



UNIVERSITÀ
DEGLI STUDI
FIRENZE

PhD in
Industrial Engineering

CYCLE XXXVIII

Condition Monitoring of Railway Tracks Using Instrumented Wheelsets: Development of Physics-Based Models and Time-Frequency/Data-Driven Methods

Academic Discipline (SSD) ING-IND/13

Doctoral Candidate
Dr. Bellacci Giovanni

Supervisor
Prof. Pugi Luca

Coordinator
Prof. Vangi Dario

Years 2022/2025

Alerta!

Ezio Raspanti

I'm a selfish person so this goes to me.

DECLARATION

I hereby declare that this submission is my own work and, to the best of my knowledge and belief, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at University of Florence or any other educational institution, except where due references are provided in the thesis itself.

Any contribution made to the research by others, with whom I have been working at the University of Florence or elsewhere, is explicitly acknowledged in the thesis.

Giovanni Bellacci

[A long time ago, in a galaxy far away... 31st October 2025]

ACKNOWLEDGEMENTS

A special thanks goes to my supervisors Luca Pugi, Mani Entezami and Paul Francis Weston for their guidance, mentorship, and constructive feedback along this thorny journey. I'm glad it's over and thanks to you it has been one of the most interesting experiences I've ever done.

To all the people I have collaborated with, Richard Horridge, Luca di Carlo, Giulio Rosano, Marco Fiaschi, Jenny Hillingsworth, and Egidio di Gialleonardo. Thank you for all your assistance and efforts.

To the University of Florence, Italcertifer S.P.A., University of Birmingham, and Politecnico di Milano, thank you for the support and the resources you've provided.

To my family, always ready to lend me a hand and to all my dearest friends, thank you from the bottom of my hearts. It would have been extremely tough without you.

I acknowledge the use of OpenAI's ChatGPT (GPT-5) as a writing and language editing assistant during the preparation of this thesis. All conceptual and analytical content remains my own.

ABSTRACT

The increasing need for reliable and cost-effective railway maintenance demands monitoring solutions capable of capturing infrastructure degradation between traditional in-situ inspections. Instrumented or in-service diagnostic vehicles can fill this gap, yet current systems rely predominantly on inertial and optical sensors, whose performance may be limited in detecting specific defect classes or in harsh environmental conditions. In contrast, the use of instrumented wheelsets (IWS) equipped with strain-gauge-based dynamometric systems, although mature and standardised for vehicle acceptance, has not been fully explored for condition-monitoring purposes.

This thesis investigates whether IWS can serve as an effective and standalone diagnostic technology for track condition assessment, clarifying their capabilities, limitations, and the type of information they can provide beyond conventional sensors. A combination of finite-element modelling, state-space simulations, advanced signal processing, and machine learning is adopted to characterise the sensitivity of wheel–rail contact forces to long-, short-, and very-short-wavelength rail irregularities.

The work demonstrates that contact-force measurements can detect and localise defects such as dipped welds, squats, corrugation, and ballast voids, and can provide meaningful insights into vertical alignment. A validated numerical model confirms an effective bandwidth above 200 Hz for single-bridge configurations, enabling the detection of fast-evolving dynamic phenomena. Time–frequency and wavelet analyses show that IWS outperform optical systems in capturing high-frequency irregularities, while a simplified 3-DOF model successfully reconstructs mean vertical geometry with good agreement to Track Recording Vehicle (TRV) reference data. Furthermore, a machine-learning framework based on Isolation Forest proves effective for railhead fault detection when using both space- and space–frequency-domain features.

These results indicate that IWS can complement, and in some scenarios partially substitute, traditional inertial and optical solutions, offering actionable diagnostic information for infrastructure managers and certification bodies. In particular, the findings can be readily applied to TRVs already equipped with force-measurement systems, such as Diamante, enabling enhanced defect detection without hardware modifications.

Overall, this research establishes a solid methodological foundation for the broader adoption of IWS in predictive maintenance workflows. The proposed hybrid diagnostic framework supports future developments involving deep learning, multisensor data fusion, and large-scale validation under real operating conditions, contributing to more robust, safe, and efficient railway monitoring systems.

Keywords: Condition Monitoring, Railway, Data Analysis, Fault Detection, Signal Processing

PROPOSITIONS

Motivation

- **Limited exploitation of instrumented wheelsets for track condition monitoring.** Although instrumented wheelsets (IWS) are well-established tools for vehicle dynamics studies and wheel–rail interaction analysis, their potential for direct rail condition monitoring remains largely unexplored. Developing methodologies to extract track-related information from contact force measurements could open new perspectives for non-intrusive, continuous, and train-based diagnostics.
- **Industrial and operational interest in data-driven, train-borne monitoring solutions with instrumented wheelsets.** The growing need for predictive maintenance and cost-efficient infrastructure management has stimulated the interest of railway operators and industrial partners in on-board monitoring technologies. The research aims to provide scientific foundations and practical tools to support the industrial development of IWS-based monitoring systems.

Original contribution

- **Development of a simplified model-based approach for track vertical geometry estimation.** A simplified vertical dynamic model of the railway vehicle was developed to estimate the mean vertical alignment of the track using contact force measurements from instrumented wheelsets. This represents a novel model-based approach for rail condition monitoring.
- **Design of a time-domain feature-based anomaly detection framework.** A complete pipeline for the detection of wheel–rail contact anomalies was implemented, including a graphical interface for data exploration. Statistical time-domain features (mean, RMS, kurtosis, skewness, crest factor, etc.) were used as inputs to several one-class classifiers (OCSVM, Isolation Forest, LOF), whose performances were systematically compared for detecting tread surface defects.
- **Extension of the detection framework to frequency-domain features and feature fusion.** Frequency-domain features derived from the FFT (spectral centroid, spread, skewness, kurtosis, flatness, slope, etc.) were analysed and compared with time-domain features. A feature fusion strategy combining both domains was proposed and shown to improve detection performance by increasing true positives and reducing false positives.
- **Preliminary exploration of deep learning approaches for contact-force-based condition monitoring.**
Initial investigations on deep learning architectures (Autoencoder, Convolutional

Autoencoder, Conditional Variational Autoencoder) were carried out using time–frequency images derived from the STFT and Hilbert–Huang Transform of the contact force signals, paving the way for future developments in data-driven rail diagnostics.

Open questions

- **Trustworthiness of data-driven approaches.** Given the limited ground truth and the lack of a complete track characterization, the true effectiveness of the proposed data-driven methodologies remains uncertain. To what extent can machine learning techniques be considered trustworthy for anomaly detection, fault identification, and classification in this context?
- **Sensitivity of contact forces to shorter-wavelength track irregularities.** Can contact force measurements be effectively used to estimate track geometry irregularities with wavelengths shorter than the D1 range (3–25 m), such as local defects or corrugations?
- **Reliability of simplified vehicle models for track geometry estimation.** Can a simplified vehicle model (with up to six degrees of freedom) provide sufficiently accurate estimations of the track vertical alignment using only contact forces, without relying on complex multibody simulation tools?
- **Potential of deep learning methods for feature extraction and defect recognition.** Can deep learning models, applied to time–frequency representations of contact force signals (e.g., STFT or Hilbert–Huang Transform), outperform traditional feature-based approaches in identifying and classifying track or wheel defects?

INDEX

Declaration	I
Acknowledgements	III
Abstract	V
Propositions	VII
Index	IX
List of Figures	XIII
List of Tables	XIX
Synopsis	XXIII
1 Introduction	26
1.1 Background and Motivation	
1.2 Problem Statement	
1.3 Research Objectives	
1.4 Research Questions And Hypotheses	
1.5 Methodological Approach	
1.6 Thesis Structure	
1.7 Expected Contributions	
1.8 References	
2 Architecture of instrumented wheelsets	32
2.1 Introduction	
2.1.1 Alternative Measurement Approaches	
2.1.2 Requirements of the Measuring Chain	
2.1.3 Practical implications for design and calibration	
2.2 Axle bending strain measurements for contact forces inverse identification	
2.3 Wheel-web based strain measurements	
2.3.1 Gauge placement and layout strategies	
2.3.2 Decoupling of Q, Y and contact position d_R	
2.3.3 Industrial implementations	
2.4 Combined approach: axle and wheel-web strain measurements	
2.4.1 Telemetry, synchronisation and inversion	
2.4.2 Sampling strategies: constant time and constant space	
2.4.3 Bandwidth considerations	

2.5	References	
3	Validation Of A Wheelset Finite Element Model For Static Force Inverse Identification	48
3.1	Introduction	
3.2	Wheelset Finite Element Model	
3.3	Finite Element Static Analysis	
3.4	Comparison with experimental data	
3.5	Inverse force identification: further FE model validation	
3.6	Conclusions and Future Developments	
3.7	References	
4	Frequency Response Analysis Of The Instrumented Wheelset	62
4.1	Introduction	
4.2	Tested strain gauges layout	
4.3	Wheelset FE simulation frequency response analysis	
4.4	Comparison between FEM, state-space-model and experimental transfer functions	
4.5	References	
5	Condition Monitoring With Time-Frequency Analysis And Wavelet Transform	75
5.1	Chapter Summary	
5.2	Introduction	
5.3	The Wavelet Analysis	
5.4	Results Discussion	
5.5	Conclusions and Future Developments	
5.6	References	
6	Model-Based Railway Condition Monitoring With Contact Force Measurements: Track Geometry	85
6.1	Chapter Summary	
6.2	Track geometry estimation: Introduction	
6.3	State-of-the-art review of equipment and methods for railway condition monitoring	
6.4	Review of models and methods for on-board railway condition monitoring	
6.5	Instrumented wheelsets and data acquisition	
6.5.1	Model description and simulation methods	
6.5.2	Linear state space representation of the half-a-coach vehicle model	
6.6	Results and Discussion	
6.7	Error Analysis	
6.8	Conclusions and future developments	
6.9	Future developments: 6-DOF coach model concept	
	Abbreviations	108
6.10	References	

7	Data-driven methods: Novelty Detection Using Contact Force Space Features and Machine Learning	113
7.1	Introduction: big data analysis	
7.2	State of the art of data-driven methods	
7.3	Datasets and pre-processing	
7.3.1	Spatial resampling and windowing	
7.3.2	Data Normalization	
7.3.3	Novelty Detection Models	
7.4	Features extraction	
7.4.1	Hjorth Parameters	
7.5	Anomaly Detection Framework: Setup and Evaluation	
7.5.1	Framework Setup	
7.5.2	Evaluation metrics	
7.6	Results Discussion	
7.6.1	Normalization Influence	
7.6.2	Influence of signal windowing	
7.7	Graphical User Interface (GUI)	
7.8	Conclusions	
	Abbreviations	154
7.9	References	
8	Data-driven methods: Novelty Detection Using Contact Force Spectra Features and Machine Learning	159
8.1	Introduction	
8.2	State of the art and literature review	
8.3	Time-Frequency analysis	
8.3.1	Definition of wavelength bands	
8.4	Features extraction	
8.5	Data Preparation	
8.6	Machine learning and novelty detection pipeline	
8.7	Results and discussion	
8.7.1	Influence of signal windowing	
8.8	Conclusions and future developments	
	Abbreviations	173
8.9	References	
9	Data-driven methods: Combining One-Class novelty detection Frameworks	177
9.1	Introduction	
9.2	Limitations	
9.3	Results and Discussion	
9.3.1	Considerations over inference time	
9.4	Conclusions	
10	Original contribution	185

10.1 List of Publications

11 Conclusions 187

11.1 Main findings

11.1.1 Finite Element Modelling and Frequency Response Characterisation

11.1.2 Finite Element Modelling and Frequency Response Characterisation

11.1.3 Model-Based Approach

11.1.4 Data-Driven Approaches

11.2 Overall Contributions

11.3 Future Developments

11.4 Final Remarks

LIST OF FIGURES

Figure 2.1: High-speed wheelset. CAD rendering.	33
Figure 2.2: Wheelset nomenclature.	33
Figure 2.3: Measuring sections of the instrumented axle method [12].	37
Figure 2.4: Axle instrumentation detail: journal quadrature bridges typical of measuring sections 2, and 7 (referring to Figure 2.3).	38
Figure 2.5: Possible full-bridge strain gauge configurations [9].	40
Figure 2.6: Full-bridge strain gauge configurations arrangements insensitive to thermal and centrifugal deformation [12].	41
Figure 2.7: Telemetry system.	42
Figure 3.1: Schematic representation of Italcertifer s.p.a's test-bench for instrumented wheelsets testing. [15]	49
Figure 3.2: stress distribution on the wheel-web field side due to a) compression only, b) compression and bending moment.	51
Figure 3.3: Wheel mesh.	52
Figure 3.4: Axle mesh.	52
Figure 3.5: Brake disk sub-assemblies. a) sub-assembly 1: spacer; b) sub-assembly 2: sleeve; c) sub-assembly 3: disk.	53
Figure 3.6: Full FE model of the instrumented wheelset.	53
Figure 3.7: Wheelset FE model used for the static analysis.	54
Figure 3.8: Constraint scheme with applied loads directions: side A roller (instrumented wheel), side B pinned (non-instrumented wheel), Q vertical and Y lateral contact forces.	54
Figure 3.9: Detail of the adopted constraint solution.	54
Figure 3.10: Instrumented wheelset FE model. Load case Q = 50 kN and Y = 20 kN for various inclination angles. Probe points on the wheel-web field side are highlighted in blue (180° position).	55
Figure 3.11: Schematic representation of sensors positioning for the static test ("survey"), covering most of the wheel-web radius for an extensive evaluation of the surface strain.	56

Figure 3.12: Comparison between numerical and experimental results for two loading configurations: (a) $Q = 50$ kN and $Y = 0$ kN and (b) $Q = 50$ kN and $Y = 20$ kN. The strain gauge test is aligned with the z-axis of the absolute reference system. An undeformed section of the wheel is displayed to facilitate the visualization of the points of interest on the wheel web field side.	57
Figure 4.1: Schematic representation of the instrumented wheelset tested strain gauges layout.	63
Figure 4.2: Wheatstone bridge schematic representation: a) full-bridge configuration [6], b) quadrature sine (left) and cosine (right) Wheatstone bridges.	63
Figure 4.3: Wheelset suspended configuration.	64
Figure 4.4: Instrumented hammer and amplifier used for the hammer test.	65
Figure 4.5: Picture of the tested instrumented wheelset in the laboratory.	66
Figure 4.6: Acquisition board: imc CRONOScompact.	66
Figure 4.7: Hammer test excitation signal (top) and example of a single strain gauge output (bottom).	66
Figure 4.8: Recorded strain gauge output after hammering the wheel.	67
Figure 4.9: Strain gauge output spectra.	67
Figure 4.10: Wheelset modal and frequency response analysis results: axle bending. a) mode 2; b) mode 4; c) mode 6.	69
Figure 4.11: Wheelset modal and frequency response analysis results: Wheels' modal shapes. a) mode 8; b) mode 9.	69
Figure 4.12: Comparison of the state-space (blue) and FEM (red) transfer functions.	70
Figure 4.13: Comparison of the experimental (blue) and state space (red) transfer functions with Coherence.	72
Figure 5.1: Initiation of a top irregularity. Comparison of track geometry longitudinal level and recorded vertical forces scaled with vehicle speed, for two wheelsets on the same bogie. Top: leading wheelset. Middle: trailing wheelset. Bottom: normalized longitudinal level with maximum admissible value. Red lines indicate the left side, blue lines indicate the right side of the TRV.	76
Figure 5.2: Vertical wheel-rail force under impact excitation: train passing over a dipped-joint [8].	77
Figure 5.3: Scalogram of the vertical contact force. Top: scalogram; bottom: vertical contact force. (a) Initial 14 s recording for a broader view; (b) and (c) zoomed-in views of the scalogram and forces showing expected irregularities.	79
Figure 5.4: Normalized scalogram magnitude of vertical contact forces at 87 Hz, corresponding to 31 cm wavelength irregularities.	79
Figure 5.5: Identification of force impulses through scalogram magnitude selection. (a) 87 Hz scalogram row corresponding to 31 cm wavelength irregularities. (b) 282 Hz scalogram row corresponding to 9.5 cm wavelength irregularities.	80
Figure 5.6: Identification of force impulses through higher scalogram magnitudes in the 87–282 Hz frequency range. (a) Identified irregularities in the whole recording; (b) zoom of impulses.	81

Figure 6.1: An example of a Rete Ferroviaria Italiana's (RFI) TRV for conventional lines. [4]	87
Figure 6.2: RFI diagnostic fleet and their characteristics. [4].	87
Figure 6.3: Schematic representation of available sensors mounted on the RFI's Diamante TRV for high-speed lines diagnosis [17].	88
Figure 6.4: Schematic representation of a railway coach.	93
Figure 6.5: Schematic representation of the system. Simplified 3 DOF $\frac{1}{2}$ a vehicle model, for vertical displacement estimation only.	94
Figure 6.6: Example of geometric filtering of the vertical alignment [50], λ refers to wavelength.	96
Figure 6.7: Schematic representation of the process.	97
Figure 6.8: D1 wavelength range band-pass filtered data. Comparison between reference measured and simulated vertical displacements. Blue line - measured data; red line – simulation; a) dataset 1; b) dataset 2; c) dataset 3; d) dataset 4.	102
Figure 6.9: Spectra comparison of a) dataset 1; b) dataset 2; c) dataset 3; d) dataset.	102
Figure 6.10: Map of the railway infrastructure in Europe, grouped by maximum line speed. Green shades represent 90-200 km/h lines [54].	103
Figure 6.11: Expected behaviour of the R value for the investigated train speed range, showcasing the applicability range of the model.	104
Figure 6.13: Schematic representation of the model concept. Simplified 5-DOF vehicle model, for vertical displacement estimation only.	107
Figure 6.14: Schematic representation of one-dimensional model often used to represent rubber springs and Vampire models [56].	108
Figure 7.1: C12 and C14 curvature data.	121
Figure 7.2: C12 and C14 speed data.	121
Figure 7.3: Examples artificial numerical faults waveform created from the literature. a) weld, b) squat, c) joint and/or welds.	122
Figure 7.4: Figure illustrates global anomalies (x_1 , x_2), a local anomaly x_3 and a micro-cluster c_3 [16].	125
Figure 7.5: Detail example of raw force signal (top), with its calculated Mobility (middle) and Complexity (bottom), showing the noisiness and low frequency fluctuations of the parameters.	129
Figure 7.6: Detail example of raw force signal (top), with LP filtered Mobility (middle) and LP filtered Complexity (bottom), showing the high reactivity of these parameters to data fluctuations.	129
Figure 7.7: Unrealistic anomalies detected by the LOF algorithm. Effect of imposed zeros at the beginning and at the end of the feature matrix, due to feature calculation procedure with sliding windows.	133
Figure 7.8: PR curves obtained varying the window length L_W for the OCSVM classifier. a) $L_W=10$ elements; b) $L_W=20$ elements; c) $L_W=50$ elements; d) $L_W=100$ elements.	135

Figure 7.9: PR curves obtained varying features selection. Features are summarised in Table 7.7.	139
Figure 7.10: Confusion matrices of the statistical space-features method. a) OCSVM; b) IF; c) LOF classifier. Percentages are shown for total retrieved instances (“total”); percentage over the GT, and percentage over the retrieved instances after applying the optimal threshold for the FP (“thresh”).	144
Figure 7.11: PR curves of the statistical space-features method. a) OCSVM; b) IF; c) LOF classifier.	146
Figure 7.12: Classifiers outliers detection output: a) OCSVM (threshold: 0.2408881); b) IF (threshold: 0.0510494); c) LOF (threshold: 1.6000777).	148
Figure 7.13: OCSVM detection output: a) default threshold (> 0); b) optimal threshold ($tr = 0.2408881$).	150
Figure 7.14: Anomaly detection GUI.	150
Figure 7.15: Anomaly detection GUI: “Options windows” for analysis parameters selection. a) filter parameters; b) feature selection; c) Model hyperparameters; d) pre-processing parameters; e) decision function threshold.	153
Figure 8.1: Typical wavelength and frequencies for common vehicle and track components, defects and faults [14].	162
Figure 8.2: STFT of the C14 recording.	165
Figure 8.3: C12 contact force recording with ground truth anomalies.	166
Figure 8.4: Confusion matrices of the statistical frequency-features method. a) OCSVM; b) IF; c) LOF classifier. Percentages are shown for total retrieved instances (“total”); percentage over the GT, and percentage over the retrieved instances after applying the optimal threshold for the FP (“thresh”).	168
Figure 8.5: PR curves of the statistical frequency-domain features method. a) OCSVM; b) IF; c) LOF classifier.	170
Figure 8.6: Classifiers outliers detection output: a) OCSVM (threshold: 0.045730750); b) IF (threshold: 0.038239348); c) LOF (threshold: 0.709274613).	171
Figure 8.7: OCSVM detection. a) threshold: 0.0457307504553889, proposed discrete value; b) threshold: 0.0457307504553900, interpolated value.	172
Figure 9.1: PR curves of a) OCSVM; b) IF; c) LOF classifiers with combined set of features.	181
Figure 9.2: Confusion matrices of a) OCSVM; b) IF; c) LOF classifiers. Statistical space-domain-feature method combined with statistical frequency-domain-feature method. Percentages are shown for total retrieved instances (“total”); percentage over the GT, and percentage over the retrieved instances after applying the optimal threshold for the FP (“thresh”).	182
Figure 9.3: Classifiers outliers detection output: a) OCSVM (threshold: 0.0231815); b) IF (threshold: 0.0436101); c) LOF (threshold: 1.6995327).	184
Figure 11.1: Summary of classifiers’ performance. Comparison among the tested methods.	189
Figure 11.2: Retrieved instances with optimal decision function threshold applied for each classifier and case. Comparison among the tested methods.	189

LIST OF TABLES

Table 3.1: Experimental and FE strains, for $Q = 50$ kN and $Y = 0$ kN, relative to Fig. 3(a).	58
Table 6.1: EN 13848 – 1 standardised wavelength ranges.	88
Table 6.2: Optimized model parameters for the vehicle speed of 160 km/h and D1 range.	98
Table 6.3: Nomenclature and characteristics of parsed datasets.	98
Table 6.4: Error and correlation analysis for the recorded and simulated mean vertical alignment.	104
Table 7.1: State of the art summary.	118
Table 7.2: Training and fitting dataset characteristics.	120
Table 7.3: Extracted spatial features.	126
Table 7.4: Hjorth Parameters [33].	128
Table 7.5: List and description of Python libraries used for the data-driven analysis.	130
Table 7.6: OCSVM classifier comparison varying L_W . Detection performance summary.	136
Table 7.7: OCSVM classifier comparison varying features. Detection performance summary.	139
Table 7.8: OCSVM comparison varying features scaling method. Detection performance summary.	141
Table 7.9: IF classifier comparison varying features. Detection performance summary.	142
Table 7.10: Classifiers comparison and detection performance summary.	143
Table 8.1: Wavelength bands definition.	163
Table 8.2: Extracted spectral features.	164
Table 8.3: Classifiers comparison and detection performance summary.	167
Table 9.1: Classifiers comparison and detection performance summary.	179

SYNOPSIS

Ensuring the safety, reliability, and availability of railway networks remains a continuous challenge for infrastructure managers. Traditional maintenance strategies rely on periodic on-site inspections, which are often unable to match the actual rate of track degradation. As a result, defects may develop between two successive inspections, posing safety and operational risks. The increasing adoption of diagnostic and in-service monitoring technologies has therefore become crucial to move from reactive or preventive maintenance to predictive maintenance (PdM) and condition-based maintenance (CBM).

In this context, instrumented wheelsets (IWS), wheelsets equipped with strain gauges capable of directly measuring the contact forces exchanged at the wheel–rail interface, represent a promising yet underexplored technology. While IWS are widely used for homologation, derailment risk assessment, and static load verification, their potential use for in-service condition monitoring has not yet been fully investigated. This doctoral research aims to assess whether IWS can be effectively employed for condition monitoring purposes, to identify their strengths and limitations, and to explore how their data can complement or even replace more traditional sensing technologies such as inertial or optical systems.

Research Objectives and Structure

The main objective of this PhD work is to demonstrate that contact-force measurements obtained from IWS can provide meaningful information about track condition and geometry, potentially enabling fault detection and degradation assessment. To achieve this, the research combines finite element modelling, signal processing, and machine learning (ML) techniques within an integrated framework for data analysis and interpretation.

The thesis is structured into three main parts:

- Modelling and Validation of the Instrumented Wheelset
- Traditional and Model-Based Data Analysis Techniques
- Data-Driven and Machine Learning Approaches for Anomaly Detection

Part I – Modelling and Experimental Validation

The first phase focuses on developing and validating a finite element (FE) model of the instrumented wheelset. The model was calibrated using material properties and geometrical parameters consistent with the experimental setup provided by the industrial partner. Static analyses were performed to validate the stress and strain fields of the wheel-web induced by vertical and lateral loads, confirming the accuracy of the strain gauge positioning on the wheel web and axle.

Subsequently, the FE model was linearised into a state-space (SS) model to perform dynamic analyses. The numerical transfer functions obtained from the SS model were compared with those derived from experimental modal tests using an instrumented hammer. The comparison showed good agreement up to approximately 260 Hz, confirming that the simplified linear model

correctly reproduces the dominant dynamic behaviour of the wheelset in this range. Although the experimental setup (small deformations and low excitation levels) limits the generalisation of the results to real operational conditions, the analysis provided an indicative passband of around 200 Hz for a single strain-gauge bridge. This represents a realistic upper limit for dynamic measurements before significant signal attenuation occurs. The model validation stage therefore provides a solid foundation for the subsequent simulation-based and data-driven investigations, proving that the adopted modelling approach captures the essential features of the IWS dynamic response.

Part II – Traditional and Model-Based Analysis

The second part of the research explores classical data analysis techniques for fault localisation and track geometry reconstruction.

Wavelet Transform (WT) analysis was applied to the vertical contact-force signals acquired from IWS during test runs. The WT successfully identified transient, impulsive, and short-wavelength irregularities such as squats, dipped welds, and ballast voids within the 9–31 cm wavelength range. Although a quantitative validation was not possible due to the absence of ground-truth data, a qualitative visual comparison demonstrated that most major vertical-force impulses corresponded to the expected irregularities.

Complementing the signal-based methods, a reduced-order 3-DOF model of a half-coach vehicle was developed to estimate the mean vertical alignment (D1 range: 3–25 m wavelength) using contact-force data as inputs. The simulated displacements were compared with reference data from Track Recording Vehicles (TRVs), showing good agreement in waveform and wavelength with correlation coefficients (R) up to 0.84 for speeds between 100 and 200 km/h. Despite the model's simplifications, these results demonstrate that vertical contact forces can be used to estimate track geometry features, particularly when combined with model-based estimators or data-fusion strategies.

Part III – Data-Driven and Machine Learning Frameworks

The final part of the thesis introduces data-driven anomaly detection approaches applied to contact-force signals. Three one-class classifiers—One-Class Support Vector Machine (OCSVM), Isolation Forest (IF), and Local Outlier Factor (LOF)—were trained using spectral and statistical features extracted from the contact-force data. The models were benchmarked using both space-domain and space-frequency-domain features, with varying window lengths, feature combinations, and scaling strategies. Results indicate that the performance of the classifiers strongly depends on the selected feature set. When using combined spatial and spectral features, the IF achieved the most balanced and robust detection performance, retrieving 78 true positives (out of 81) with only four false positives, and showing the highest overall Precision, Recall, and F1-score among the tested models. Although the OCSVM reached a comparable number of true positives, it exhibited a higher false-positive rate, particularly when using spectral features alone, which reduced its overall reliability. The LOF consistently showed the lowest performance, confirming its limited suitability for this application due to high computational demand and poor precision. These findings demonstrate that combining spatial and spectral features improves the general detection robustness, and that the Isolation Forest, thanks to its low false-positive rate and computational efficiency, represents the most promising candidate for future implementation in real-time or embedded condition-monitoring frameworks.

Main Findings and Contributions

The major contributions of this doctoral research can be summarised as follows:

Model Validation and Bandwidth Characterisation

A validated FE-based and state-space model of the IWS was developed and correlated with experimental modal test data, identifying an indicative passband above 200 Hz for the single Wheatstone bridge and confirming the model's predictive capability up to 260 Hz.

Extension of Traditional Signal Processing Techniques

Wavelet and time–frequency analyses were successfully applied to contact-force data, showing that IWS can detect and localise short-wavelength rail irregularities beyond the capability of conventional optical systems.

Model-Based Estimation of Track Geometry

A simplified 3-DOF model was proposed for reconstructing mean vertical alignment (D1 range) using only measured contact forces, achieving good correlation with TRV reference measurements.

Machine Learning Framework for Fault Detection

Data-driven anomaly detection pipelines were implemented and compared, demonstrating that IF provides the most reliable framework for railhead fault detection when combined with both space and space-frequency-domain features.

Hybrid Methodology and Future Applications

The combination of model-based, traditional, and data-driven methods represents a robust, hybrid framework for condition monitoring, supporting future integration with deep learning architectures and real-world field data.

Conclusions and Outlook

This research demonstrates that instrumented wheelsets, when properly modelled, calibrated, and processed, are capable of providing valuable diagnostic information about the condition of railway tracks. Their sensitivity to dynamic phenomena at the wheel-rail interface makes them suitable for detecting both long- and short-wavelength defects, bridging the gap between inertial and optical monitoring systems.

The proposed methodologies, ranging from physical modelling to machine learning, form a coherent diagnostic workflow that can support predictive maintenance strategies and infrastructure asset management.

Future research will focus on validating the proposed approaches in real operational environments, expanding datasets to include true ground-truth information, and exploring deep learning and data fusion methods for automated defect classification and localisation.

Ultimately, this work provides a solid foundation for the broader adoption of IWS technology as a core component of next-generation railway monitoring systems, contributing to safer, more efficient, and cost-effective railway operations.

1 INTRODUCTION

1.1 Background and Motivation

Ensuring the safety, reliability, and availability of railway networks remains a constant challenge for infrastructure managers and operators. Modern railways are required to meet increasingly stringent performance and safety standards while simultaneously minimising maintenance costs and downtime. This challenge is particularly relevant in the context of predictive and condition-based maintenance, which aims to replace scheduled interventions with strategies driven by the actual condition of the assets.

Track degradation and faults contribute to the deterioration of track quality, ride comfort, and vehicle integrity. Track degradation prediction can generally be divided into long-term and short-term approaches. Long-term models aim to study the progressive deterioration of track components over their service life, while short-term predictions focus on faults that emerge more abruptly and require immediate intervention.

Long-term degradation has been successfully modelled using data from Track Recording Vehicles (TRVs) through various techniques such as Bayesian inference, Monte Carlo Markov Chains, and Data Mining approaches.

Short-term degradation prediction, on the other hand, has increasingly relied on Machine Learning (ML) techniques, with Random Forest regression and Artificial Neural Networks (ANNs) yielding the most promising results.

When considering track faults, research has largely concentrated on short-wavelength defects, as these are typically more disruptive and demand more frequent inspection and maintenance. Such faults include track geometry irregularities caused by ballast loss, rail breaks, fatigue, or local misalignment, often addressed through ML frameworks combining models such as Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks, or regressive neural networks coupled with nearest-neighbour searches [1].

The collection of data from both diagnostic and in-service vehicles significantly enhances asset condition assessment, supporting both diagnosis (current state) and prognosis (future state) of infrastructure components. Traditionally, asset condition was assessed through visual inspections and manual measurements. Diagnostic vehicles provide high-accuracy track

geometry data, but their temporal resolution remains limited due to the low frequency of inspection runs. Nevertheless, these specialised vehicles remain indispensable for measuring parameters that cannot be inferred indirectly from in-service vehicles. A certified reference of the track condition is, in fact, essential to ensure the validation and reliability of complementary measurement datasets [2].

Modern TRVs, equipped with visual, radar, ultrasound, and induction sensors, are used worldwide to measure track geometry and other infrastructure parameters dynamically. For example, Network Rail's New Measurement Train in the UK monitors twist, cyclic top, and gauge defects using laser and camera systems. In Japan, the Dr Yellow inspection vehicles operate at speeds exceeding 250 km/h on the Shinkansen network, measuring gauge, cross-level, alignment, longitudinal level, and twist using inertial and optical sensors. In the United States, ENSCO operates a fleet of TRVs dedicated to track inspection and maintenance, while China's Comprehensive Inspection Trains are high-speed platforms used for both dynamic and static analyses of track geometry [1]. In Italy, Rete Ferroviaria Italiana (RFI) deploys a large fleet of TRVs and TRCs tailored to different track types. Among these, the Diamante 2.0 series features a wide range of sensors monitoring all aspects of the infrastructure—from the track and overhead contact line to data transmission systems and electromagnetic, acoustic, and vibration emissions for high-speed lines.

Although TRVs deliver accurate and comprehensive measurements, their use is constrained by limited fleet size, inspection frequency, and the occasional need for line closures. Between successive inspections, defects may develop unnoticed.

Traditionally, railway infrastructure has therefore relied on periodic inspections and preventive maintenance, often carried out manually or through visual checks. This approach is labour-intensive, time-consuming, and prone to human error, particularly under unfavourable environmental conditions. With increasing traffic density, such traditional methods are becoming increasingly inefficient and costly [3].

In recent years, track geometry monitoring has increasingly depended on TRVs equipped with optical, inertial, GNSS, radar, ultrasound, LiDAR, and inductive sensors, as well as high-speed camera systems [1][2][3][4].

While these technologies offer high measurement accuracy and repeatability, they are limited by their low operational frequency, high maintenance costs, and inability to capture dynamic interactions at the wheel-rail interface. Conversely, onboard sensors such as accelerometers and gyroscopes mounted on in-service vehicles allow for more frequent monitoring across extensive networks, offering higher temporal resolution. The acquired data can be processed to reconstruct track geometry or predict its evolution. Combining signals from multiple onboard sensors mitigates individual weaknesses, making data fusion techniques a standard practice in railway diagnostics. Degradation prediction is thus essential to optimise maintenance schedules and reduce operational costs. Artificial Intelligence (AI) plays a pivotal role when large, heterogeneous sensor datasets are available, yet a major challenge lies in extracting actionable insights from such complex information. Each sensing system has distinct advantages and limitations [2].

Acceleration-based approaches, particularly those using passenger vehicles, often suffer from noise, drift, and model dependence when estimating absolute quantities such as displacements or forces, even when using certified TRV data. In this context, Instrumented Wheelsets (IWS) have emerged as a promising alternative, capable of directly measuring the contact forces exchanged at the wheel-rail interface [4][5]. By recording the vertical, lateral, and longitudinal

components of the contact force, IWS provide valuable insight into the dynamic interaction between vehicle and infrastructure. These measurements can reveal track and wheel defects that remain undetectable by conventional sensors, offering a dynamic, high-frequency perspective on vehicle-track behaviour.

As highlighted by Peinado et al. [1]: “Dynamic load, while difficult to measure, is a major contributor to track geometry degradation, and several sensors and algorithms are suitable for studying it.” An interesting question posed by the authors—“Can track geometry be autonomously determined from onboard monitoring data?”—provides valuable insights and points of reflection for the present research, namely: which technologies and techniques can enable autonomous monitoring, and can contact forces be used for that purpose?

Although IWS technology is mature in applications such as vehicle homologation, derailment risk assessment, and load verification, its use for continuous condition monitoring remains limited. The complexity of the instrumentation, calibration challenges, and the high data volume involved have hindered widespread adoption. However, recent advances in computational power and data-driven analytics have renewed interest in IWS applications, enabling the development of hybrid diagnostic frameworks that integrate physical models, signal processing, and machine learning (ML).

In the field of AI-driven maintenance, two main objectives can be identified:

- Maintenance optimisation, which formulates maintenance scheduling as an optimisation problem aimed at minimising operational costs or achieving target Track Quality Index (TQI) values, and
- Fault detection and prediction, where ML methods such as ANNs and Genetic Algorithms (GAs) are most commonly employed [1].

Digital tools and adaptive models form the foundation of Predictive Maintenance (PdM), as they can automatically adjust to system operation, make informed decisions, and suggest strategies based on real-time data. PdM represents a key enabler of Structural Health Monitoring (SHM), particularly in the railway sector, where the transition from traditional periodic maintenance to predictive approaches can significantly reduce intervention time and cost. To improve the availability, capacity, and quality of railway services, investment in modern SHM technologies and PdM systems is essential. These tools optimise resource allocation, enabling more efficient and sustainable infrastructure management [3].

Data-driven PdM and digital models [6], supported by the rapid growth of AI, represents one of the most promising directions in modern railway maintenance. By leveraging ML and Deep Learning (DL) algorithms capable of modelling system behaviour, identifying trends, and predicting faults, AI enhances system reliability and supports data-driven decision making.

In this context, the regular collection of inspection data, coupled with advanced analytics, enables the prediction of track degradation and fault evolution, ultimately improving maintenance scheduling and contributing to greater safety and passenger comfort.

1.2 Problem Statement

Railway infrastructure degradation is an inevitable process caused by the repeated passage of trains, environmental effects, and the accumulation of minor defects ranging from design phase to realisation and maintenance. While traditional on-board monitoring techniques can identify large-scale geometry defects, they often fail to capture short-wavelength irregularities such as squats, corrugation, or dipped welds. These high-frequency defects, though localised, can have significant long-term impacts on ride quality, wheel and rail wear, and maintenance costs. The main challenges that this thesis seeks to address are therefore:

1. the lack of validated methodologies for identifying track faults from contact-force measurements;
2. the need for models that can transform raw force data into meaningful indicators of track condition; and
3. the difficulty of integrating physics-based and data-driven frameworks for reliable anomaly detection.

1.3 Research Objectives

The overarching aim of this thesis is to explore and validate methodologies for railway track condition monitoring based on instrumented wheelset (IWS) force measurements, combining model-based, traditional, and data-driven approaches.

The specific objectives are as follows:

4. Develop and validate a finite element (FE) model of the instrumented wheelset for inverse estimation of wheel-rail contact forces.
5. Characterise the frequency response and measurement bandwidth of the IWS system through experimental testing.
6. Apply traditional signal analysis techniques (e.g., Wavelet Transform, STFT) to detect and localise dynamic irregularities and short-wavelength faults in the railhead.
7. Formulate and validate a simplified model-based approach (3-DOF) for estimating track vertical alignment from measured contact forces.
8. Design and implement data-driven anomaly detection pipelines based on one-class ML classifiers (OCSVM, IF, LOF) using both space- and space–frequency-domain features.
9. Evaluate and compare the performance of the proposed methods and assess the potential of hybrid solutions that integrate physical interpretability with data-driven adaptability.

1.4 Research Questions And Hypotheses

This work is guided by several key research questions:

- **RQ1:** Can contact forces measured by IWS be effectively used to detect and classify track defects over a broad frequency range?
- **RQ2:** How accurately can simplified physical models estimate track geometry and dynamic behaviour from force data?
- **RQ3:** Are ML-based novelty detection algorithms reliable and interpretable when trained on unlabelled or artificially modified datasets?

- **RQ4:** What is the added value of combining space-domain and frequency-domain features for anomaly detection, applied to railway monitoring?

The working hypothesis is that contact-force-based monitoring, when enhanced by suitable modelling, feature extraction, and ML algorithms, can significantly improve the accuracy and robustness of track fault detection compared to acceleration-based or purely optical methods.

1.5 Methodological Approach

The research follows a hierarchical structure:

10. Physical Modelling and System Identification:

- a. Development of a validated FE model of the instrumented wheelset for inverse force estimation and bandwidth determination (characterisation of the measurement chain).

11. Traditional and Model-Based Analysis:

- a. Application of time-frequency methods (Wavelet, STFT) for anomaly localisation.
- b. Creation of a reduced-order 3-DOF dynamic model for estimating vertical track alignment.

12. Data-Driven Frameworks:

- a. Design of ML pipelines for unsupervised (semi-supervised) anomaly detection (and novelty detection), including feature extraction in both space and space–frequency domains.

1.6 Thesis Structure

The thesis is organised as follows:

- **Chapter 1:** introduces the research motivation, objectives, and methodological framework.
- **Chapter 2:** presents an overview of instrumented wheelsets (IWS), describing the measurement principles, system architecture, and relevant state of the art.
- **Chapter 3:** details the development and validation of the finite element model and the static and dynamic characterisation of the IWS used to measure experimental reference data.
- **Chapter 4:** provides the experimental frequency response analysis of such IWS and discusses the operational bandwidth of the instrumentation.
- **Chapter 5:** explores traditional signal processing techniques, including Wavelet-based analysis, for detecting localised track irregularities.
- **Chapter 6:** introduces the model-based approach for estimating vertical track alignment using a simplified 3-DOF vehicle model.
- **Chapter 7-9:** describe and compare data-driven anomaly detection frameworks using one-class ML classifiers and hybrid feature domains.
- **Chapter 10:** summarises the main findings, contributions, and perspectives for future research.

1.7 Expected Contributions

This thesis aims to provide both scientific and practical contributions to the field of railway diagnostics:

- A validated modelling framework for contact-force-based condition monitoring.
- A comprehensive analysis combining physics-based, traditional, and data-driven methodologies.
- A novel hybrid pipeline that integrates spectral features and ML algorithms for unsupervised fault detection.
- Insights into the operational limits and calibration requirements of IWS for track condition monitoring.
- A step toward the digitalisation of railway maintenance, enabling predictive diagnostics and smarter asset management.

1.8 References

1. Peinado Gonzalo, A., Horridge, R., Steele, H., Stewart, E., & Entezami, M. (2022). Review of data analytics for condition monitoring of railway track geometry. *IEEE Transactions on Intelligent Transportation Systems*, 23(12), 22737–22754. <https://doi.org/10.1109/TITS.2022.3214121>.
2. Hoelzl, C., Dertimanis, V., Landgraf, M., Ancu, L., Zurkirchen, M., & Chatzi, E. (2022). On-board monitoring for smart assessment of railway infrastructure: A systematic review. In *The Rise of Smart Cities: Advanced Structural Sensing and Monitoring Systems* (pp. 223–259). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-817784-6.00015-1>.
3. Bianchi, G., Fanelli, C., Freddi, F., Giuliani, F., & La Placa, A. (2025). Systematic review of railway infrastructure monitoring: From classic techniques to predictive maintenance. *Advances in Mechanical Engineering*, 17(1). <https://doi.org/10.1177/16878132241285631>.
4. Rosano, G., Massini, D., Bocciolini, L., Zappacosta, C., Di Gialleonardo, E., Somaschini, C., et al. (2024). Diagnostics of the railway track - Possibility of development through the measurement of accelerations and contact forces [JB]. *Ingegneria Ferroviaria*, 81–102.
5. Riggall, G., & Santos, G. F. M. (2011). Use of IWS to determine best bogie configuration for Vale under 37.5 tonne axle loads. In *Proceedings of the 2011 International Heavy Haul Association Conference* (pp. 1–9). Calgary, Canada.
6. Strano, S., & Terzo, M. (2019). Review on model-based methods for on-board condition monitoring in railway-vehicle dynamics. *Advances in Mechanical Engineering*, 11(2). <https://doi.org/10.1177/1687814019826795>.

2 ARCHITECTURE OF INSTRUMENTED WHEELSETS

2.1 Introduction

The interaction between wheel and rail can be expressed through three orthogonal components of force: vertical, lateral, and longitudinal, as illustrated in Figure 2.1. The vertical wheel-rail force is mainly influenced by irregularities and imperfections in both the wheel and the rail, such as wheel out-of-roundness, rail corrugation, joints, and track misalignments. This component is of primary importance for assessing derailment risk, particularly in cases of wheel unloading or uneven track support.

The lateral force acts transversely to the direction of motion, either towards or away from the wheel flange. It is mainly affected by creep forces, vehicle hunting, and curve negotiation, and is closely related to the risk of rail rollover and the dynamic stability of the vehicle. The lateral-to-vertical force ratio (Y/Q) is a key indicator of safety against derailment.

The longitudinal force, on the other hand, acts parallel to the rail and arises primarily from traction and braking actions. Its magnitude depends on wheel–rail adhesion conditions, which can be significantly influenced by factors such as tread wear, contamination by water, ice, sand, or other debris.

Because the wheel-rail contact patch continuously moves along the rail as the wheel rotates, direct measurement of contact forces at the interface is practically impossible. Instead, these forces are transmitted through the wheel and axle to the bogie, and, by virtue of Newton's third law, also to the rail and track substructure. Consequently, the measurement of wheel–rail interaction forces is generally achieved indirectly, by evaluating the strain response of the wheel or axle. These approaches are commonly known as instrumented wheelsets (IWS) or dynamometric wheelsets [1-5].



Figure 2.1: High-speed wheelset. CAD rendering.

According to the measurement location, three principal methods can be identified [6-8]:

- Axle bending strain method, where contact forces are derived from the bending deformation of the axle.
- Wheel web strain method, which relies on the strain field within the wheel web to reconstruct the applied loads.
- Contact point estimation method, which determines the instantaneous position of the contact patch and its associated force components.

