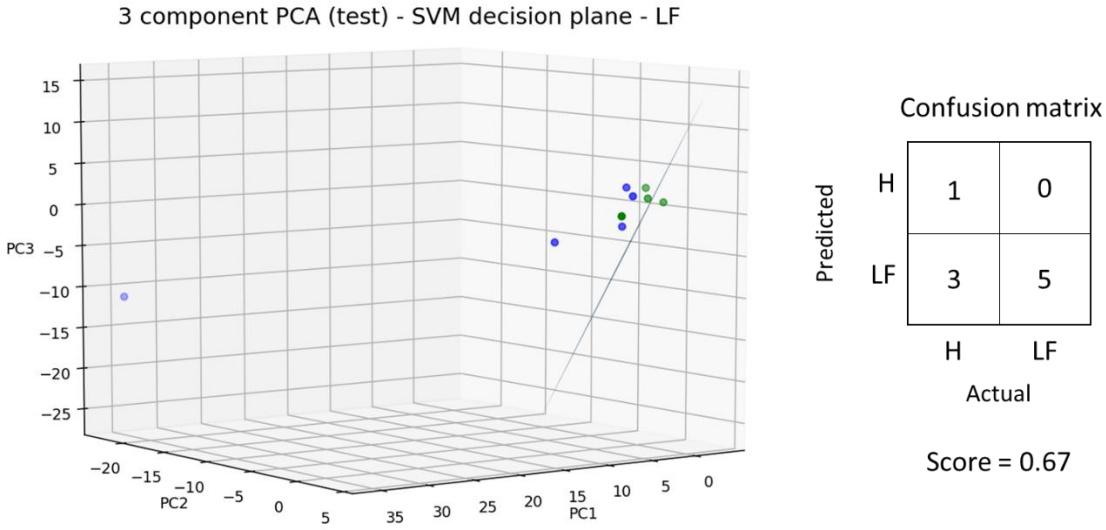


Figure X2 | *Test results of SVM lipping failure*

Appendix Y | Validation results SVM BHF

For validation purposes, the training dataset and test dataset were switched. So, the SVM is constructed based on the test dataset rather than the training dataset. ***Figure Y1*** shows the decision plane trained on the test dataset. A performance score of 0.89 is obtained which is slightly lower than the first classifier. However, ***Figure Y2*** shows the decision plane tested on the training dataset. It can be seen that the performance score increased from a mere 0.67 to 0.78. Moreover, the decision plane is totally different (almost perpendicular) from the first decision plane. Therefore, it must be concluded that the SVM is not stable and unreliable.