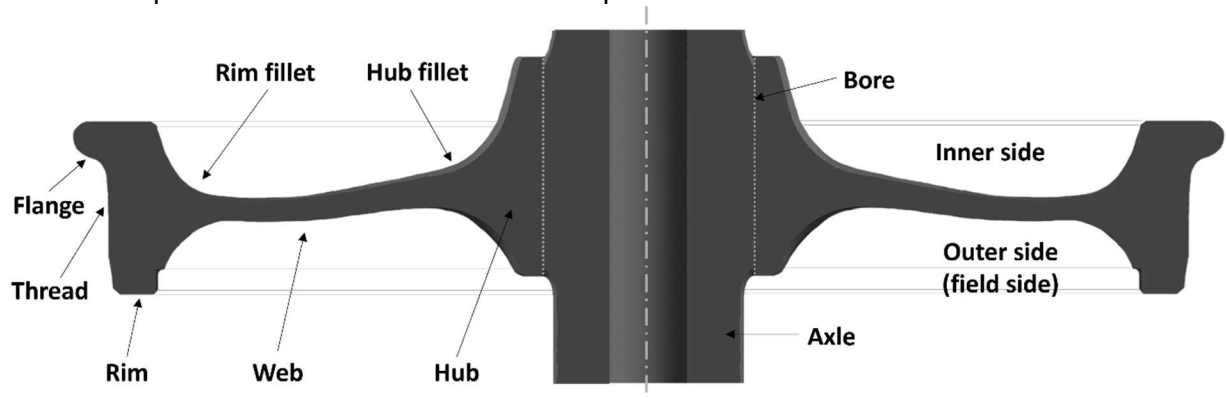


Figure 2.2: Wheelset nomenclature.

Different radial layouts of strain gauges can be adopted depending on the measurement objectives. Gauges are usually positioned in regions where the coupling effect (cross-sensitivity) between vertical and transverse forces is minimal, or where strain sensitivity to specific components is maximised [9]. Force decoupling is then achieved via numerical post-processing. In some research configurations, holes are drilled in the wheel web to locally amplify strain through stress concentration, improving the sensitivity to the vertical load component. Although such drilled-web methods enable precise vertical load estimation, they are rarely applied to high-speed wheelsets due to the resulting stress concentrations, which may initiate fatigue cracks and compromise structural integrity. For safety reasons and homologation issues, wheels designed

for high-speed service are therefore manufactured without drilled holes for measurement purposes.

Industrial solutions have evolved significantly. For instance, Alstom employs strain gauges positioned at multiple radial and circumferential locations, using up to 32 gauges (16 per side of the wheel web) [10]. This layout ensures high signal-to-noise ratio and a frequency response up to about 100 Hz, albeit at substantial manufacturing cost. Similarly, Lucchini, in collaboration with the University of Rome and the Politecnico di Milano [11], developed an advanced IWS configuration in which both the wheel and axle are instrumented. Finite Element Modelling (FEM) is used during the design phase to determine the optimal gauge placement. Each force component (vertical, lateral, and longitudinal) is measured through specific strain bridges, and decoupled through dedicated algorithms. The system also integrates wireless signal transmission, improving reliability and reducing the need for complex cabling. Other configuration, mainly used for test bench calibration and measurement of stress and strain field of the web surfaces, one wheel radius can be instrumented with approximately 40 strain gauges, applied on both the inner and outer faces of the web. In addition to varying the vertical and lateral loads and the relative position between the wheel, rail, and roller, the instrumented radius is also set to various angular positions, allowing the reconstruction of the strain distribution over the entire wheel web. [7,8].

IWS offer several distinct advantages providing direct and continuous measurement of wheel-rail forces in real time, including all three force components for both wheels of an axle. Unlike wayside monitoring systems, which measure forces only at specific track locations, IWS enables continuous onboard monitoring, allowing the evaluation of vehicle and track conditions over the entire route. The resulting data enable real-time assessment of vehicle dynamics, operational safety, and infrastructure condition, supporting intelligent maintenance and predictive diagnostics.

Despite their high precision, IWS systems have several drawbacks. Their manufacturing, calibration and maintenance costs are considerable. The installation and configuration of strain gauges require skilled labour, and the optimisation of sensor layouts often depends on proprietary industrial know-how accumulated over decades. Some layouts require modifications to the wheel, such as drilling, which is unacceptable for long-term operation on high-speed lines due to reduced fatigue strength. IWS systems typically capture low-frequency components of the wheel-rail force spectrum (up to about 200-300 Hz) [6,12-14]. Their quasi-static calibration procedures do not account for the modal vibrations of the wheelset, limiting their ability to measure high-frequency dynamic interactions caused by track irregularities. Calibration procedures are complex and time-consuming. Laboratory calibration involves stepwise application of known forces to derive calibration matrices, while on-track validation compares measured and theoretical loads during static and quasi-static conditions, such as when passing through curves.

2.1.1 Alternative Measurement Approaches

A number of indirect force estimation methods have been developed to address some of these limitations. These approaches infer wheel-rail forces from acceleration or displacement measurements at various points of the vehicle (typically the axle box or bogie frame). Examples include [6]:

- Linear system identification techniques, which estimate forces through inverse modelling based on measured accelerations.
- Reverse dynamic models that use body acceleration signals to reconstruct contact forces [15].
- Non-contact sensing, where eddy current and displacement sensors measure deformations without directly instrumenting the wheel.
- Frequency-response-based approaches, correlating axle-box acceleration spectra with track irregularities and high wheel-rail forces.
- Dynamic inversion and optimisation models using finite element and multi-body simulations to estimate forces under various operating conditions.
- Data-Driven methods for force identification (or estimation) from vehicle accelerations [16,17].

While these inverse or model-based approaches avoid the need for structural modification of the wheelset, they require accurate dynamic models and extensive calibration, and their accuracy typically decreases with increasing system non-linearity.

2.1.2 Requirements of the Measuring Chain

The reliability of an IWS lies on the overall measuring chain, which comprises the transducers (strain gauges and ancillary sensors), the bridge circuitry, signal conditioning, telemetry, A/D conversion, and the downstream calibration and processing. Because contact forces are reconstructed from strains measured on rotating components, specific performance requirements and limitations apply [12,18]. Core performance requirements are:

- Sensitivity and resolution. Sensitivity is set by the structural compliance at the gauge location(s), the gauge factor, bridge configuration, and excitation voltage; resolution is limited by electrical noise and A/D quantisation. Together they determine the signal-to-noise ratio (SNR) and the smallest detectable change in contact force. High sensitivity should not come at the expense of excessive cross-talk or structural weakening.
- Bi-uniqueness (one-to-one mapping). Within the calibrated operating domain, the vector of measured bridge outputs must correspond uniquely to the vector of contact force components. Poorly conditioned calibration matrices (e.g. due to non-optimal gauge placement) jeopardise identifiability, inflate uncertainty, and make decoupling unstable. This requirement is essential, while other specifications may allow wider tolerances for cost reasons.
- Bandwidth and ripple. Because gauges rotate with the wheel, even steady loads appear as amplitude-modulated signals. The resulting ripple depends on gauge angular position and harmonic content. Layouts in quadrature and/or distributed bridges reduce modulation, while synchronous demodulation and low-pass filtering further suppress ripple. The useful bandwidth is bounded above by the wheel/wheelset modes and by anti-aliasing; as a rule, the chain should provide a flat response at least to the target analysis band (typically ≥ 40 –100 Hz for on-board monitoring).
- Mechanical and electrical cross-talk, or cross-sensitivities, arise from structural coupling (e.g. Y affecting Q channels) and from imperfect bridge balance or wiring. They must be minimised by FEM-guided placement, full-bridge topologies, careful routing/shielding, and corrected through the calibration coefficient matrix with appropriate regularisation.

- Rotation centrifugal effects induces speed-dependent apparent strains (bias terms) and small geometry changes. These must be characterised and compensated (e.g. by symmetric full bridges, speed-indexed zeroing, or explicit centrifugal terms in the calibration model).
- Thermal effects. Temperature and gradients produce apparent strain and drift. Mitigations include matched self-temperature-compensated gauges, full-bridge configurations on isothermal paths, active temperature measurements for post-correction, and stable excitation. Thermal transients should be included in uncertainty budgets.
- Noise (electromagnetic and telemetry). EMI from traction equipment, telemetry link noise, and ground loops degrade SNR.
- Linearity and hysteresis. Adhesive creep, substrate non-linearity, and bridge imbalance can introduce hysteresis and non-linear response. High-grade gauges, controlled surface preparation, and mechanical fixtures with adequate stiffness reduce these effects; residuals are captured during multi-level calibration.

2.1.3 Practical implications for design and calibration

Full Wheatstone bridges (rather than quarter/half) maximise sensitivity, suppress thermal/centrifugal common-mode terms, and inherently improve SNR. Quadrature pairs and distributed layouts are preferred where ripple must be minimised. Radial and circumferential positions should be optimised (typically via FEM) to maximise sensitivity to targeted components while keeping the calibration matrix well-conditioned. Avoid regions prone to local stress raisers unless purposely used in research; drilled-web solutions are normally excluded for high-speed service. Anti-alias filters and sampling rates must cover the required band with margin. When possible, use synchronous (angle-referenced) acquisition to stabilise demodulation and reduce ripple; otherwise, ensure accurate tachometry for post-processing.

Bench calibration (multi-axis, multi-level) establishes the force-strain coefficient matrix; line verification confirms performance under static and quasi-static conditions (straight track loads, low-speed curves). Accuracy targets for high-quality IWS are typically on the order of $\pm 5\%$ for the reconstructed forces within the calibrated domain, contingent on proper assembly and environmental compensation. Uncertainty must be managed, providing confidence intervals by propagating measurement noise, model uncertainty (e.g. contact point position), thermal/centrifugal corrections, and cross-talk residuals. Monitor health via redundancy (e.g. mirrored bridges) and drift checks.

In summary, an effective measuring chain balances sensitivity, bandwidth, and bi-uniqueness while actively mitigating ripple, cross-talk, and environmental biases. Careful bridge design and rigorous calibration are indispensable to meet accuracy specifications without compromising wheelset integrity.

2.2 Axle bending strain measurements for contact forces inverse identification

The feasibility of estimating wheel-rail contact forces from strains measured on the axle was first demonstrated at the Deutsche Bundesbahn research centre [19]. The approach rests on a linear mapping between applied forces/moments at the wheel-rail interface and the bending/torsional strain state at instrumented axle sections. Under the simplifying assumption

that the axle behaves as a Saint-Venant beam, the relationship can be inverted to recover the vertical (Q), lateral (Y) and, where torsion is instrumented, the longitudinal components of the contact force. In practice, the proportionality terms are not known a priori and must be identified through calibration, yielding the coefficient (influence) matrix for the instrumented wheelset [20,21]. Despite its conceptual economy and ease of installation, axle-based reconstruction is affected by several systematic issues. The lateral and longitudinal positions of the contact patch vary with wear, steering, and track geometry. Because the lever arms linking contact forces to axle bending depend on this position, unmodelled contact-point shifts introduce bias into the reconstructed forces unless the position is estimated and compensated [7,8]. The inverse estimation method typically ignores the inertial forces of the wheel, brake discs and other components located between the contact patch and the strain sections. These terms scale with speed and excitation frequency and become non-negligible for dynamic events, thereby degrading accuracy, especially relative to wheel-web methods where the sensors are closer to the load path [9,22]. One mitigation is to co-measure wheelset accelerations (e.g. via axle-box accelerometers) and incorporate inertia corrections in the estimator. Because the gauges rotate with the axle, even steady loads appear as sinusoidally modulated strains (rotational modulation, ripple). The dominant carrier is synchronous with the rotational frequency, with higher harmonics depending on geometry and load distribution. Compared with wheel-web gauges, the axle signal is closer to a pure sinusoid, making it amenable to synchronous demodulation; nevertheless, residual ripple must be filtered or modelled.

Typical layouts employ full Wheatstone bridges on diametrically opposite fibres of the axle to sense bending in two orthogonal planes; an additional bridge in a $\pm 45^\circ$ grid or on a dedicated section senses torsion for longitudinal force estimation. To stabilise the output against angular dependence and provide quadrature components for demodulation, paired bridges are mounted 90° apart around the circumference (Figure 2.3). The resulting in-phase and quadrature signals are processed to obtain a slowly varying estimate of the underlying load components, while suppressing the rotational carrier.

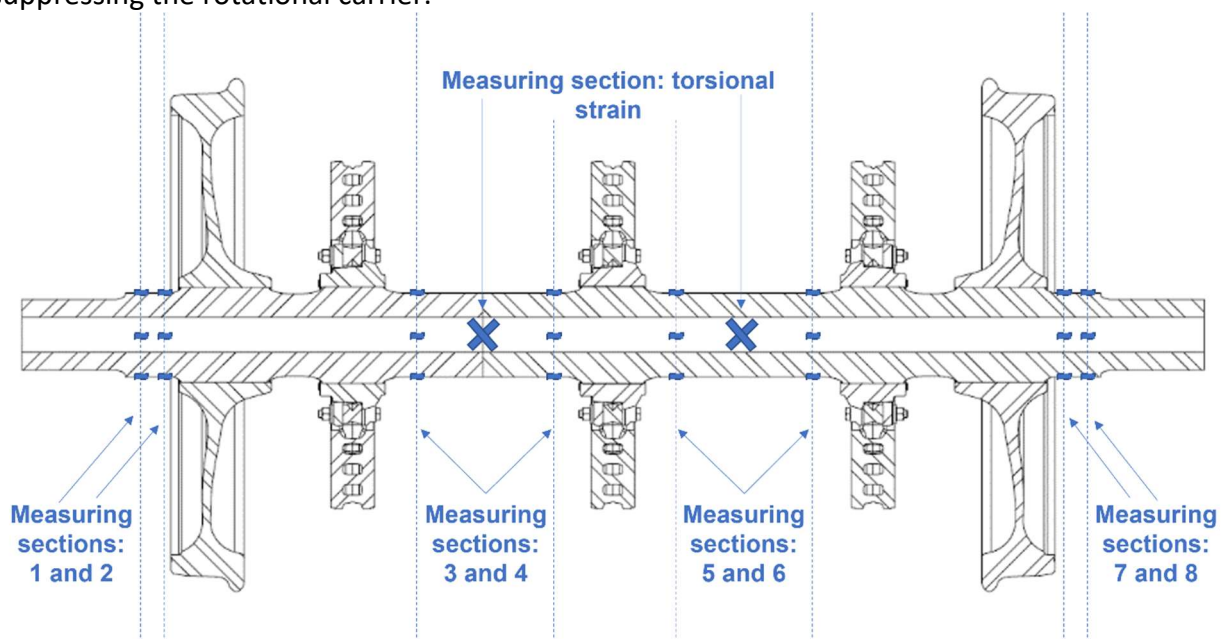


Figure 2.3: Measuring sections of the instrumented axle method [12].

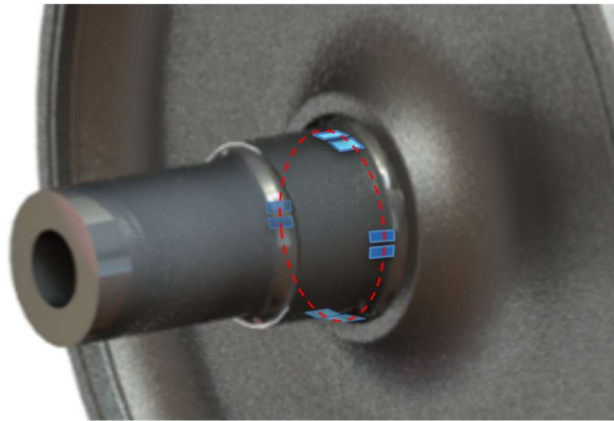


Figure 2.4: Axle instrumentation detail: journal quadrature bridges typical of measuring sections 2, and 7 (referring to Figure 2.3).

Axle-based methods require a two-step validation process. The first validation is a bench calibration to determine the force-strain coefficient matrix under controlled horizontal and vertical load steps (and, where applicable, torsional inputs). This matrix also captures cross-sensitivities between channels (cross-talk). The second level of validation takes place on-track, under static and quasi-static conditions (straight-line wheel loads, low-speed curves) to check the mapping in situ, including the effect of contact-point offsets. The conditioning of the calibration matrix is critical: poorly separated sensitivities between Q and Y, lead to noise amplification during inversion. Regularisation, carefully chosen gauge locations, and independent temperature/centrifugal compensation help preserve bi-uniqueness within the operating domain.

In summary, axle-bending approaches provide a structurally unobtrusive and well-established route to reconstruct wheel-rail forces, particularly effective at low to mid frequencies. Their performance ultimately depends on accurate calibration, reliable demodulation of rotating signals, and explicit treatment of contact-point variability and interposed inertia-aspects that motivate the use of hybrid sensing when higher dynamic fidelity is required.

2.3 Wheel-web based strain measurements

Early attempts to reconstruct wheel-rail forces from wheel instrumentation exploited spoked wheels, where stress in the spokes could be mapped almost orthogonally to vertical and lateral load components. The approach offered negligible cross-talk and relative insensitivity to changes in the lateral contact position for the lateral/longitudinal bridges. However, it required wheels whose geometry differs markedly from contemporary monobloc designs, with potential reductions in sensitivity and appreciable cost. More importantly, the altered structural behaviour of the wheelset rendered the method impractical for modern service and it is no longer used in routine applications.

By contrast, wheel-web instrumentation on standard monobloc wheels has matured into a versatile family of techniques [19]. Of particular interest is the “CML” method developed by Sapienza University of Rome [23,24]. With only minor modifications (and typically none for high-speed operation), strain gauges are bonded to the inner and/or outer surfaces of the web at carefully selected radial and circumferential positions. Compared with axle-based methods,

wheel-web sensing places the transducer closer to the load path, thereby reducing inertial effects between the contact patch and the measurement points and extending the usable bandwidth.

Among the advantages over axle-based approaches, the following are worth mentioning. The accessibility of the wheels and web surfaces is much improved compared to the axle, being generally free from interference by transmission or brake components, easing installation and maintenance [19]. Lower contribution of interposed inertia improves fidelity to dynamic phenomena, given the proximity to the contact patch area. Whereas axle bending approaches are effectively constrained below the first wheelset bending mode (≈ 80 Hz), web measurements are responsive up to the first wheel bending mode (≈ 200 Hz), making analysis to ~ 100 Hz practical for superstructure-related dynamics. The absence of a central axle section suitable for bending gauges does not preclude web-based force reconstruction. This makes wheel-web strain measurements applicable to independent wheels.

2.3.1 Gauge placement and layout strategies

Designing the measuring chain entails choosing radial and angular positions to maximise sensitivity, minimising cross-talk and ripple. For what concerns radial positioning, several strategies exist [9,12,22-26]. Strain gauges can be placed at radii where sensitivity to an undesired component is near zero. In such way, vertical and lateral components are already decoupled, and all the measured strain by the particular Wheatstone bridge can be attributed entirely to one of the two force components. A different strategy places strain gauges at radii with similar sensitivity profiles, then separating components through calibrated linear combinations (this solution requires the definition of the calibration matrix with linear coefficients of influence) [20,21]. In research contexts, where durability and structural integrity of the components are not a primary concern, stress concentration can be exploited by drilling holes in the web surface to enhance vertical-load sensitivity [27]. This approach has been historically adopted mainly in research and experimental contexts, for example during the calibration of instrumented wheelsets or in laboratory investigations aimed at validating numerical models. However, in operational wheelsets intended for regular service, especially on high-speed trains equipped with solid wheels, the introduction of additional holes is generally avoided. Moreover, The EN 13260, EN 13261, and EN 13262 standards, together with the ERA regulations, prohibit any unauthorised geometrical modifications to wheels and axles. Drilling new holes may compromise the structural integrity and fatigue strength of the wheel, which are critical parameters governed by standards such as EN 13262 and by railway safety regulations. In practical applications, strain gauges may occasionally be mounted by taking advantage of existing geometrical features, such as the bolt holes used for brake disc attachment on the inner face of the wheel, provided that no modification of the original geometry or material condition is introduced. This solution allows the use of instrumented wheelsets without affecting the wheel's mechanical integrity. Regarding the angular positioning, three canonical layouts are used (see Figure 2.5) [12]:

- Distributed bridges (a), in which multiple gauges arranged around a circumference; signals are summed to approximate the DC component of radial strain, inherently reducing rotational modulation (ripple).
- Sample-at-angle (point) methods (b). Strain is read only when the instrumented radius passes a reference angle. Although this suppresses angular dependence, it yields more than one

sample per revolution and an impractically low effective bandwidth; it is generally not recommended.

- Quadrature pairs (c). Two gauges separated by 90° provide in-phase and quadrature signals. Demodulation (e.g. $(\epsilon_c^2 + \epsilon_F^2)^{1/2}$) removes the carrier but loses the sign of the force. More sophisticated linear combinations preserve sign yet remain sensitive to higher harmonics when the contact point migrates.

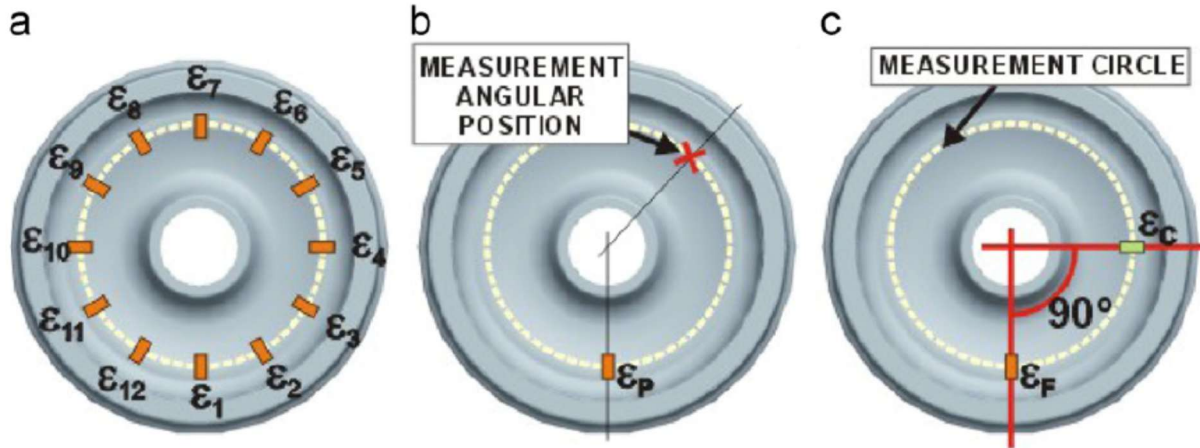


Figure 2.5: Possible full-bridge strain gauge configurations [9].

Full Wheatstone bridges are preferred in all cases to maximise sensitivity and cancel axisymmetric disturbances (centrifugal and thermal).

2.3.2 Decoupling of Q, Y and contact position d_R

In wheel-web measurements the lateral force Y often dominates the strain field on the web, while the vertical component Q and the lateral contact offset d_R superimpose additional patterns. This means that the web is most sensitive to Y component, especially close to the hub (still far from the thickening area). Sensitivity is inverted in the opposite face of the web. Remarkable is the presence of a narrow area of the wheel-web, approximately a circular ring, characterised by insensitivity to Y component variations. [4,7,8]

Some contemporary systems measure only web strains and recover forces via a calibration matrix that assigns influence coefficients to each bridge, rather than attempting a direct “single-bridge per component” mapping [19]. This requires strict bench calibration and careful management of cross-sensitivities. A notable method developed at Sapienza University of Rome [7,8] exploits the strain contrast between specific web regions: an inner zone near the hub exhibits maximum sensitivity to lateral force, while a zone closer to the tread features zero-crossing sensitivity to lateral force and is therefore suitable for estimating vertical load and d_R . By instrumenting both inner and outer web surfaces and applying a pseudo-inverse-based optimisation (typically minimising mean squared error) [19], the method separates Y and Q while concurrently estimating the lateral contact position. The arrangement uses full bridges configured to minimise axisymmetric effects of centrifugal forces and thermal effects, and to optimise signal-to-noise ratio. Because centrifugal and thermal fields are approximately axisymmetric, they can be strongly attenuated by symmetry in bridge wiring and by pairing inner/outer gauges (Figure 2.6). Residual biases are modelled as speed- and temperature-

dependent terms within the calibration and subtracted during processing. This approach avoids reliance on drilled webs while maintaining robustness across operating conditions.

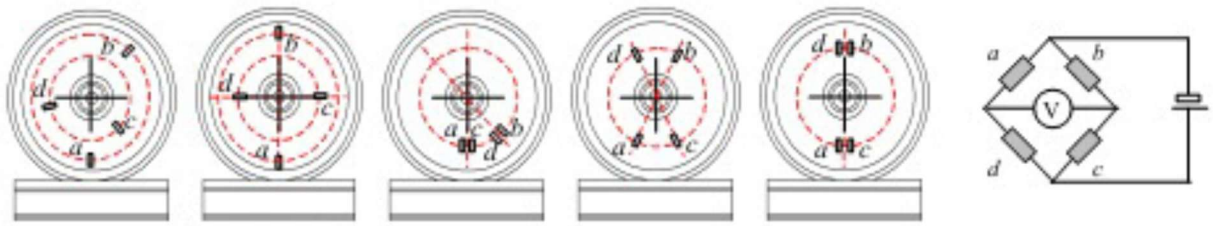


Figure 2.6: Full-bridge strain gauge configurations arrangements insensitive to thermal and centrifugal deformation [12].

2.3.3 Industrial implementations

An evolution of the discrete/quadrature concept is the “continuous” layout popularised industrially (e.g. Alstom [10]): 32 gauges (16 per side) are distributed over several circumferences so that the summed bridge outputs are nearly constant with angle, after compensating for centrifugal effects. The resulting high SNR is sustained up to roughly 100 Hz, at the cost of increased complexity and expense.

To summarise, wheel-web instrumentation offers a high-fidelity, structurally unobtrusive route to reconstruct wheel-rail forces, with superior bandwidth and reduced inertial contamination compared with axle-only approaches. Successful application hinges on FEM-guided placement, fully symmetric bridge topologies, rigorous calibration, and tailored demodulation to manage ripple, contact-point variability, and environmental biases.

2.4 Combined approach: axle and wheel-web strain measurements

A combined approach instrumenting both the axle and the wheel web was reported by the Politecnico di Milano and Lucchini RS [10]. In this method, the instrumented wheelset is calibrated on a roller rig and the unknown mapping between contact forces/moments and measured strains is obtained by minimising a suitable error functional, effectively identifying the influence (coefficient) matrices while quantifying uncertainty [20,21]. Prior to hardware implementation, finite-element modelling (FEM) is used to optimise sensor locations with respect to sensitivity, cross-talk, and conditioning of the calibration system.

In typical configurations the wheel is equipped with 32 strain gauges arranged at a constant radius, 16 per side (inner and outer for each wheel), at angular steps of 22.5°. Other configurations see just the field side of the wheels instrumented. On the axle, multiple independent sections are instrumented to resolve bending in the vertical and lateral planes; furthermore, two torsion bridges may be added to recover the longitudinal component. When the objective is to reconstruct all six components of the wheel-rail contact action (three forces and three moments), at least six axle sections are selected to ensure a sufficiently rich and mutually independent measurement set.

The knowledge of the instantaneous contact-point location, particularly its lateral offset, which strongly affects the lever arms for moments balance, is crucial. Since the neutral section

assumption on the axle cannot be relied upon in general, the contact position must be estimated concurrently, either numerically by profile coupling or via independent vision/odometry cues and fed into the reconstruction algorithm.

Where detailed knowledge of distributed and concentrated masses (e.g. brake discs) is available, the number of bending sections can be reduced (e.g. from four to two) and torsion may be captured at a single section. Nevertheless, retaining the full set improves robustness to modelling errors and operational variability.

2.4.1 Telemetry, synchronisation and inversion

Data can be transmitted by wireless telemetry (Figure 2.7), with either synchronous or asynchronous acquisition. In synchronous operation, a proximity probe provides the angular trigger for demodulation and consistent vector decomposition. In asynchronous mode, the same angular reference is still recorded to allow post-hoc angle-based processing. Reconstruction proceeds by inverting a linear system whose coefficients come from calibration.



Figure 2.7: Telemetry system.

2.4.2 Sampling strategies: constant time and constant space

The strain-gauge chain is analogue, and signals must be anti-alias filtered and digitised. Two sampling strategies are commonly employed [7,8,14,18,28].

The first one is known as constant-time sampling ($C\Delta t$), or “continuous”. Signals are sampled at fixed time intervals, meeting the spatial-resolution requirement of the standards at all speeds (e.g. $\Delta s < 0.5$ m per). Sensitivity and accuracy are modest and rotational ripple is present, but this approach captures higher-frequency content with relatively few channels and is the default for continuous web measurements.

The second strategy is called constant-space sampling (CΔs), or “discrete”. Signals are sampled at fixed angular increments determined by the tachometric pulses per revolution. Sensitivity and theoretical accuracy improve, but more channels are needed to represent continuous phenomena over a revolution. A CΔs approach has been proposed in [23], sampling as the instrumented radius passes directly over the contact zone; increasing the number of instrumented radii raises the effective sampling rate but complicates the measuring chain.

In practice, many wheel-web methods employ CΔt, with rotating carriers requiring careful demodulation. Electrical radial connections can introduce angle-dependent gain variations; compensation typically combines calibration-based gain correction with multi-bridge combinations to suppress the non-sinusoidal carrier. These solutions tend to be wheel-specific and less portable across designs, sometimes requiring drilling the web surface [27].

For longitudinal alignment and comparability across runs, signals can be resampled to a constant space grid (CΔs) using odometry. Odometric measurements, however, are prone to drift and require periodic reset at known reference points along the line, either during a diagnostic run or in post-processing. Resets introduce discontinuities that must be handled carefully during interpolation and data alignment. GNSS can aid absolute positioning but is affected by outages and accuracy limitations, particularly in tunnels and cuttings.

2.4.3 Bandwidth considerations

Sampling frequency should reflect both the system’s natural dynamics and the intrinsic low-pass filtering introduced by the distance between the contact patch and the sensor location [28]. While standards typically require a minimum effective cut-off of ≈ 40 Hz, practical deployments use sampling at ≥ 200 Hz to provide adequate margin. As a guide, the first wheelset bending mode lies around 80 Hz, and the first wheel bending mode around 200 Hz [6,12-14]; consequently, extending faithful reconstruction much beyond ~ 100 Hz is challenging and must be justified by calibration evidence. Careful anti-alias filtering is essential.

In summary, simultaneous axle and wheel-web instrumentation delivers a redundant, well-conditioned measurement ensemble capable of reconstructing all components of wheel–rail action with improved robustness. Success hinges on rigorous FEM-guided placement, reliable angle referencing, judicious sampling strategy (CΔt and/or CΔs), and explicit estimation of the contact patch location, together with thorough calibration and uncertainty management.

2.5 References

1. Luo, R., & Shi, H. (2018). Railway vehicle system dynamics and applications. Chengdu: Southwest Jiaotong University Press.
2. Riggall, G., & Santos, G. F. M. (2011). Use of IWS to determine best bogie configuration for Vale under 37.5 tonne axle loads. In Proceedings of the 2011 International Heavy Haul Association Conference (pp. 1–9). Calgary, Canada.
3. Cervello, S., & Bracciali, A. (1997). Development of a vibroacoustical methodology for the design and validation of low-noise railway wheels. In Proceedings of the World Congress on Railway Research (pp. 16–19). Florence, Italy.
4. Licciardello, R. V., Broggiato, G. B., Bruner, M., & colleagues. (2006). A comparative study of the results of field tests carried out with different contact force measurement methods. In Proceedings of the 7th World Congress on Railway Research (WCRR) (pp. 1–10). Montreal, Canada.
5. Braghin, F., Bruni, S., Cervello, S., & colleagues. (2003). A new method for the measure of the wheel–rail contact forces. In Proceedings of the 6th International World Congress on Contact Mechanics and Wear of Rail/Wheel Systems (CM2003) (pp. 313–319). Gothenburg, Sweden.
6. Wu, P., Zhang, F., Wang, J., Wei, L., & Huo, W. (2023). Review of wheel–rail forces measuring technology for railway vehicles. *Advances in Mechanical Engineering*, 15(3). <https://doi.org/10.1177/16878132231158991>.
7. Corazza, G. R., Licciardello, R., Malavasi, G., & Marcone, M. (1999, November 3–5). La ruota come sensore d’interazione ruota–rotaia. In Atti del Convegno Nazionale Conclusivo del PFT2. Università degli Studi di Roma, Rome, Italy.
8. Corazza, G. R., Licciardello, R., & Malavasi, G. (2000). The wheel as sensor of interaction between wheel and rail. In Proceedings of the World Congress on Railway Research. Università degli Studi di Roma “La Sapienza”.
9. Gómez, E., Giménez, J. G., & Alonso, A. (2011). Method for the reduction of measurement errors associated to the wheel rotation in railway dynamometric wheelsets. *Mechanical Systems and Signal Processing*, 25(8), 3062–3077. <https://doi.org/10.1016/j.ymssp.2011.05.008>.
10. Crosio, P. F. (2011). Stima delle forze di contatto tra ruota e rotaia (Doctoral dissertation). Politecnico di Milano.
11. Benigni, E., Braghin, F., Cervello, S., Cigada, A., & Resta, F. (2002). Determinazione delle forze di contatto tra ruota e rotaia a partire da misure di deformazione dell’assile. *Ingegneria Ferroviaria*, 57(12), 1059–1075.
12. Bracciali, A., Cavaliere, F., & Macherelli, M. (2014). Review of instrumented wheelset technology and applications. In J. Pombo (Ed.), Proceedings of the Second International Conference on Railway Technology: Research, Development and Maintenance. Civil-Comp Press. <https://doi.org/10.4203/ccp.104.1>.
13. Bellacci, G., Neri, F., & Pugi, L. (2024). Preliminary frequency response analysis of a contact force measurement system for rail applications. In F. Bellotti et al. (Eds.), Applications in electronics pervading industry, environment and society (ApplePies 2023) (Lecture Notes in Electrical Engineering, Vol. 1110, pp. 240–247). Springer. https://doi.org/10.1007/978-3-031-48121-5_20
14. Bracciali, A., & Rissone, P. (1994). Analisi modale sperimentale di una sala portante del convoglio ETR500. *Ingegneria Ferroviaria*, 7/8, 394–407.

15. Pieringer, A., & Kropp, W. (2022). Model-based estimation of rail roughness from axle box acceleration. *Applied Acoustics*, 193, 108760. <https://doi.org/10.1016/j.apacoust.2022.108760>.
16. Walther, S., Müller, S., Renggli, R., Ünlü, F., & Fuerst, A. (2022). Prediction of wheel–rail contact forces using simple onboard monitoring system and machine learning. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 237(5), 553–562. <https://doi.org/10.1177/09544097221122006>.
17. Zhu, T., Wang, X., Wu, J., Zhang, J., Xiao, S., Lu, L., Yang, B., & Yang, G. (2025). Comprehensive identification of wheel–rail forces for rail vehicles based on the time domain and machine learning methods. *Mechanical Systems and Signal Processing*, 222, 110667. <https://doi.org/10.1016/j.ymssp.2024.111635>.
18. Licciardello, R. V., Broggiato, G. B., Bruner, M., Corazza, G. R., Cosciotti, E., & Malavasi, G. (2006, June 4-8). A comparative study of the results of field tests carried out with different contact force measurement methods. In *Proceedings of the 7th World Congress on Railway Research (WCRR '06)*, Montréal, Canada.
19. Broggiato, G.B., Cosciotti, E. Ottimizzazione numerica della disposizione di ponti estensimetrici sulle sale ferroviarie utilizzate nella misura delle forze di contatto ruota-rotaia. *Ingegneria Ferroviaria*, 11, pp. 879-889 (2006).
20. Diana, G., Resta, F., Braghin, F., Bocciolone, M., Di Gialleonardo, E., & Crosio, P. (2012). Methodology for the calibration of dynamometric wheelsets for the measurement of the wheel–rail contact forces. *Ingegneria Ferroviaria*, 67(1), 9–21.
21. Bellacci, G., Pugi, L., & Baldanzini, N. (2024). Validation of a wheelset finite element model for static structural analysis and inverse force identification. In F. Concli, L. Maccioni, R. Vidoni, & D. T. Matt (Eds.), *Latest advancements in mechanical engineering (ISIEA 2024) (Lecture Notes in Networks and Systems, Vol. 1124, pp. 236–244)*. Springer. https://doi.org/10.1007/978-3-031-70462-8_23.
22. Gómez, E., Giménez, J. G., & Alonso, A. (2012). Railway dynamometric wheelsets: A comparison of existing solutions and a proposal for the reduction of measurement errors. In Y. Q. Ni & X. W. Ye (Eds.), *Proceedings of the 1st International Workshop on High-Speed and Intercity Railways (Lecture Notes in Electrical Engineering, Vol. 148, pp. 261–284)*. Springer. https://doi.org/10.1007/978-3-642-27963-8_25.
23. Velletrani, F., Licciardello, R., & Bruner, M. (2020). Intelligent wheelsets for the trains of the future: The role of in-service wheel–rail force measurement. *Ingegneria Ferroviaria*, 75(7–8), 701–710.
24. Alessandria, M., Dotta, B., & Licciardello, R. (2011). Long-term contact force measurements with the CML method. *Ingegneria Ferroviaria*, 66(11), 929–947.
25. Ronasi, H., Johansson, H., Larsson, F. Identification of wheel-rail contact forces based on strain measurements, an inverse scheme and a finite-element model of the wheel. In: *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*. 228, pp. 343-354 (2013).
26. Ronasi, H., Nielsen, J.C.O. Inverse identification of wheel–rail contact forces based on observation of wheel disc strains: an evaluation of three numerical algorithms. *Vehicle System Dynamics: International Journal of Vehicle Mechanics and Mobility*, 51:1, pp. 74-90 (2013).
27. Kanehara, H., & Fujioka, T. (2002). Measuring rail/wheel contact points of running railway vehicles. *Wear*, 253(1-2), 275–283. [https://doi.org/10.1016/S0043-1648\(02\)00114-X](https://doi.org/10.1016/S0043-1648(02)00114-X).

28. Licciardello, R. V. (2002). Il rilievo in linea dei parametri di interazione ruota/rotaia come strumento di ricerca (Doctoral dissertation). University of Rome "La Sapienza".

3 VALIDATION OF A WHEELSET FINITE ELEMENT MODEL FOR STATIC FORCE INVERSE IDENTIFICATION

3.1 Introduction

Maintenance activities are of strategic economic importance for railway managers. Reducing the associated costs of such activities can lead to significant savings for infrastructure companies. Recently, the use of diagnostic or in-service vehicles, equipped with various sensors to assess the condition of the railway, has become a popular topic of research for railway condition monitoring [1]. These technologies are employed to achieve goals in terms of passenger comfort and safety, transportation efficiency and energy savings. Obtaining precise insights about the railway status from force and acceleration measurements or laser scanned displacements depends strictly on measured and processed physical quantities, which must be as reliable and representative of the infrastructure's state as possible [2]. Therefore, understanding the performance and limitations of the equipment used in the measurement chain, plays a fundamental role. In this regard, numerical simulations represent the most advantageous and cost-effective alternative for evaluating different sensor configurations and limiting the high costs associated with experimentation. The particular advantage lies in the ability to rapidly set up new models and obtain data. However, their reliability is often questionable and must be verified by comparing and validating synthetic results with a minimum number of experimental tests or known data. This is especially true for the estimation of contact forces exchanged between wheels and rails. Forces are physical quantities that must be monitored not only to meet standards requirements [3] for static loads assessment or vehicles and new components homologations, but also for condition monitoring purposes. Numerous publications in the literature discuss the measurement of contact forces and the estimation of strains in railway wheelsets through field tests, numerical, and analytical models [4-7]. Over the years, techniques have been proposed to optimize sensor positioning and reduce the measurement error of acquired signals [8-11]. It is well-documented in the literature [12][13] that the estimation of contact forces for condition monitoring purposes or homologation, can be achieved through the measurement of quantities

derived from the deformation of the axle and wheel web. In the first case, by measuring the bending induced by forces on the axle, it is possible to estimate the vertical component Q of the contact force. In the second case, measuring vertical forces is more challenging, as the surface portions of the wheels (inner and outer sides) where the strain gauges are attached, are subject to both bending and compression/tension. Decoupling the effects of lateral forces from vertical ones is not straightforward, especially for signals derived from train motion and self-excited vibrations (like in case of corrugation or excitation of natural frequencies deriving by consecutive impulses). However, it is possible to determine the lateral component of contact forces through appropriate signal processing, as indicated in [7]. To dissociate research activities from equipment and laboratory spaces availability, a finite element model of a railway axle has been developed to study and characterize the measurement chain of an existing instrumented wheelset, based on strain gauge measurements for inverse forces identification [14]. Validation of the results obtained through simulations has been deemed necessary to confirm the accuracy and reliability of the developed model by comparing it with the reference instrumented wheelset tested on the static calibration bench available at Osmannoro facility, managed by Italcertifer s.p.a. (see Figure 3.1) [15], during previous laboratory activities. For the purpose of validating the FE model, an error analysis has been conducted for a finite number of known sample points common to both the experimental and numerical tests summarized in Table 3.1, providing a first attempt indicator of the uncertainty associated with the measurement chain. The capability of inverse-identifying applied loads has been pursued as a further validation step for the model by solving the strains equations system, applying superposition techniques, and pseudo-inverse Moore-Penrose method in MATLAB.

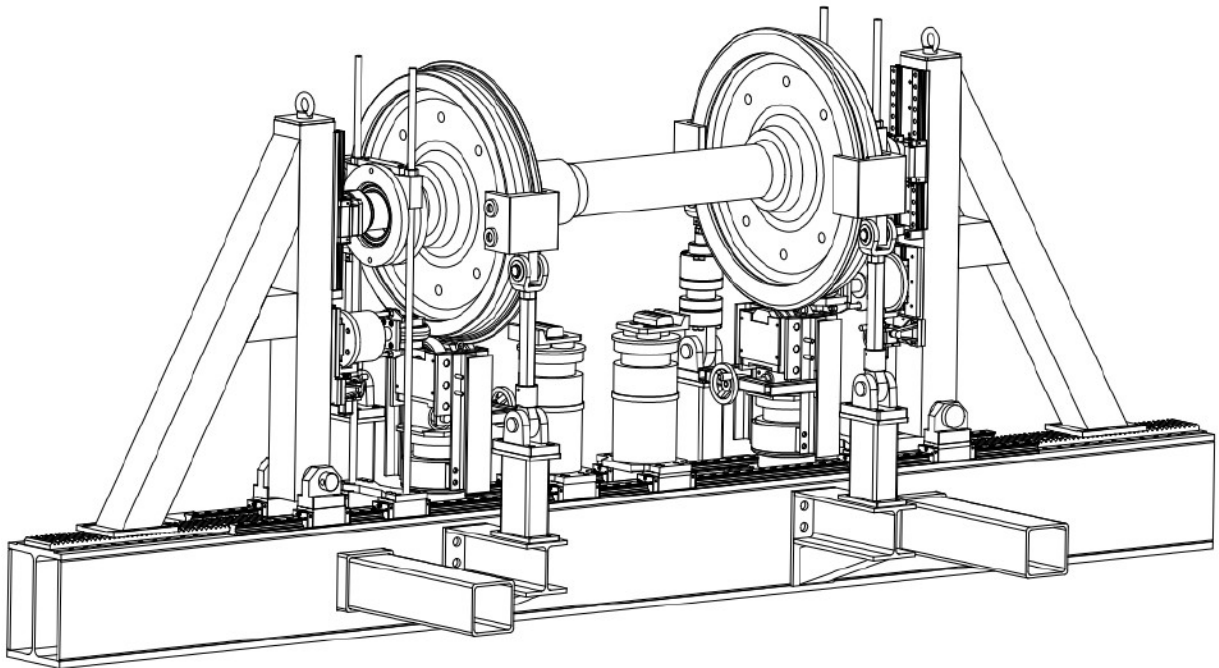


Figure 3.1: Schematic representation of Italcertifer s.p.a.'s test-bench for instrumented wheelsets testing. [15]

On the FEM side, the uncertainties in the results can be attributed to:

- Positioning of the roller contact areas for the application of the vertical (Q) and lateral (Y) forces: these are represented as elliptical projections on the running surface and on the wheel flange. Their placement significantly affects the bending moments, as it changes the lever arm of the applied force with respect to the neutral planes of the wheel, thus leading to different levels of deformation;
- Imported axle geometry: the presence of residual lines generated by the CAD software for the blending of the curved surfaces of the web introduces irregularities in the finite element model. These irregularities lead to numerical noise in the results, manifesting as fluctuations in the deformation relative to the expected trend of the curve (in the region of the web thickening, near the wheel flange).
- Use of a simplified reference system for rapid data extraction: the deformations of the wheel web surface measured by strain gauges refer to the actual curvature of the outer face of the wheel plate, which, with respect to the radial direction, forms tangential angles varying from 0° (perfectly aligned) up to approximately 20° (in absolute value) in the hub transition region. The results obtained by the two reference systems – one locally rotated according to the surface tangent and the other aligned with the global radial Z direction – were compared for a finite number of representative samples. The error introduced by approximating the extraction direction was found to be one order of magnitude lower than that observed when comparing experimental and numerical data. It was therefore considered acceptable (though still present), given the simplification of the model and the time saved in extracting and organising the results.

It is worth noting that, in the ideal case of applying only the vertical load Q , the wheel radial stress in the web surface (σ_{zz}) assumes the configuration shown in Figure 3.2, with a compression region near the rail contact zone and a diametrically opposite tensile region [1][6][7]. When a lateral component Y of the contact force is also applied, the deformation pattern changes, as the contribution of the lateral force to the bending moment is much greater than that of the vertical force. This leads to an inversion of the web's behaviour: it is subjected to tension in the lower lobe and experiences increased compression immediately above the flange, in the wheel-rail contact region. In the presence of the lateral force component, the deformations are therefore characterised by a larger magnitude due to the higher applied bending moment.

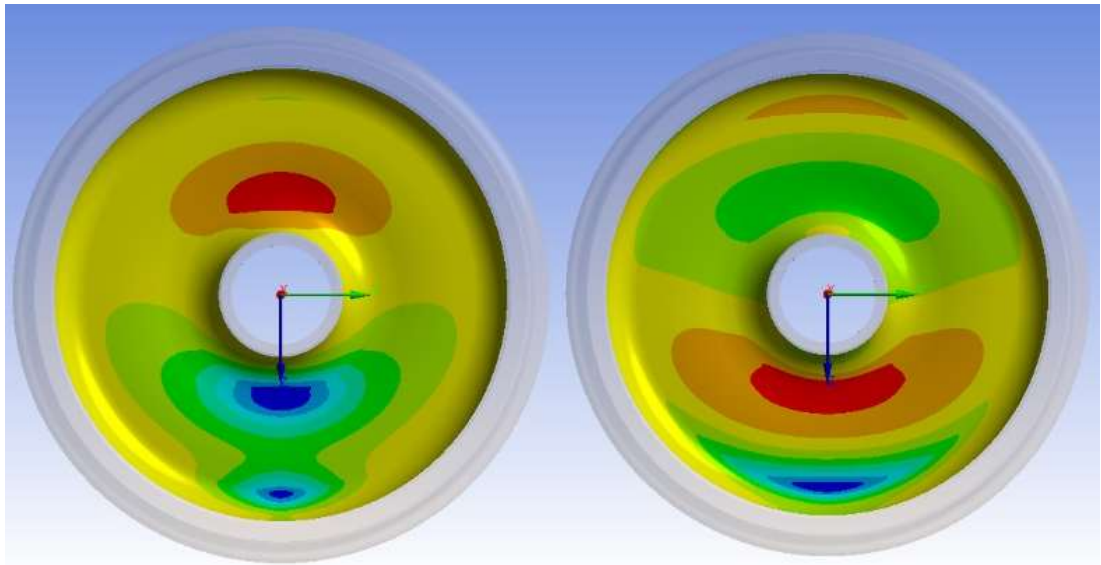


Figure 3.2: stress distribution on the wheel-web field side due to a) compression only, b) compression and bending moment.

3.2 Wheelset Finite Element Model

A finite element model, comprehensive of all the wheelset main components, was developed to represent and study the wheelset under different load cases. For the static analysis hereafter described the brake discs were removed, since their contribution to the overall bending stiffness of the axle is limited. Their effective load-carrying cross-section is small, as the external loads are transmitted through the wheel rim and the rolling surface. Consequently, the brake discs have a negligible influence on the deformation field of the wheel web. In the modal and frequency-response analyses of chapter 4, the brake discs were retained, as they provide a non-negligible bending and torsional inertia. Their presence significantly affects the natural frequencies and mode shapes, and therefore cannot be omitted in dynamic analyses. Given the general lower interest on brake disks strains for both analyses, a second order tetra mesh type was chosen for sub-assembly 2 and 3. For sub-assembly 1 an hexa mesh type was used, with element dimension 0.01 m, 0.02 and 0.03 for the three sub-assembly components respectively. For the wheel a second order tetra mesh was used with 0.02 m element size. Some reference lines necessary for the strain gauge patch definition were preventing the mapping of the surface, making it impossible to use an hexa mesh type. The tetra mesh was then employed. The same goes for the axle. Although the regular geometry would normally allow to use an hexa mesh, reference lines for the strain gauges patch definition made impossible the mapping of the surface. Thus, the same element type and dimension of the wheels was used. The final mesh showed 1528756 nodes and 1000686 elements. This represents a compromise between calculation time and accuracy. Indeed, finer meshes were iteratively tested on the same static analysis, reducing the dimension of the elements till convergence of the results: an additional reduction couldn't bring any improvement in terms of accuracy of the results, severely penalising the calculation performance. It was determined that the chosen mesh sizing adequately approximated reference results, and no further refinements were necessary to enhance the precision and accuracy of calculated outcomes. To summarise, a finer mesh is employed for the instrumented components

of the assembly such as the wheels and shaft (Figure 3.3 and Figure 3.4), while a coarser mesh is used for the non-instrumented parts, including the brakes sub-assembly (Figure 3.5).

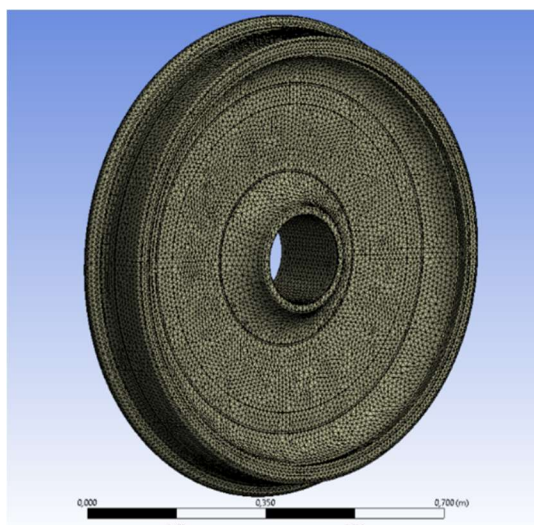


Figure 3.3: Wheel mesh.

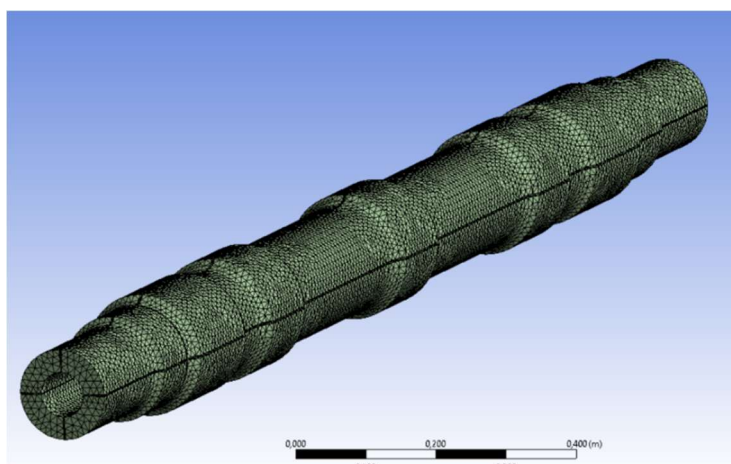
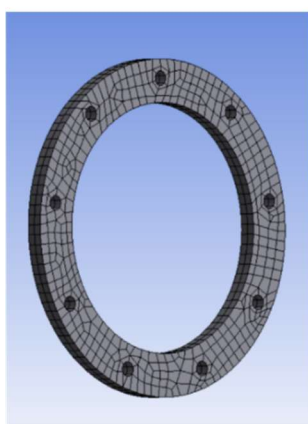
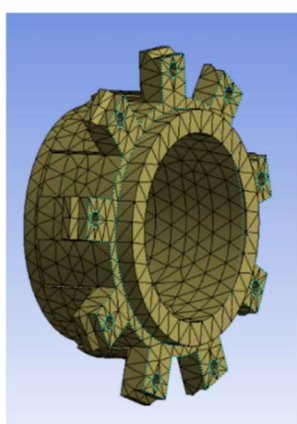


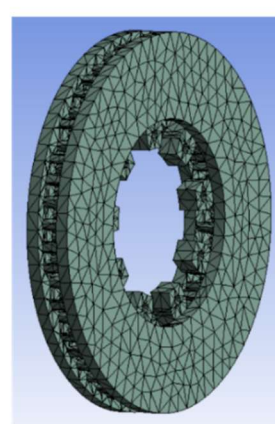
Figure 3.4: Axle mesh.



a)



b)



c)

Figure 3.5: Brake disk sub-assemblies. a) sub-assembly 1: spacer; b) sub-assembly 2: sleeve; c) sub-assembly 3: disk.

The deformation of the strain gauges was modelled as linear monoaxial strain according to Equation 3.1, as the distance difference between two points along the same line of the sensor patch, which represents the effective action line of the strain gauge divided for the undeformed initial distance. This was due to the impossibility of directly extracting deformations as outputs from the frequency analysis software used.

$$\epsilon = \frac{l - l_0}{l_0} \quad 3.1$$

The full IWS model is shown on Figure 3.6.

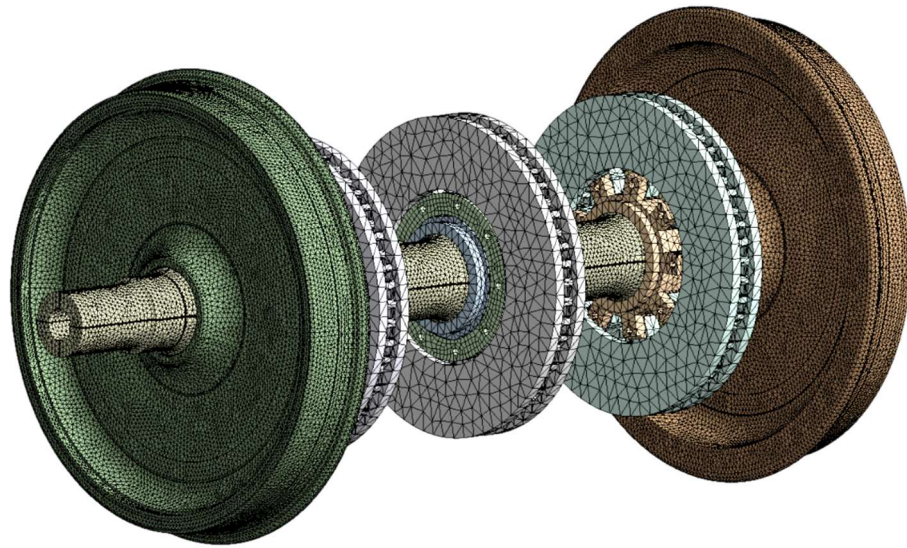


Figure 3.6: Full FE model of the instrumented wheelset.

3.3 Finite Element Static Analysis

The model was created using a commercial FE software and designed to observe deformations in the outer surface of the wheel web induced by the static application of contact forces. Brake disks assemblies, not subject to loads or constraints, have been omitted from the simulation to lighten the computational load, as shown in Figure 3.7.

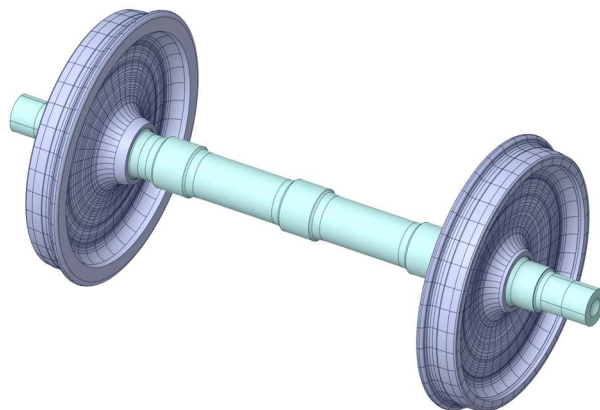


Figure 3.7: Wheelset FE model used for the static analysis.

Figure 3.8 shows the FE setup of the wheelset used for the static simulation. The axle has been fixed at the journals with ideal constraints, without imposing contacts or nonlinearities on the system, while the wheels have been subjected to various loading conditions, combining the vertical component Q , lateral component Y of the contact force and wheel rotation angle:

- $Q = 50$ kN, $Y = 0, 20, 45$ kN: maintaining a constant Q and varying Y , which includes the condition of only a constant vertical component applied.
- $Q = 10$ kN, $Y = 45$ kN, where the Q component is reduced while maintaining the maximum Y .

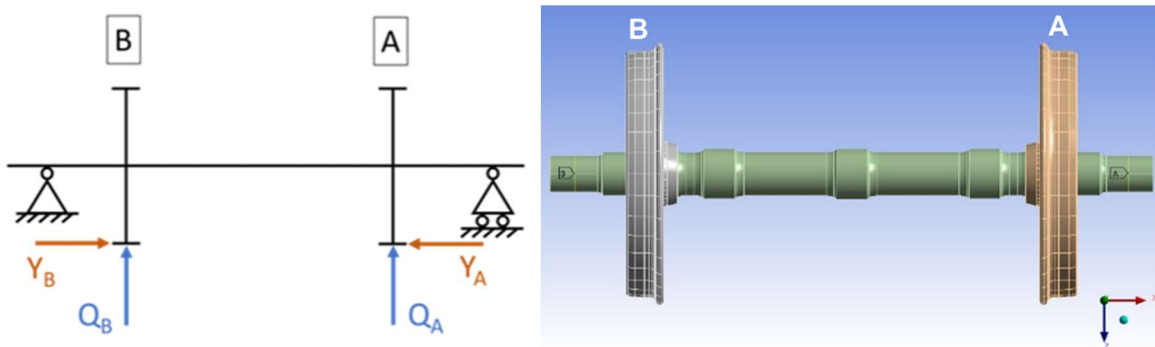


Figure 3.8: Constraint scheme with applied loads directions: side A roller (instrumented wheel), side B pinned (non-instrumented wheel), Q vertical and Y lateral contact forces.

A detail of the constraint is shown in Figure 3.9.

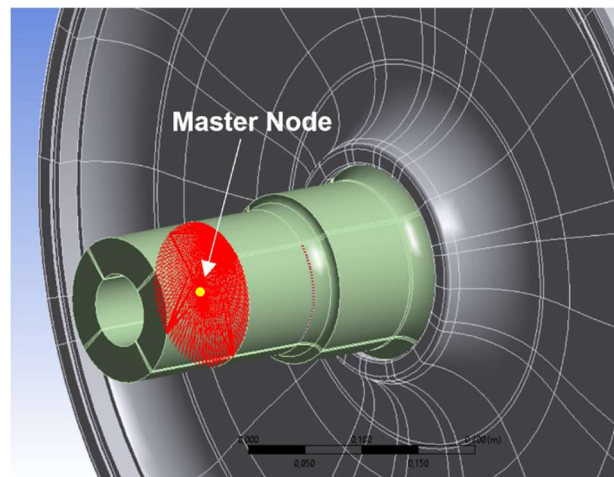


Figure 3.9: Detail of the adopted constraint solution.

The vertical load Q has been applied to the rolling surface at a distance of 70 mm from the wheel rim, while the lateral load Y , if present, has been applied directly to the wheel rim at a height of 26.8 mm from the fillet edge. The contact between the rollers and the wheel has been represented by an elliptical projection that reproduces the imprint left by the rollers on the wheels after experimental tests. The applied constraints aimed to replicate an ideal pinned-roller isostatic scheme illustrated in Figure 3.8, adopted for the experimental tests.

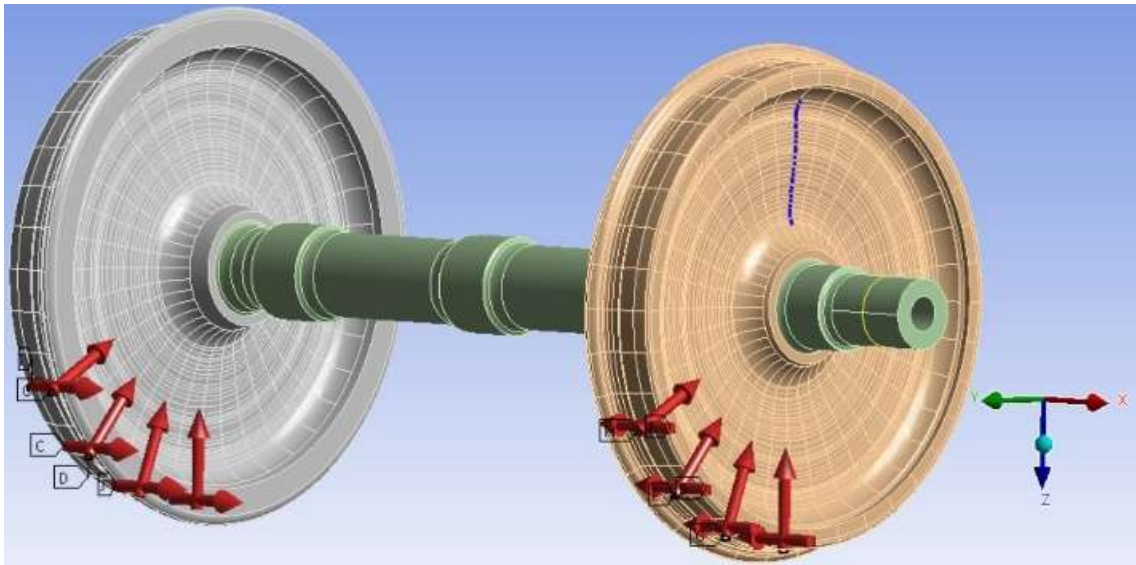


Figure 3.10: Instrumented wheelset FE model. Load case $Q = 50$ kN and $Y = 20$ kN for various inclination angles. Probe points on the wheel-web field side are highlighted in blue (180° position).

Experimental reference strain data used to validate the FE model, have been previously obtained from laboratory tests performed by the partner company on a static calibration test bench for railway wheelsets previously showed. In the calibration bench [15], two bushings per side support the journals, restraining rotation around the shaft axis (x-axis in red Figure 3.10) and allowing small rotations around the other axes (vertical z in blue and transverse y in green). Meanwhile, two hydraulic actuators per side (vertical and lateral) press the rollers onto the rolling surface and wheel rim, applying the loads and preventing transverse movements [15]. Uncertainty in the experimental data arises from imperfect sensors positioning, presence of adhesive between the strain gauge and the wheel, which inherently does not have the same mechanical characteristics as the metal object under examination, and amplifications or attenuations of signals. The error obtained from the error analysis can be used as a corrective coefficient for the synthetic data to account for the uncertainty in the measurement chain. From the FE perspective, uncertainties in the results may be due to an imperfect positioning of the roller contact areas for the application of Q and Y forces, which generates significantly different bending moments depending on the lever arm relative to the neutral planes of the wheel, resulting in strains of different magnitudes. The imported geometry of the wheelset can also introduce irregularities in the FE model, translating into numerical noise in the results. Lastly, the use of a simplified reference system for results extraction potentially leads to discrepancies between real and synthetic data. Real strains measured by sensors refer to the actual curvature of the outer surface of the wheel web, which, relative to the radial direction, assumes tangent angles ranging from 0° (perfectly aligned) to approximately 20° in absolute value in the hub region. The results obtained through the two reference systems, one rotated point by point according to the tangent to the surface and the simplified one aligned with the radial direction z of the absolute reference, have been compared for a finite number of samples of interest. The error introduced in approximating the direction is an order of magnitude lower than the calculated error between experimental and numerical data. It was deemed acceptable, though present.

3.4 Comparison with experimental data

Only the results of two loading configurations are presented in this thesis, corresponding to cases $Q = 50 \text{ kN}$, $Y = 0 \text{ kN}$, and $Q = 50 \text{ kN}$, $Y = 20 \text{ kN}$. The strain gauges are aligned along the z-axis of the absolute reference system (as reported in Figure 3.10), which is the vertical line connecting the centre of the force application imprint to the axis of the wheel (referring to Figure 3.11, when $\theta = \pi$ or 0).

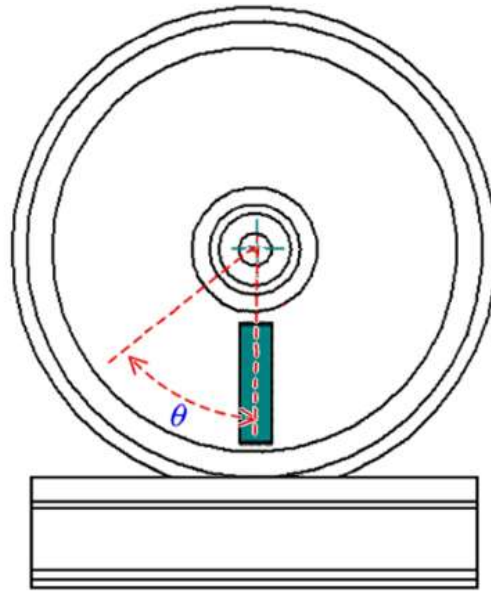


Figure 3.11: Schematic representation of sensors positioning for the static test (“survey”), covering most of the wheel-web radius for an extensive evaluation of the surface strain.

The graphs in Figure 3.12 reveal that the FE model accurately reproduces the experimental deformation state of the wheel web. Table 3.1, summarizes the results related to the first shown test (Figure 3.12-a), where the relative error varies from 1% to 12%, with an average of 5%. Interesting insights into strain gauge positioning can be deduced from the results, highlighting the radii where the sensitivity of sensors to applied loads is most pronounced. While the central area of the wheel web exhibits the lowest strain values in each test, areas close to the hub and rim demonstrate higher sensitivity, showcasing the highest strains magnitudes. Figure 3.12-b illustrates the difference in sensitivity across the entire wheel web to applied loads: lateral loads have a significant impact on the strain state of the wheel, and the variation in magnitude among all strain values is a clear indicator that, although the wheel can measure both vertical and lateral forces, it is more sensitive to the latter.

Due to a lack of resolution in the reference data, certain regions of the experimental curve lack a well-defined deformation trend that could be accurately captured by a simple linear interpolation, which is instead used for reconstructing the curve derived from numerical simulation results. Consequently, only the known samples are reported to avoid any misinterpretation of the curve.

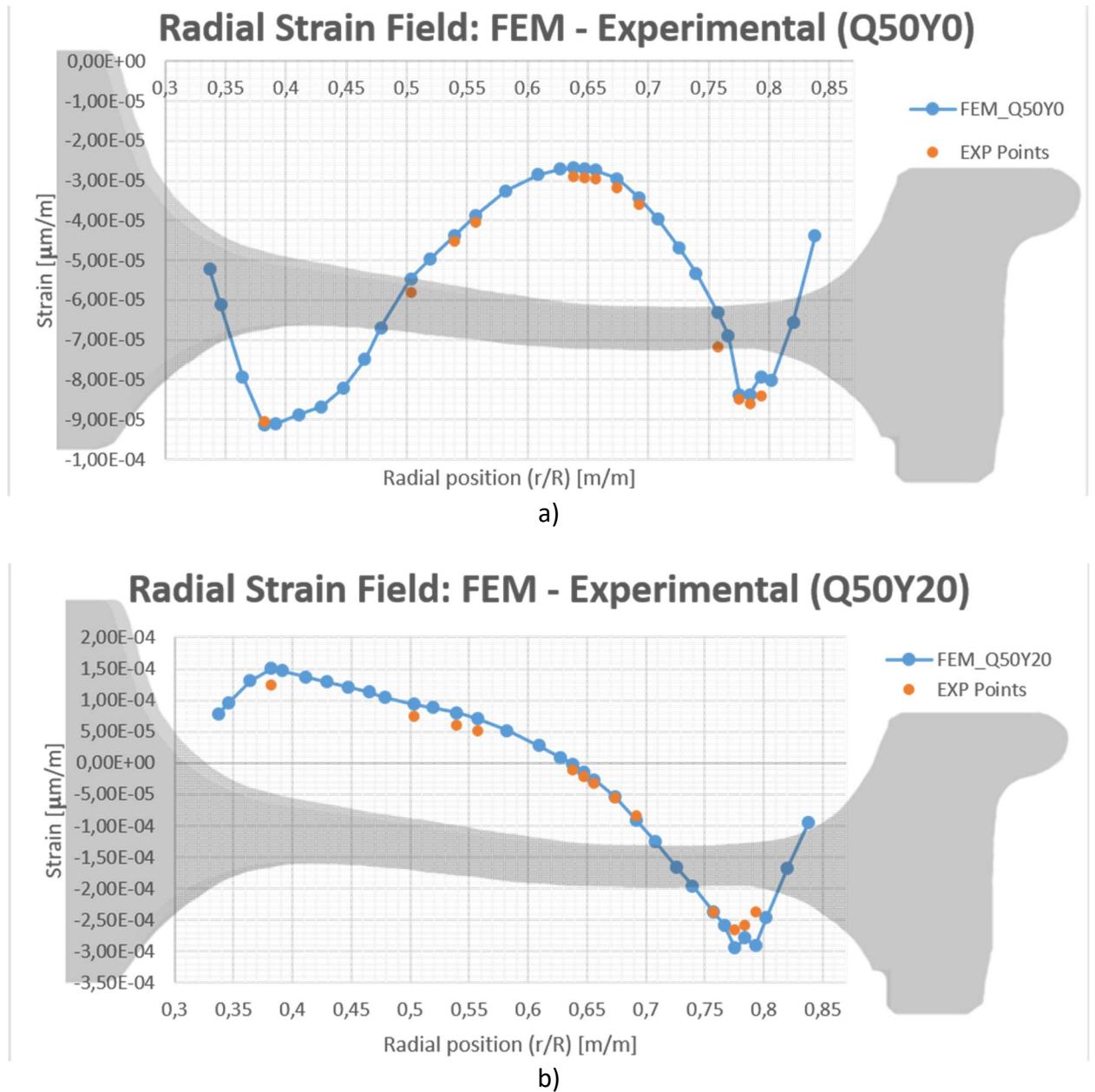


Figure 3.12: Comparison between numerical and experimental results for two loading configurations: (a) $Q = 50 \text{ kN}$ and $Y = 0 \text{ kN}$ and (b) $Q = 50 \text{ kN}$ and $Y = 20 \text{ kN}$. The strain gauge test is aligned with the z-axis of the absolute reference system. An undeformed section of the wheel is displayed to facilitate the visualization of the points of interest on the wheel web field side.

Table 3.1: Experimental and FE strains, for Q = 50 kN and Y = 0 kN, relative to Fig. 3(a).

Reference position	r/R [m/m]	Experimental [$\mu\text{m/m}$]	FEM [$\mu\text{m/m}$]	Abs. error [$\mu\text{m/m}$]	Rel. error [%]
R1	0,382	-9,1E-05	-9,14E-05	-8,51E-07	1%
R2	0,503	-5,8E-05	-5,47E-05	3,37E-06	-6%
R3	0,539	-4,5E-05	-4,38E-05	1,27E-06	-3%
R4	0,557	-4,1E-05	-3,87E-05	1,82E-06	-4%
R5	0,638	-2,9E-05	-2,68E-05	2,08E-06	-7%
R6	0,647	-2,9E-05	-2,69E-05	2,41E-06	-8%
R7	0,656	-3E-05	-2,73E-05	2,30E-06	-8%
R8	0,674	-3,2E-05	-2,94E-05	2,22E-06	-7%
R9	0,692	-3,6E-05	-3,41E-05	1,83E-06	-5%
R10	0,757	-7,2E-05	-6,30E-05	8,80E-06	-12%
R11	0,775	-8,5E-05	-8,37E-05	1,03E-06	-1%
R12	0,784	-8,6E-05	-8,37E-05	2,19E-06	-3%
R13	0,793	-8,4E-05	-7,93E-05	4,80E-06	-6%

3.5 Inverse force identification: further FE model validation

To identify forces, coefficients of influence for each strain gauge placed on the wheel web field side radius have been estimated. Considering the identification of the two forces (Q and Y) applied to the wheels, a total of two equations could be sufficient to determine the coefficients. Given the possibility to add other two equations because of the multiple load test cases performed at the static calibration test bench, a redundant 4-equation system was solved for each strain gauge. Expressing the system in its synthetic matrix form, the problem can be summarized in equation 3.2:

$$\varepsilon_i = [QY](k_i) \quad 3.2$$

Where ε_i represents the strain vector of the i-th strain gauge for the various load cases $L = (L_1, L_2, L_3, L_4)$, $[QY]$ is the load matrix, containing all the applied loads combinations (the four load cases previously mentioned) and k_i is the vector of the coefficients of influence for the strain gauge, respect to forces Q and Y [16]. The expanded form of the system is given in equation 3.3. While redundancy enhances the robustness of the method, it results in a system rank lower than the number of equations. This indicates a linear dependency in the system's equations, which is solved using the pseudo-inverse Moore-Penrose method. This method provides coefficients k_Q

and k_Y for both the Q and Y forces respectively, for each strain gauge. The procedure was applied to both the FE simulation results and the reference experimental data. An error value was calculated to assess the results, revealing a generally good agreement, albeit with some out-of-scale values in the relative error, possibly due to near-zero divisions.

$$\begin{cases} \varepsilon_{L1} = k_{Qi}Q_{L1} + k_{Yi}Y_{L1} \\ \varepsilon_{L2} = k_{Qi}Q_{L2} + k_{Yi}Y_{L2} \\ \varepsilon_{L3} = k_{Qi}Q_{L3} + k_{Yi}Y_{L3} \\ \varepsilon_{L4} = k_{Qi}Q_{L4} + k_{Yi}Y_{L4} \end{cases} \quad 3.3$$

Influence coefficients offer an estimation of the sensors' sensitivity to the specific measured force. Higher magnitudes indicate higher sensitivity, while lower magnitudes denote less sensitivity. Based on this, a preliminary positioning procedure can be devised by selecting strain gauges with maximum coefficients for a more precise identification of contact forces. It's noteworthy that some strain gauges may have influence coefficients approximatively zero. For instance, strain gauge at reference position R7, shown an extremely small k_Y value (0.050492) compared to other coefficients, suggesting insensitivity to lateral Y force for the specific wheel radius $r/R=0.656$. Although the error analysis comparing FE and experimental data revealed a relative error of 8%, slightly higher than the average 5%, it has been deemed worthwhile and interesting to explore if, by considering such specific strain gauge, contact forces could be inversely identified. At least two sets of equations are required to calculate Q and Y force (variables of the inverse identification problem) from the strain gauge measurements on the wheel web field side. An example is given in equation 3.4, where strain gauges in position R7 and R10 were selected for the calculation. Criteria for selection is: insensitivity to lateral force Y (strain gauge R7), lowest obtained errors compared to experimental data for k_Q and k_Y coefficients, while maintaining relatively high values for both coefficients (strain gauge R10).

$$\begin{cases} \varepsilon_7 = k_{Q7}Q + k_{Y7}Y \\ \varepsilon_{10} = k_{Q10}Q + k_{Y10}Y \end{cases} \quad 3.4$$

From which the variables can be obtained through straightforward algebraic steps:

$$\begin{cases} Q = \left(\frac{k_{Y10}\varepsilon_7 - k_{Y7}\varepsilon_{10}}{k_{Q7}k_{Y10} - k_{Y7}k_{Q10}} \right) \\ Y = \left(\frac{\varepsilon_{10} - k_{Q10}Q}{k_{Y10}} \right) \end{cases} \quad 3.5$$

By substituting FE strains into equation 3.5 for the various analysed load cases, all forces have been identified with remarkable precision, with a maximum absolute error of ~850 N (2% relative error) in the Y force for load cases L_3 and L_4 ($Y = 45$ kN). The bias may depend also on the position uncertainty of the contact area, which is not accounted for in the current method.

3.6 Conclusions and Future Developments

In this chapter the validation of the finite element model of an instrumented wheelset designed for railway homologation and condition monitoring purposes is presented. The static analysis results of the simulation have been compared with experimental data obtained from

previously conducted static tests. A general good agreement is observed, with an average relative error of 5% between synthetic and experimental strains. This value serves as an initial estimator of the measurement chain uncertainty for the real wheelset setup. The results highlight areas on the wheel web field side that are more sensitive to lateral forces, although the system can measure both vertical and lateral loads applied to the wheel. Superposition techniques are used to determine the strains based on pairs of applied loads. By solving the inverse system of obtained deformations, coefficients of influence of forces for the strain gauges have been determined. Subsequently, an inverse identification of the loads has been performed to further validate the model. To manage numerical noise in results generated during Finite Element Method application, the Moore-Penrose pseudo-inverse method is utilized for calculating the coefficients of influence. This technique, based on least squares procedures, enhances the accuracy of the obtained coefficients within an overdetermined system, allowing for optimal model calibration and reducing the impact of noise and uncertainties in the results. By substituting synthetic strains obtained from FE simulations and using the calculated coefficients, applied loads have been inversely identified with a maximum absolute error of ~850 N (relative error of 2%), demonstrating the reliability of the simulated results. Considering these premises, a digital twin of the measurement chain may be developed with the specific purpose of generating synthetic data for the calibration of the real instrumented wheelset, or to conduct virtual tests reproducing realistic running conditions, for studying optimal sensor layouts for inverse-force identification. Possible future developments will focus on enhancing the accuracy of the FE model to reduce overall strain magnitude errors, primarily by improving the imported geometry. Optimal strain gauge layout selection, organized in Wheatstone bridges for in-field applications, will be conducted based on the obtained coefficients of influence. The same procedure used for the single sensors presented here will be applied to a selection of Wheatstone bridges, comparing the different sensors layouts. To enhance the completeness of the model, additional probes, especially located on the journals and axle, will be added. This will enable the estimation of the position of the contact patch area of the wheel-rail interface, which is not accounted in the investigated configuration.

3.7 References

1. Velletrani, F., Licciardello, R., Bruner, M. Intelligent Wheelsets for the trains of the future: the role of in-service wheel-rail force measurement. *Ingegneria Ferroviaria*. 701 (2020).
2. Brkić, R., Adamović, Ž., Bukvić, M. Modeling of Reliability and Availability of Data Transmission in Railway System. *Advanced Engineering Letters*, 1(4), pp. 136-141 (2022).
3. EN 14363 2005 Railway Applications - Testing for the Acceptance of Running Characteristics of Railway Vehicles - Testing of Running Behavior and Stationary Tests. (Brussels: European Committee for Standardization, CEN) (2005).
4. Ronasi, H., Johansson, H., Larsson, F. Identification of wheel-rail contact forces based on strain measurements, an inverse scheme and a finite-element model of the wheel. In: *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*. 228, pp. 343-354 (2013).
5. Ronasi, H., Nielsen, J.C.O. Inverse identification of wheel–rail contact forces based on observation of wheel disc strains: an evaluation of three numerical algorithms. *Vehicle*

- System Dynamics: International Journal of Vehicle Mechanics and Mobility, 51:1, pp. 74-90 (2013).
6. Broggiato, G.B., Bruner, M., Licciardello, R. Studio numerico-sperimentale delle deformazioni di una sala ferroviaria di misura. AIAS (Associazione Italiana Analisi delle Sollecitazioni), XXXII Convegno Nazionale (2003).
 7. Corazza, G.R., Licciardello, R., Malavasi, G., Marcone, M. La ruota come sensore di interazione ruota-rotaia. *Ingegneria Ferroviaria*. - ISSN 0020-0956. - 3, pp. 119-131 (1999).
 8. Papini, S., Pugi, L., Rindi, A. An integrated approach for the optimization of wheel–rail contact force measurement systems. *J. Mod. Transport*. 21, pp. 95–102 (2013).
 9. Alessandria, M., Dotta, B., Licciardello, R. Long-term contact force measurements with the CML method. *Scienza e Tecnica, Ingegneria Ferroviaria*, pp. 929-947, 11 (2011).
 10. Broggiato, G.B., Cosciotti, E. Ottimizzazione numerica della disposizione di ponti estensimetrici sulle sale ferroviarie utilizzate nella misura delle forze di contatto ruota-rotaia. *Ingegneria Ferroviaria*, 11, pp. 879-889 (2006).
 11. Gómez, E., Giménez, J.G., Alonso, A. Method for the Reduction of Measurement Errors Associated to the Wheel Rotation in Railway Dynamometric Wheelsets. *Mech. Syst. Signal Process.*, 25(8), pp. 3062–3077 (2011)
 12. Licciardello, R., Broggiato, G.B., Bruner, M., Corazza, G.R., Malavasi, G. A Comparative Study of the Results of Field Tests carried out with different Contact Force Measurement Methods. In: *Proceeding of the 7th World Congress on Railway Research (WCRR)*, pp.1–10, Montreal, Canada (2006).
 13. Bracciali, A., Cavaliere F., Macherelli, M. Review of Instrumented Wheelset Technology and Applications. In: Pombo J., *Proceedings of the Second International Conference on Railway Technology: Research, Development and Maintenance*, Civil-Comp Press, Stirlingshire, Scotland (2014).
 14. Uhl, T. The inverse identification problem and its technical application. *Archive of Applied Mechanics*. 77, pp. 325–337 (2007).
 15. Macherelli, M., Braccesi, C., Delprete, C., Bracciali, A. Analisi delle forze ruota-rotaia mediante sale montate strumentate. *Progettazione e validazione sperimentale di sistemi basati su strumentazione estensimetrica*. PhD thesis, Università degli studi di Firenze, dottorato di ricerca in ingegneria industriale (2016).
 16. Diana, G., Resta, F., Braghin, F., Bocciolone, M., Gialleonardo, E.D., Crosio, P. Methodology for the calibration of dynamometric wheelsets for the measurement of the wheel-rail contact forces. *Ingegneria Ferroviaria*. 67, pp. 9–21 (2012).

4 FREQUENCY RESPONSE ANALYSIS OF THE INSTRUMENTED WHEELSET

4.1 Introduction

To ensure a proper diagnosis of the under-investigation vehicle or infrastructure, two main acquisition techniques can be employed: wayside monitoring systems and onboard monitoring systems. In the case of onboard monitoring, portable, non-invasive devices, such as GNSS, accelerometers, strain gauges, etc., can be installed on the vehicle to acquire data on both track and train. This enables a reliable diagnosis of the system's state. When such devices are equipped on the rolling stock, they form an instrumented wheelset which functions as a sensing system for the train. These technologies are used to measure wheel-rail contact forces for several purposes, such as monitoring of rolling stock to evaluate ride quality of the tested vehicle. Conversely some measurement system can be used to locate defects and identify the state of the line contributing to a faster and safer management of maintenance activities [1]. It is well-known from numerous studies [2,3] that signals originating from wheel/rail irregularities and defects, such as squats, roughness, wheel flats, etc., can be captured. These signals provide information to estimate their position along the track and ensure accurate and targeted predictive maintenance of the infrastructure or rolling stocks. To achieve this, understanding the frequency response of the wheelset plays a crucial role in inverse-forces identification from diagnostic and in-service high-speed trains. Conventionally, a bandwidth of 20÷40 Hz limits data acquisition to low-speed information for running dynamic tests [4], while no bandwidth is still specified for diagnostic purposes on the standards. Improvements can be made to enable the consideration of the high-frequency content related to defects. While various sensors can be installed on wheelsets (e.g. accelerometers, displacement sensors, hall sensors, etc.), this chapter focuses on a specific case where an existing wheelset equipped with strain gauges is replicated and simulated using the finite element method (FEM). The model and its frequency analysis results are then compared

with a reduced state-space model of the same wheelset, extracted from the FEM model, using MATLAB.

4.2 Tested strain gauges layout

To test the current setup, the instrumented wheelset was considered equipped with 16 strain gauges placed on the wheel web field sides of the two wheels. This configuration included a total of 2 Wheatstone bridges for each side of the axle shaft [5-6]. Figure 4.1 illustrates the positioning of the strain gauges on the wheel web. The arrangement was as follows: 8 strain gauges (4 + 4), denoted as quadrature bridge "Q_x_s", symmetrically positioned relative to the vertical and horizontal reference axes of the web face, where x indicates the orientation (vertical or horizontal, sine and cosine respectively), and s denotes the side (A or B for right and left wheel, respectively).

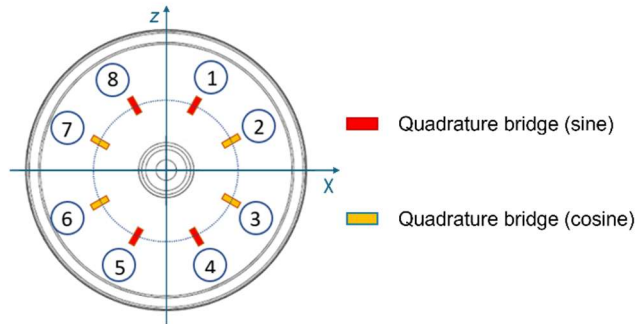


Figure 4.1: Schematic representation of the instrumented wheelset tested strain gauges layout.

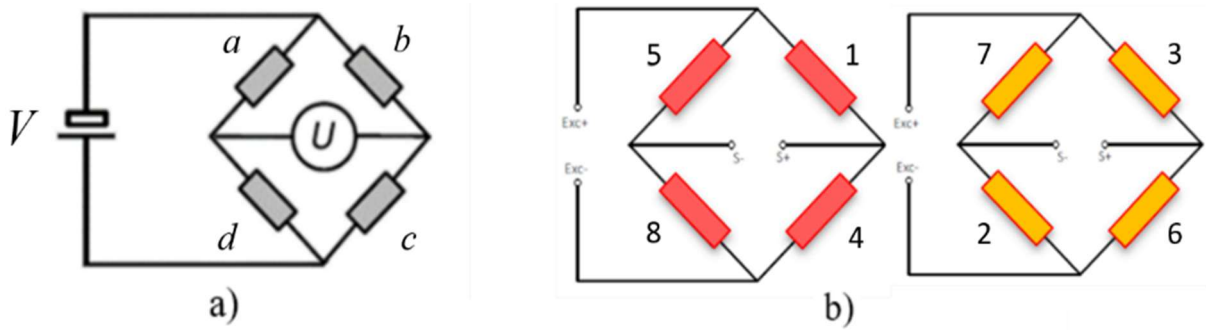


Figure 4.2: Wheatstone bridge schematic representation: a) full-bridge configuration [6], b) quadrature sine (left) and cosine (right) Wheatstone bridges.

Referring to the scheme in Figure 4.1 and Figure 4.2, strain signals originating from the wheel-rail interaction are treated as follows from equation (4.1):

$$U = \frac{VG}{4} (\varepsilon_a - \varepsilon_b + \varepsilon_c - \varepsilon_d) \quad 4.1$$

where U is the Wheatstone bridge's output signal, V denotes the input tension, G is the gauge factor, and $\varepsilon_{(a-d)}$ are the strains of the respective strain gauges (or equivalent strain in case of multiple series/parallel configurations of strain gauges in the same leg circuit leg). The root sum square (RSS) of the sine and cosine quadrature bridge signals is calculated to obtain a continuous,

nearly flat signal (otherwise sinusoidal-like, due to the alternating nature of the acquisition), that is independent of the wheel rotation angle [7]. The sign of the signal is lost due to this operation, but the main advantage is to reduce the waviness of the signal, which can be better interpreted. Equation (4.4) provides an example of the combined signal.

$$Q_{sin} = \frac{VG}{4} (\varepsilon_1 - \varepsilon_4 + \varepsilon_8 - \varepsilon_5) \quad 4.2$$

$$Q_{cos} = \frac{VG}{4} (\varepsilon_3 - \varepsilon_6 + \varepsilon_2 - \varepsilon_7) \quad 4.3$$

$$Q = \sqrt{Q_{sin}^2 + Q_{cos}^2} \quad 4.4$$

Given the significant complexity of the actual system mounted on a bogie during operation, a preliminary study aimed to approximate the behaviour of a free-free system was undertaken. A 'free-free condition' denotes a state where the system is entirely devoid of external mechanical constraints or boundaries. This allows it to move and behave naturally without any imposed limitations or restrictions. Such an ideal condition permits the observation of the natural frequencies of the studied object without disturbances arising from interactions with mechanical constraints. It was also chosen to avoid the modelling of the supports used in other performed tests, with rubber and plastic spacers, and other components with unknown characteristics.

The closest constrain setup for real world tests is a suspended configuration with belts as shown on Figure 4.3. All the dynamic tests were performed using this constrained configuration.