Training

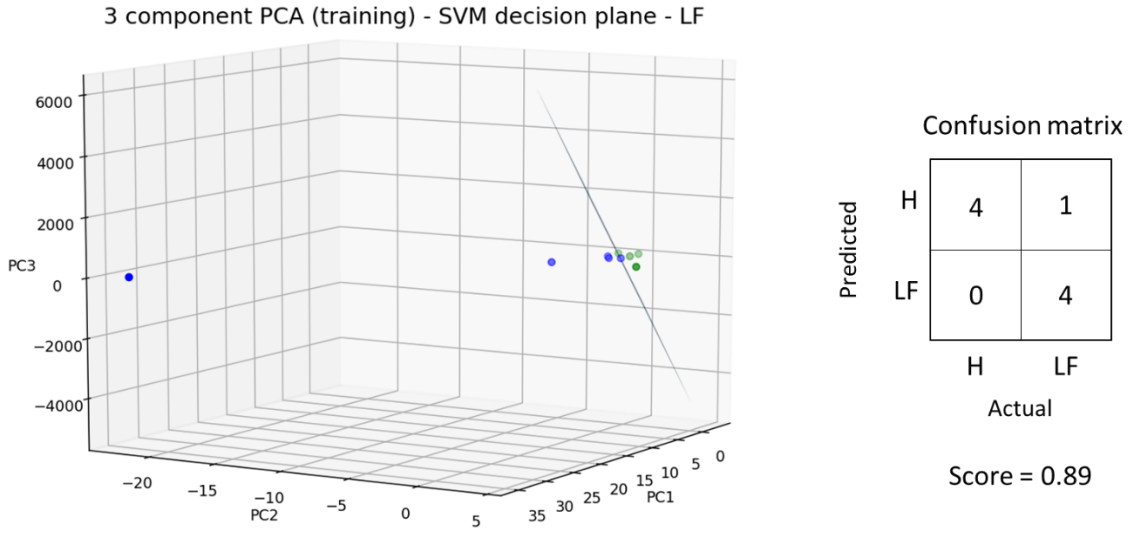


Figure Y1| *3D plot PCA and SVM decision plane including the confusion matrix and training score. The confusion matrix presents the results in terms of correctly predicted healthy (H) or lipping failed (LF) IRJs.*

Test

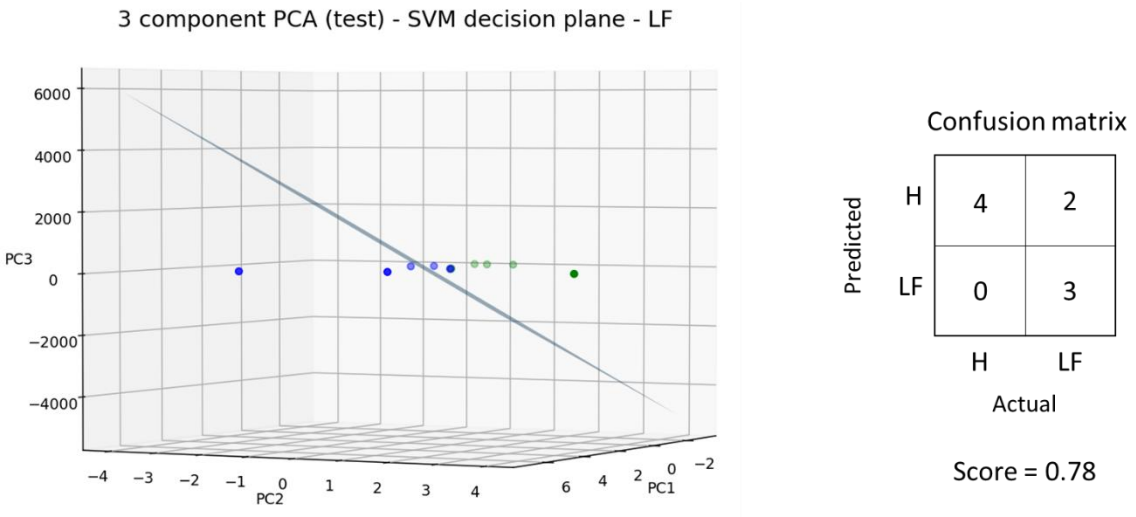


Figure Y2 | *3D plot PCA and SVM decision plane including the confusion matrix and test score. The confusion matrix presents the results in terms of correctly predicted healthy (H) or lipping failed (LF) IRJs.*

Appendix Z | Peak frequency disadvantage

Peak frequency is the frequency at which the peak power occurs and is characteristic of the IRJ. Deviating peak frequencies therefore might indicate a change of health condition. Firstly, there are two attention points, regarding the interpretation of the peak frequency graphs. First of all, the peak frequency should be interpreted as an approximation of the actual peak frequency, due to inaccuracies of the WT. As explained in the methodology (subchapter Wavelet Transform) two parameters of the wavelet had to be chosen, which define a trade-off between time precision and frequency precision. Because of the need for time information and some practical reasons, among them limited calculation power, the parameters are chosen to result in an intermediate frequency resolution. The higher the frequency resolution, the more precise the peak frequency would have been, but at the cost of time resolution and with that the resolution in space. For this reason, a less accurate frequency resolution has been sufficient and therefore, the peak frequency should be interpreted as an approximation of the precise peak frequency. Below an example of a WPS is given in which the peak frequency shows a little error due to the low frequency resolution. ***Figure Z1*** gives an example of IRJ 9 with a peak frequency of 375 Hz, which could also have been 400 Hz or 440 Hz if the frequency resolution of the wavelet transform would have been higher. Therefore, the peak frequency should be interpreted as an approximation of the precise peak frequency.