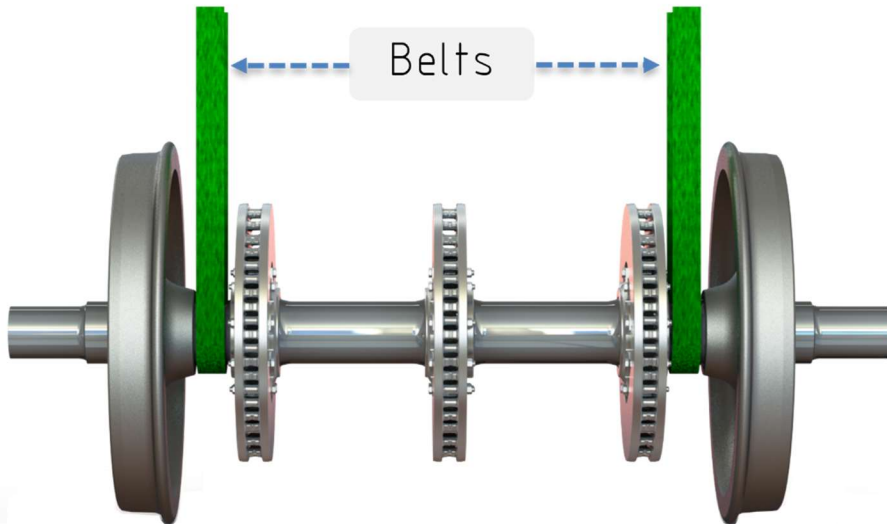


Figure 4.3: Wheelset suspended configuration.

To evaluate the performance of the measurement chain, the focus is directed to a specific Wheatstone bridge, namely Q_{sin_A} (quadrature vertical Wheatstone bridge), which is bonded to the wheel web field side, located on wheel "A". To begin the study, a modal analysis was

conducted on the FE model to determine the eigenvectors and eigenvalues (mode shapes and natural frequencies) of the model with no frequency limits. Subsequently, a frequency response analysis was performed to examine the behaviour of the wheelset in the frequency range from 0 to 1500 Hz, under various load cases and with different positions of input forces [8-9]. The FE model employed in this study yielded a total of 32 outputs (2 points for each strain gauge) and 15 inputs, representing the forces applied in each direction (X, Y, and Q: longitudinal, lateral, and vertical, respectively) at five different locations on the wheel tread. For the test an instrumented hammer (impulse hammer model PCB Piezotronics inc 086B20, with accelerometer 308B, and amplifier 480B06) with heads characterised by different softness was used. Hammer and amplifier instrumentation are shown on Figure 4.4. The tested IWS mounted on pedestals is shown Figure 4.5.



Figure 4.4: Instrumented hammer and amplifier used for the hammer test.



Figure 4.5: Picture of the tested instrumented wheelset in the laboratory.

Signals from strain gauges were acquired with an imc CRONOScompact acquisition board, shown in Figure 4.6.



Figure 4.6: Acquisition board: imc CRONOScompact.

The following figures (from Figure 4.7 to Figure 4.9) show some of the recorded signals (Figure 4.7, Figure 4.8) and processed strain gauge spectra (Figure 4.9).

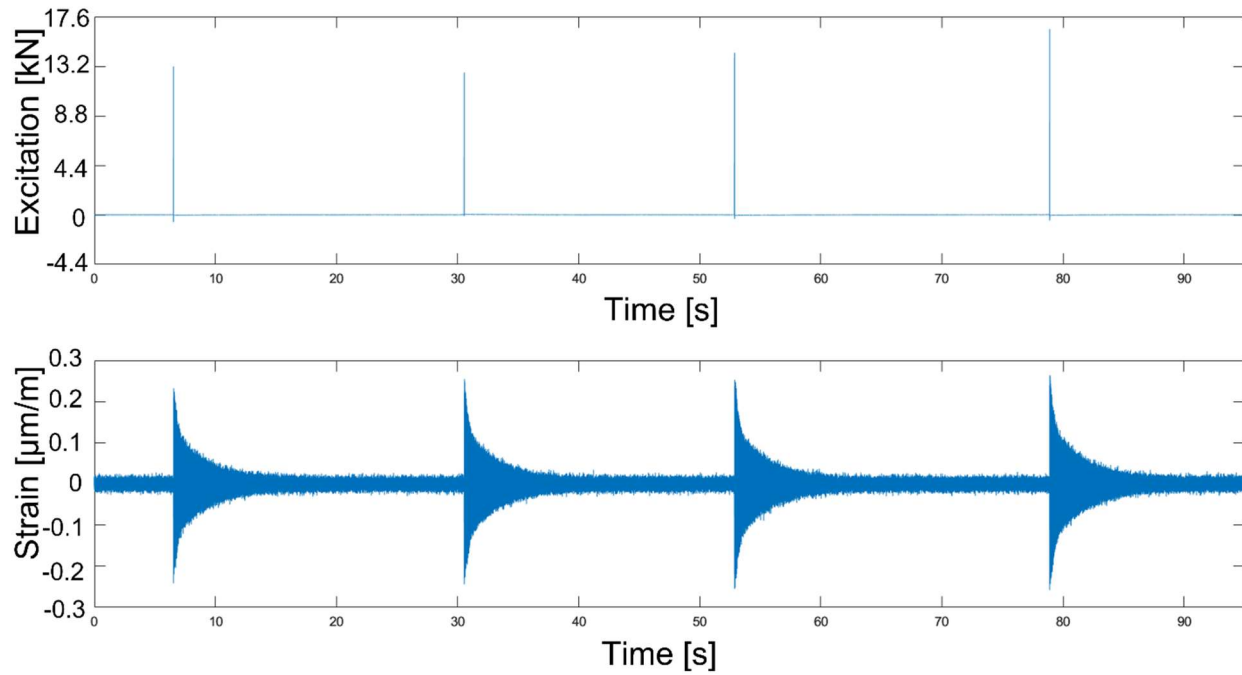


Figure 4.7: Hammer test excitation signal (top) and example of a single strain gauge output (bottom).

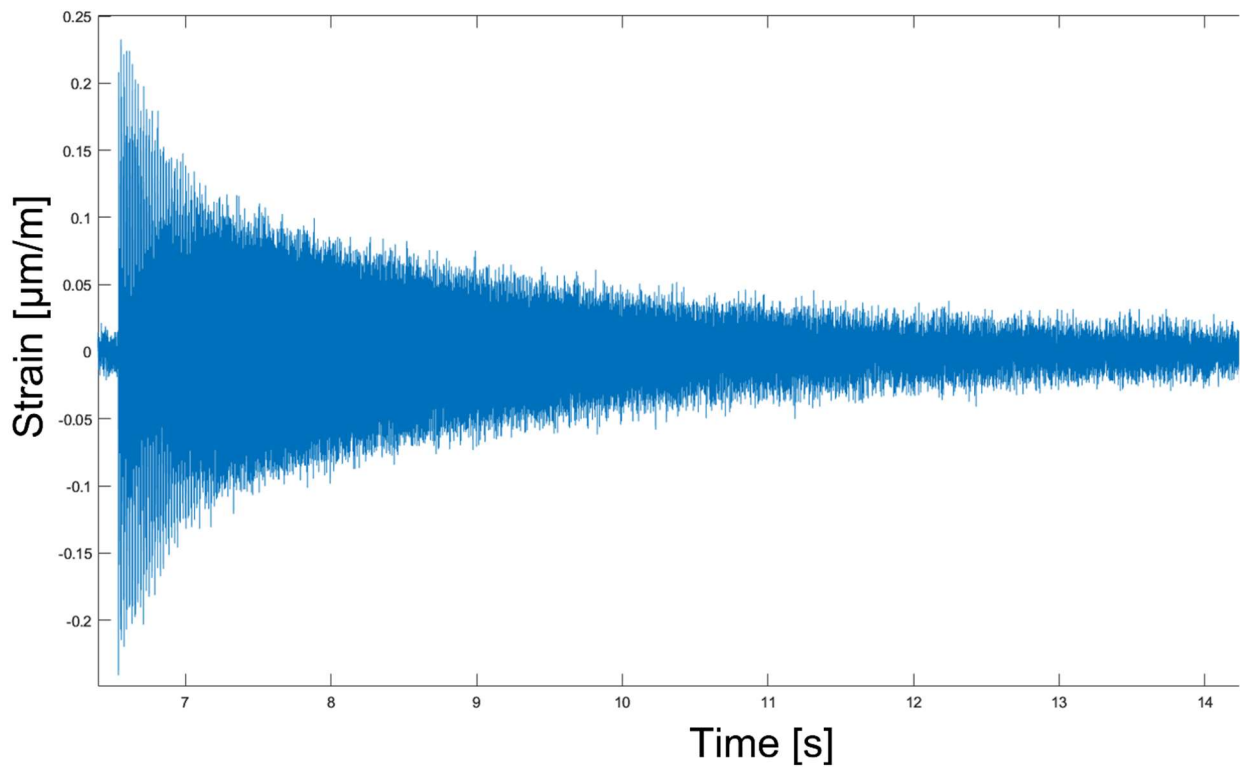


Figure 4.8: Recorded strain gauge output after hammering the wheel.

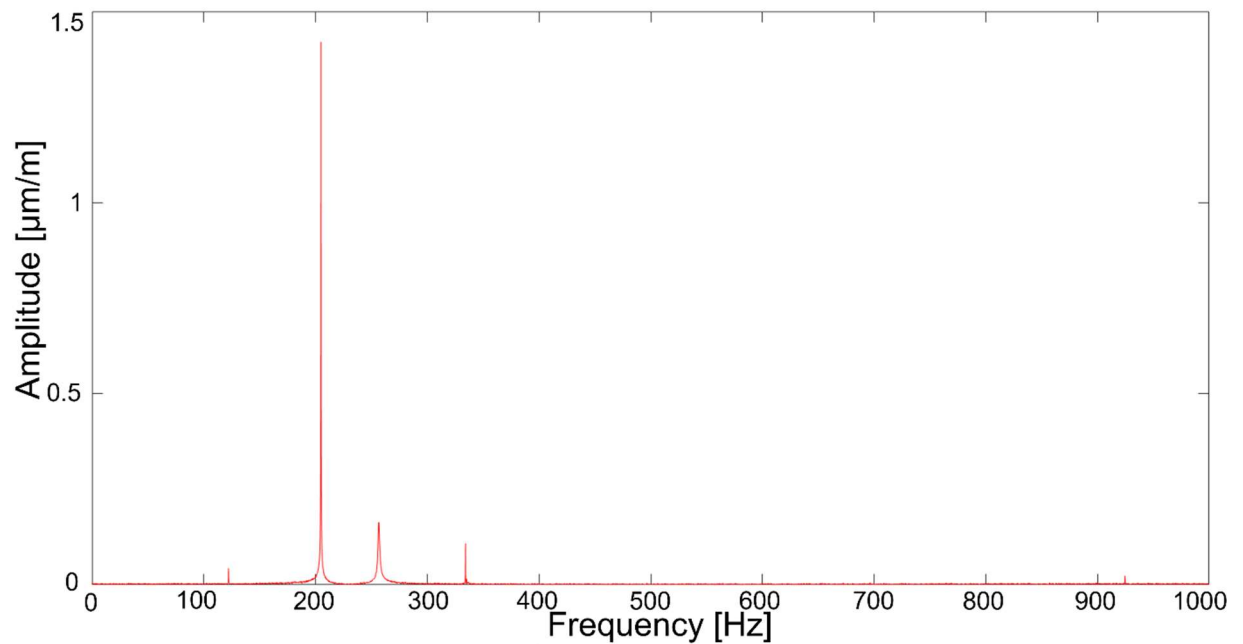
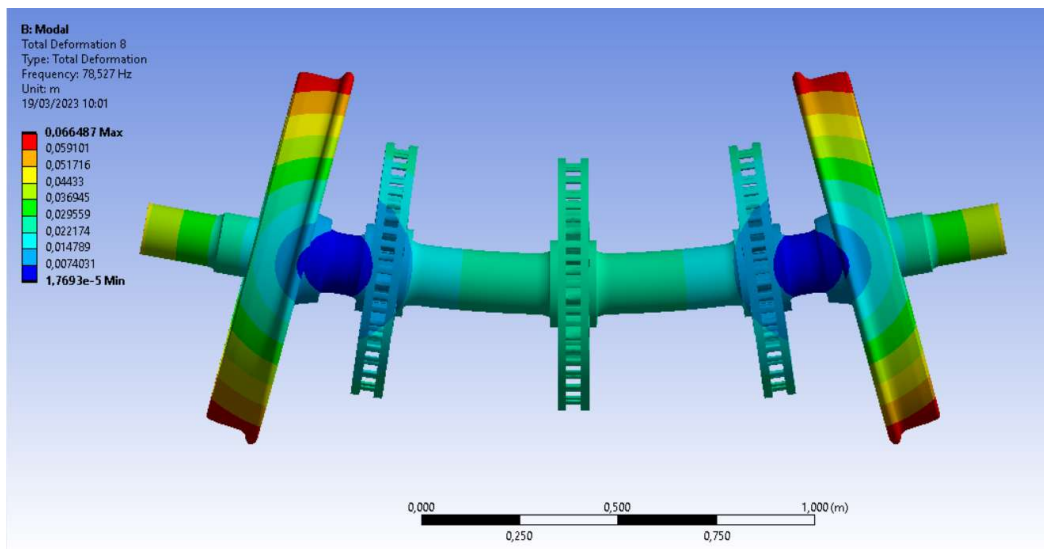


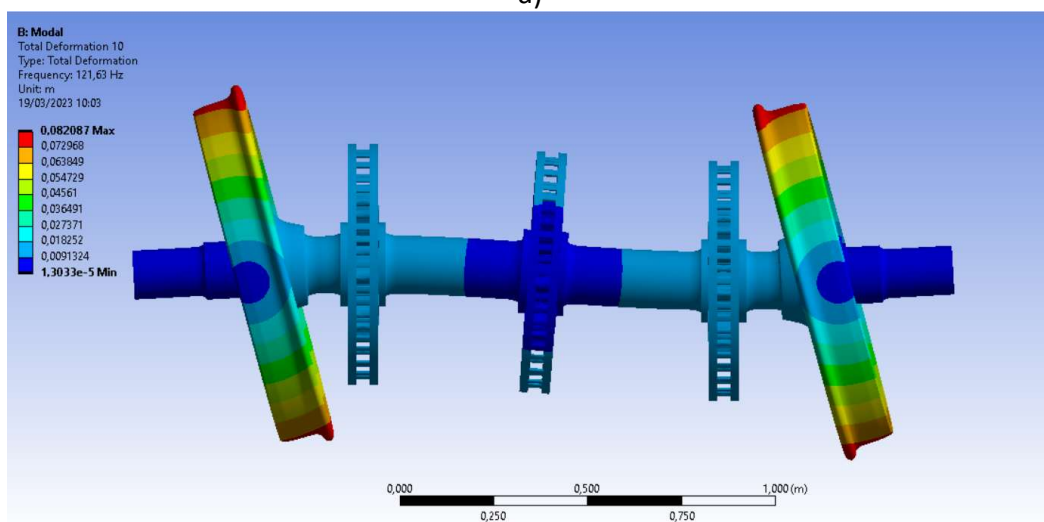
Figure 4.9: Strain gauge output spectra.

In the FE model (see chapter 3), which includes the brake disks, the bending-type modal deformations of the axle and the wheel web were investigated. The comparison between experimental and numerical analyses shows good agreement of the results. The 1st bending mode (mode 2 in Figure 4.10(a)) exhibits a resonance frequency of 78 Hz, the 4th mode (Figure 4.10(b)) occurs at 121.63 Hz, and the 6th mode (Figure 4.10(c)) at 185 Hz. The reference signals

from the quadrature bridges at radii $\Phi 1$ and $\Phi 3$ (anonymised for NDA restrictions) display two clear resonances at 72.5 Hz and 122.22 Hz, similar to those observed for the quadrature bridges, although the peak amplitudes at these frequencies differ. The 8th and 9th mode shapes are also characterised by a bending state in the wheels in phase and counterphase (Figure 4.11). The quadrature bridges located on the wheel plate are sensitive to bending stress states, unlike the distributed bridges, whose strain components in tension and compression tend to cancel each other out, yielding an approximately null output signal. Conversely, the distributed bridges are particularly sensitive to deformation modes known as “umbrella” or “bell” shapes, in which a uniform tensile or compressive state is present along the wheel radii. In such cases, all the strain gauges of a bridge located on the same radius experience nearly identical deformations, producing a coherent and high-amplitude output signal. In contrast, the quadrature bridges provide a nearly zero output, since the signals from strain gauges on adjacent arms subtract with equal magnitude.



a)



b)

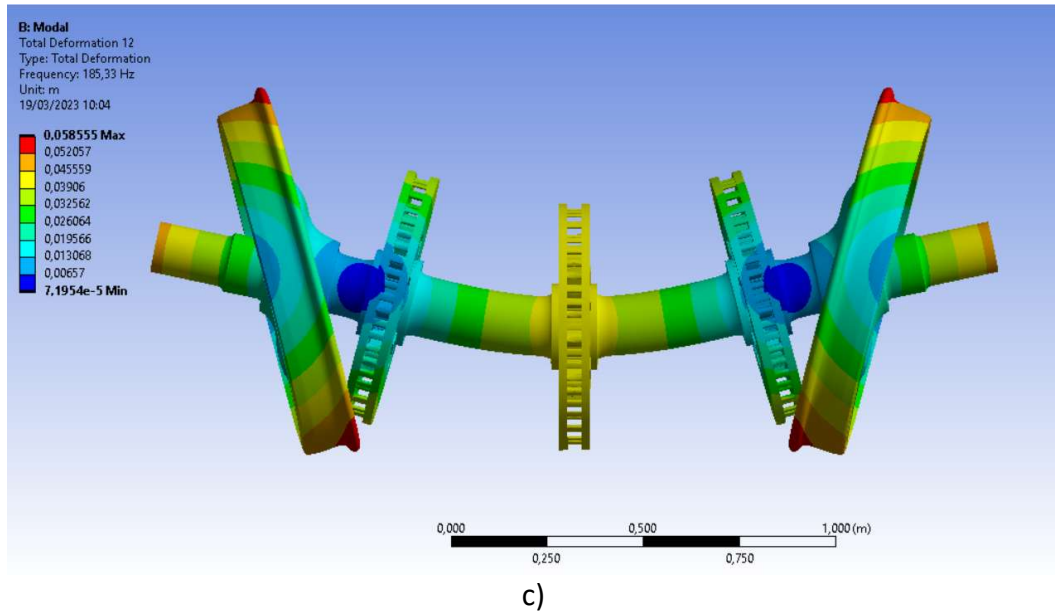


Figure 4.10: Wheelset modal and frequency response analysis results: axle bending. a) mode 2; b) mode 4; c) mode 6.

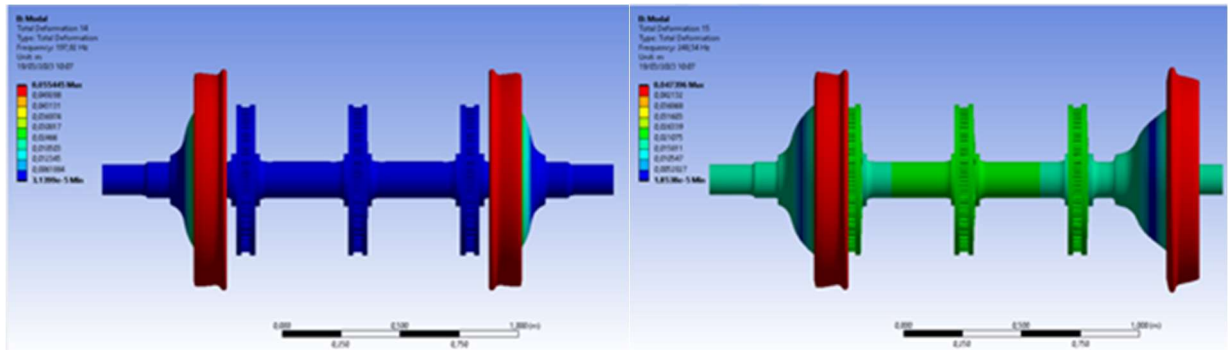


Figure 4.11: Wheelset modal and frequency response analysis results: Wheels' modal shapes. a) mode 8; b) mode 9.

4.3 Wheelset FE simulation frequency response analysis

Referring to the general theory of state-space (SS) reduced models, the n second-order equations representing and approximating the dynamic system, can be transformed into $2n$ pairwise-coupled first-order equations (state equations):

$$\{x\} = \begin{Bmatrix} \{z\} \\ \{\dot{z}\} \end{Bmatrix} \quad 4.5$$

$$\begin{cases} \dot{x} = Ax + Bu \\ y = Cx + Du \end{cases} \quad 4.6$$

Where x represents the state vector, z and $\dot{z}$ are the modal coordinates and velocities, A , B , C , and D correspond to the system (or state), input, output, and feedthrough matrices, respectively, and u is the input vector. The matrices and vectors were derived from simulations conducted in FE software and processed using MATLAB. Output signals y , which indicate Wheatstone bridges deformation, were obtained by solving the model equations. Transfer functions between the input u and output signals y were calculated for each Wheatstone bridge and input force. To analyse the Q_sin_A bridge, the initial MIMO (Multiple-Input Multiple-Output) setup was decomposed into a set of SISO (Single-Input Single-Output) state-space models. This involved considering one input force at a time for each of the four Wheatstone bridges and performing linear matrix operations on the state-space model. A transfer function was then obtained for the Q_sin_A bridge and compared with the results of the FEM simulation.

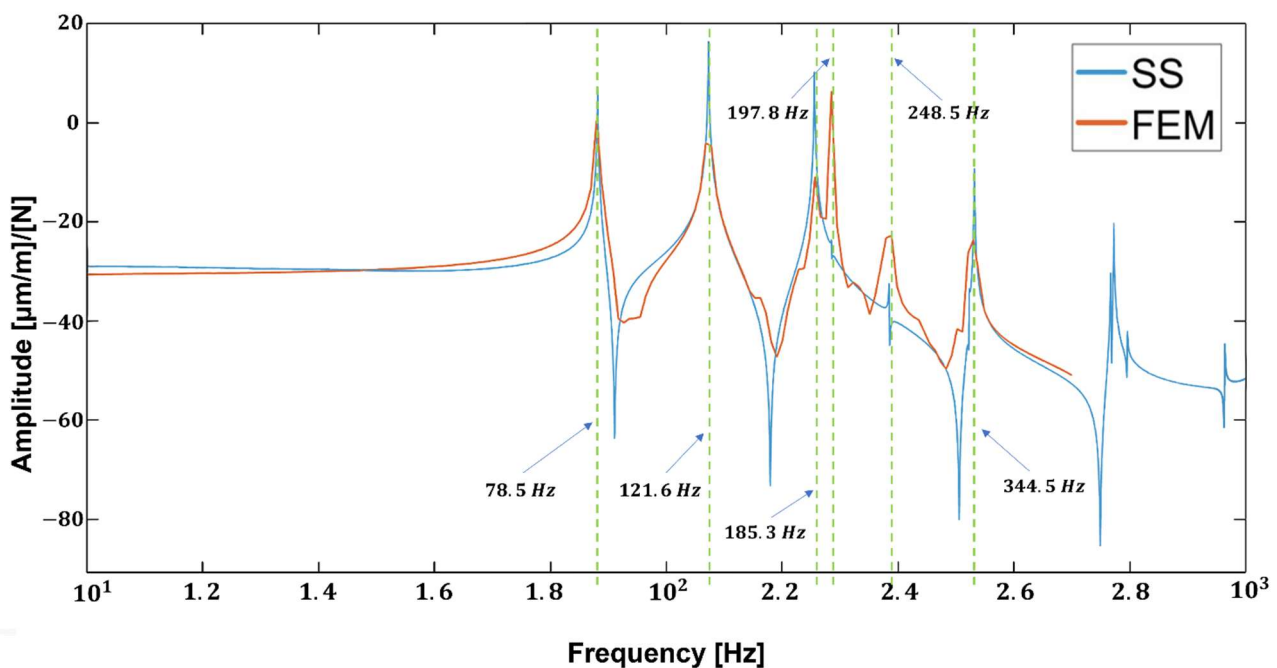


Figure 4.12: Comparison of the state-space (blue) and FEM (red) transfer functions.

The two models are in good agreement regarding the low frequency response, up to approximately 500 Hz. The state-space calculated transfer function accurately depicts the first three represented natural frequencies in Figure 4.12 (corresponding to 78.5 Hz, 121.6 Hz and 185.3 Hz respectively), relative to the bending deformation of the wheel web, and maintaining a correspondence with the fifth and sixth natural frequencies. For the Nyquist–Shannon sampling theorem, the FE model response was trimmed at 750 Hz (displayed 500 Hz). The amplitude is of minor interest due to the use of a first try damping factor of 0.001. It can be seen from Figure 4.12 that the FEM-calculated transfer function trims resonance peaks due to resolution limitations, necessary to maintain the lowest computational cost and simulation time, while the state-space representation offers higher resolution, depending on a user input parameter on the MATLAB script, fully highlighting resonance peaks and anti-resonances.

Coherence was calculated for further validating the model. Let $x(t)$ be the input and $y(t)$ the output of a linear time-invariant (LTI) system. The relationship between these two signals can be

modelled through the impulse response $h(t)$. In the frequency domain, this relation is expressed as:

$$y(f) = h(f)x(f) \quad 4.7$$

where $h(f)$ denotes the system transfer function and represents the ratio between the output and input signals. In the frequency domain is defined as:

$$H(f) = \frac{Y(f)}{X(f)} = \frac{P_{yx}(f)}{P_{xx}(f)} \quad 4.8$$

where $X(f)$ and $Y(f)$ are the Fourier transforms of the input and output signals, respectively, and $P_{yx}(f)$ and $P_{xx}(f)$ denote the cross- and auto-power spectral densities. The TF describes how the dynamic system (in this case, the wheel) transmits the excitation from the input (applied force) to the measured output (strain or acceleration). Peaks in the magnitude of $|H(f)|$ correspond to the system's natural frequencies.

For a SISO (Single Input Single Output) system, two spectral estimators of the transfer function can be defined:

$$H_1(f) = \frac{P_{yx}(f)}{P_{xx}(f)} \quad 4.9$$

where $P_{yx}(f)$ is the cross power spectral density between y and x , and $P_{xx}(f)$ is the power spectral density of the input signal. This estimator is valid when the noise is uncorrelated with the input.

Similarly, one can define:

$$H_2(f) = \frac{P_{yy}(f)}{P_{xy}(f)} \quad 4.10$$

where $P_{yy}(f)$ is the power spectral density of the output, and $P_{xy}(f) = P_{yx}^*$ is the complex conjugate of the cross spectrum. This estimator is appropriate when the noise is uncorrelated with the output.

Figure 4.13 compares the experimental and numerical transfer functions (TFs) obtained for the Q_01_sin_A bridge.

The coherence function $\gamma^2(f)$, plotted in the lower panel, quantifies the linear correlation between input and output in the frequency domain:

$$\gamma^2(f) = \frac{|P_{yx}(f)|^2}{P_{xx}(f)P_{yy}(f)} = \frac{H_1(f)}{H_2(f)} \quad 4.11$$

Its value ranges between 0 and 1.

A coherence close to unity indicates a strong linear relationship and reliable transfer-function estimation at that frequency, whereas a low coherence highlights measurement noise or nonlinear behaviour.

In this case, the experimental and numerical TFs show good agreement up to approximately 265 Hz, with corresponding peaks indicating coincident resonance modes (see Figure 4.13). Beyond

this frequency, differences become more evident, as coherence progressively decreases, revealing reduced signal correlation and increasing influence of measurement noise and model simplifications.

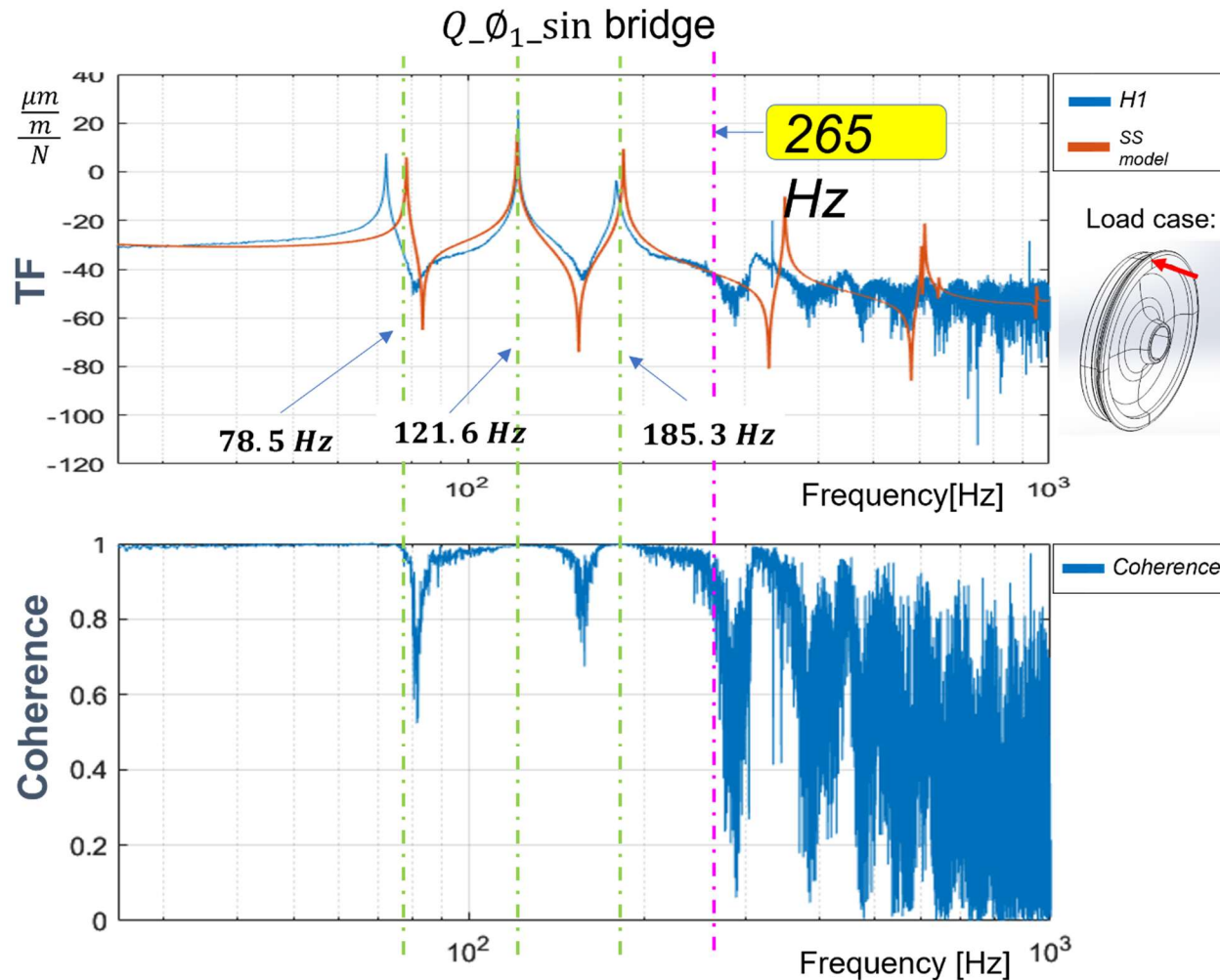


Figure 4.13: Comparison of the experimental (blue) and state space (red) transfer functions with Coherence.

4.4 Comparison between FEM, state-space-model and experimental transfer functions

In this chapter, a comparison between the frequency response analysis results of a state-space model derived from a FEM-simulated instrumented wheelset and the FEM-simulated frequency response itself is performed. The real system's approximate linearity enables its linear state-space representation and the subsequent application of superposition techniques. By leveraging linearity and modal superposition, displacements are evaluated as a linear combination of eigenmodes [10]. Output signals are treated as engineering strain due to the impossibility of directly extracting strains from the FEM simulation, which may affect result accuracy. A strong dependence was observed between the output nodes positions on the FE model, corresponding to the strain gauges, and their respective output signals. A strategic

selection of output nodes can enhance strain values and, consequently, the simulation's accuracy. The study demonstrates the state-space model's capability to represent the dynamic behaviour of the FE wheelset under free-free constrained conditions, significantly reducing computational time. The calculation time has been reduced from 16 hours using FEM to just 45 seconds using the state-space model in MATLAB. Additionally, it improves memory usage, reducing it from 80 Gb in the low-resolution FEM simulation to just 79 kb in the MATLAB state-space script. Furthermore, the resolution of the results has been significantly enhanced, allowing users to manage the calculated transfer function with a user-defined number of points for more effective post-processing. The damping factor of 0.001 utilized in the FE model will be further calibrated in further validation activities. Future improvements of the model will focus on directly extracting strain output and developing a refined non-physics state-space model in the MATLAB environment. This will involve using the FE-obtained state-space as initial conditions for the definitive state-space identification. Subsequently, different constrained configurations will be modelled to replicate more realistic operating conditions, and the study will be extended to include all Wheatstone bridges installed on the IWS.

Further comparison between the experimental FRF and the state-space FRF derived from the FE model shows good agreement of resonance/anti-resonance locations up to approximately 260 Hz. The experimental input-output coherence is high over the same range (with expected local drops near notches), indicating adequate signal-to-noise ratio and quasi-linear behaviour under small-deformation, impact-hammer excitation. These findings support the validity of the linearised SS model in the low-to-mid frequency range and suggest an indicative usable bandwidth of at least ~200 Hz for a single Wheatstone bridge under the present test conditions. However, because the modal tests employed an instrumented hammer with limited high-frequency content and boundary conditions different from in-service operation, Figure 4.13 plot should not be interpreted as a definitive in-line passband of the IWS. In practice, the effective bandwidth can be further extended or refined by combining multiple bridges through calibration matrices and applying proprietary post-processing, as typically done in operational measurements. A formal bandwidth verification will require controlled excitation (e.g., shaker or harder hammer tips), MIMO testing, and SNR/linearity checks across bridges to decouple sensor limits from excitation constraints. Further bandwidth improvements may be investigated under real operating conditions, which strongly influence the effective frequency limits, as noted by Baeza et al. [11]. According to their findings, the effective bandwidth of a non-rotating, object, subjected to a single point excitation is higher than that of the same object when rotating and constrained under actual operational conditions. This highlights the need for dedicated field tests to reliably assess the effective bandwidth of the instrumented wheelset, providing valuable insights into the capabilities and limitations of the system under realistic service environments.

4.5 References

1. Bracciali A., Cavaliere F., Macherelli M., "Review of Instrumented Wheelset Technology and Applications", in J. Pombo, (Editor), "Proceedings of the Second International Conference on Railway Technology: Research, Development and Maintenance", (2014).
2. Nielsen J.C.O., "High-frequency vertical wheel-rail contact forces-validation of a prediction model by field testing", *Wear*, 2008; 265:1465–71. (2008).
3. Gullers P., Dreik P., Nielsen J.C.O., Ekberg A., Andersson L., "Track condition analyser: identification of rail rolling surface defects, likely to generate fatigue damage in wheels, Using instrumented wheelset measurements", *Proceedings of the institution of mechanical engineers*, vol. 225. Part F: J Rail and Rapid Transit; 2011. p. 1–13.
4. Licciardello R.V., Broggiato G.B., Bruner M., Corazza G.R., Malavasi G.: "A Comparative Study of the Results of Field Tests carried out with different Contact Force Measurement Methods", *Proceeding of the 7th World Congress on Railway Research (WCRR)*, Montreal, Canada, pp.1–10, (2006).
5. Alessandria M., Dotta B., Licciardello R.V.: "Long-term contact force measurements with the CML method", *Scienza e Tecnica, Ingegneria Ferroviaria*, p.929-947, 11, (2011).
6. Broggiato G.B., Cosciotti E.: "Ottimizzazione numerica della disposizione di ponti estensimetrici sulle sale ferroviarie utilizzate nella misura delle forze di contatto ruota-rotaia", *Ingegneria Ferroviaria*, 879-889, 11/ (2006).
7. Gómez E., Giménez J.G., Alonso A., "Method for the Reduction of Measurement Errors Associated to the Wheel Rotation in Railway Dynamometric Wheelsets", *Mech. Syst. Signal Process.*, 25(8), pp. 3062–3077 (2011)
8. Bracciali A., Rissone P., "Analisi modale sperimentale di una sala portante del convoglio ETR500", *Ingegneria Ferroviaria* 7/8, 394-407, (1994).
9. Ewins D. J., "Modal Testing: Theory, Practice and Application, 2nd Edition", *Mechanical Engineering Research Studies: Engineering Dynamics Series*, (2009).
10. Pugi L., Abati A., "Design optimization of a planar piezo-electric actuation stage for vibration control of rotating machinery", *Meccanica*, 55 (3), pp. 581-596, (2020).
11. Baeza, L., Giner-Navarro, J., Knuth, C., Thompson, D. J. Comprehensive Model of a Rotating Flexible Wheelset for High-Frequency Railway Dynamics. *Mechanical Systems and Signal Processing* 2023, 200, 110592. doi:10.1016/j.ymssp.2023.110592.

5 CONDITION MONITORING WITH TIME-FREQUENCY ANALYSIS AND WAVELET TRANSFORM

5.1 Chapter Summary

Unlike rail geometry measurements, force-based measurements are particularly sensitive to impulsive irregularities and defects, including rail head issues such as squats, voids, dipped joints, corrugation, and weld irregularities. Consequently, data from instrumented wheelsets can offer valuable information on both the magnitude and location of such irregularities along the track, for example, through traditional data analysis techniques, such as wavelet transform. In this chapter a method based on continuous wavelet transform (CWT) for identifying and localizing faults along the track using unconventional force recordings is proposed.

5.2 Introduction

Instrumented wheelsets are well-established technology in the railway sector, primarily used for measuring contact forces at the wheel-rail interface. Widely adopted for component homologation and static (or quasi-static) verification of contact forces, they provide a reliable tool for assessing the running characteristics of railway vehicles [1], as regulated by the current European standard EN14363 [2]. While dynamometric wheelsets have proven mature for these applications, their use in condition monitoring remains limited. Expanding their role to support condition-based and predictive maintenance, by complementing optical devices commonly used

for track geometry monitoring [3], offers new opportunities for rail infrastructure assessment. Although there is documented evidence of mutual influence between track geometry and contact forces [4], no analytical formula or comprehensive mathematical model currently defines this relationship. As a result, the use of instrumented wheelsets in condition monitoring has been largely restricted to onboard sensing for derailment assessment [5]. An additional challenge is linking contact forces to specific track geometry parameters such as top, alignment, gauge, twist, and superelevation, a task that is neither simple nor straightforward. This limitation restricts the potential of modern contact force measurement systems, which can capture data at an observable bandwidth often exceeding 100 Hz [6, 7]. Lightly damped modes of the wheelset offer an opportunity to detect impacts and other high-frequency events. These events tend to excite specific eigenfrequencies associated with poorly damped sinusoidal signals, making them easier to identify. Track inspection is primarily conducted using Track Recording Vehicles (TRVs) equipped with inertial measurement units (IMUs) and optical scanners. While optical scanners provide valuable data on rail track geometry, their spatial resolution limits the detection to geometric faults with wavelengths longer than twice the sampling spatial frequency, which is set by standards at 250 mm. Consequently, significant rail head irregularities may be over-looked. Additionally, track geometry parameters may remain within safety thresholds despite the presence of faults, particularly in cases involving defects such as dipped joints, squats, or weld irregularities. This can lead to an inaccurate interpretation of track conditions. The lack of detail in geometric measurements can be addressed by analysing force impulses or unusual load fluctuations that arise from wheel-rail interactions with rail head faults, such as squats, roughness, dipped joints, voids, and weld irregularities, as illustrated in Figure 5.1.

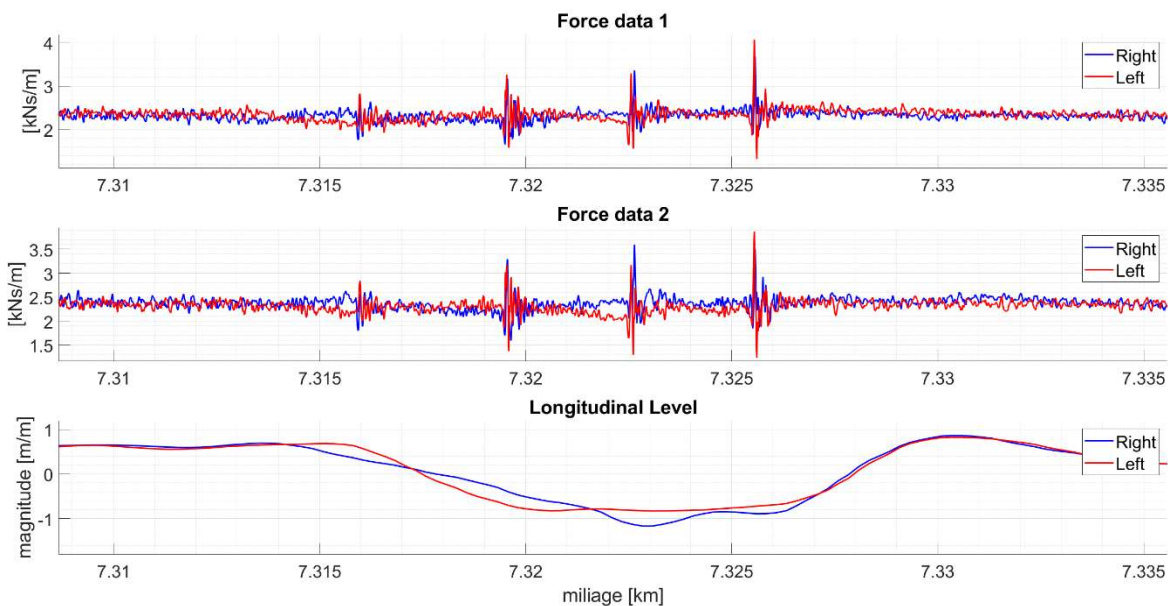


Figure 5.1: Initiation of a top irregularity. Comparison of track geometry longitudinal level and recorded vertical forces scaled with vehicle speed, for two wheelsets on the same bogie. Top: leading wheelset. Middle: trailing wheelset. Bottom: normalized longitudinal level with maximum admissible value. Red lines indicate the left side, blue lines indicate the right side of the TRV.

Time series data for these faults, particularly for squats, dipped joints, and weld defects, are well-documented in the literature [3, 8–11]. Although each fault type generates a unique response influenced by train speed and contact patch position, some characteristics remain consistently

recognizable. Typical examples are the P1 and P2 forces associated with dipped joints, squats, and welds, as shown in Figure 5.2. Other features may be represented by patterns of oscillation with local maxima and minima. These established characteristics, present in the force recordings used in this study, serve as a reference for an initial assessment of likely track faults detected by TRVs.

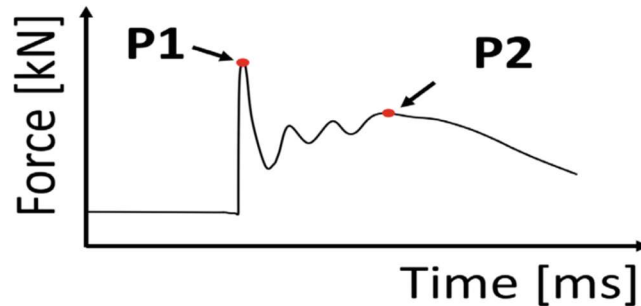


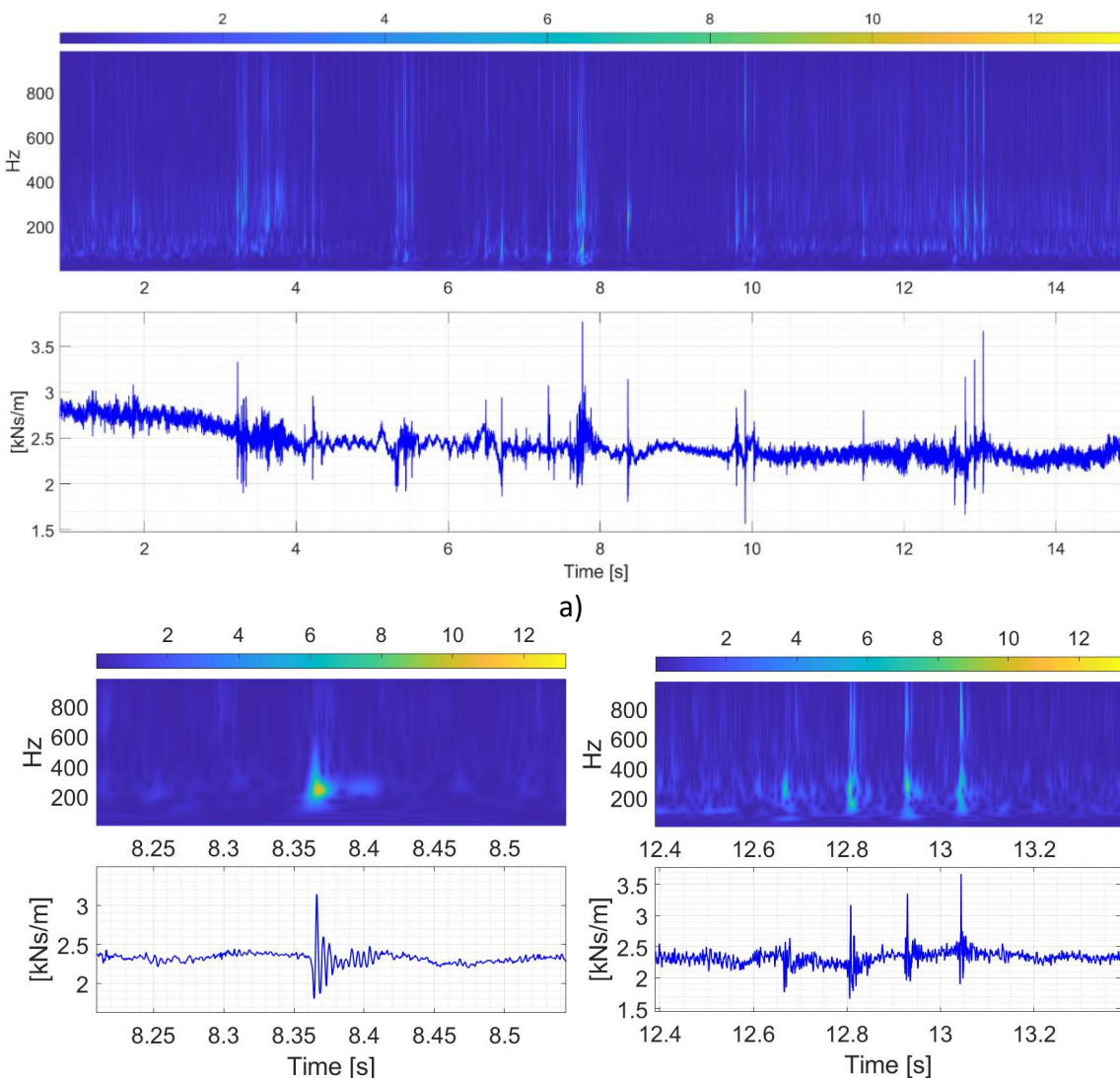
Figure 5.2: Vertical wheel-rail force under impact excitation: train passing over a dipped-joint [8].

In Figure 5.1, although the track geometry data remain within the standard's quality limits [4], a distinct depression is evident in both the left and right rails, with a maximum vertical deflection reaching the admissible value for the maximum allowed track speed. In this case the alert limit (AL) in the D1 wavelength range is exceeded, although the intervention limit (IL) is not. D1 is the smallest wavelength range specified by the standard [4] in which track geometry must be filtered in the space domain. For this purpose, the standard prescribes a second-order linear band-pass filter, with band-pass wavelengths ranging from 3 to 25 m. When viewed solely through track geometry data, this type of fault provides limited information. However, force measurements from the same track section reveal four distinct peaks and fluctuations, suggesting additional rail irregularities that optical equipment may have missed. While finer representations of rail vertical irregularities can be captured through scanned geometry, filtering in the 3–25 m wavelength range (D1 range), as in this case, smooths the signal and often masks critical details such as maximum dip amplitude, dip count, and original irregularity's wavelength. This thesis tries to address this information gap identifying faults by pinpointing their locations along the track. We propose a technique based on wavelet time-frequency analysis to identify rail head irregularities by examining variations in force scalogram magnitude. The precise identification and localization of real track faults will contribute to a more efficient and effective maintenance system. Rail transport will benefit from improvements in safety, reliability and efficiency. Furthermore, reducing unplanned service disruptions and maintenance activities will support a more sustainable transportation system for both passengers and goods.

5.3 The Wavelet Analysis

A signal that evolves over time depends on two variables: scale (or frequency) and time. Wavelets offer a powerful tool for precise time-frequency localization by decomposing such signals into frequency components or bands, analysing each with a resolution that matches its scale. Unlike the Discrete Fourier Transform (DFT), which requires extremely narrow windowing for finer time localization, wavelets adapt their time-widths based on frequency. This means that high-frequency wavelets are very narrow, while low-frequency wavelets are much broader. As a result, wavelet transforms can effectively “zoom in” on short, high-frequency phenomena, such as transients or singularities [13, 14], which, in this study, are associated with rail head irregularities

and defects. To locate anomalous forces resulting from the interaction between wheels and rail head faults, a time-frequency analysis was performed using the Continuous Wavelet Transform (CWT) and Bump mother wavelet, in MATLAB [15]. For this analysis, two kilometres of force and track geometry data, recorded by a dedicated TRV over a 75 s duration, were examined. Given the nature of the faults under investigation, it was anticipated that middle to high frequencies would be predominantly excited by these impulsive contact forces. Referring to Figure 5.3-a, focusing solely on the low-frequency content would not be effective for fault detection. Low-frequency forces generally arise from slowly varying or long-wavelength rail irregularities, which differ from the short, localized faults under consideration. Prioritizing low frequencies would divert the focus from the intended goal of the localization technique, which aims to detect impulsive irregularities. To enhance fault detection, the lower portion of the scalogram, corresponding to frequencies below 60–70 Hz, was excluded. This effectively acts as a high-pass filter, simplifying the identification of the desired faults and minimizing positioning errors. This filtering is evident in Figure 5.3 (b) and Figure 5.3 (c), where broader time spans can be observed for high-magnitude values at lower frequencies. Frequencies below this threshold tend to provide imprecise time localization, leading to a less reliable detection and localization method. Conversely, focusing on frequencies above 60–70 Hz allows for a more stretched and accurate representation of the impulsive events.



b)

c)

Figure 5.3: Scalogram of the vertical contact force. Top: scalogram; bottom: vertical contact force. (a) Initial 14 s recording for a broader view; (b) and (c) zoomed-in views of the scalogram and forces showing expected irregularities.

Narrow vertical bands can be identified each time an abnormal vertical force impulse is generated by a rail head defect interacting with the wheels. In fact, defects along the track influence the system dynamics across the entire frequency range under analysis. However, for the purpose of accurately identifying a localized defect, it is more effective to focus on the higher-frequency band, which is typically excited by impulsive phenomena and characterised by very short decay times. Damping (or decay) is measured in oscillation periods. Low-frequency events exhibit longer decay times, since they attenuate over multiple periods associated with those low frequencies. As a consequence, they appear in the scalogram as features with a larger time span. Conversely, high-frequency events produce short-duration impulsive responses and therefore become more easily distinguishable when the analysis or algorithm is designed to emphasise the high-frequency content while filtering out or de-emphasising the lower-frequency components. Figure 5.3 b and c present a zoom-in of the initial seconds of the dataset, highlighting these significant vertical lines, which indicate high spectral magnitudes associated with impulsive events. For an average stationary forward speed of 96 km/h (~ 26.74 m/s), frequencies in the range of 230–530 Hz correspond to the passage over faults with wavelengths of 5–11.5 cm, such parsing these magnitude variations theoretically allows pinpointing the exact moment when a force impulse occurs due to a rail head fault interacting with the wheel. By setting a threshold on the normalized scalogram magnitude, such as 40% of the maximum magnitude as the lower limit, events associated with irregularities (e.g., impacts and impulses) can be detected. This approach excludes features with power levels below 60% of the maximum magnitude, which are considered part of events corresponding to the selected wavelength irregularity and excitation frequency. As shown in Figure 5.4, high force amplitudes do not necessarily correlate with higher scalogram magnitudes, which are more influenced by the frequency content and waveform of the signal.

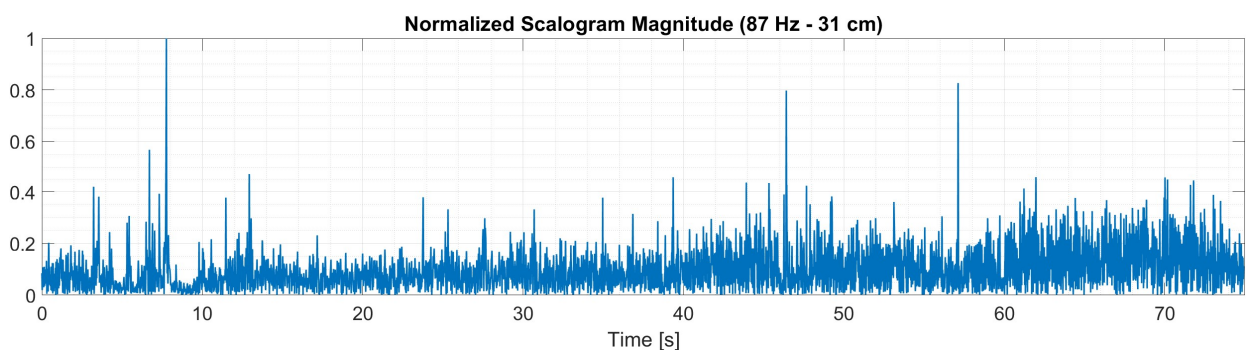


Figure 5.4: Normalized scalogram magnitude of vertical contact forces at 87 Hz, corresponding to 31 cm wavelength irregularities.

The scalogram magnitudes were extracted within the 87–282 Hz frequency range, corresponding to a 9.5–31 cm wavelength range, which includes severe squats, dipped joints, welds, and long-pitch corrugation irregularities. The selection of normalised magnitudes exceeding 40% of the maximum power was performed iteratively for all available scalogram

frequencies. The detected points that fell outside the threshold (outliers) were summed into a single array after each iteration. Figure 5.5 illustrates two steps of this iterative process, showing the points identified at different frequencies as squats [15–18], welds, corrugation, and dip-joints [12]. Longer wavelength faults, like long-wave corrugation or severely degraded dipped joints and squats, typically ranging between 10–30 cm, are expected to appear in the 80–110 Hz frequency band.

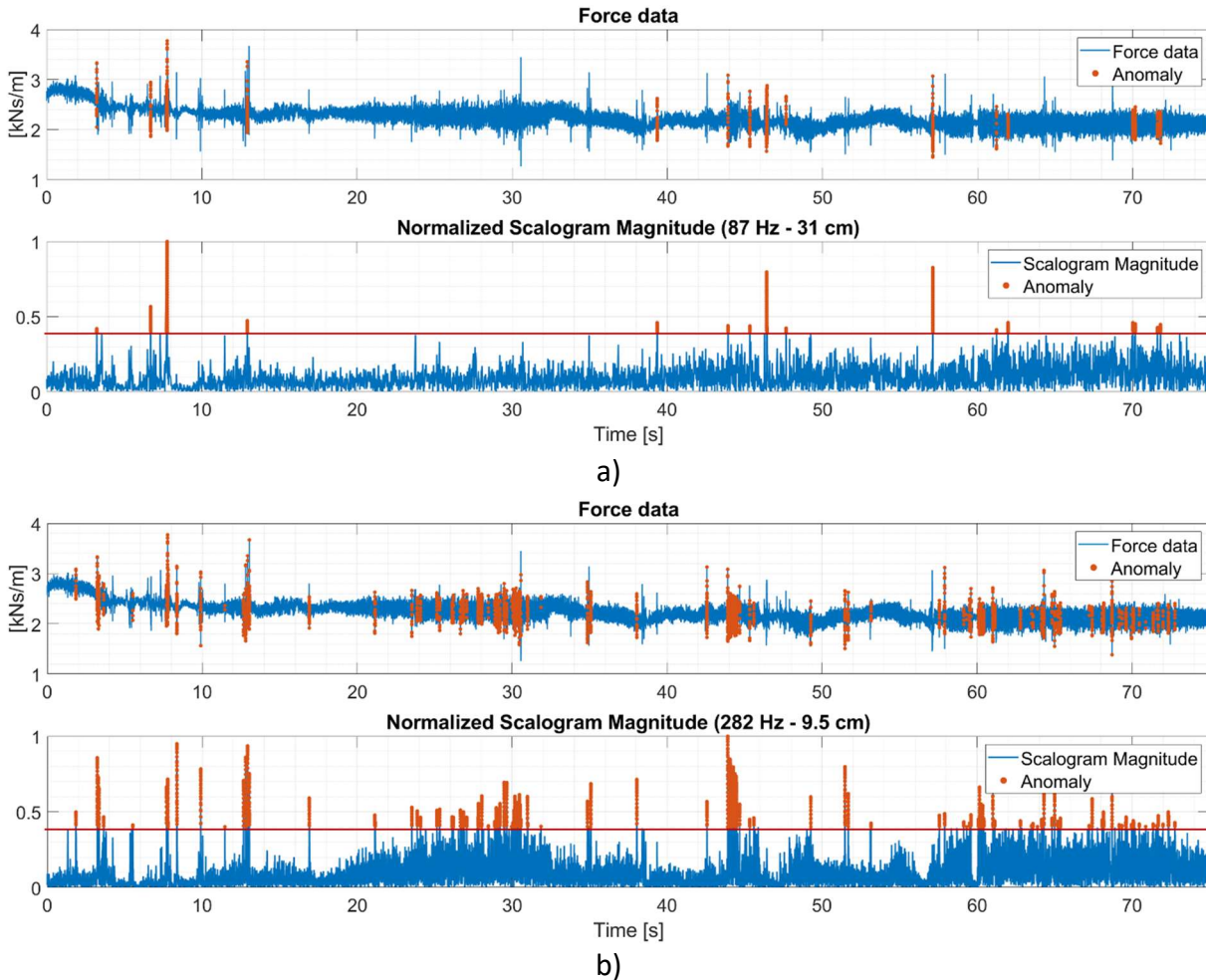


Figure 5.5: Identification of force impulses through scalogram magnitude selection. (a) 87 Hz scalogram row corresponding to 31 cm wavelength irregularities. (b) 282 Hz scalogram row corresponding to 9.5 cm wavelength irregularities.

By extracting the scalogram magnitudes at these critical frequencies, the maximum signal power at a given time and frequency can serve as an indicator for locating faults and irregularities.

5.4 Results Discussion

The identified irregularities along the investigated track section are presented in Figure 5.6. Given that a complete characterization of the rail head condition is not available, meaning that the faults are unknown, a quantitative evaluation of the method's precision and reliability is not possible. However, it is still worthwhile noting that, from a visual comparison, almost all the most prominent vertical force impulses and irregularities have been detected, except for those with

wavelengths outside the investigated range. The lack of comparison data for the validation of the method further explains the narrow frequency band parsed, which excludes many early-stage and short-wavelength irregularities like light squats, welds, and the initiation of dipped joints. Below a 9 cm wavelength, distinguishing false positives from actual rail head defects becomes extremely challenging. Consequently, the upper limit of the frequency range was kept well below 500 Hz (~5 cm wavelength irregularity), which is characteristic of the aforementioned faults at the available vehicle speed. The threshold lower limit plays a critical role in the detection's success. An excessively low threshold would lead to an overestimation of irregularities, capturing those with minimal contributions to the specific parsed frequency. Conversely, a high threshold would underestimate the number of faults. The 40% limit was chosen as a compromise between these aspects. With this method, the temporal (and spatial) position of rail head faults can be estimated. For a complete track condition characterization, the time localization, force waveform time-history, and spectra could be used to identify which defect corresponds to the detected irregularity by comparing its characteristics with known defects from literature or recorded data. This could be further developed into a machine learning approach for automatic identification, classification, and position estimation of short-wavelength track faults. Using vectorized scalogram figures, image recognition techniques could enhance this process.

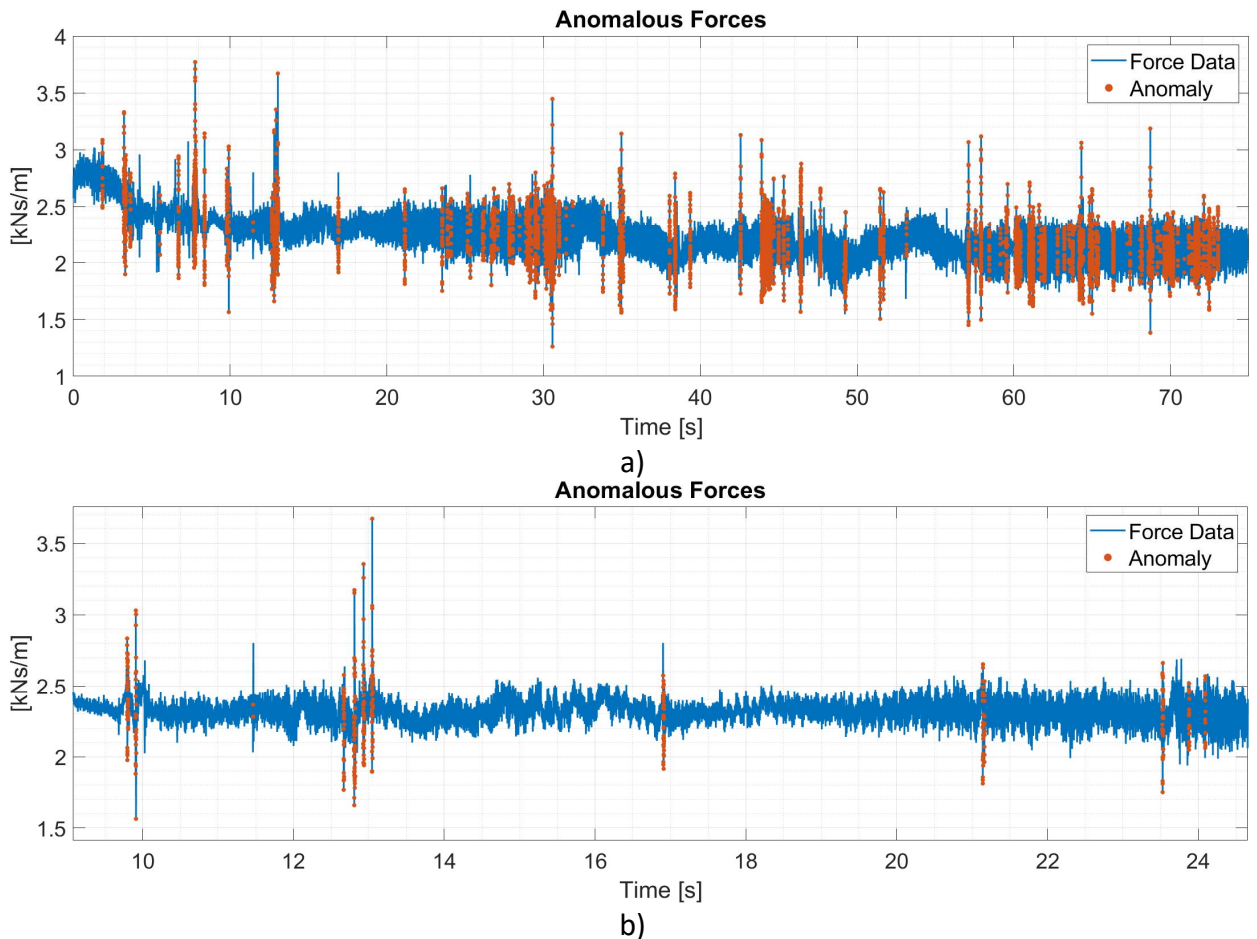


Figure 5.6: Identification of force impulses through higher scalogram magnitudes in the 87–282 Hz frequency range. (a) Identified irregularities in the whole recording; (b) zoom of impulses.

5.5 Conclusions and Future Developments

This study has demonstrated that time and position identification of localized, impulsive, or short-wavelength rail faults can be achieved through wavelet analysis of vertical contact forces measured by instrumented wheelsets on a TRV. Unlike optical equipment, which operates at a fixed spatial sampling frequency of 0.25 m, instrumented wheelsets leverage a higher sampling rate, capturing additional critical information. Irregularities in the 87–282 Hz frequency range, corresponding to a 9.5–31 cm wavelength, including developed squats, dipped joints, welds, ballast voids, and long-pitch corrugation, were successfully detected and their timing accurately estimated. This detection process should be further refined by adding a fault-severity estimation, which will be the subject of future developments. This will involve defining an objective threshold linked to a suitable evaluation metric for both fault occurrence and severity, supported by targeted experimental campaigns including a larger set of well-known and pre-characterised defects.

Although a full characterization of the rail condition is unavailable, a visual comparison of the results indicates that most significant vertical force peaks and impulses were detected, underscoring the method's effectiveness. Future work will focus on integrating wavelet-derived insights with machine learning approaches. Specifically, the vertical force signal waveform and spectral data will be utilized to train algorithms capable of analysing and recognizing scalogram magnitude variations to pinpoint the exact moments of force impulses. Additionally, by parsing the sequence of local maxima and minima in the vertical force time history and comparing spectral characteristics with documented defects, further insights into fault types and characteristics can be obtained. This combined detection and classification framework would enhance maintenance strategies by addressing short-wavelength irregularities and faults, which are often overlooked by conventional optical equipment used in onboard track geometry monitoring.

5.6 References

1. Bracciali, A., Cavaliere, F., Macherelli, M.: Review of instrumented wheelset technology and applications. In: Pombo, J. (ed.) Proceedings of the Second International Conference on Railway Technology: Research, Development and Maintenance (2014)
2. EN14363. Standard. Railway applications-Testing and Simulation for the acceptance of running characteristics of railway vehicles-Running behavior and stationary tests (2022)
3. Hoelzl, C., Dertimanis, V., Landgraf, M., Ancu, L., Zurkirchen, M., Chatzi, E.: Chapter 9-on-board monitoring for smart assessment of railway infrastructure: a systematic review. In: Amir, H., Alavi, M., Feng, Q., Jiao, P., Sharif-Khodaei, Z. (eds.) The Rise of Smart Cities, pp. 223–259. Butterworth-Heinemann (2022)
4. EN13848–5. Standard. Railway applications-Track-Track geometry quality-Part 1: Characterization of track geometry (2019)
5. Bernal, E., Spiryagin, M., Cole, C.: Onboard condition monitoring sensors, systems and techniques for freight railway vehicles: a review. IEEE Sens. J. 19(1), 4–24 (2019)
6. Bellacci, G., Neri, F., Pugi, L., Giachetti, A., Barlacchi, E., Baldanzini, N.: Preliminary frequency response analysis of a contact force measurement system for rail applications. Lecture Notes in Electrical Engineering, vol. 1110 LNEE, pp. 139–145 (2024)
7. Bellacci, G., Pugi, L., Baldanzini, N.: Validation of a wheelset finite element model for static structural analysis and inverse force identification. In: Concli, F., Maccioni, L., Vidoni, R., Matt, D.T. (eds.) Latest Advancements in Mechanical Engineering. ISIEA 2024. Lecture Notes in Networks and Systems, vol. 1124. Springer, Cham (2024)
8. Iwnicki, S., Spiryagin, M., Cole, C., McSweeney, T. (eds.) Handbook of Railway Vehicle Dynamics, Second Edition (2nd ed.). CRC Press (2019)
9. Zhao, X., Li, Z., Dollevoet, R.: The vertical and the longitudinal dynamic responses of the vehicle-track system to squat-type short wavelength irregularity. Veh. Syst. Dyn. 51(12), 1918–1937 (2013)
10. Dukkupati, R.V., Dong, R.: The dynamic effects of conventional freight car running over a dipped-joint. Veh. Syst. Dyn. 31(2), 95–111 (1999)
11. Esveld, C.: Modern Railway Track 3rd edition, MRT-Productions (2017)
12. Lewis, R., Olofsson, U.: Wheel Rail Interface Handbook-CRC Press (2009)
13. Daubechies, I.: The what, why, and how of wavelets. In: Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, pp. 1–16 (1992)
14. Caprioli, A., Cigada, A., Raveglia, D.: Rail inspection in track maintenance: a benchmark between the wavelet approach and the more conventional fourier analysis. Mech. Syst. Signal Process. 21(2), 631–652 (2007)
15. Molodova, M., Li, Z., Núñez, A., Dollevoet, R.: Automatic detection of squats in railway infrastructure. IEEE Trans. Intell. Transp. Syst. 15–5, 1980–1990 (2014)
16. Li, Z., Dollevoet, R., Molodova, M., Zhao, X.: Squat growth—some observations and the validation of numerical predictions. Wear 271(1–2), 148–157 (2011)
17. Molodova, M., Oregui, M., Núñez, A., Li, Z., Dollevoet, R.: Health condition monitoring of insulated joints based on axle box acceleration measurements. In: Engineering Structures, vol. 123 (2016)
18. Rosano, G., et al.: Diagnostics of the railway track-possibility of development through the measurement of accelerations and contact forces. Ingegneria Ferroviaria 79(2), 81–102 (2024)

6 MODEL-BASED RAILWAY CONDITION MONITORING WITH CONTACT FORCE MEASUREMENTS: TRACK GEOMETRY

6.1 Chapter Summary

In this chapter, a model-based method for the estimation of the rail mean vertical alignment is presented. The purpose is achieved through double integration of wheel-rail contact forces measured using dynamometric wheelsets on a dedicated Track Recording Vehicle (TRV). A simplified three-Degrees-Of-Freedom (DOF) linear model of half a train coach has been developed for this purpose. The model's ability to simulate the average left and right longitudinal level has been tested using vertical contact force recordings from a constant-speed track section, as measured by the TRV. The results are compared with available track geometry (TG) data, recorded by the optical system of the same vehicle, used for condition monitoring of the Italian railway infrastructure. Model parameters, such as masses, stiffness and damping of the suspensive system have been optimized. An error analysis has been conducted on results. A good agreement is found between reconstructed and recorded vertical alignment at the D1 level, suggesting the feasibility of using contact forces measured with instrumented wheelsets for railway TG condition monitoring. This computationally efficient approach highlights the potential of strain gauges and instrumented wheelsets as alternative or complementary technologies to the widely adopted accelerometers, rate gyros and optical devices for railway condition

monitoring. Given its low computational cost, embedded and real-time TG estimation could be further investigated.

6.2 Track geometry estimation: Introduction

Rail transportation in Europe accounts for a significant share of total passenger and freight transport (11.9% in 2022) [1], with rapid growth following the COVID-19 pandemic. Ensuring fast, safe, and comfortable journeys is key to increasing rail transport's market share against competitors such as cars, trucks, buses, air, and maritime transport. The condition of railway tracks is critical for the safe and efficient operation of mass passenger and freight transport. Track degradation is a primary cause of service disruptions, delays, and accidents, often linked to derailments or un-scheduled maintenance. To ensure rail travel remains fast, safe, and comfortable, aligning with Sustainable Development Goal (SDG) 9: "Industry, Innovation, and Infrastructure", continuous monitoring and inspection of railway infrastructure are essential. This enables predictive maintenance by assessing real-time conditions and tracking the deterioration trends of railway components. Proper assessment of rail integrity relies on well-defined condition indices, facilitating the adoption of new predictive maintenance strategies. It is crucial to measure physical quantities that directly or indirectly indicate infrastructure degradation. The measurement of wheel-rail contact forces and axle boxes, bogies, and car bodies accelerations onboard railway vehicles plays a fundamental role in this process. A growing trend in the railway sector involves continuous monitoring of infrastructure using sensors installed both on vehicles and trackside. This approach enables condition-based maintenance, triggered when track degradation reaches non-negligible levels. Key parameters, such as the catenary, superstructure, and track foundation, can be monitored using onboard sensors [2-3] typically installed on dedicated TRVs. These vehicles facilitate long-range rail inspections, allowing many simultaneous measurements such as pantograph-catenary interaction, wheel-rail interaction, and track geometry (TG) [4]. Other inspection methodologies, involving the use of portable trolleys, are costly and time-consuming, requiring track occupation for extended periods and the involvement of highly specialized workers and technicians. In the Italian railway infrastructure, each type of railway line is regularly inspected by different TVR equipped with dedicated sensors. Commonly adopted technologies include inertial sensors (accelerometers, rate gyros), video inspection and contactless systems (laser scanners, cameras), force/strain measurement devices (strain gauges, load cells), as well as eddy current and ultrasound sensors [4,5]. Although instrumented wheelsets (IWS) are installed onboard, their use is typically limited to running dynamics assessment rather than TG monitoring for their relatively higher cost.



Figure 6.1: An example of a Rete Ferroviaria Italiana's (RFI) TRV for conventional lines. [4]

Given their direct contact with the rail, without the interposition of a suspension system to mitigate interactions, contact forces measured by these devices can provide valuable insights into rail conditions, including both running surface integrity and TG anomalies. They are primarily employed to evaluate maximum allowable contact forces and to monitor the derailment index (Y/Q ratio), particularly in freight operations. To the author's knowledge, academic research on the use of IWS for railway infrastructure condition monitoring is scarce. While numerous studies focus on optimal strain gauge positioning, layout optimization, and finite element (FE) modelling for assessing strain states for inverse force identification or frequency responses, very few explore the potential of IWS for direct track condition monitoring [6-8].

	Type 1 (n. 5)	Type 2 (n. 7)	Type 3 (n. 2)	Type 4 (n. 15)	DIC-80 US (n. 1)	Falco (n. 2)	Sirter (n. 1)	Aldebaran 2.0 (n. 1)	K12 (n. 3)
									
Track Geometry	✓		✓	✓	✓	✓	✓	✓	✓
Telecommunications	✓		✓	✓				✓	
Overhead line	✓		✓	✓		✓	✓	✓	
Pantograph video inspection	✓		✓	✓					
Signalling	✓		✓	✓					
Minimum Clearance Gauge	✓			✓					
Signalling Video Inspection	✓		✓	✓					
Track Video Inspection	✓		✓	✓					✓
Electric arcs, voltage, Current, Overhead Line Dynamics	✓		✓	✓					
Wheel/Rail interaction	✓		✓	✓					
Running Dynamics	✓		✓	✓					
Switch Parameters				✓		✓			✓
Ultrasonics & Eddy current		✓			✓				
 RFI RETE FERROVIARIA ITALIANA	Conventional lines	All lines	High Speed lines	Nodes, yards and junctions	All lines	Nodes, yards and junctions	Nodes, yards and junctions	Conventional lines	Yards and conventional lines

Figure 6.2: RFI diagnostic fleet and their characteristics. [4].

The chapter is structured as follows. The state-of-art review of TVRs and in-service vehicle technologies for railway condition monitoring is presented in section 6.3 and 6.4. A detailed description of the 3-DOF lumped model developed and used for the analysis is presented in Section 6.5. Results discussion and error analysis are presented in section 6.6 and 6.7 respectively. Conclusions and future research perspectives are given in Section 6.8.

6.3 State-of-the-art review of equipment and methods for railway condition monitoring

A significant portion of academic research in railway condition monitoring focuses on kinematic and inertial measurement. Vibrations and kinematics of specific vehicle components such as car bodies, bogies, or axle boxes are recorded as specified by the European standards EN 13848-1 and EN 14363 [9,10]. These measurements are used to assess TG quality, which includes parameters such as track gauge, lateral alignment, longitudinal level (also referred to as vertical alignment, or top), twist and cross level (also known as cant or superelevation). These

parameters are typically analysed within specific wavelength ranges (D1, D2, and D3), as defined by the standards and summarized in Table 6.1.

Table 6.1: EN 13848 – 1 standardised wavelength ranges.

Range	Wavelength value	Unit
D0	$1 < \lambda \leq 5$	[m]
D1	$3 < \lambda \leq 25$	[m]
D2	$25 < \lambda \leq 70$	[m]
D3	$70 < \lambda \leq 150$ (200)	[m]

In onboard monitoring systems, TG is either measured directly or estimated using redundant inertial or optical sensors. Displacements are typically derived from acceleration data through direct integration or by combining angular velocities with other TG parameters, such as horizontal curvature for cross-level estimation [11,12], or vertical and lateral alignment assessment [13,14]. For these applications, accelerometers and rate gyros are the most common sensors. Additionally, other technologies, such as inertial and chord-based laser scanners, measure optical displacements from the rail centreline, generating high-resolution point cloud mappings of the track [15,16]. A schematic representation of available sensors mounted on the RFI's Diamante TRV [17] for inspection of high-speed lines is shown in Figure 6.3.

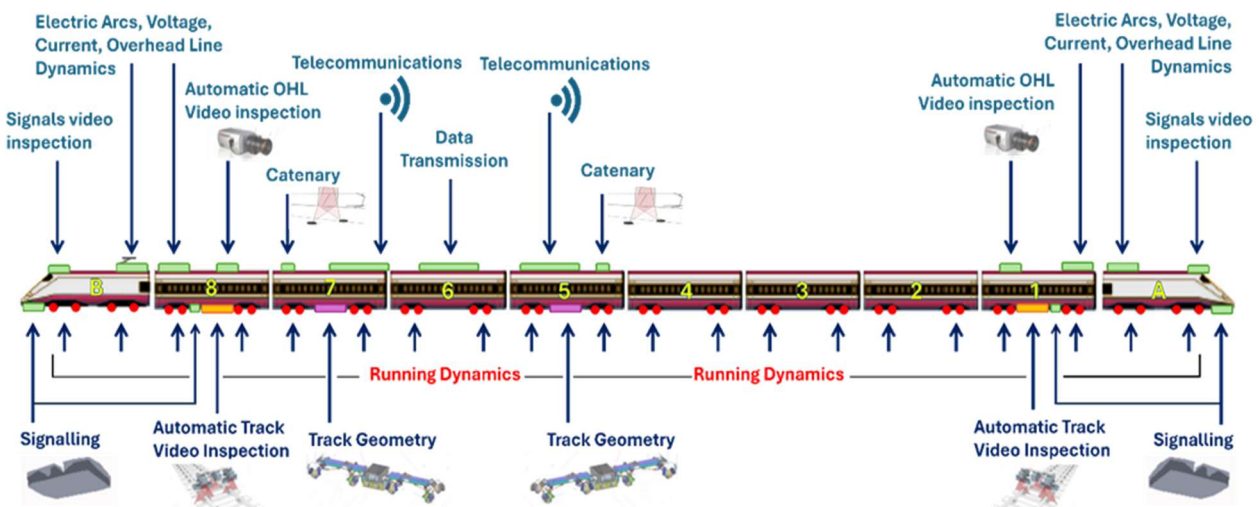


Figure 6.3: Schematic representation of available sensors mounted on the RFI's Diamante TRV for high-speed lines diagnosis [17].

The ability to estimate track geometry parameters, including alignment, longitudinal level, and cross-level, makes these sensors and systems well-suited for railway infrastructure condition monitoring. Parsing maximum lateral and vertical accelerations enables the assessment of ride comfort of train journeys. Furthermore, it serves for vehicles and new components acceptance under specific testing conditions. Despite this, accurately estimating wheel-rail contact forces using only accelerometers remains a challenge for onboard monitoring. In contrast, IWS provide an alternative method for force estimation. These wheelsets employ strain gauges to measure the deformation of structural components, providing a reliable method for an inverse

measurement of wheel-rail contact forces. When properly installed on a component's surface, strain gauges generate an electrical signal proportional to the experienced strain [17-22]. Also referred to as dynamometric wheelsets, IWS are primarily used for specialized applications, such as testing and homologation of vehicles and components. However, their application in condition monitoring remains limited, mainly due to high costs, complex installation procedures, and the need for highly specialized labour. Despite these challenges, they provide high-precision contact force estimations and are considered an established technology for this purpose [23].

Wear of railway components has a significant impact on exchanged contact forces. Contact conditions vary considerably depending upon the shape of the worn wheel and rail profiles, determined by the amount of wear, which in turn affects the overall ride quality [24-26]. A comprehensive review of green tribology and wear evaluation methods is presented in [27, 28], highlighting techniques to quantify wear in engineering surfaces under operational conditions with valuable examples for the maintenance of wheels and rails. However, in the present study, the analysis focuses on a relatively precise bandwidth of measured contact forces, filtered within the selected range and used to derive a reliable estimation of the rail's vertical alignment.

Even under strict measurement standards, small estimation errors can arise during test campaigns [29]. Given the various sources of error and uncertainty in real-world applications, even minor uncertainties in the measurement chain can introduce small offsets and drifts, resulting in non-zero mean terms. These terms become problematic when double-integrating acceleration data to obtain displacements, as they amplify errors throughout the integration process. If a measurement contains spurious terms or offsets, often introduced during sensor calibration or taring, non-zero mean terms emerge in the time-series data. While these small errors may seem negligible in direct comparisons, their impact grows significantly when integrated over time, leading to deviations from the expected physical behaviour of the system. Thus, estimated quantities may exhibit unrealistic or even diverging displacements, inconsistent with the physical model or observed system behaviour. Moreover, integrating noise causes the root mean square (RMS) value of the output to increase over time, even when no actual changes occur in the sensor's state (e.g., motion in accelerometers or strain in strain gauges) [11]. While, in theory, integrating measured data can yield displacements and positions, in practice, this process is often compromised by such errors. A common solution is the application of high-pass filters to mitigate drifts and stabilize the integration of acceleration data derived from measured forces [12]. Offset identification and compensation in real-world measurements are often performed using direct and indirect Kalman filtering [30,31]. However, these filters require a well-defined model of the sensing process, along with knowledge of the covariance matrix and the errors associated with the acquisition instrumentation. In this study, a second order Butterworth high-pass filter has been applied as a pre-processing step to the acquired data, along with the use of a simplified system model. This model could potentially serve as a foundation for the future development of a Kalman filter. For the current work, the focus is on applying a time-domain filter and a simplified model to estimate equivalent displacements, and thereby detect track defects, within the wavelength range specified by the standards. These wavelengths align with the domain of validity of the proposed model.

6.4 Review of models and methods for on-board railway condition monitoring

Lewis and Richards [12] proposed a method for estimating track cross-level using a reduced sensor set. They combined high-frequency vibration data from horizontally oriented accelerometers with low-frequency data from a rate gyro. Using a carefully selected crossover frequency, they employed a third-pole Butterworth filter to reconstruct rail superelevation from onboard measurements.

Weston et al. [13,14] developed methods for estimating mean vertical and lateral track alignment using sensors mounted on in-service vehicle bogies. Their approach relied on bogie-mounted vertical and lateral accelerometers, along with yaw rate sensors, to estimate mean lateral track irregularities without requiring optical or contact sensors. They demonstrated that a yaw rate gyro can serve as an alternative to a roll-compensated lateral accelerometer. Additionally, they introduced an inverse model based on a Kalman filter to improve estimation accuracy at specific vehicle speeds. Their method treated lateral track irregularities as an integrated random walk, with yaw rate serving as the system's observation input.

Sun and Dhanasekar [32], and Zhai and Cai [33], developed a dynamic railway vehicle model to examine vertical interactions between the track and wagon system and for simulating train-track interaction at the wheel-rail interface. Their 10-DOF lumped mass system consists of a carbody, two bogies (including the respective inertial properties), and four unsprung wheelset masses. The track is modelled as an infinitely long beam discretely supported at rail-sleeper junctions, with the contact patch represented by Hertzian contact. Using this model, they investigated the track's dynamic response at rail joints and quantified the relationship between unsprung mass and dynamic wheel-rail forces.

Kawasaki and Youcef-Toumi [34] developed an inverse model for estimating vertical, lateral, and cross-level track irregularities using carbody acceleration measurements. Their approach was validated with data collected onboard a Shinkansen train. While vertical and lateral irregularities were accurately estimated, level irregularities exhibited more bias compared to reference data. The authors highlighted that vehicle mass and speed significantly influence estimation accuracy, requiring different model parameters to accommodate variations.

Mori et al. [35] developed a portable track-condition-monitoring system for high-speed railways. The detection and identification of rail irregularities from in-service vehicles uses acceleration data and carbody roll angle. Rail corrugation is detected using cabin noise and spectral peak calculation. The location of identified faults is pinpointed along the track by a GPS and a map-matching algorithm. Field tests were conducted to determine the performance of the equipment and method.

Odashima et al. [36] proposed a track condition monitoring technique using carbody acceleration measured by in-service vehicles. Authors demonstrated that estimating track irregularities of conventional railway tracks using carbody acceleration only is possible with a bias between experimental TVR data and estimated by the model of 1 mm. Their method uses inverse dynamics by applying a Kalman filter. For the latter track geometry is expressed in a random walk model, incorporating it into a state equation. Impulse response is incorporated as measurement equation supporting variations in vehicle speed.

Alfi and Bruni [37] proposed a method for estimating long-wavelength track irregularities using vehicle acceleration measurements and a 6-DOF model comprising three rigid bodies: the carbody and two bogies. Each body was assigned two degrees of freedom, accounting for vertical displacement and pitch rotation, while all other motions were neglected. The model assumed that the wheels followed track irregularities without separation and that the vehicle moved at a constant forward speed. Acceleration data were processed using a frequency-domain model-based identification procedure, decomposing track irregularities into harmonic components that were identified separately. The method was validated and compared with measurements from a TRV on high-speed lines.

Miwa et al. [38] investigated the influence of mass variation on dynamic wheel-rail load fluctuations through numerical simulations and test rig experiments. Their findings showed that unsprung mass has the most significant impact on wheel dynamic load variations. Reducing wheel mass effectively decreases the acting load, particularly at higher vehicle speeds. In contrast, variations in sprung masses (bogies and carbody) had only a minor influence on dynamic load variations.

Recent research has focused on machine learning (ML) and data-driven methods for railway condition monitoring. Both supervised and unsupervised anomaly detection algorithms, enable rapid processing of large datasets. These methods uncover hidden relationships between input features and key track parameters, even when traditional analytical relationships are difficult to establish. Typically, they classify track sections into "anomalous" or "healthy" categories, making them particularly useful for detecting faults in large railway networks.

Tsunashima [39] implemented a track condition monitoring system using an onboard sensing device and machine learning techniques to diagnose track conditions based on carbody vibrations. The equipment and method were used to automatically detect and classify track irregularities on regional railway lines in Japan.

La Paglia et al. [40] developed a condition monitoring system based on bogie acceleration measurements from in-service high-speed trains. Using prebuilt multiple regression models, the system estimated track longitudinal level. Synthetic indicators, such as the root mean square (RMS) of bogie vertical acceleration, were computed and correlated with the maximum rail top level. This methodology distinguished between isolated and distributed defects and tracked vertical alignment evolution over time.

De Rosa et al. [41] introduced an ML-based classifier for monitoring lateral and cross-level track irregularities using lateral and roll bogie frame accelerations. The classifier was trained on synthetic data and validated using real measurements from a diagnostic vehicle operating at 300 km/h on a straight high-speed rail line. Three ML algorithms were tested: linear support vector machine (LSVM), Gaussian support vector machine (GSVM), and decision tree (DT). While all performed well, GSVM yielded the highest accuracy for identifying lateral track faults, whereas DT had the lowest performance.

Ghiasi et al. [42] proposed an artificial intelligence (AI)-based framework for TG diagnosis using acceleration data from bogies and carbody of a dedicated high-speed TRV. The anomaly detection system involved two stages: feature selection and a data-driven anomaly detection algorithm based on the One-Class Support Vector Machine (OCSVM). The study compared OCSVM against Isolation Forest (IF), Local Outlier Factor (LOF), and Robust Mahalanobis Distance (RMD) and found that OCSVM provided the highest detection accuracy, outperforming approaches that relied on raw acceleration data alone.

Gullers et al. [43] developed a track condition analyser (TCA) based on ML techniques and instrumented wheelset force measurements. The system classified track conditions using supervised learning and labelled datasets of faulty and healthy track force recordings. It identified short-pitch rail corrugation, rolling surface defects with isolated or periodic force transients, and stiff rail supports. Characteristic features, including mean values, percentiles, and standard deviations of vertical contact forces (Q), were extracted and filtered across different frequency intervals. To the authors' knowledge, this study represents one of the first applications of IWS and force recordings for rail superstructure condition monitoring.

This chapter presents a model-based method that utilizes wheel-rail contact forces measured with IWS and a simplified 3-DOF train coach system to estimate the mean vertical rail longitudinal level within the D1 wavelength range. The model was tested using experimental data from a dedicated TRV and validated across a speed range of 100 km/h to 200 km/h, making it applicable to most conventional railway lines. One key advantage of this approach is its ability to rapidly process several minutes of recorded data, making it suitable for embedded real-time and online applications. However, further research is needed to improve estimation accuracy and extend the model's applicability to non-constant speed track sections. The use of contact force recordings for track geometry condition monitoring is a promising development in rail-way maintenance. This study highlights the potential for expanding the role of IWS beyond their traditional applications.

6.5 Instrumented wheelsets and data acquisition

All input data for the model, specifically excitation vertical forces and reference track geometry data (longitudinal level) were recorded by a dedicated TRV used for Italian's railway tracks and whole infrastructure condition monitoring. Due to NDA, these datasets are protected and not publicly available. The instrumented wheelset strain gauges layout is made available in Chapter 2 and strictly protected. Some further bridges configurations of specific application for inverse force identification in the railway sector can be found on literature (Gomez et al. [6], Bellacci et al. [19], Ronasi et al. [20]). Usually, for this kind of applications, sensors are installed on both the inner and the outer surface of the wheel-web and on the axle. This allows a finer and more precise estimation of the vertical and lateral contact forces by applying a calibration matrix to the recorded data, as reported in Bellacci et al. [19] and Diana et al. [44], adding the possibility to estimate the contact patch position [45], as already discussed. Power and data are exchanged between the rotor (wheels) to the stator (carbody or bogie frame), by an antenna, (static antenna) which powers the strain gauges' Wheatstone bridges, filters, and converts the digitalised signal received from the rotor antenna into analog. The rotor antenna instead amplifies and digitalises the acquired data before sending them to the static antenna. Studies on equipment's frequency response and passband were previously conducted, as reported in Chapter 4, although with severe limitation on the impact test hammer, which was introducing strong noise to the measurements [22].