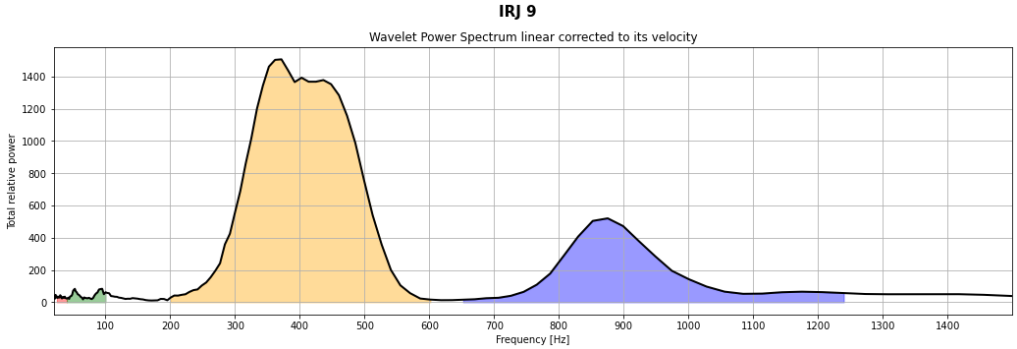


Figure Z1 | *Example of WPS with an uncertain peak frequency*

The second attention point concerns the edge frequencies. As the boundaries of the bandwidths are fixed, the peak frequency could be beyond the bandwidth boundaries. Therefore, all the individual IRJs which coincide with the edges of the frequency bands should be interpreted as a possible lower or higher peak frequency rather than the exact value of the edge frequency. The second scenario is described in ***Figure Z2*** in which a peak value actually belongs to another frequency band that is not defined. The peak frequency of IRJ 51 is, according to the definition of the frequency at the high peak relative power, 200 Hz within frequency band 200-600 Hz. However, as can be seen, the peak actually belongs to a new frequency band 100-300 Hz or even more accurate 150-250 Hz. The most likely peak frequency within frequency band 200-600 Hz is 450 Hz. This explains the inaccurate approach of peak frequency for predefined frequency bands. Also, it explains a more accurate study has to be done in order to investigate the frequency behaviour of IRJs more precisely.

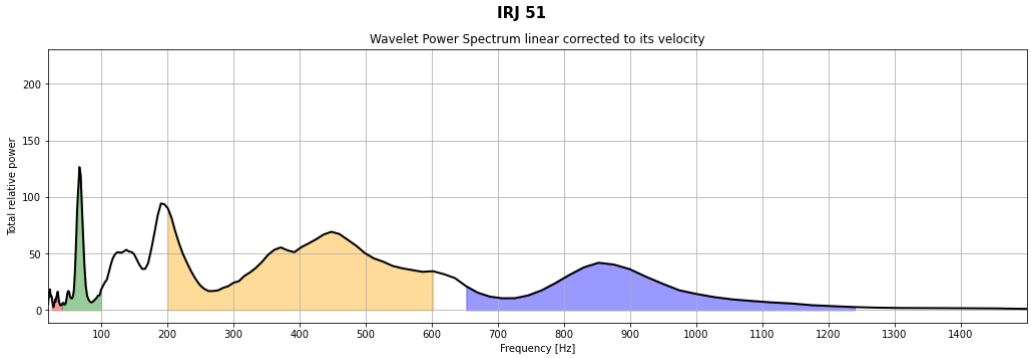


Figure Z2 | *Example of WPS of an unreliable peak frequency within frequency band 200-600 Hz*