6.5.1 Model description and simulation methods

In this section a brief overview of the main components of a railway vehicle is presented to introduce the simplified model hereafter adopted. A conventional passenger coach features a

two-level suspension system. The vertical configuration consists of four unsuspended wheelsets, two at the front and two at the rear; two bogies, each housing a pair of wheelsets, providing steering capability and an intermediate level of suspension; finally, the carbody, which is the main suspended mass. The suspension system is structured in two levels. The first level is the primary suspension, and it's located between wheelsets and bogies. It is composed of four vertical springs and four dampers. The second level is the secondary suspension, located between bogies and carbody. It consists of two vertical springs and two dampers per bogie. For high-speed vehicles the architecture may differ slightly [24].

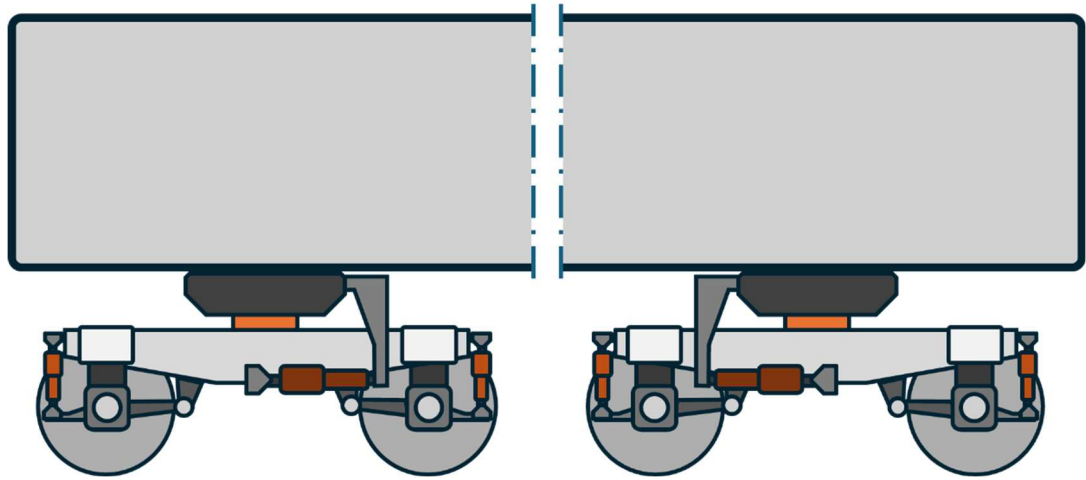


Figure 6.4: Schematic representation of a railway coach.

Excluding the simplest dot-mass (1-DOF) model [46], multi degrees of freedom (MDOF) models are often necessary to accurately capture the dynamic behaviour of a railway coach. In this research, a 3-DOF half-coach model is adopted [47]. A detailed schematic representation is available in Figure 6.5. It consists of an equivalent unsuspended mass, representing the combined mass of two wheelsets acting on a single bogie; a suspended bogie mass; and a suspended half-carbody mass. These masses are interconnected by equivalent springs and dampers, representing the primary and secondary suspension systems. The model is specifically designed for the simulation of vertical vehicle dynamics and the estimation of equivalent wheel vertical displacement using contact forces as input instead of accelerations, which are instead commonly used in similar applications. Thus, only vertical motion is considered.

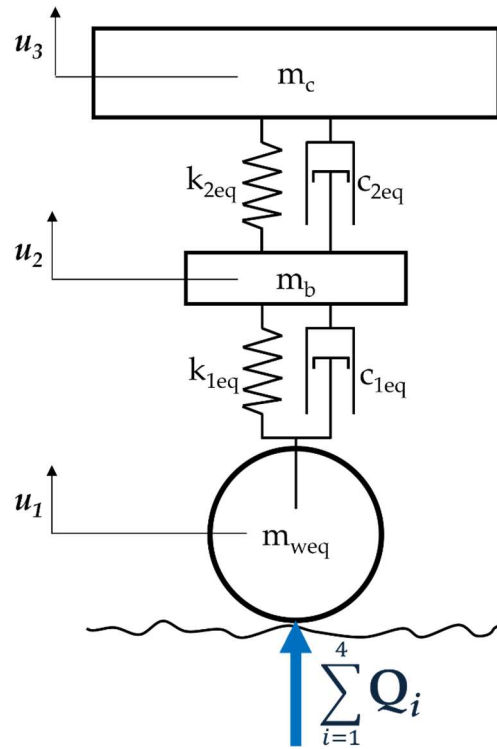


Figure 6.5: Schematic representation of the system. Simplified 3 DOF 1/2 a vehicle model, for vertical displacement estimation only.

For the contrary, out-of-plane displacements and rotational inertia effects are neglected. As a result, pitch, roll, and yaw motions are not included. Unlike the classic bicycle model commonly used for symmetric road vehicles [48], this model considers half of the railway coach, divided by a vertical and transversal plane passing through the carbody's centre of gravity. The model's structure is designed to exploit the existing strain gauges layout which couldn't be modified.

The model takes as input vertical forces recorded by the dedicated TRV used in the Italian Railway track condition monitoring. The outputs are the vertical displacements u_i (with $i = 1, 2$ and 3) of the three masses. Since the primary goal is to estimate the mean vertical displacement of the left and right rail (longitudinal level), only the equivalent wheel movement is analysed. In the model m_{weq} , m_b and m_c denote the equivalent wheel, bogie and half-carbody mass respectively; k_{1eq} and k_{2eq} represent the equivalent stiffness of primary and secondary suspensions, calculated as a parallel of 4 and 2 springs respectively; c_{1eq} and c_{2eq} represent the equivalent damping of primary and secondary suspensions, obtained as a parallel of 4 and 2 dampers respectively; Q_i are the 4 vertical wheel-rail contact forces acting on each wheel of the same bogie. The proposed 3-DOF model can reproduce the vertical dynamics of the three masses within a certain confidence range, as demonstrated in the following sections. This is achieved by focusing on track segments with a constant forward speed and in the context of track geometry monitoring within the D1 wavelength range, which is required by the standards [9].

Setting up a high-DOF model requires considerable effort, making it impractical for certain applications, especially for embedded and real time. Moreover, when the objective is to estimate vertical rail displacements filtered according to a wavelength range, the advantages of a high-DOF model become less relevant, considering the drawback of the high computational cost. In such cases, the performance of simpler models outweighs the need for absolute precision, particularly due to their significantly faster computation time. The comparison between the 3-

DOF model and actual TRV measurements highlights the impact of data filtering on accuracy. Filtering the data according to standardized wavelength ranges (such as D1) effectively reduces the bias between measured and simulated displacements. As shown in the results and discussion section, pre-filtering the data makes TRV measurements closely resemble the simulated results, whereas raw data exhibits a more noticeable discrepancy with simulated values. This demonstrates that achieving accuracy comparable to a multi-body model is unnecessary in this context. The proposed model inherently introduces geometric filtering effects [49], like those identified by Yeo et al. [50] and shown in Figure 6.6. These effects must be considered when evaluating the model's overall performance in representing measured vertical displacements of the track. Geometric filtering such as the bogie spacing and wheelbase, affects the model's output. Specifically, for wavelengths close to the wheelbase and double wheelbase length (approximately 3 and 6-7 meters respectively), discrepancies may arise between the model's signals and those measured by the TRV. This occurs because the wheelbase naturally acts as a geometric low-pass filter in the spatial domain, attenuating vertical variations for track features smaller than 3 meters. Examples of these implications are illustrated in Figure 6.6. Moreover, because the sum of the four forces is adopted, the total applied load acting on the equivalent wheel averages the actions on each wheel. Furthermore, rotations are neglected by hypothesis, potentially leading to other mismatching between simulation and measurements.

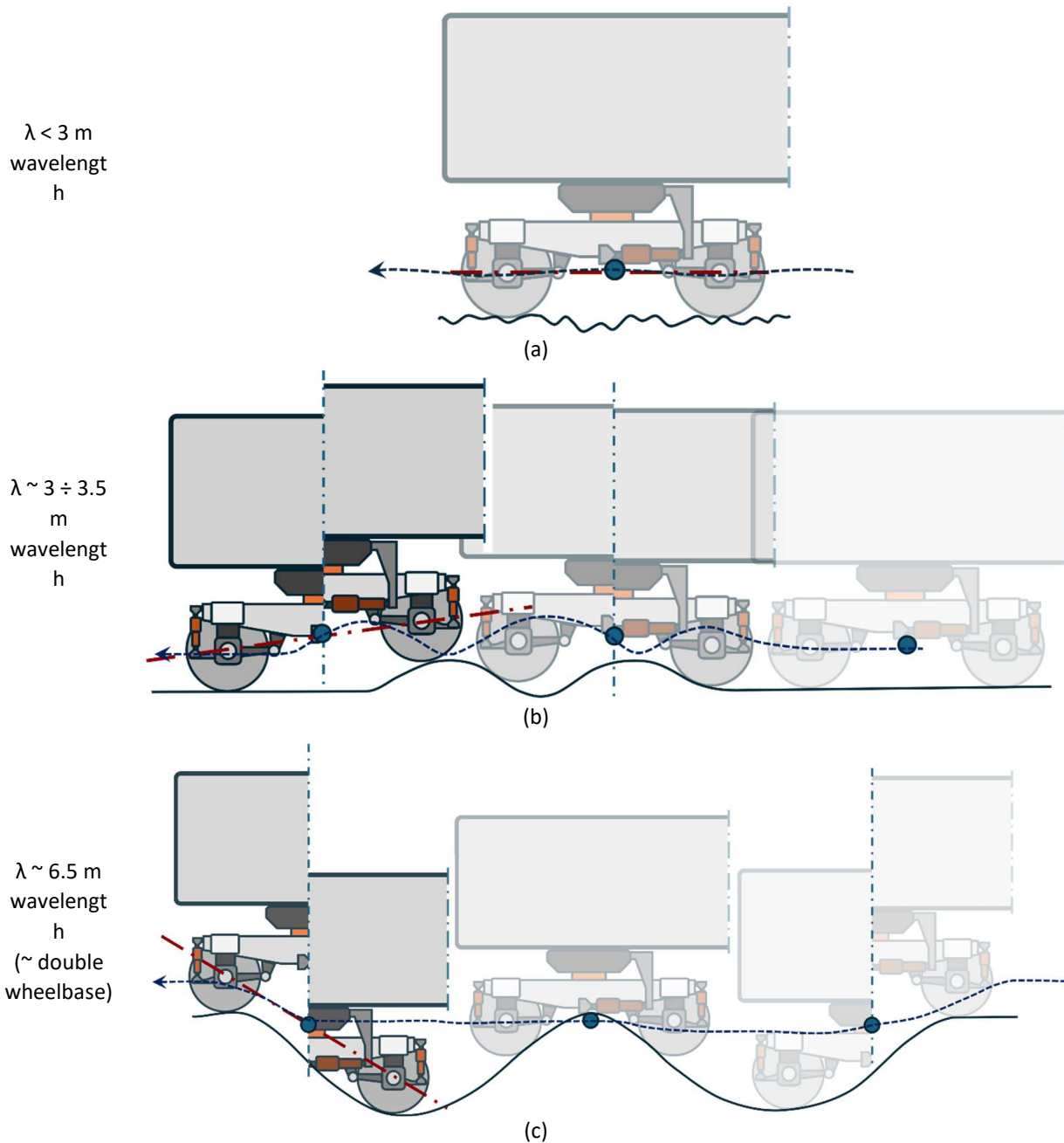


Figure 6.6: Example of geometric filtering of the vertical alignment [50], λ refers to wavelength.

With these assumptions, such a simplified 3-DOF model of half-a-coach is used to simulate the dynamic behaviour of the vehicle's suspension system: the corresponding implementation is shown in the scheme of Figure 6.7. All the simplifications are reasonable considering portion of track with approximately constant speed, which is true for most of the cruising sections of many ordinary and high-speed lines. To address low-frequency noise, a second order linear Butterworth high-pass filter with 0.1 Hz cut off frequency is applied to the recorded input forces to prevent output signal drift caused by slow-varying or non-zero mean terms during the integration process, while a second order linear Butterworth bandpass filter, with a passband from 3 m to 25 m, corresponding to the D1 range, is applied in the space domain to the estimated

equivalent wheel, and measured track vertical displacement for comparison, as prescribed by the standards [9].

Filtering also improves the robustness of proposed estimator with respect of measurement errors of the instrumented wheelset:

- Unavoidable offsets in force measurements are removed by the high-pass filter (Figure 6.7).
- Gain errors on measured forces are small. Their near influence is almost negligible.
- High-frequency errors introduced by noise and bandwidth limits of the measurement system are rejected by the bandpass filter (see Figure 6.7) and by the intrinsic nature of the 3-DOF model implemented by the estimator (a double stage integration associated with additional poles at few Hertz).

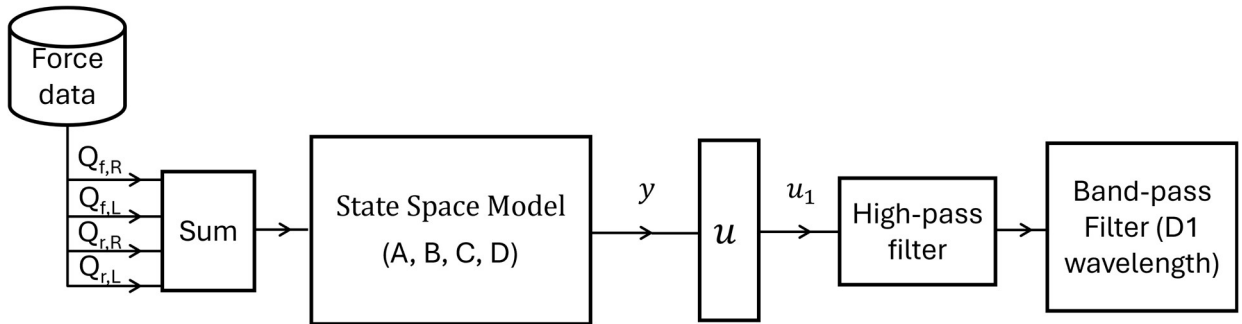


Figure 6.7: Schematic representation of the process.

In Figure 6.7 $Q_i \in Q_{ext}$ $Q_i \in Q_{ext}$ are front right and left, and rear right and left vertical contact forces; M , C and K are mass, damping and stiffness matrix respectively; $\ddot{u}_i \in \ddot{u}$, $\dot{u}_i \in \dot{u}$ and $u_i \in u$ are the 3 masses' vertical accelerations, speeds and displacements. By neglecting the rotational inertia of the bodies and considering only vertical displacements, the equations of motion for the three masses are solved according to the scheme in Figure 6.7, and the wheel displacement is obtained by double-integrating the accelerations over time. More details on the numerical integration are given in appendix A. The resulting displacements are then compared to the mean vertical rail displacement obtained by averaging the longitudinal level between the left and right rails of a line survey. Line longitudinal level is included in the datasets. As shown in Figure 6.1, the TRV is equipped with optical measurement systems that acquire real-time track-geometry data, which serve as reference for validating the proposed method.

A track section of nearly 30 km (dataset 1 in Table 6.3), consisting of both curved and straight track, and average constant train speed of 160 km/h has been used for the optimization of the model's parameters, obtaining the best fitting values for suspensions and masses, meaning stiffnesses, damping and masses for the coach model. Parameters resulting from the optimization are summarized in Table 6.2.

Optimization is performed using the Matlab-Simulink Response Optimizer™, a tool that essentially applies a robust gradient-based search to minimize the least-squares error between a reference signal (experimental data in our case) and a target signal (numerical model results). It is noteworthy that, at the beginning of the calibration process, the actual characteristics of the measurement train were deliberately withheld by the industrial partner that provided the experimental force measurements. Consequently, the performance achieved by the proposed estimator was obtained through self-calibration on the experimental dataset. This represents an

important result for the application of the proposed methodology in operational scenarios where vehicle parameters are unknown or significantly perturbed with respect to their nominal values.

Table 6.2: Optimized model parameters for the vehicle speed of 160 km/h and D1 range.

Parameter	Value	Unit
m_{weq}	2300	[kg]
m_b	2180	[kg]
m_c	19000	[kg]
k_{1eq}	2.0e+06	[N/m]
k_{2eq}	2.16e+06	[N/m]
C_{1eq}	64.9e+03	[kNs/m]
C_{2eq}	38.2e+03	[kNs/m]

Other three mixed track sections of 100, 160 and 200 km/h constant train speed have been used for validating the model and the parameters in a wider range of speeds.

Corresponding datasets are described in Table 6.3. Available calibration and validation datasets correspond to extended runs in which the recorded speed of the train is substantially constant with coasting or moderate longitudinal forces are applied. Datasets are referred to different lines with modest curvatures and slopes that are statistically common on the Italian network. Once calibrated, model parameters as reported in in Table 6.2 were subsequently utilized for all the validation process.

Table 6.3: Nomenclature and characteristics of parsed datasets.

Name	Train speed and tolerance [km/h]	Track type	Length [km] (Duration [s])
Dataset 1 (calibration)	160±5	Mixed ¹	29.4 (661)
Dataset 2 (validation)	160±5	Mixed ²	30.5 (686)
Dataset 3 (validation)	200±5	Mixed ³	33.5 (603)
Dataset 4 (validation)	100±5	Mixed ⁴	13.2 (488)

Features of Mixed Mission Profiles

¹Max long. acc. $\pm 0.1 [m/s^2]$. Min curv. radius 5500 [m]. Max slope 5‰

²Max long. acc. $\pm 0.15 [m/s^2]$. Min curv. radius 5000 [m]. Max slope 10‰

³Max long. acc. $\pm 0.15 [m/s^2]$. Min curv. radius 5500 [m]. Max slope 5‰

⁴Max long. acc. $\pm 0.1 [m/s^2]$. Min curv. radius 5500 [m]. Max slope 10‰

It is acknowledged that a substantial body of research exists on the use of acceleration measurements as input for estimating rail-track geometry through double integration of the equations of motion. However, to the best of our knowledge, no previous study has employed measured contact forces as the primary input for this purpose. Therefore, IWS, although a well-established technology for homologation and testing, represents a novel approach to track-

condition monitoring. The present work aims to explore the feasibility of using IWS as an alternative or complementary tool to existing diagnostic technologies, such as accelerometers, rate gyros and laser scanners, by assessing their capabilities for condition-monitoring tasks. A key aspect of this study is the adoption of a reduced-order vehicle model, progressively extending it to the current 3-DOF setup, in order to seek an effective trade-off between simplicity, reliability and accuracy. As highlighted in the literature, most related works rely on multibody simulations for vehicle–track interaction modelling, and only few employ reduced-DOF models, rarely with fewer than five degrees of freedom. Although a comprehensive experimental comparison between IWS-based and accelerometer-based approaches is still required, an activity outside the scope of the present study, this work provides a first assessment of the potential of IWS for this application. Certain features of IWS deserve attention: unlike accelerometer-based approaches, IWS measurements are not affected by the double filtering effect of primary and secondary suspensions, which is inherent in bogie- and carbody-mounted acceleration data. On the other hand, axle-box (and, more generally, bogie and carbody) acceleration signals are known to exhibit low signal-to-noise ratios in service conditions [51, 52], which necessitate advanced filtering and identification strategies, such as Vold–Kalman filtering (VKF), band-pass (BP) or high-pass (HPF) filtering, before meaningful interpretation. Therefore, it would be premature to assume that ABA measurements are inherently more accurate than IWS data without a direct, controlled comparison.

6.5.2 Linear state space representation of the half-a-coach vehicle model

The model has been optimized on MATLAB Simulink’s Parameter Optimization App, by minimizing the trade-off between the model state-space output and the reference data. Hereafter the equivalent state space representation used to solve the linear model of half-a-coach vehicle is presented for completeness, check the results and obtain the plots.

$$[\dot{x}] = [A][x] + [B][q], \quad 6.1$$

$$[\dot{x}] = [A][x] + [B][q], \quad 6.2$$

$$x = [x_1 \ x_2 \ x_3 \ x_4 \ x_5 \ x_6]' = [u_1 \ \dot{u}_1 \ u_2 \ \dot{u}_2 \ u_3 \ \dot{u}_3]', \quad 6.3$$

$$\dot{x} = [\dot{x}_1 \ \ddot{x}_1 \ \dot{x}_2 \ \ddot{x}_2 \ \dot{x}_3 \ \ddot{x}_3]', \quad 6.4$$

$$q = \left[0 \ \sum_{i=1}^4 Q_i \ 0 \ 0 \ 0 \ 0 \right]', \quad 6.5$$

$$Q = [Q_1 \ Q_2 \ Q_3 \ Q_4]' = [Q_{f,R} \ Q_{f,L} \ Q_{r,R} \ Q_{r,L}]', \quad 6.6$$

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{k_{1eq}}{m_1} & -\frac{c_{1eq}}{m_1} & \frac{k_{1eq}}{m_1} & \frac{c_{1e}}{m_1} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{k_{1eq}}{m_2} & \frac{c_{1eq}}{m_2} & -\frac{(k_{1e} + k_{2eq})}{m_2} & -\frac{(c_{1eq} + c_{2eq})}{m_2} & \frac{k_{2e}}{m_2} & \frac{c_{2eq}}{m_2} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{k_{2eq}}{m_3} & \frac{c_{2eq}}{m_3} & -\frac{k_{2eq}}{m_3} & -\frac{c_{2eq}}{m_3} \end{bmatrix}, \quad 6.7$$

$$B = \begin{bmatrix} 0 & -\frac{1}{m_1} & 0 & 0 & 0 & 0 \end{bmatrix}, \quad 6.8$$

$$C = [1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0], \quad 6.9$$

$$D = 0. \quad 6.10$$

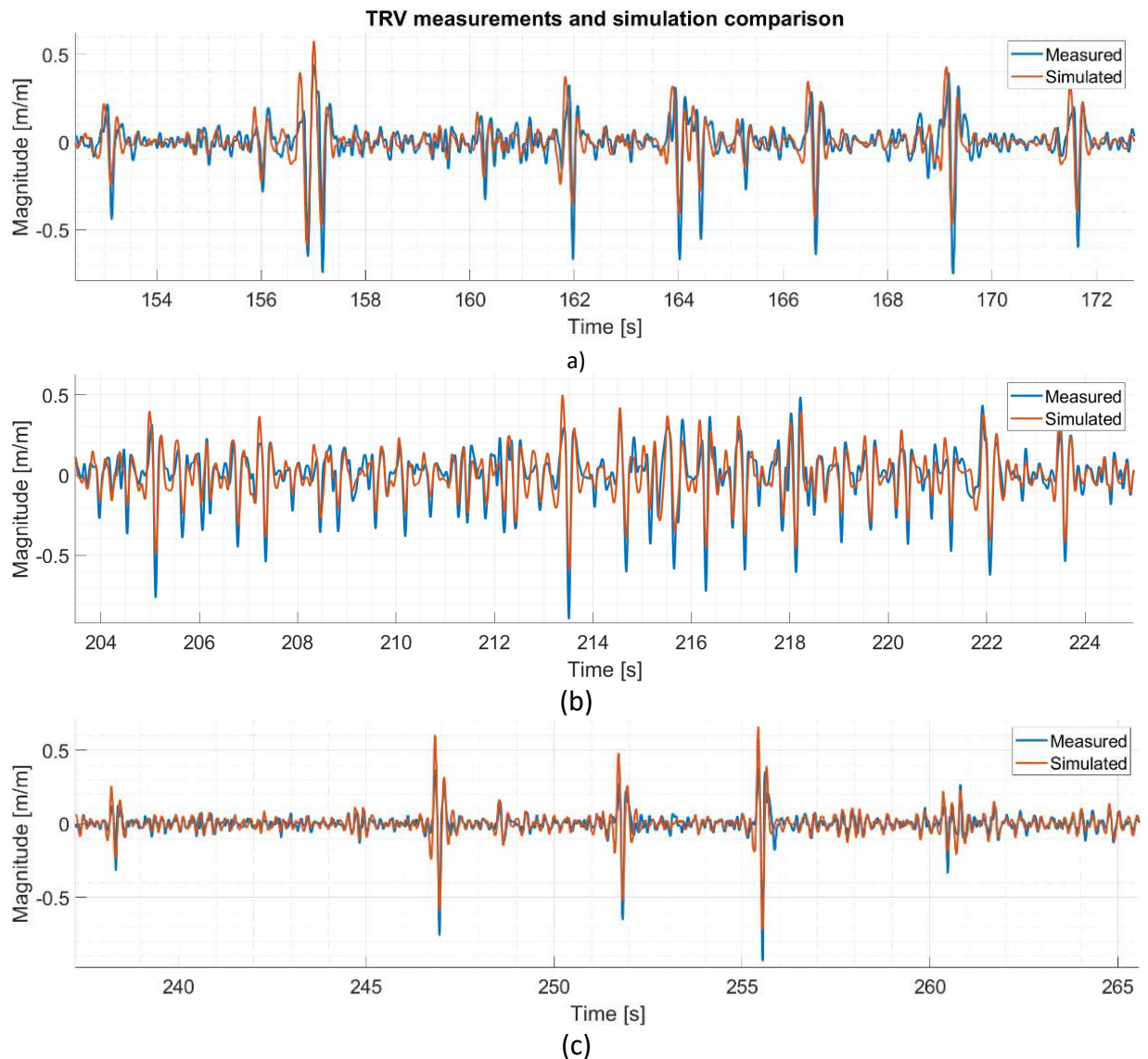
In the state space representation x is the state vector and $\dot{x}$ its first derivative; y is the output vector; q is the input vector; of which Q is the load vector containing vertical forces Q_i ; Subscript f and r denotes front and rear respectively, while subscript R and L denotes right and left side respectively; A is the system matrix, B is the input matrix, C is the output matrix and D is the feedthrough matrix. For this representation, given the linearity of the system, the second order linear Butterworth high-pass filter with cut-off frequency of 0.1 Hz is applied to the output signal instead of the input contact forces used for the simulation. Obtained vertical displacements are then filtered in the D1 wavelength range with a second order linear Butterworth band-pass filter, with cut of wavelengths of 25 and 3 meters respectively (1.78 and 14.81 Hz) for the vehicle speed of 160 km/h (44.44 m/s).

6.6 Results and Discussion

As illustrated in Figure 6.8, showcasing a visual comparison of an extract of recorded and simulated time series, the model closely mimics data measured by the TRV. By solving the linear system which mathematically represents the model, vertical wheel's displacements have been obtained and filtered, as previously discussed, using Matlab. For data confidentiality, all results and measurements have been normalized with the maximum value of the measured mean vertical alignment, which is dependent on the specific dataset. Although a clear underestimation of the output's magnitude is obtained for negative displacements, and a light overestimation of positive peaks can be identified, both waveform and wavelength of the simulated displacements are closely related to the reference data. The highest errors are observed at lower speeds (dataset 4). The primary reason for this behaviour is likely related to the filtering of static force contributions, which are unavoidable when estimating vertical displacements from contact force or acceleration measurements. This is reasonable because the proposed procedure involves the double integration of real-world recorded signals, which are affected by small offsets and low-

frequency dynamics associated with unmodelled degrees of freedom (DOFs) of the actual vehicle. In the proposed methodology, these contributions are currently filtered out using high-pass and band-pass filters operating purely in the time (space) domain. These filters will be replaced by more sophisticated disturbance estimators, which are the subject of ongoing and future research activities. Furthermore, low-frequency load transfers could be better captured by increasing the number of modelled DOFs, thus achieving an improved trade-off between model complexity and estimation accuracy. The solution proposed in the present work achieves good accuracy despite the extreme simplification of the vehicle dynamics and the limited number of parameters and data required for its implementation or calibration.

A Fast Fourier Transform (FFT) of the track irregularities measured by the TRV and simulated vertical displacements has been calculated for a comparison of data spectra. The FFT (Figure 6.9) confirms the close relation between experimental and simulated data. In the validity field of the D1 range, for a train speed of 160 km/h (44.4 m/s), the inferior limit corresponds to 1.78 Hz, while the superior limit corresponds to 14.81 Hz for dataset 1.



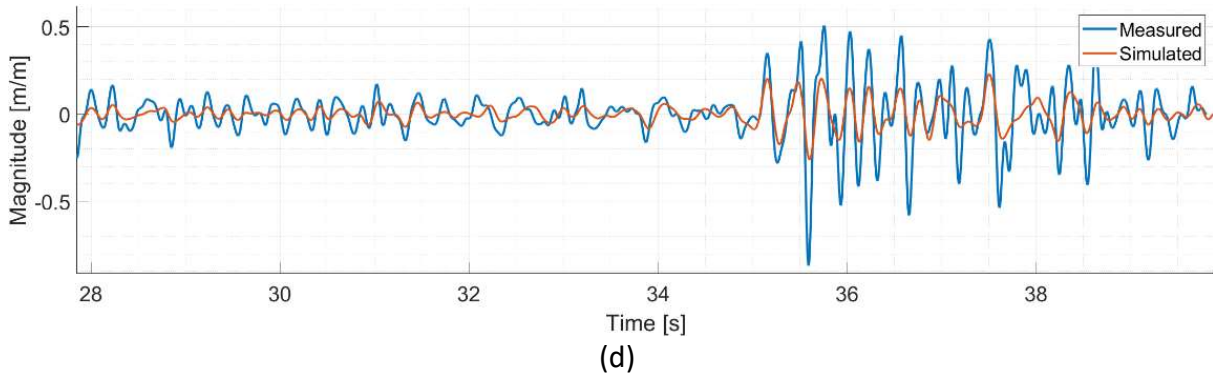


Figure 6.8: D1 wavelength range band-pass filtered data. Comparison between reference measured and simulated vertical displacements. Blue line - measured data; red line – simulation; a) dataset 1; b) dataset 2; c) dataset 3; d) dataset 4.

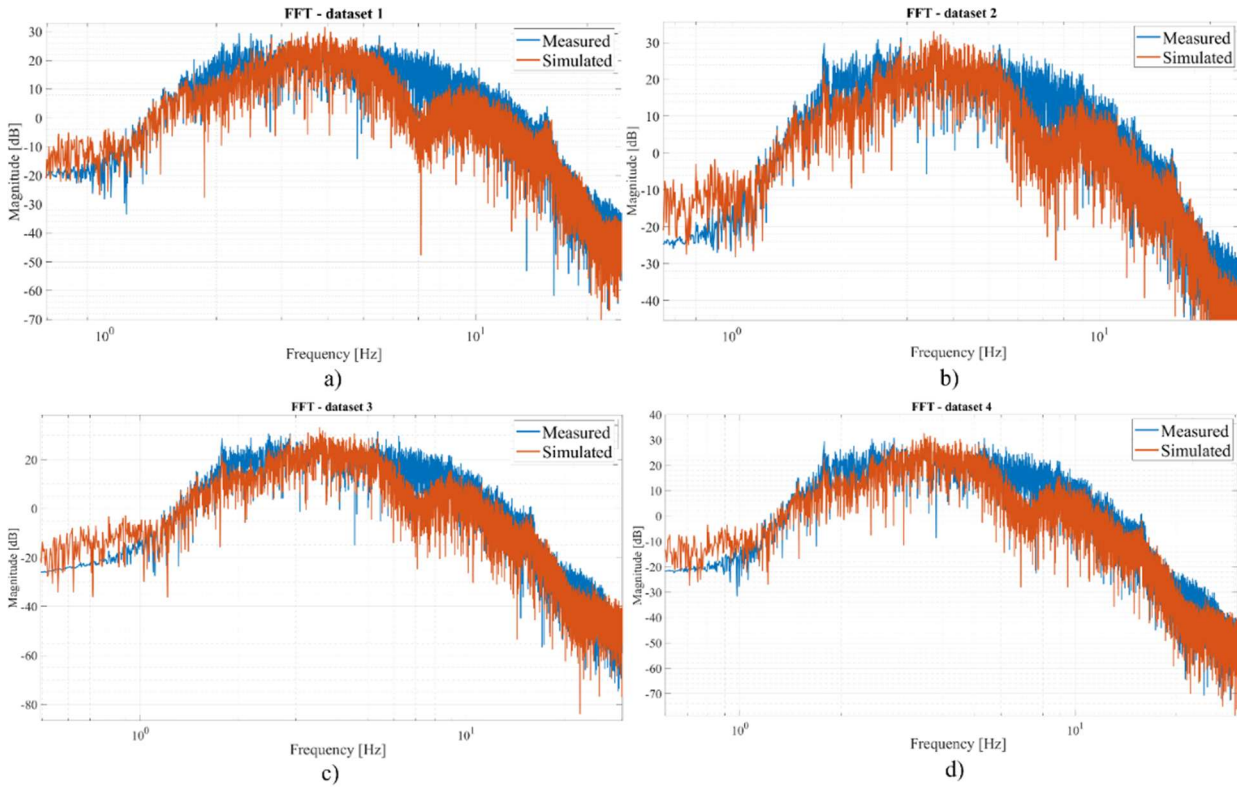


Figure 6.9: Spectra comparison of a) dataset 1; b) dataset 2; c) dataset 3; d) dataset.

6.7 Error Analysis

An error and correlation analysis has been conducted to evaluate the performance of the model. As the first metrics, absolute residuals (6.11) and root mean squared error (RMSE) (6.12) have been calculated between the measured and simulated time series:

$$res_i = |u_{meas_i}| - |u_{sim_i}| \quad 6.11$$

$$abs_res_i = \left| |u_{meas_i}| - |u_{sim_i}| \right| \quad 6.12$$

Subscript i implies element-wise operation between the two arrays. The results for each parsed track section are summarized in Table 6.4.

Residuals and RMSE metrics highlight differences in signal magnitude, but they do not capture the similarity between the simulated and measured average longitudinal level. To address this, a correlation analysis was performed to further validate the model. Pearson correlation coefficient (R) and coefficient of determination (R^2) were calculated according to equation 6.13 [53]:

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{j=1}^n (Y_j - \bar{Y})^2 \right]^{1/2}} \quad 6.13$$

with X and Y measured and simulated arrays of n elements, $\bar{X}$ and $\bar{Y}$ their mean values respectively. Model's performance has been evaluated at two other available train speeds in an attempt of extending the validity range of the proposed model, parsing available datasets 2, 3 and 4, which complies with most of the conventional European railways (fast regional, intercity, cross country and some high-speed connections), has depicted on the map available on openrailwaymap.org [54] (see Figure 6.10).

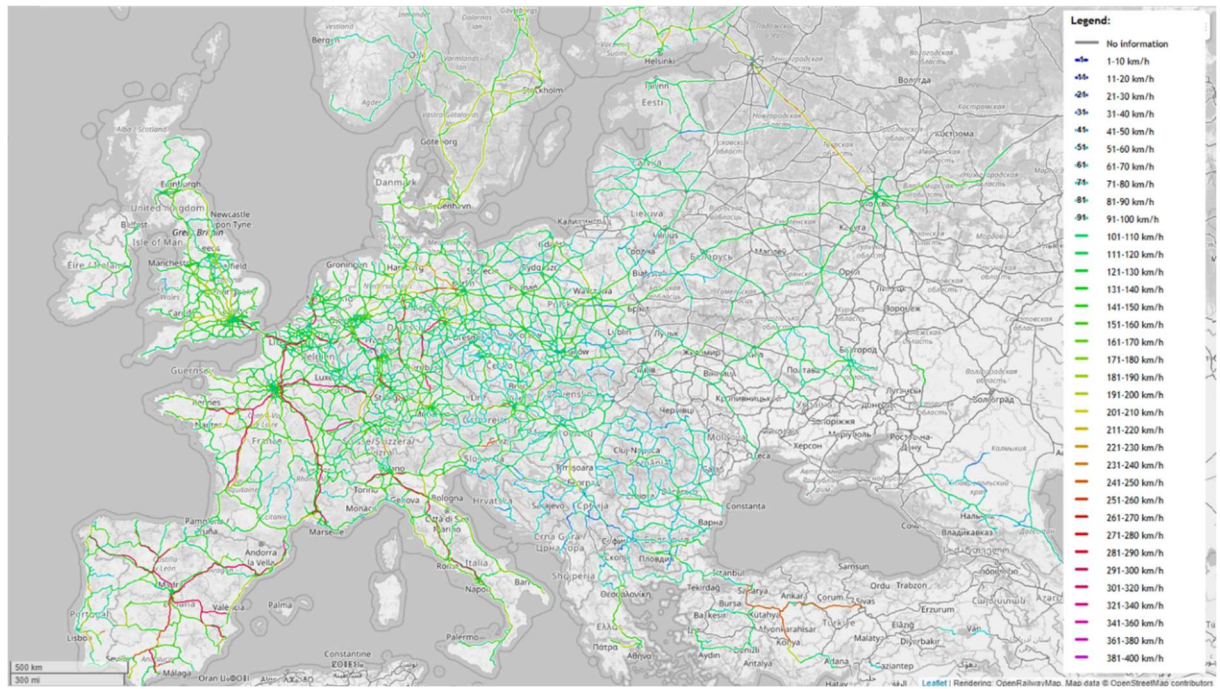


Figure 6.10: Map of the railway infrastructure in Europe, grouped by maximum line speed. Green shades represent 90-200 km/h lines [54].

Table 6.4: Error and correlation analysis for the recorded and simulated mean vertical alignment.

Property	Dataset 1: 160 km/h	Dataset 2: 160 km/h	Dataset 3: 200 km/h	Dataset 4: 100 km/h
Residuals [mm]	0.323	0.323	0.282	0.588
RMSE [mm]	0.578	0.705	0.497	0.994
R	0.827	0.843	0.798	0.579
R^2	0.684	0.711	0.636	0.336

*Mean values are showcased for residual and RMSE. R and R^2 are scalar values obtained for the entire recordings.

The high R -value for datasets 1, 2 and 3, confirms the strong correlation between the time series, consistent with the visual comparison and Fourier analysis. Although relatively high, the R^2 values suggest areas where the model's output still deviates from the measured data, particularly in amplitude and frequency, rather than wavelength. These discrepancies, combined with the time delay introduced by the filters, likely contribute to the lower R and R^2 scores for datasets 4, despite the overall similarity between the simulated and measured signals. The high RMSE values could be related to the underestimation of the vertical displacement by the model and localized mismatching in the waveform of the signals. Results of datasets 1, 2 and 3 are promising, exhibiting good repeatability. As the travelling speed of the train decreases (dataset 4), the fitting capability of the model also decreases, but it's expected to slowly fall after the 200 km/h limit of dataset 3 too. No further data are available over 200 km/h to the authors. Based on these outcomes, the expected range of applicability of the model with respect to train speed is illustrated in Figure 6.11, where the values before 100 km/h and above 200 km/h are extrapolated.

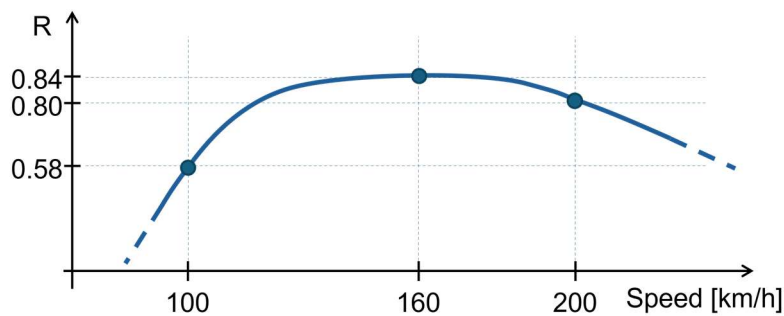


Figure 6.11: Expected behaviour of the R value for the investigated train speed range, showcasing the applicability range of the model.

6.8 Conclusions and future developments

In this Chapter a simplified 3-DOF linear model of half-a-coach train for estimating the vertical displacement of the wheels, and consequently the longitudinal level of the track in the D1 range (3-25 m wavelength), using force recordings is proposed. Despite to its simplicity the method has proven to provide significant results for the estimation of D1 vertical irregularities in a range of

speed comprise between 100 and 200 km/h when tested with real experimental data. The performed analysis is focused solely on vertical motion. High-pass filters have been applied to remove low-frequency inaccuracies in the force recordings, which arise from uncertainties in the measurements. This approach effectively filtered out non-zero mean terms and retained only the high-frequency content of the excitation signals. The equations of motion for the 3-DOF system were solved by double integrating the accelerations, and the resulting mean displacements were compared with a reference mean vertical alignment calculated from the measurements of TRV used for track geometry monitoring and diagnosis. A visual comparison of the time series and spectra shown high degree of correlation between the two datasets. The error analysis indicated the presence of inaccuracies by the means of residual values, attributable to amplitude and phase mismatches between the measured and simulated displacements. The error analysis revealed inaccuracies in terms of the maximum residual value, mainly attributable to amplitude and phase mismatches between the measured and simulated displacements. These discrepancies arise from the reduced number of modelled DOFs which, while being the main source of error, also represents the strength of the proposed approach, as it requires negligible computational resources and only limited technical information regarding track, wheel-rail contact, and suspensive system. For this reason, the proposed functionality can be considered a low-cost add-on that can be easily integrated into measurement setups where wheel-rail contact force measurements are performed, to further refine the outputs of optical devices.

The neglected inertias and rotational effects, particularly pitch motion, are believed to be the primary cause of the reduced performance of the proposed estimator, especially in scenarios where dynamic load transfers between wheelsets are expected to increase (e.g., narrow curves, high longitudinal accelerations, and decelerations). Additionally, the applied filters to the force recordings and the geometric ones as shown in Figure 6.6, introduce a further degree of approximation. These filters are currently implemented in the time (or space) domain, with their poles continuously adjusted according to train speed in order to capture the desired wavelengths of the railway track irregularities. While this approach is suitable for smooth and slowly varying vehicle speeds, it leads to an increase in estimation errors when significant longitudinal accelerations or decelerations occur.

The discrepancies in waveform and amplitude between the simulated and recorded data, along with the time delay introduced by the filters, may also account for lowering the R and R_2 coefficients obtained for dataset 3 and dataset 4, used for the validation of the model. However, results are promising, showing good agreement with experimental data, making the proposed method suited for the analysis of data recorded between 100 and 200 km/h. The high R coefficient for datasets 1 and 2, and the still relatively high value obtained for dataset 3, highlights the good performance and the potential of the simplified model in a wide range of train speed, accounting for many of the ordinary railway lines and some high-speed interconnections. Minor adjustments and improvements to the model could lead to the development of a rapid and straightforward tool for the estimation of the mean vertical alignment based solely on force recordings. Given the non-stationary speed of the train, future work will focus on correcting the delay and distortion introduced in the simulation results by the filters, aiming to improve the correlation between simulated and reference signals. By extending the model to 6-DOF (5+1 constrained), accounting for 2 bogies with 2 equivalent wheels and an entire coach, considering its rotational inertia and pitch motion, a more precise model can be implemented with the introduction of a Kalman filter that should be useful for multiple purposes:

- Calibrate model parameters when uncertain or partially known

- Self-calibration and optimal filtering with respect of speed, line conditions on other external disturbances
- Data fusion with additional measurements from inertial or optical sensors further improving measurement precision and reliability.

Additionally, the study could be extended to investigate the lateral motion of the wheelsets for rail alignment estimation, following a similar methodology as discussed herein. These efforts will contribute to the broader adoption of IWS for condition monitoring and diagnostics. When combined with optical devices, this approach may yield significant benefits for the railway maintenance sector and track geometry condition monitoring, utilizing both TRVs and in-service vehicles. The potential to use force recordings for reconstructing the mean vertical alignment of the track is evident, and the adopted method is consistent with reference data, as supported by error and correlation analysis. However, it is important to note that existing standards impose limits on the maximum permissible values for geometric parameters. Therefore, the mean value derived using the proposed method may lead to an underestimation of the actual track condition. This aspect will be carefully considered in the development of a vertical motion estimation algorithm based on the proposed simplified model.

6.9 Future developments: 6-DOF coach model concept

As previously mentioned, the model can be improved in several aspects, starting with the integration of the neglected degrees of freedom (DOFs) and rotational inertias.

In fact, to account for load transfer during acceleration and braking phases, or in the presence of a non-negligible track gradient, carbody rotation must be included. Consequently, the model should be extended to represent an entire coach (rather than only half, as in the previous formulation), in which not only the vertical motion of the carbody, the two bogies, and the four wheels is considered, but also the rotational motion of the carbody. Since the rear wheel is expected to follow the vertical motion of the front wheel with a temporal (or spatial) delay, the degree of freedom associated with the vertical displacement of the rear wheels becomes constrained by that of the front wheels. In this way, a 5-DOF model (effectively 6 DOFs, one of which is constrained, i.e. $5 + 1$) is obtained, allowing load transfers to be captured more accurately. This enhancement would likely improve the estimation capability of the model, making it more reliable in critical scenarios where variations in speed are non-negligible, conditions under which the 3-DOF vertical model exhibited estimation shifts in both wavelength and phase. A schematic representation of the 5-DOF model concept is shown in Figure 6.12.

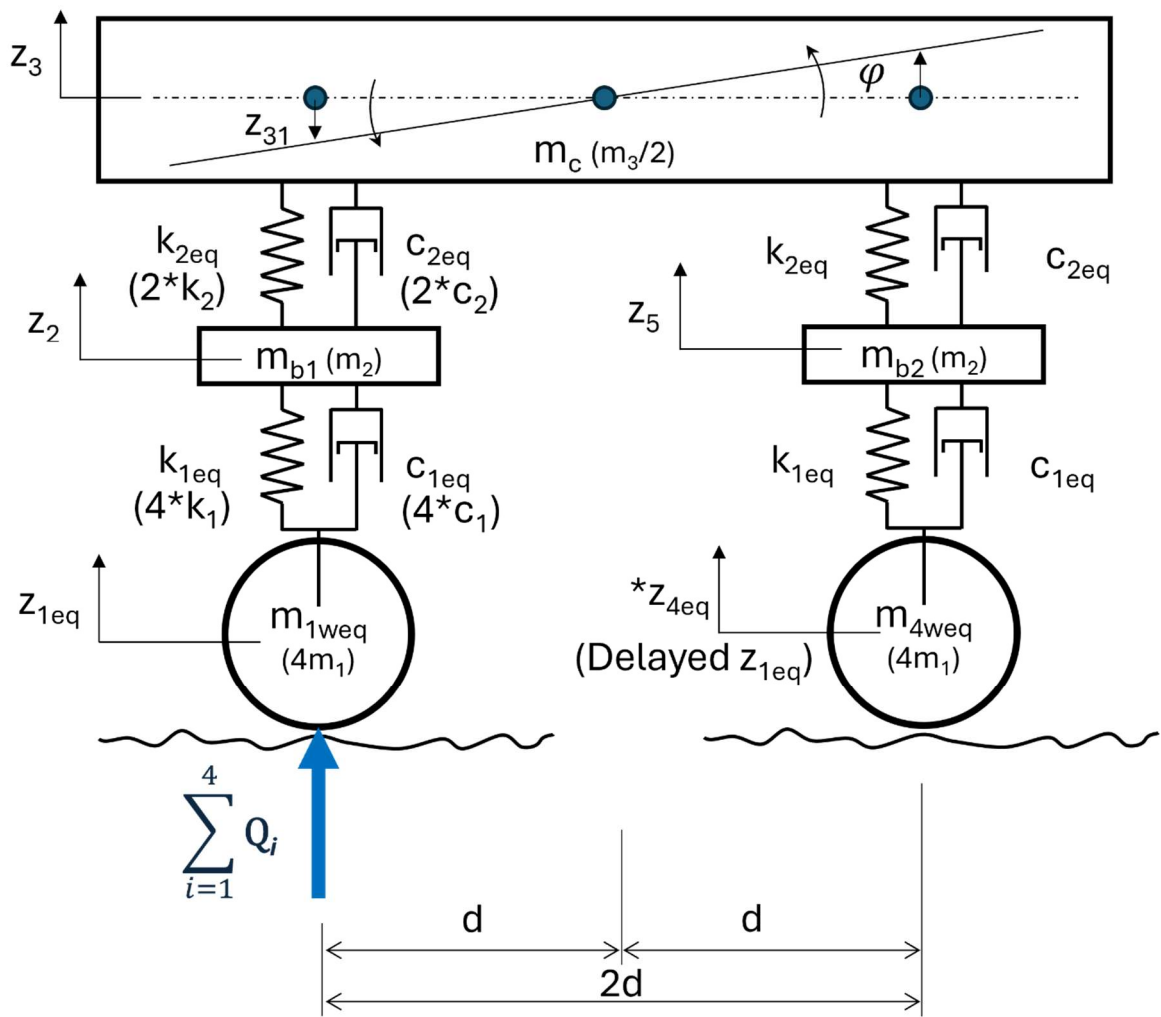


Figure 6.12: Schematic representation of the model concept. Simplified 5-DOF vehicle model, for vertical displacement estimation only.

Further improvements may regard the introduction of non-linearities in the suspensions, also considering the introduction of rubber springs or air springs especially in the secondary suspension, which provide damping as well as increased stiffness with increased excitation frequencies and decreased amplitude. Alternatively, a spring element in series with the damper may be added, introducing the Oda-Nishimura model [55], or the Vampire vertical model [56]. Schematic examples of such solutions are shown in Figure 6.13.

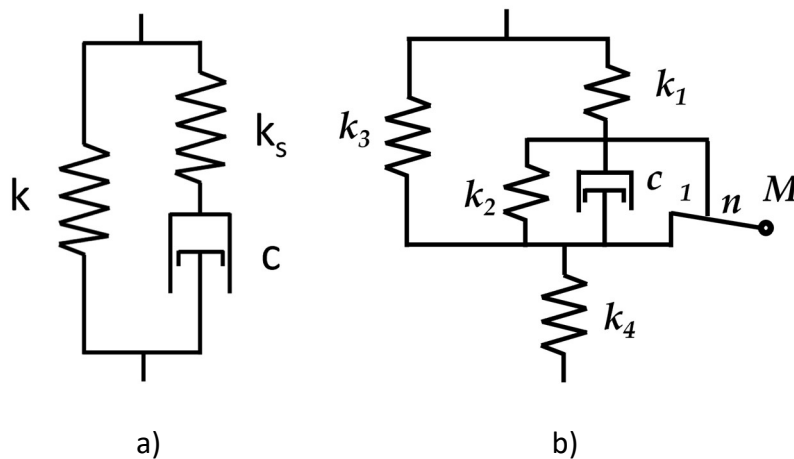


Figure 6.13: Schematic representation of one-dimensional model often used to represent rubber springs and Vampire models [56].

The Nishimura model enables the reproduction of the transition between the low-frequency and high-frequency stiffness characteristics of pneumatic suspensions. The vertical model implemented in Vampire, shown in Figure 6.13(b), is based on the Nishimura formulation, with the addition of a lumped mass M and an auxiliary stiffness element K_4 arranged in series.

The spring K_4 represents the emergency spring and can be assumed to have an infinitely large stiffness. The lumped mass M accounts for the inertia of the air contained in the surge pipe.

It should also be noted that in the Vampire model the damping force can be defined according to a quadratic law, which provides a more realistic representation than the conventional linear damping assumption [56].

Abbreviations

The following abbreviations are used in this chapter:

BP	Band-Pass
DOF	Degree Of Freedom
FE	Finite Element
FFT	Fast Fourier Transform
HPF	High-Pass Filter
IWS	Instrumented Wheelsets
MDOF	Multi Degree Of Freedom
TG	Track Geometry
TRV	Track Recording Vehicle
VKF	Vold–Kalman Filtering

6.10 References

1. European Commission. EU Transport in Figures: Statistical Pocketbook 2024; Mobility and Transport, Publications Office of the European Union: Luxembourg, 2024; ISBN 978-92-68-19812-4. <https://doi.org/10.2832/16593>.
2. Ward, C.P.; Weston, P.F.; Stewart, E.J.C.; Li, H.; Goodman, C.J.; Roberts, C. Condition Monitoring Opportunities Using Vehicle-Based Sensors. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* 2011, 225, 202–218. <https://doi.org/10.1177/09544097JRRT406>.
3. Weston, P.; Roberts, C.; Yeo, G.; Stewart, E. Perspectives on railway track geometry condition monitoring from in-service railway vehicles. *Vehicle System Dynamics*, 2015, 53(7), 1063–1091. <https://doi.org/10.1080/00423114.2015.1034730>
4. CIFI. La diagnostica mobile e le misure per la certificazione. Webinar, 31 January 2022. Available online: https://www.cifi.it/UplDocumenti/Roma31012022/Presentazione%20Bonafe_Palmiott%20CIFI%202022-01-31.pdf (accessed on 7 January 2025).
5. MERMEC Group. I rotabili di RFI—Il treno Archimede. Available online: <https://www.mermecgroup.com/press-room/media-coverage/420/i-rotabili-di-rfi-il-treno-archimede.php> (accessed on 7 January 2025).
6. Gomez, E.; Giménez, J.G.; Alonso, A. Method for the Reduction of Measurement Errors Associated to the Wheel Rotation in Railway Dynamometric Wheelsets. *Mech. Syst. Signal Process.* 2011, 25, 3062–3077. <https://doi.org/10.1016/j.ymssp.2011.05.006>.
7. Nielsen, J.C.O. High-Frequency Vertical Wheel–Rail Contact Forces—Validation of a Prediction Model by Field Testing. *Wear* 2008, 265, 1465–1471. <https://doi.org/10.1016/j.wear.2008.02.038>.
8. Ronasi, H.; Johansson, H.; Larsson, F. Identification of Wheel–Rail Contact Forces Based on Strain Measurements, an Inverse Scheme and a Finite-Element Model of the Wheel. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* 2014, 228, 343–354. <https://doi.org/10.1177/0954409712473961>.
9. European Committee for Standardization (CEN). EN 13848-1:2019 Railway Applications—Track—Track Geometry Quality—Part 1: Characterization of Track Geometry; CEN: Brussels, Belgium, 2019.
10. European Committee for Standardization (CEN). EN 14363:2022 Railway Applications—Testing and Simulation for the Acceptance of Running Characteristics of Railway Vehicles—Running Behaviour and Stationary Tests; CEN: Brussels, Belgium, 2022.
11. Thong, Y.K.; Woolfson, M.S.; Crowe, J.A.; Hayes-Gill, B.R. Numerical Double Integration of Acceleration Measurements in Noise. *Measurement* 2004, 36, 73–92. <https://doi.org/10.1016/j.measurement.2004.03.010>.
12. Lewis, R.B.; Richards, A.N. A Compensated Accelerometer for the Measurement of Railway Track Crosslevel. *IEEE Trans. Veh. Technol.* 1988, 37, 174–178. <https://doi.org/10.1109/25.2896>.
13. Weston, P.F.; Ling, C.S.; Goodman, C.J.; Roberts, C.; Li, H.; Goodall, R.M. Monitoring Vertical Track Irregularity from In-Service Railway Vehicles. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* 2007, 221, 75–88. <https://doi.org/10.1243/0954409JRRT98>.
14. Weston, P.F.; Ling, C.S.; Goodman, C.J.; Roberts, C.; Li, H.; Goodall, R.M. Monitoring Lateral Track Irregularity from In-Service Railway Vehicles. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* 2007, 221, 89–100. <https://doi.org/10.1243/0954409JRRT99>.

15. Zschiesche, K.; Reiterer, A. Optical Measurement System for Monitoring Railway Infrastructure—A Review. *Appl. Sci.* 2024, 14, 8801. <https://doi.org/10.3390/app14198801>.
16. Soilán, M.; Sánchez-Rodríguez, A.; del Río-Barral, P.; Pérez-Collazo, C.; Lorenzo, H.; Arias, P. Review of Laser Scanning Technologies and Their Applications for Road and Railway Infrastructure Monitoring. *Infrastructures* 2019, 4, 58. <https://doi.org/10.3390/infrastructures4040058>.
17. CIFI. CIFI Documents and Webinars. Available online: <https://www.cifi.it/UplDocumenti/WebinarCIFI23112020-Dalla%20manutenzione%20ciclica%20alla%20predittiva.pdf> (accessed on 7 January 2025).
18. Rosano, G.; Massini, D.; Bocciolini, L.; Bracciali, A. Diagnostics of the Railway Track—Possibility of Development through the Measurement of Accelerations and Contact Forces. *Ing. Ferro.* 2024, 79, 81–102. <https://doi.org/10.57597/IF.02.2024.ART.1>.
19. Bellacci, G.; Pugi, L.; Baldanzini, N. Validation of a Wheelset Finite Element Model for Static Structural Analysis and Inverse Force Identification. In *Latest Advancements in Mechanical Engineering. ISIEA 2024. Lecture Notes in Networks and Systems*; Concli, F., Maccioni, L., Vidoni, R., Matt, D.T., Eds.; Springer: Cham, Switzerland, 2024; Volume 1124, pp. 236–244. https://doi.org/10.1007/978-3-031-70462-8_23.
20. Ronasi, H.; Nielsen, J.C.O. Inverse Identification of Wheel–Rail Contact Forces Based on Observation of Wheel Disc Strains: An Evaluation of Three Numerical Algorithms. *Veh. Syst. Dyn.* 2012, 51, 74–90. <https://doi.org/10.1080/00423114.2012.713498>.
21. Bracciali, A.; Cavaliere, F.; Macherelli, M. Review of Instrumented Wheelset Technology and Applications. In *Proceedings of the Second International Conference on Railway Technology: Research, Development and Maintenance*; Pombo, J., Ed.; Civil-Comp Press: Stirlingshire, Scotland, 2014.
22. Bellacci, G.; Neri, F.; Pugi, L.; et al. Preliminary Frequency Response Analysis of a Contact Force Measurement System for Rail Applications. In *Applications in Electronics Pervading Industry, Environment and Society. ApplePies 2023*; Bellotti, F., et al., Eds.; *Lecture Notes in Electrical Engineering*, Vol. 1110; Springer: Cham, Switzerland, 2024; pp. 240–247. https://doi.org/10.1007/978-3-031-48121-5_20.
23. Italcertifer. Estensimetria, sale e pantografi strumentati. Available online: <https://www.italcertifer.com/it/laboratori/estensimetria--sale-e-pantografi-strumentati-.html> (accessed on 7 January 2025).
24. Iwnicki, S.; Spiryagin, M.; Cole, C.; McSweeney, T. *Handbook of Railway Vehicle Dynamics*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2019. <https://doi.org/10.1201/9780429469398>.
25. Abeidi, A.S.; Bosso, N.; Gugliotta, A.; Somà, A. Numerical Simulation of Wear in Railway Wheel Profiles. In *Proceedings of the ASME 8th Biennial Conference on Engineering Systems Design and Analysis (ESDA 2006)*; Volume 3: Dynamic Systems and Controls, Symposium on Design and Analysis of Advanced Structures, and Tribology; ASME: Torino, Italy, 4–7 July 2006; pp. 963–977. <https://doi.org/10.1115/ESDA2006-95571>.
26. Bosso, N.; Magelli, M.; Zampieri, N. Simulation of Wheel and Rail Profile Wear: A Review of Numerical Models. *Railw. Eng. Sci.* 2022, 30, 403–436. <https://doi.org/10.1007/s40534-022-00279-w>.
27. Logozzo, S.; Valigi, M. C. Green Tribology: Wear Evaluation Methods for Sustainability Purposes. *International Journal of Mechanics and Control* 2022, 23(1), 23–34.

28. Logozzo, S.; Valigi, M. C. Wear Assessment and Reduction for Sustainability: Some Applications. In *Mechanisms and Machine Science*, Vol. 108; Springer: Cham, 2022; pp. 395–402. doi:10.1007/978-3-030-87383-7_43.
29. European Committee for Standardization (CEN). EN 13848-2:2020 Railway Applications—Track—Track Geometry Quality—Part 2: Measuring Systems—Track Recording Vehicles; CEN: Brussels, Belgium, 2020.
30. Hernández, W. Improving the Responses of Several Accelerometers Used in a Car Under Performance Tests by Using Kalman Filtering. *Sensors* 2001, 1, 38–52. <https://doi.org/10.3390/s10100038>.
31. Locorotondo, E.; Lutzemberger, G.; Pugi, L. State-of-Charge Estimation Based on Model-Adaptive Kalman Filters. *Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng.* 2021, 235, 1272–1286. <https://doi.org/10.1177/0959651821110xxxx>.
32. Sun, Y.Q.; Dhanasekar, M. A Dynamic Model for the Vertical Interaction of the Rail Track and Wagon System. *Int. J. Solids Struct.* 2002, 39, 1337–1359.
33. Zhai, W.; Cai, Z. Dynamic Interaction between a Lumped Mass Vehicle and a Discretely Supported Continuous Rail Track. *Comput. Struct.* 1997, 63, 987–997. [https://doi.org/10.1016/S0045-7949\(96\)00401-4](https://doi.org/10.1016/S0045-7949(96)00401-4).
34. Kawasaki, J.; Youcef-Toumi, K. Estimation of Rail Irregularities. In *Proceedings of the 2002 American Control Conference*, Anchorage, AK, USA, 8–10 May 2002.
35. Mori, H.; Tsunashima, H.; Kojima, T.; et al. Condition Monitoring of Railway Track Using In-Service Vehicle. *J. Mech. Syst. Transp. Logist.* 2012, 3, 154–165.
36. Odashima, M.; Azami, S.; Naganuma, Y.; et al. Track Geometry Estimation of a Conventional Railway from Car-Body Acceleration Measurement. *Mech. Eng. J.* 2017, 4, 16-00498.
37. Alfi, S.; Bruni, S. Estimation of Long Wavelength Track Irregularity from Onboard Measurement. In *Proceedings of the 4th IET International Conference on Railway Condition Monitoring (RCM 2008)*, London, UK, 18–20 June 2008; pp. 1–6. <https://doi.org/10.1049/ic:20080323>.
38. Miwa, M.; Kawasaki, Y.; Yoshimura, A. Influence of Vehicle Unsprung-Mass on Dynamic Wheel Load. In *Computers in Railways XI—Computer System Design and Operation in the Railway and Other Transit Systems*; Allan, J., Arias, E., Brebbia, C.A., Goodman, C., Rumsey, A.F., Sciutto, G., Tomii, N., Eds.; WIT Press: Southampton, UK, 2008; Volume 103, pp. 715–724.
39. Tsunashima, H. Condition Monitoring of Railway Tracks from Car-Body Vibration Using a Machine Learning Technique. *Appl. Sci.* 2019, 9, 2734. <https://doi.org/10.3390/app9132734>.
40. La Paglia, I.; Di Gialleonardo, E.; Facchinetti, A.; Braghin, F. Acceleration-Based Condition Monitoring of Track Longitudinal Level Using Multiple Regression Models. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* 2024, 238, 479–488. <https://doi.org/10.1177/09544097231201513>.
41. De Rosa, A.; Kulkarni, R.; Qazizadeh, A.; et al. Monitoring of Lateral and Cross Level Track Geometry Irregularities through Onboard Vehicle Dynamics Measurements Using Machine Learning Classification Algorithms. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* 2021, 235, 107–120. <https://doi.org/10.1177/0954409720906649>.
42. Ghiasi, R.; Khan, A.M.; Sorrentino, D.; et al. An Unsupervised Anomaly Detection Framework for Onboard Monitoring of Railway Track Geometrical Defects Using One-

- Class Support Vector Machine. *Eng. Appl. Artif. Intell.* 2024, 133, 108167. <https://doi.org/10.1016/j.engappai.2024.108167>.
43. Gullers, P.; Dreik, P.; Nielsen, J.C.O.; et al. Track Condition Analyser: Identification of Rail Rolling Surface Defects Likely to Generate Fatigue Damage in Wheels Using Instrumented Wheelset Measurements. *Proc. Inst. Mech. Eng. Part F J. Rail Rapid Transit* 2011, 225, 1–13. <https://doi.org/10.1243/09544097JRRT398>.
44. Diana, G.; Resta, F.; Braghin, F.; Bocciolone, M.; Di Gialleonardo, E.; Crosio, P. Methodology for the Calibration of Dynamometric Wheelsets for the Measurement of the Wheel–Rail Contact Forces. *Ing. Ferrovi.* 2012, 67, 9–21.
45. Velletrani, F.; Licciardello, R.; Bruner, M. Intelligent Wheelsets for the Trains of the Future: The Role of In-Service Wheel–Rail Force Measurement. *Ing. Ferrovi.* 2020, 75, 701–725.
46. Bellacci, G.; Entezami, M.; Weston, P.; Baldanzini, N.; Pugi, L. Evaluating the Feasibility of Estimating Rail Longitudinal Level through Double Integration of Vertical Contact Forces. In *Proceedings of I4SDG Workshop 2025—IFTToMM for Sustainable Development Goals (I4SDG 2025)*; Carbone, G., Quaglia, G., Eds.; Mechanisms and Machine Science, Vol. 180; Springer: Cham, Switzerland, 2025; pp. 737–744. https://doi.org/10.1007/978-3-031-91179-8_63.
47. Evans, J.R. The Modelling of Railway Passenger Vehicles. *Veh. Syst. Dyn.* 2007, 20, 144–156.
48. Wong, J.Y. *Theory of Ground Vehicles*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 2001.
49. Dragoş, S.; Madalina, D.; Mihai, L. The Geometric Filtering Effect on Ride Comfort of the Railway Vehicles. *UPB Sci. Bull. Ser. D Mech. Eng.* 2021, 83, 137–154.
50. Yeo, G.J.; Weston, P.F.; Roberts, C. The Utility of Continual Monitoring of Track Geometry from an In-Service Vehicle. In *Proceedings of the 6th IET Conference on Railway Condition Monitoring (RCM 2014)*, Birmingham, UK, 17–20 November 2014; pp. 1–6.
51. Hoelzl, C.; Dertimanis, V.; Ancu, L.; Kollros, A.; Chatzi, E. Vold–Kalman Filter Order Tracking of Axle Box Accelerations for Track Stiffness Assessment. *Mechanical Systems and Signal Processing* 2023, 204, 110817. doi:10.1016/j.ymssp.2023.110817.
52. Pieringer, A.; Kropp, W. Model-Based Estimation of Rail Roughness from Axle Box Acceleration. *Applied Acoustics* 2022, 193, 108760. doi:10.1016/j.apacoust.2022.108760.
53. Howell, D.C. *Statistical Methods for Psychology*, 7th ed.; Wadsworth, Cengage Learning: Belmont, CA, USA, 2009.
54. OpenRailwayMap. Available online: <https://www.openrailwaymap.org/> (accessed on 7 January 2025).
55. Oda, N., & Nishimura, S. (1970). Vibration of air suspension bogies and their design. *Bulletin of the Japan Society of Mechanical Engineers*, 13(55), 43–50. <https://doi.org/10.1299/jsme1958.13.43>.
56. Bruni, S., Vinolas, J., Berg, M., Polach, O., & Stichel, S. (2011). Modelling of suspension components in a rail vehicle dynamics context. *Vehicle System Dynamics*, 49(7), 1021–1072. <https://doi.org/10.1080/00423114.2011.586430>.

7 DATA-DRIVEN METHODS: NOVELTY DETECTION USING CONTACT FORCE SPACE FEATURES AND MACHINE LEARNING

7.1 Introduction: big data analysis

Traditional data analysis techniques may not suffice in analysing the huge amount of data recorded during a single or various trips of a train along a journey through its destination. Usually, for faults detection in rotary machines, bearings monitoring, etc, classic statistical analysis with time features variation like mean, standard deviation, RMS, skewness and kurtosis, is used to find incipient variation of the component's status, degradation, wear and predict component's failure. More complex time-frequency (or space-wavelength) techniques such as FFT (STFT) and wavelet, are commonly employed to study changes in the frequency content of the system to address normal and abnormal behaviour by observing variations in the spectrogram (or scalogram, depending on the technique adopted), S/N ratio, I/O ratio as in the case of frequency response analysis.

When applied to our necessities, which are parsing mission recordings to find track faults, such techniques start to show severe limitations, in terms of information we can extract. Usually, because rotary components are involved in the reference studies (motors, bearings, etc.) and in our research (wheelsets), a certain degree of similarity is expected in the results of the analysis.

The main difference lies in the nature of the signal, which is not periodic and easily subjected to impulses, undergoing not negligible fluctuations in the case of wheel-rail interaction. Many factors like curved track, longitudinal and lateral load transfer, track geometry variability, presence of faults in the rail head, subgrade and just the presence of healthy or degraded track features like switches and crossings (S&C), frogs, joints and insulated joints, bridges, etc., affect the signal in many different and complex ways, distorting it from the pure periodic one. Periodicity is expected inside the single revolution of the wheel, but such cycle is repeated countless times during a mission and only if none of the previously mentioned external factors are present. Faults can be identified by observing distortions from a reference “clean” signal arising from straight or curved track section, free of faults. Many times, it is not trivial to distinguish true track faults from track features, which instead are always present and propaedeutic for the rail transportation. Some “abnormal” contact force fluctuations can be easily detected, while others less easily, but the main drawback of traditional data analysis is that they cannot really distinguish from sane to damaged, or from healthy to worn track feature. To do so, the observation of various recordings of the same track over a long monitoring period to evaluate track evolution and force time-series variation, is the only solution, for example when observing rail wear through spectrogram variations for the detection of growing roughness. Alternatively to traditional methods, machine learning is gaining more interest in both the scientific academic and the industrial communities. Analysing big data becomes particularly troublesome when relying only on the experience of a human operator, and visual comparison. Automated anomaly detection (or novelty detection) is becoming every year more reliable and precise. Many algorithms have been created for all kinds of purposes.

7.2 State of the art of data-driven methods

This section surveys data-driven approaches that infer the rail condition as railhead surface defects (e.g., head checks, squats, corrugation) and geometry-related indicators (e.g., longitudinal level for EN 13848), from on-board dynamic responses, namely instrumented wheelset forces, axle-box/bogie accelerations, and carbody accelerations. The emphasis is on models trained from measured signals rather than physics-only estimators, and on outputs that map directly to actionable maintenance judgements.

For what regards sensing and mounting, three families dominate the field: Instrumented wheelsets (IWS); axle-box (ABA)/bogie accelerometers; and carbody accelerations [1,2]. Instrumented wheelsets (IWS) provide direct estimates of wheel-rail contact forces up to high frequency, enabling sensitivity to rail rolling-surface defects; they are powerful but costlier and operationally demanding. Pioneering work by Gullers [3,4] established the Track Condition Analyser (TCA) to identify rail rolling-surface defects likely to trigger wheel fatigue using IWS measurements and documented high-frequency contact-force behaviour in the field. The TCA system based on Machine Learning (ML), classified track conditions using supervised learning and labelled datasets of faulty and healthy track force recordings. It identified short-pitch rail corrugation, rolling surface defects with isolated or periodic force transients, and stiff rail supports. Characteristic features, including mean values, percentiles, and standard deviations of vertical contact forces (Q), were extracted and filtered across different frequency intervals.

Axle-box/bogie accelerometers are a pragmatic compromise, closer to the primary suspension and therefore more rail-sensitive than carbody sensors; recent studies regress rail longitudinal level from bogie accelerations with strong agreement to instrumented inspections [5]. Carbody accelerations are easiest to retrofit and have matured into operational systems on regional lines, provided that position correction and speed-robust processing are handled carefully.

Typical pipelines include precise geo-referencing (GNSS with map-matching/yaw-rate correction), detrending and speed normalisation; spatial windowing (10-100 m) consistent with maintenance practice; and multi-domain features (time, frequency, and time-frequency). Time-frequency representations (e.g., HHT/CWT) are particularly informative for localised, rail-surface-driven transients (squats, welds, and joints) and corrugation [1,2].

In terms of learning paradigms, four complementary strands can be distinguished. First, feature-based regression and classification models use statistical indicators of vibration (e.g., RMS, peak, crest factor, kurtosis) to estimate geometric indicators such as the longitudinal level D1 defined by EN 13848. Regression analyses on bogie-mounted accelerations have demonstrated high determination coefficients, enabling direct comparison to standardised maintenance thresholds [5,6]. Complementary to these deterministic regressions, classification frameworks have also been explored for the lateral and cross-level irregularities of the track. De Rosa et al. (2021) trained several supervised machine-learning classifiers, decision trees, SVMs, and ensemble models, on vibration features measured on board to detect and distinguish different irregularity types. Their results confirmed that vehicle-borne dynamic responses can effectively discriminate geometry-related track defects, providing a data-driven alternative to geometry car inspections [7]. Second, probabilistic approaches and uncertainty-aware inference. Gaussian Process Regression (GPR) has been used to estimate longitudinal-level irregularities from carbody acceleration, with the advantage of returning confidence bounds alongside point estimates, useful for risk-aware maintenance triggers [8]. In parallel, Malekjafarian et al. (2023)

proposed a two-stage learning-based framework combining an artificial neural network (ANN) trained on the energy of acceleration signals from healthy runs with a Gaussian fitting stage applied to prediction errors. The resulting deviation index proved sensitive to localised rail or track damage, illustrating a robust hybrid between data-driven prediction and statistical damage detection [9]. Third, supervised decision models such as compact neural networks have been employed to classify maintenance needs (e.g., tamping or surfacing) from in-service vibration data. These frameworks, often using ANOVA-based feature selection, shift the focus from identifying defects to predicting the most appropriate intervention [10,11]. Finally, unsupervised frameworks have emerged for label-free anomaly detection. In scenarios where labelled data are scarce or unavailable, particularly for rare or emerging defect types, unsupervised anomaly detection offers a complementary path. Ghiasi (University College Dublin) [12] proposed an unsupervised framework for on-board detection of track geometrical defects based on a One-Class Support Vector Machine (OC-SVM) trained on “healthy” vibration signatures. The method identifies deviations from nominal behaviour without explicit defect labels, providing early warning indicators and improved generalisation across lines and vehicle types. This approach underscores the growing interest in self-adaptive, label-free monitoring capable of continuous learning from in-service data streams.

Building on Gullers’ IWS groundwork, a 2024 Italian survey by Rosano [13] argues for integrating accelerations and contact-force measurements to improve detection and prediction of permanent-way defects, particularly rail-surface degradation that may not be fully captured by low-frequency geometric indicators. This supports mixed-sensor strategies where accelerations provide coverage and IWS sampling campaigns provide high-fidelity labels and transfer learning anchors.

For instance, Tsunashima and colleagues have deployed operational carbody monitoring systems on regional lines, incorporating time-frequency analysis and position correction to track condition changes over time. La Paglia et al. demonstrated that bogie accelerations can accurately estimate longitudinal level on high-speed lines using multiple regression, validated over multi-year datasets. De Rosa et al. (2021) extended this perspective by applying machine-learning classification algorithms to vehicle-borne vibration data, successfully identifying lateral and cross-level irregularities through statistical and spectral features extracted from bogie dynamics. Building on these geometry-oriented efforts, Tsunashima (2024) introduced Gaussian Process Regression (GPR) with explicit uncertainty quantification, bridging simulation and field data for non-linear estimation of track irregularities. Malekjafarian et al. (2023) proposed a two-stage monitoring framework, in which an artificial neural network trained on the energy of healthy acceleration signals predicts expected responses, and a Gaussian fitting stage applied to the prediction errors enhances sensitivity to localised track damage. Complementary work by Ghiasi and Malekjafarian (2025) applied a fully connected neural network to classify maintenance needs with over 90 % balanced accuracy, supported by explainability through SHAP values. At the other end of the sensing spectrum, Gullers’ instrumented wheelset studies remain foundational for linking contact-force dynamics to rail rolling-surface defects, while Rosano et al. (2024) have highlighted the diagnostic potential of combining accelerations and contact-force measurements to provide a more holistic view of track condition. State of the art review is summarised in Table 7.1.

ML methods were also used to identify wheel-rail contact forces from onboard measurements, as reported in recent activities by Zhu et al. [14], and Walther et al. [15]. In Zhu et al., the authors combine traditional time-domain identification methods for dynamic loads with machine-learning models to correlate vehicle component responses (such as

accelerations of axle boxes and frame) with both vertical and lateral wheel-rail forces. They demonstrate high correlation coefficients for straight and curved track conditions.

In Walther et al., the focus is on a cost-effective onboard monitoring system using simple sensors (accelerometers or deflections) combined with an ML algorithm (Random Forest regressor) to predict lateral and vertical wheel-rail contact forces in tilting trains. They report correlation levels of about 95% for relevant track sections, showing that ML can capture the nonlinear dynamics of the vehicle.

During our research semi-supervised anomaly detection pipelines were employed and tested. In semi-supervised anomaly detection, both training and test datasets are used; however, the training data include only normal instances, without any anomalies. The core concept is that the model first learns the characteristics of the normal class, and subsequently detects anomalies as data points that deviate from this learned representation [16]. This principle is commonly referred to as one-class classification [17]. This approach is acceptable in railway applications because the railway line can be considered scarcely populated by faults if compared to normal instances.

Table 7.1: State of the art summary.

Study (year)	Sensor / Mounting	Rail-specific target	Representation & features	Learning model	Output & metric	Setting / notes
Gullers (2011), <i>Proc. IMechE Part F</i>	Instrumented wheelsets, high-freq contact forces	Rail rolling-surface defects → wheel fatigue risk	High-frequency force spectra and transients	Rule-/threshold-based logic	Defect flags; sensitivity to rail surface irregularities	Seminal IWS field work; rail-surface focus.
Gullers (2008), <i>Wear</i>	Instrumented wheelsets	High-freq vertical contact forces (0–2000 Hz)	Frequency-domain characterisation	— (measurement study)	Baseline contact-force behaviour	Foundational IWS methodology.
Rosano et al. (2024), <i>Ingegneria Ferroviaria</i>	Review of accelerations + contact forces	Integrated diagnostics of track defects	Bibliographic synthesis and conceptual framework	— (survey)	Mixed-sensor strategy roadmap	Highlights complementarity of IWS and accelerometers.
De Rosa et al. (2021), <i>Proc. IMechE Part F</i>	Bogie accelerometers	Lateral and cross-level irregularities	Statistical and spectral vibration features	Supervised classifiers (SVM, Tree, Ensemble)	Classification accuracy of defect type	Validated geometry-defect identification via on-board dynamics.
La Paglia et al. (2024), <i>Proc. IMechE Part F</i>	Bogie accelerometers	Longitudinal level (EN 13848 D1)	RMS/peak/crest/kurtosis over 100 m windows	Multiple regression	High R ² vs reference geometry	High-speed lines; multi-year validation.
Facchinetti et al. (2024) (POLIMI report)	Bogie accelerometer (centre)	Longitudinal level (D1)	Filtered acceleration + double integration	Deterministic estimator	Geometry profile estimate vs reference	In-service validation; single-sensor setup.

Tsunashima (2023), <i>Applied Sciences</i>	Carbody accelerometers (+ GNSS, yaw-rate)	Condition change over time	Time–frequency (HHT/CWT); position correction	Feature-based analytics	Condition indices over runs	Operational regional deployment.
Tsunashima (2024), <i>Vibration</i>	Carbody accelerations	Longitudinal-level irregularity (+ uncertainty)	Simulation-aided feature sets	Gaussian Process Regression	Estimate + confidence interval	Hybrid simulation/field study.
Malekjafarian et al. (2023), <i>Sensors</i> 23(17): 7568	On-board accelerations (fleet data)	Track damage / degraded condition	Energy of accelerations; prediction-error features	Two-stage: ANN + Gaussian fitting	Detection via prediction-error increase	Hybrid data-driven/statistical; validated on fleet data.
Ghiasi & Malekjafarian (2025), <i>Railway Eng. Science</i>	In-service accelerations (vehicle-mounted)	Maintenance needs (tamping/surfacing)	Time/freq features; ANOVA selection	Feed-forward NN (CFNN)	Balanced accuracy $\approx$ 93 %	Decision-oriented framework; explainability included.
Ghiasi (2024), [12]	In-service accelerations	Track geometrical anomalies	Time/freq features; healthy reference data	One-Class SVM (unsupervised)	Anomaly score for defect segments	Label-free; domain-robust; early-warning potential.

7.3 Datasets and pre-processing

The effectiveness of any data-driven diagnostic framework strongly depends on the quality, representativeness, and reliability of the input data. In this study, all analyses were based on in-service measurement data collected from an instrumented Track Recording Vehicle (TRV) operating on the Italian railway network. The selected datasets contain high-resolution measurements of wheel-rail contact forces and complementary track geometry information, which together provide a comprehensive description of the dynamic interaction between the vehicle and the track. These data serve as the foundation for both feature extraction and the subsequent machine learning-based novelty detection analysis.

All input data used for the model, specifically the vertical wheel-rail contact forces and the corresponding reference track geometry parameters, were recorded by a dedicated TRV. The data acquisition system sampled the contact forces at a frequency of 2 kHz, ensuring adequate temporal resolution to capture dynamic phenomena associated with wheel-rail interaction. To enable spatial analysis and remove speed dependence, the time-domain data were resampled into the spatial domain at a wavenumber of 40 cycles/m, corresponding to a spatial resolution of 0.025 m per sample. Due to non-disclosure agreements (NDA) with the partner company, these datasets are confidential and therefore not publicly available. However, the datasets include both the raw vertical force signals and auxiliary track information such as curvature, cant, longitudinal level, alignment, twist, gauge, and cant deficiency, which provide essential contextual data for subsequent feature extraction and model training. Some track information are made available in Table 7.2. Curvature and train speed profile are shown in Figure 7.1 and Figure 7.2 respectively.

Table 7.2: Training and fitting dataset characteristics.

Dataset Name	Train max speed and tolerance [km/h]	Track type	Length [km] (Duration [s])
C12	320±5	Mixed ¹	112.6 (2252)
C14	330±5	Mixed ²	113.2 (2263)

Features of Mixed Mission Profiles

¹.Max long. acc. ± 0.1 [m/s²]. Min curv. radius 4000 [m]. Max slope 5‰

².Max long. acc. ± 0.1 [m/s²]. Min curv. radius 4000 [m]. Max slope 5‰

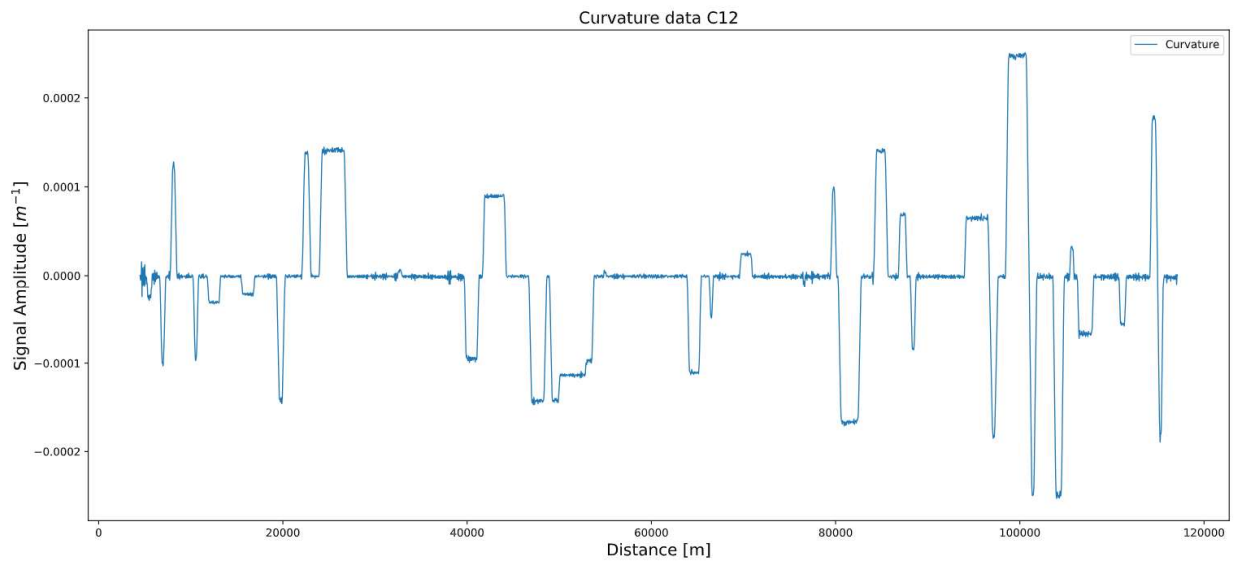


Figure 7.1: C12 and C14 curvature data.

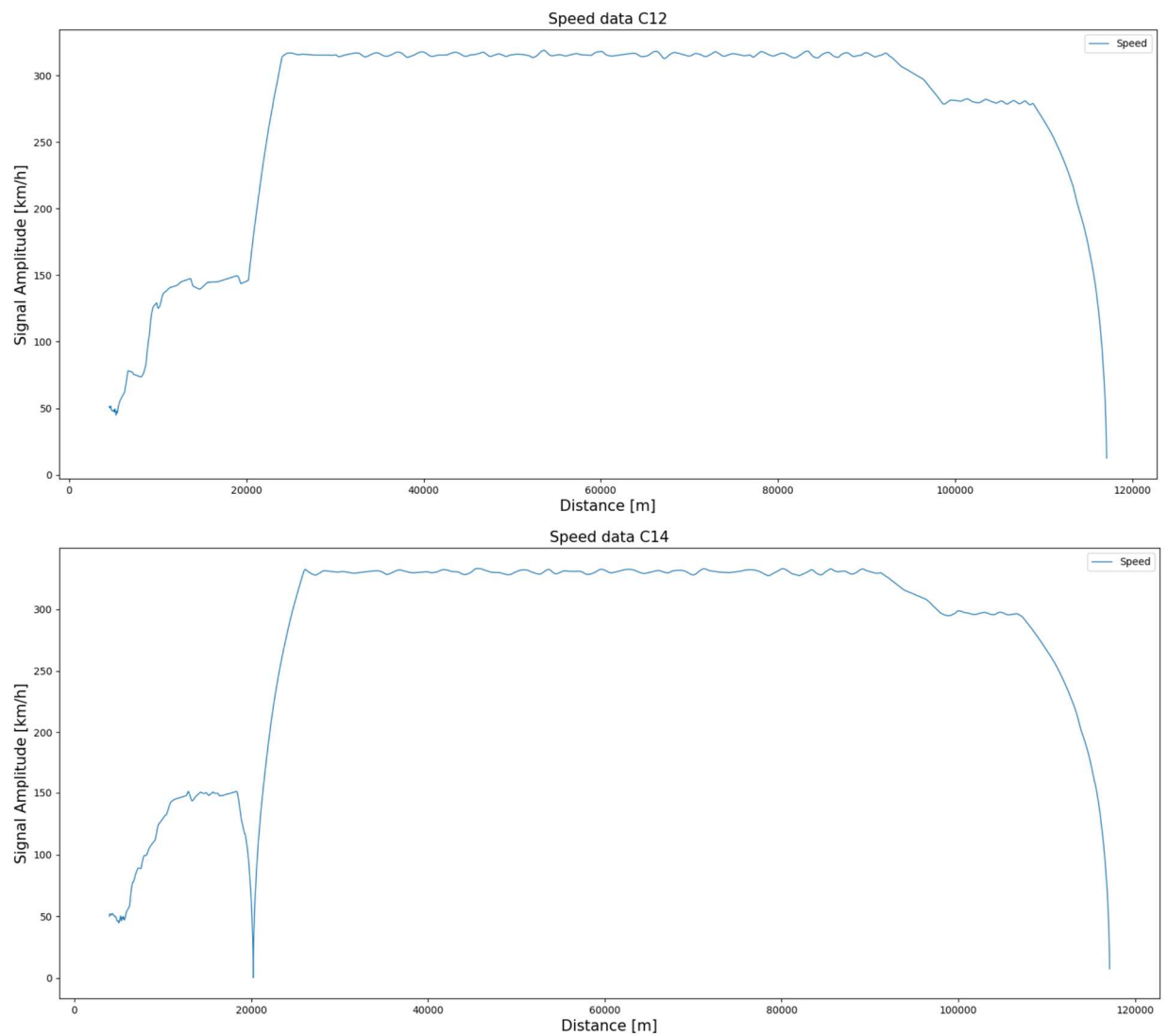


Figure 7.2: C12 and C14 speed data.

The available training dataset, namely “C14” consisted of recordings collected over runs on well-maintained high-speed lines, assumed to be predominantly free of severe defects. This dataset contains several minutes (or miles) of contact force recordings, track lateral inclination, curvature, cant deficiency, and main track geometric parameters such as longitudinal level, alignment, twist, gauge and cant. Odometry and train speed are also available.

A separate dataset recorded on the same line but at a different time, namely “C12” was used for classifiers evaluation. In this dataset, synthetic defects were numerically injected into the raw force signals. These defects were designed to emulate characteristic signatures of common rail degradations such as dipped joints, peaked or dipped welds, and squats, with waveform shapes derived from the literature [18-21]. An example of injected defects is shown in Figure 7.3.

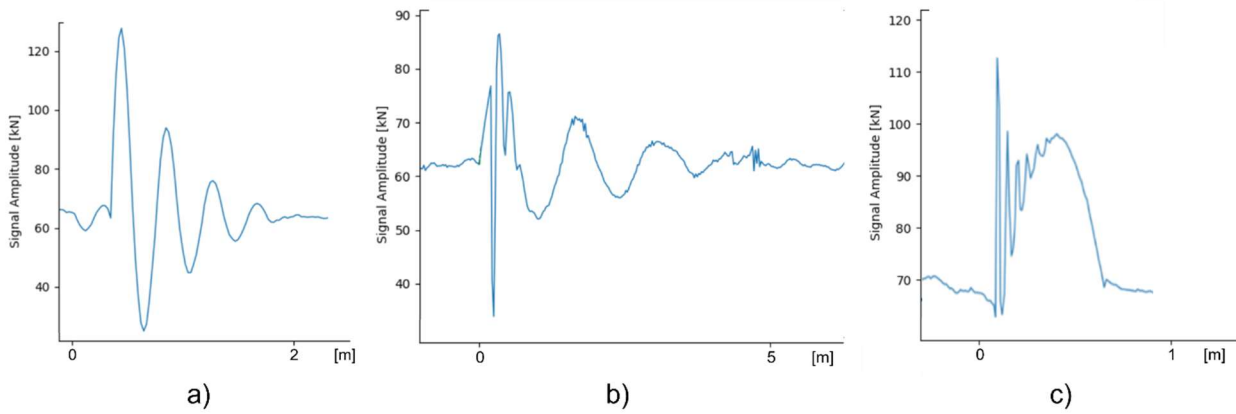


Figure 7.3: Examples artificial numerical faults waveform created from the literature. a) weld, b) squat, c) joint and/or welds.

Each defect type was introduced at multiple track positions and at three amplitude levels (100%, 70%, 55% of the nominal value), resulting in 81 annotated anomalous segments that constitute the ground truth for classifiers performance assessment. The first kind of artificial defect (Figure 7.3-a) was created with a maximum peak value of around 60 kN. Figure 7.3-b defect shows a maximum positive peak of 25 kN and a minimum of 35 kN. Figure 7.3-c type defect shows a maximum peak of 55 kN.

To mitigate the influence of train-speed variability, all further processing was carried out in the spatial domain, using the kilometre progressive coordinate (mileage) instead of time. Raw force signals were intentionally not detrended to maintain all the available information of the signal, especially slowly varying components of the contact force deriving from track curvature, slopes, vehicle acceleration and braking.

The signals were then segmented into overlapping spatial windows (window length L_w , step L_s) for feature computation. A range of window lengths was explored to capture the influence of such parameter in the detection performance of the various tested classifiers. It must be said that the signal windowing should also keep into account the wavelength of the phenomena under investigation. Particularly, short-wavelength for rail-head faults and long-wavelength for track-geometry. Given that the focus is on rail head faults of reduced extension, the wavelength was chosen accordingly as shown in the next sections. The window length should also consider the amount of information carried by the signal. In time-frequency analysis, a very short windowing of the signal, produces a not sufficiently informative spectra content, because the signal portion can't be decomposed in simpler waves (such in case of STFT, Wavelet, etc.). Thus, a compromise between georeferencing positioning precision and quantity of signal information, must be found.

7.3.1 Spatial resampling and windowing

Let $Q(s)$ denote the wheel-rail contact force as a function of the spatial coordinate s [m]. The raw time-series signal $Q(t)$ is mapped to space using.

$$s(t) = \int_0^t v(\tau) d\tau, \quad (Q(s) = Q(t(s))), \quad 7.1$$

and resampled on a uniform spatial grid.

$$s_n = n\Delta s, \quad (n = 0, \dots, N - 1). \quad 7.2$$

Feature extraction is performed over overlapping spatial windows of length $Lw = N\Delta s$. Each window is tapered with a suitable weighting function $w[n]$ (Hann with unit energy) to reduce spectral leakage.

$$x_w[n] = x[n]w[n], \quad (n = 0, \dots, N - 1). \quad 7.3$$

7.3.2 Data Normalization

Feature vectors were standardised before being passed to the novelty-detection algorithms. The effect of different scaling strategies (Standard, Robust, Min–Max) was investigated; the choice was found to influence both detection sensitivity and computational cost.

7.3.3 Novelty Detection Models

Three unsupervised one-class methods were evaluated:

- One-Class Support Vector Machine (OCSVM) [22]
- Isolation Forest (IF) [23]
- Local Outlier Factor (LOF) [24].

In the context of data-driven anomaly detection, one-class classification methods are particularly suitable when only normal data are available for training and the anomalous class is either rare or undefined. Such is the case of the present work, and especially true for railway track data, where the number of faults is (luckily) reduced when compared to normal data. Unlike traditional binary classifiers, which require labelled examples of both normal and faulty conditions, one-class algorithms learn a representation of the normal data distribution from a set of input features, generally summarised in a feature matrix, and subsequently identify observations that deviate significantly from it. A brief description of the three algorithms follows.

7.3.3.1 One-Class Support Vector Machine (OCSVM)

The OCSVM is an extension of the classical Support Vector Machine to unsupervised novelty detection, originally introduced by Schölkopf et al. [22]. It seeks to learn a decision boundary that

encloses the majority of data points in the feature space, while isolating outliers that fall outside this boundary. This allows the algorithm to distinguish only between normal and abnormal data (one single class). Mathematically, the algorithm maps input data into a higher-dimensional space using a kernel function (typically the radial basis function, RBF) and then determines the maximum-margin hyperplane that best separates the data from the origin.

The optimisation problem is governed by two hyperparameters:

- ν (ν), which defines the expected proportion of outliers and the fraction of support vectors, especially in the training dataset;
- γ (gamma), which controls the spread of the RBF kernel, meaning the smoothness of the decision boundary, which may also bring to issues like overfitting of the model, making the trained classifier performance strictly dependent on the training dataset [25].

Data points lying inside the learned hypersphere or above the decision function are classified as normal, whereas those outside are considered novelties or faults. Throughout this thesis, for practicality, the decision function sign was inverted allowing a more intuitive detection of the anomalies. The main advantages of OCSVM are its robustness to high-dimensional data and its ability to model complex, non-linear decision boundaries. However, its performance can be sensitive to kernel choice and hyperparameter tuning.

7.3.3.2 Isolation Forest (IF)

The Isolation Forest, first introduced by Liu et al. [23], is a tree-based ensemble method designed specifically for outlier detection. It assumes that anomalies are few and variegated, and can be isolated faster than normal points through recursive data partitioning. Each tree in the forest is built by randomly selecting a feature and a split value between its minimum and maximum bounds. The path length required to isolate a given data point (a leaf) across all trees becomes the basis for its anomaly score (decision function). In fact, shorter average paths correspond to higher anomaly likelihood.

IF does not rely on distance or density estimation like OCSVM or LOF, making it computationally efficient and scalable to large datasets, which is particularly interesting in the current application, with large amount of data, in which real time application become extremely appealing. Its main strengths are low memory requirements, fast training and inference, and robustness to irrelevant features. This makes IF algorithms optimal candidates for novelty detection pipelines applied to real time applications of track monitoring. However, the random nature of the splits may introduce variability in the results, and the method may struggle when anomalies are not well-separated from normal data in feature space.

7.3.3.3 Local Outlier Factor (LOF)

The Local Outlier Factor (Breunig et al. [24]) is a density-based method that identifies anomalies by comparing the local density of each sample to that of its neighbours. The main assumption is that normal points occur in regions of approximately uniform density, while anomalies are located in sparser areas. Based on the extension of the “neighbouring” area, anomalies can be roughly divided into local (if they don’t deviate significantly from the whole dataset, but only from their neighbouring points) and global, if the entire dataset is evaluated as a whole as shown in [16]. For each sample, LOF algorithm computes the local reachability density (LRD) based on the distances to its k nearest neighbours (cluster of data points), and the anomaly score (decision function) is defined as the ratio of the average LRD of the neighbours to that of

the sample itself, which in definitive is the local outlier factor. A higher LOF value indicates that the point is significantly less dense than its neighbours, and thus more likely to be anomalous. As reported by Goldstein and Uchida [16] LOF is particularly effective in identifying local anomalies, where the deviation is context-dependent rather than global. A representation of local and global anomalies is shown in Figure 7.4. Its limitations include high computational cost, strong dependence on the neighbourhood parameter k , and limited scalability to large datasets. Despite this, LOF remains a valuable benchmark method for assessing density variations in high-dimensional feature spaces.

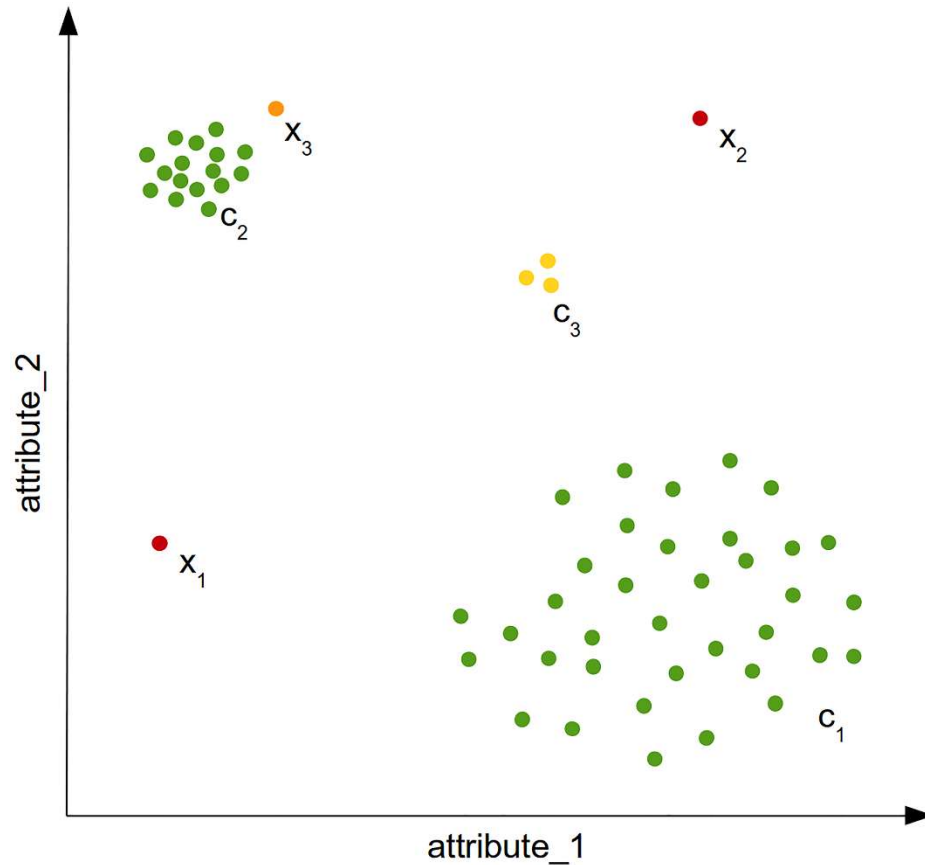


Figure 7.4: Figure illustrates global anomalies (x_1 , x_2), a local anomaly x_3 and a micro-cluster c_3 [16].

All models were trained solely on the healthy dataset (C_{14}) and were then applied to the evaluation dataset (C_{12}) with synthetic defects. Their discrimination thresholds of the decision function were varied to generate precision-recall (PR) curves. Model fitting time and prediction latency were also recorded to assess deployability in large-scale monitoring.

7.4 Features extraction

Rather than analysing the entire signal directly, the data were processed using a sliding window approach, i.e., the signal was segmented into consecutive, overlapping, spatial windows. Each window was then represented by a set of descriptive statistical features. Following the approach proposed in previous studies [3,12-32], time- or space-domain features such as RMS, crest factor, skewness, kurtosis, and others were computed and used as input

features for the classifiers, due to their ability to capture changes in signal distribution, as well as transients and irregularities associated with local defects.

For each spatial sliding window $x[n]$ of length N , the space features were then derived as described in Table 7.3 [28]:

Table 7.3: Extracted spatial features.

Parameter	Symbol	Formula
Mean	μ	$\mu = \frac{1}{N} \sum_{i=1}^N x[i]$
Peak value	x_{peak}	$\max x_i $
Standard deviation	σ	$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$
RMS	RMS_f	$RMS_f = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i)^2}$
Skewness	Sk_f	$Sk_f = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^3}{\sigma^3}$
Kurtosis	Kr_f	$Kr_f = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^4}{\sigma^4}$
Superior order moments (order $k=5, 6$)	Ord_k_f	$Ord_k_f = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^k}{\sigma^k}$
Peak factor	Pk_f	$Pk_f = \frac{x_{peak}}{\mu}$
Crest Factor	Cr_f	$Cr_f = \frac{x_{peak}}{RMS_f}$
Impulse factor	Imp_f	$Imp_f = \frac{x_{peak}}{\frac{1}{N} \sum_{i=1}^N x_i }$
Shape factor	Sh_f	$Sh_f = \frac{RMS_f}{\frac{1}{N} \sum_{i=1}^N x_i }$

Clearance Factor	Cl_f	$Cl_f = \frac{x_{peak}}{\left(\frac{1}{N} \sum_{i=1}^N \sqrt{ x_i }\right)^2}$
Margin Factor	Ma_f	$Ma_f = \frac{Cl_f}{Sh_f}$

The mean represents the average value of the data series and is a measure of central tendency. It corresponds to the static or “DC” level of the signal. In the case of force measurements, it indicates the average applied load. The standard deviation measures the dispersion of the data around the mean value. For vibration or force signals, high standard deviation values indicate greater variability or irregularity in the phenomenon. The Root Mean Square (RMS) value is related to the energy content of the signal. In force or acceleration measurements, the RMS represents the effective amplitude of the oscillations and is often used as a global indicator of signal intensity or energy. The skewness quantifies the asymmetry of the data distribution with respect to the mean.

- Skewness > 0 → long right tail (presence of positive peaks)
- Skewness < 0 → long left tail (presence of negative peaks).

This parameter highlights asymmetric events, making it useful for detecting local defects or impacts. The kurtosis measures how “peaked” or “flat” the data distribution is compared to a Gaussian distribution (which has kurtosis = 3).

- Kurtosis > 3 → “peaked” distribution (presence of impulsive events)
- Kurtosis < 3 → flatter distribution.

In diagnostic applications, kurtosis is a powerful indicator of impulsive anomalies, such as contact defects, impacts, or cracks. The higher-order moments are a generalization of statistical moments. Depending on the order k:

- k = 2 → variance
- k = 3 → non-normalized skewness
- k = 4 → non-normalized kurtosis

When $n > 4$, these parameters become increasingly sensitive to highly impulsive phenomena in the data. The peak factor indicates how much the maximum peak exceeds the mean level of the signal and is sensitive to signal asymmetry. The crest factor is the ratio between the maximum peak and the RMS amplitude. High values indicate the presence of impulsive peaks relative to the average energy level and are useful for detecting localized impacts or defects. The impulse factor is defined as the ratio between the maximum peak and the mean absolute value. It measures the degree of impulsiveness of the signal. If the signal contains isolated impulses, the impulse factor increases significantly. The margin factor quantifies the relationship between the maximum peak and a mean value weighted by the square root of the samples. It is particularly useful for characterizing intermittent impulses or anomalies. The shape factor describes the overall waveform shape. High values correspond to more impulsive signals, while low values indicate more uniform or regular signals. It is useful for distinguishing between steady noise and signals with pronounced peaks. The clearance factor is highly sensitive to localized defects, such

as irregular contacts or cracks. It is widely used in rotating machinery and railway diagnostics, where high values indicate distinct impulsive events in the signal.

7.4.1 Hjorth Parameters

Originally developed for electroencephalographic analysis, Hjorth parameters provide a compact description of a signal's variability and frequency content and have shown potential in condition-monitoring applications.

Given a windowed signal $x[n]$, let $\dot{x}[n]$ and $\ddot{x}[n]$ denote its first and second discrete derivatives. Defining the variance operator as $\text{Var}[\cdot]$, the Hjorth parameters [33] can be calculated as reported on Table 7.4.

Table 7.4: Hjorth Parameters [33].

Parameter	Symbol	Formula
Activity	A_H	$A_H = \text{Var}[x]$
Mobility	M_H	$M_H = \sqrt{\frac{\text{Var}[\dot{x}]}{\text{Var}[x]}}$
Complexity	C_H	$C_H = \sqrt{\frac{\text{Var}[\ddot{x}]/\text{Var}[\dot{x}]}{M_H}}$

Activity represents the signal power or overall amplitude variation. Mobility reflects the mean frequency content; higher mobility indicates more rapid fluctuations (higher frequency content). Complexity measures the degree of deviation of the signal from a pure sine wave; higher values correspond to more irregular or broadband patterns.

Hjorth parameters were computed and a low-pass-filtered to remove low frequency trends and fluctuations. Although these parameters tend to be noisy (see Figure 7.5), they revealed distinctive responses in the vicinity of strong contact-force anomalies. A second order linear low-pass (LP) Butterworth filter with cut-off wavelength of 2 cycles/m⁻¹ was used for the purpose. This filter is rather aggressive and trims out not only signal trends, but also many low-frequency features of the signal. Filtered details of mobility and complexity signals are shown on Figure 7.6.

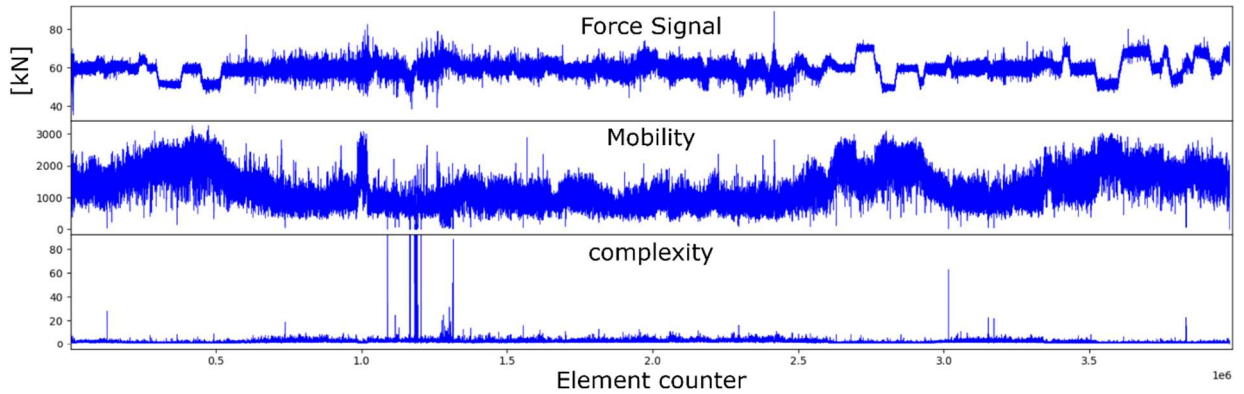


Figure 7.5: Detail example of raw force signal (top), with its calculated Mobility (middle) and Complexity (bottom), showing the noisiness and low frequency fluctuations of the parameters.

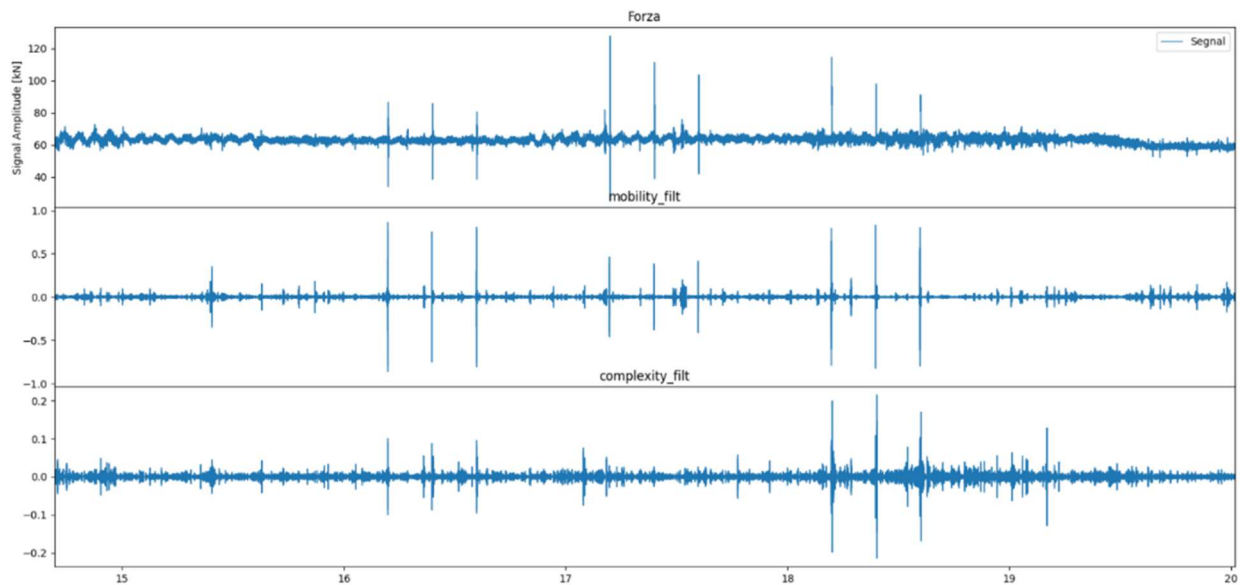


Figure 7.6: Detail example of raw force signal (top), with LP filtered Mobility (middle) and LP filtered Complexity (bottom), showing the high reactivity of these parameters to data fluctuations.

Hjorth parameters have rarely been employed for rotating machinery condition monitoring, mainly due to their reported low correlation with bearing faults [28]. However, by examining the filtered versions of mobility and complexity, a clear correspondence can be observed whenever injected defects or similar load fluctuations occur. This observation justified a deeper investigation.

Although Hjorth parameters convey information related to the signal's frequency content, they are derived from time- or space-domain characteristics rather than from the frequency domain itself (as would be the case with Fourier or wavelet-based analyses). For this reason, they are treated as space-domain features in this study.

7.5 Anomaly Detection Framework: Setup and Evaluation

7.5.1 Framework Setup

The data-driven analysis was carried out using Python. Feature extraction and novelty detection framework described in the previous sections was applied to the datasets C14 (healthy baseline) and C12 (synthetic-defect injected). All analyses were performed in the spatial domain to eliminate the influence of train-speed variability. Each force signal was resampled with respect to the travelled distance and segmented into overlapping spatial windows of length Lw and step Ls , as defined in Section 7.3. A Hann window with unit energy was used to mitigate spectral leakage.

Different combinations of window lengths were tested to evaluate their effect on classifiers performance. In particular, the window size was chosen to balance geometric localization precision and signal information content, ensuring that the wavelength of the investigated phenomena (rail-head defects) was adequately captured.

Each window was described by the space-domain statistical features listed in Table 7.3, normalized prior to model fitting. Among the tested scaling strategies, Standard; Robust; and MaxAbs, Standard scaling was selected, as it offered the best trade-off between detection performance and computational cost.

A list of the Python libraries used to run the analysis is made available in Table 7.5, with a brief description of their usage. All developed codes are made available online in a Git-hub repository.

Table 7.5: List and description of Python libraries used for the data-driven analysis.

Library	Usage description
NumPy	Basic mathematical operations
SciPy	Signal analysis, spectral analysis, FFT
Scikit-learn	Machine learning algorithms, frameworks, classifiers, evaluation metrics
Pandas	Tables and data management
Joblib	Load/save datasets, data management, features parallel computing
Tkinter	GUI design
Mdf Reader	Load/save datasets
Matplotlib	Figure plotting
Numba	High-speed parallelised features calculation

Three unsupervised one-class classifiers were implemented to perform novelty detection:

- One-Class Support Vector Machine (OCSVM)
- Isolation Forest (IF)

- Local Outlier Factor (LOF).

These classifiers are trained exclusively on clean data, with the aim of detecting anomalies in the dataset's behaviour by comparing input features. For each time-series element in the parsed dataset, the model assigns a score, referred to as the decision function or anomaly score, which discriminates between normal and abnormal (anomalous) data points. By varying the threshold applied to the decision function, different classifier performances can be obtained. The optimal threshold is determined as the value that maximises a chosen performance metric, such as the F1 score. This approach allows for a fair comparison between models under conditions where each achieves its maximum F1-score performance.

All models were trained exclusively on the healthy dataset (C14) and later applied to the evaluation dataset (C12) containing the synthetic anomalies. Model hyperparameters (kernel type, contamination ratio, number of estimators, neighbourhood size, etc.) were tuned empirically to ensure comparable sensitivity across classifiers.

7.5.2 Evaluation metrics

The performance of the novelty detection models was assessed through Precision-Recall (PR) analysis, which is more informative than Receiver Operating Characteristic (ROC) curves in the case of highly unbalanced datasets, as is typical for anomaly detection problems, where the number of defective samples is much smaller than the healthy ones.

For each classifier, the decision threshold was varied within the entire output range to generate the PR curve, from which the key metrics were derived.

The adopted evaluation metrics are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad 7.4$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad 7.5$$

$$\text{F1score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad 7.6$$

$$\text{AUC}_{PR} = \int_0^1 \text{Precision}(\text{Recall}) \cdot d(\text{Recall}) \quad 7.7$$

where:

- TP: True Positives (correctly identified defective windows)
- FP: False Positives (healthy windows misclassified as defects)
- FN: False Negatives (missed defects).

Precision quantifies the proportion of detected anomalies that are true defects, while Recall (also called Sensitivity) measures the ability of the model to identify all actual defects. The F1-score represents the harmonic mean between Precision and Recall, thus providing a single performance index that balances both aspects.

The Area Under the Precision-Recall Curve (AUC_{PR}) summarizes the classifier performance over the entire threshold range, offering an aggregate measure of robustness independent of the specific decision boundary.

In the following section, the PR-curves obtained for each classifier are presented and compared. These curves illustrate the trade-off between Precision and Recall, providing insight into each model's capability to discriminate anomalous from healthy track sections under varying decision function threshold levels.

7.6 Results Discussion

This section presents a comparative analysis between the trained classifiers and the novelty detection methodologies described in the previous section.

PR curves are shown below, illustrating the detection performance of the different classifiers in terms of Precision, Recall, and F1-score as the decision threshold varies. Once the optimal threshold is determined, which we remind as the values of decision function which maximises the F1 score, the corresponding best performances are highlighted and summarized in the results section, in Table 7.10, allowing for an overall comparison of the algorithms.

7.6.1 Normalization Influence

As previously mentioned, the feature arrays were standardized before being passed to the novelty detection algorithms. Several scaling strategies (Standard, Robust, and MaxAbs) were evaluated due to their potential impact on both detection sensitivity and computational cost. The MaxAbs scaling strategy resulted in the shortest inference time but yielded poor detection accuracy, producing the largest number of misclassified instances across all three classifiers.

Although comparable in terms of training time (with the Standard scaler being slightly faster than the Robust one), the Standard and Robust scalers provided the best detection results. For statistical space-domain features, Robust scaling was selected for OCSVM and LOF, whereas Standard scaling was preferred for IF in the final comparison of the methods. These configurations achieved the best trade-off between inference time and the number of correctly identified true positives (TP), while minimizing false positives (FP) and maximizing Precision, Recall, and F1-score at the optimal threshold. Here, the optimal threshold refers to the decision boundary value that maximizes the F1-score of the detection.

7.6.2 Influence of signal windowing

The length Lw of the moving window used for feature computation had a significant influence on detection performance. In general, reducing Lw to around 20 array elements increased the number of detected faults (TP) and consequently reduced the number of missed detections (FN). Any further reduction, however, consistently led to performance degradation across all tested algorithms. It should be noted that, in frequency-domain analyses, Lw must be sufficiently long to capture enough spectral content, making excessively small windows unsuitable. Due to the nature of the feature computation process, the sliding-window operation requires the first and last $Lw/2$ elements of each array to be set to zero. This introduces a step-like discontinuity in the signal, which the algorithm may erroneously classify as an anomaly, as illustrated in Figure 7.7.

Fortunately, the corresponding decision function values for these artificial anomalies are relatively low (particularly for OCSVM and IF), or high for LOF, and can be mitigated in two ways:

- by increasing the decision threshold, though this may suppress true positives with low anomaly scores, or
- by trimming the feature matrix after feature computation, i.e., removing $L_W/2$ elements from both the beginning and the end of the array.

This approach ensures that the remaining portion of the feature matrix corresponds to valid signal segments, excluding the artificial boundary effects introduced during recording. In the case of LOF, the issue of edge outliers can be more effectively addressed by trimming the feature matrix.

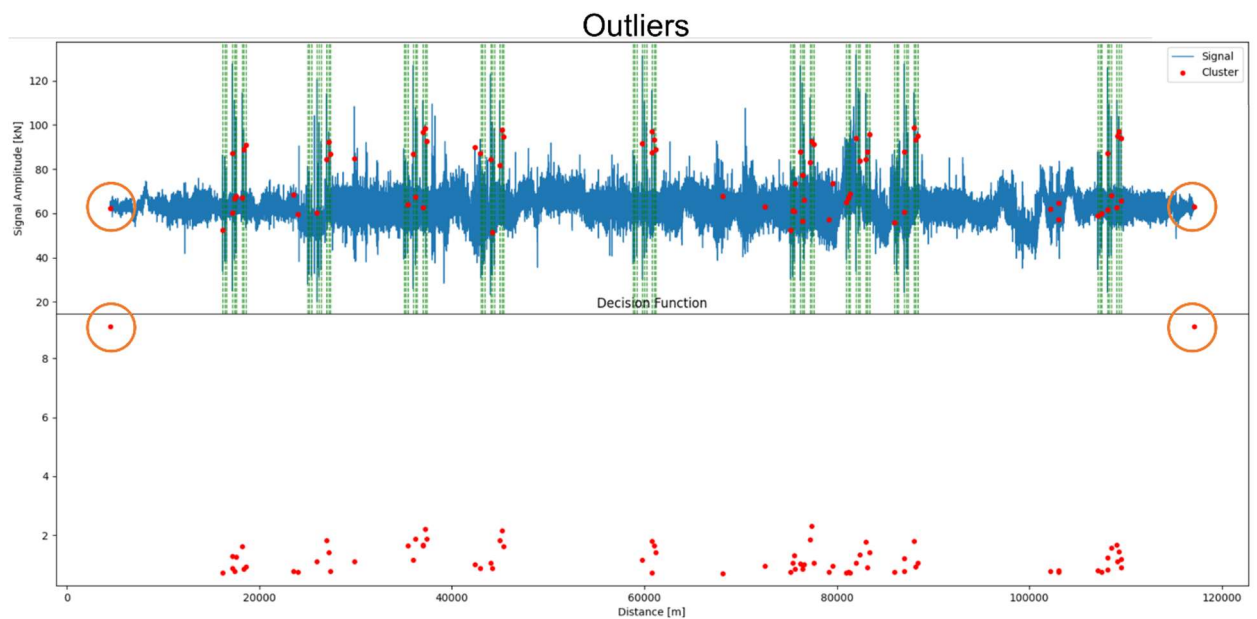
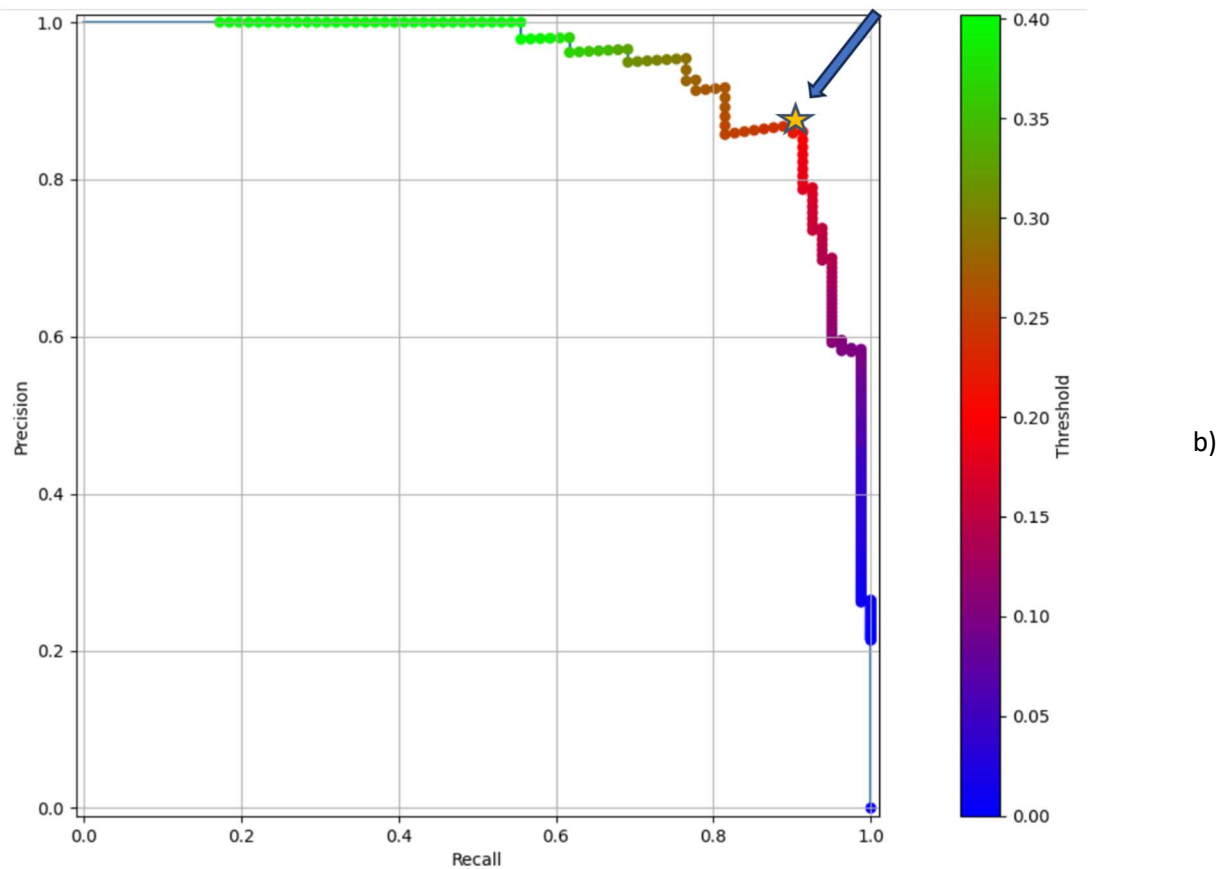
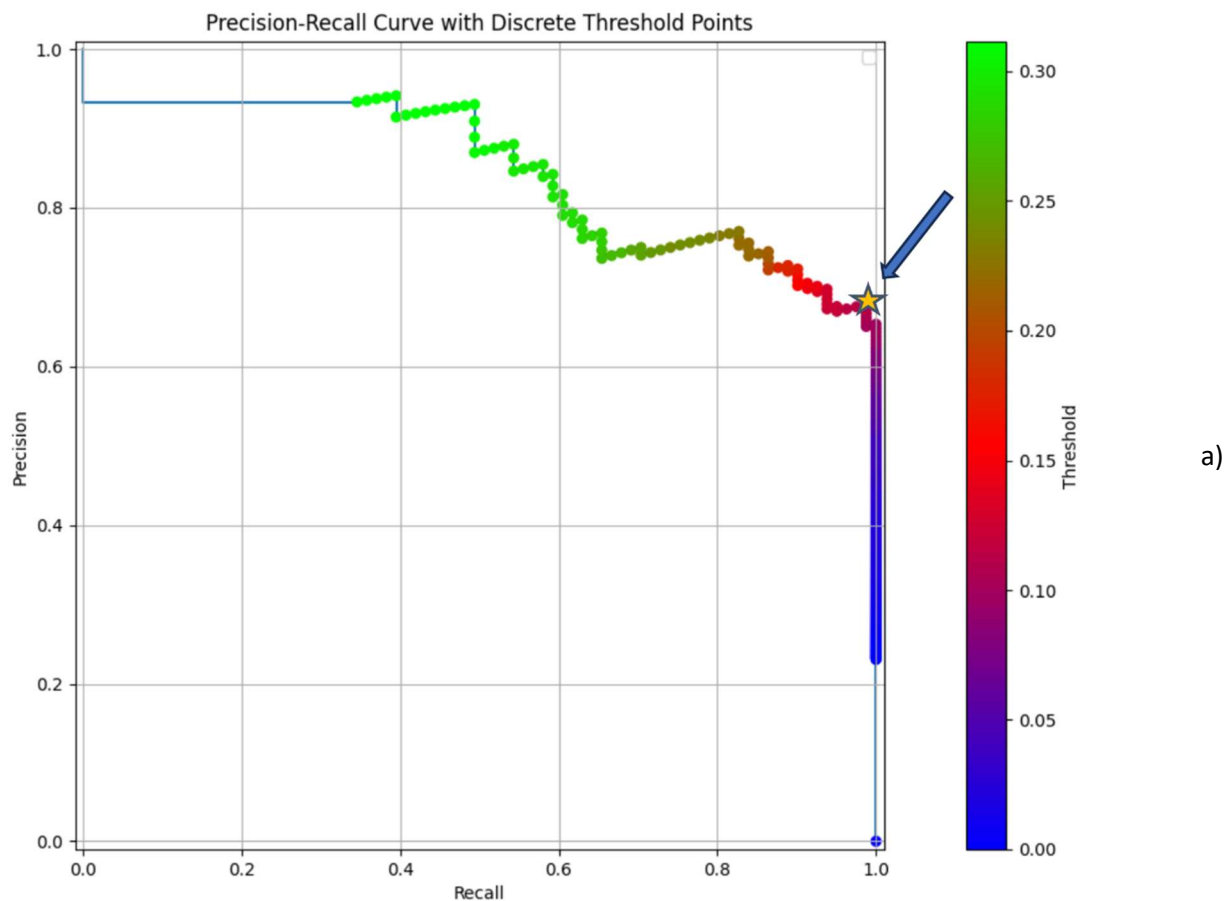


Figure 7.7: Unrealistic anomalies detected by the LOF algorithm. Effect of imposed zeros at the beginning and at the end of the feature matrix, due to feature calculation procedure with sliding windows.

In Figure 7.8 the performance of the OCSVM classifier was evaluated by varying the L_W parameter while keeping all other settings constant (i.e., ν , γ , number and type of features, and filter cut-off frequencies). Reducing L_W generally increased the number of positive instances retrieved by the classifier, both true positives (TP) and false positives (FP). When comparing the results at the threshold value that maximizes the F1-score, the best performance is obtained for a window length of 100 elements (case (d) referring to Figure 7.8), achieving the highest F1-score with a relatively high number of TPs and a lower number of FPs. However, it should be noted that this configuration would lead to a considerable number of undesired missed detections and was therefore not selected. A window length of 20 elements, which exhibited performance comparable to the $L_W = 50$ case, but with improved detection capabilities, was ultimately adopted as the reference configuration for subsequent analyses and comparisons. Results varying L_W are reported in Table 7.6



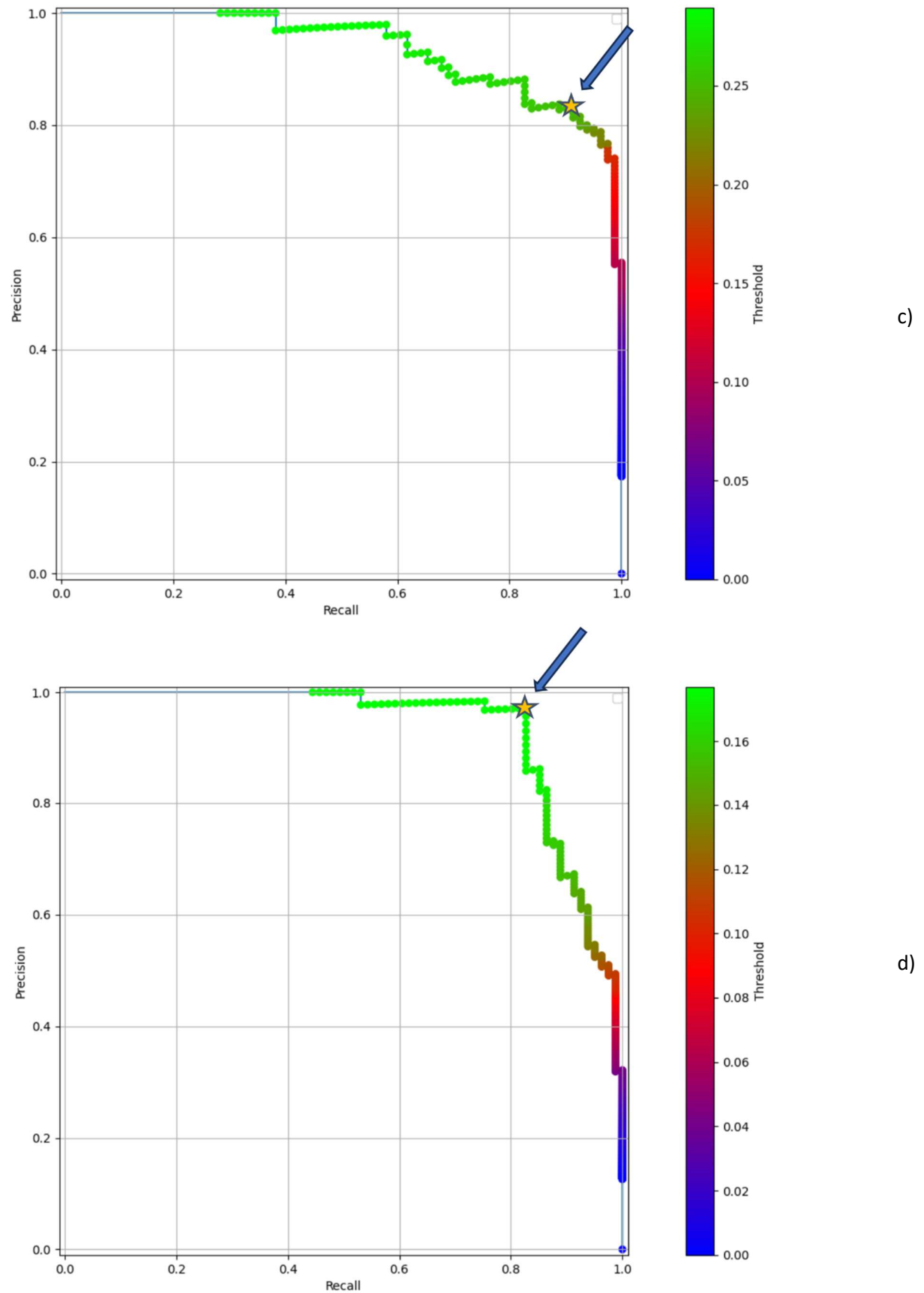


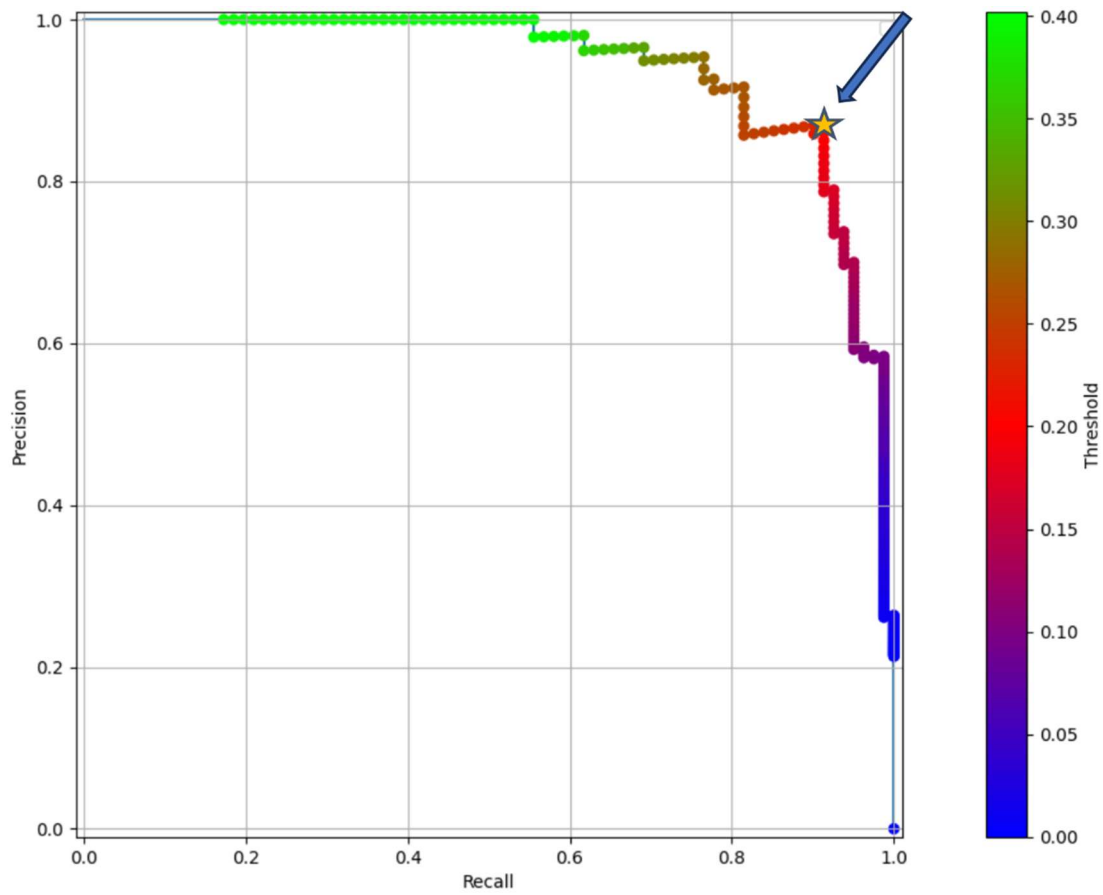
Figure 7.8: PR curves obtained varying the window length L_W for the OCSVM classifier. a) $L_W = 10$ elements; b) $L_W = 20$ elements; c) $L_W = 50$ elements; d) $L_W = 100$ elements.

Table 7.6: OCSVM classifier comparison varying L_w . Detection performance summary.

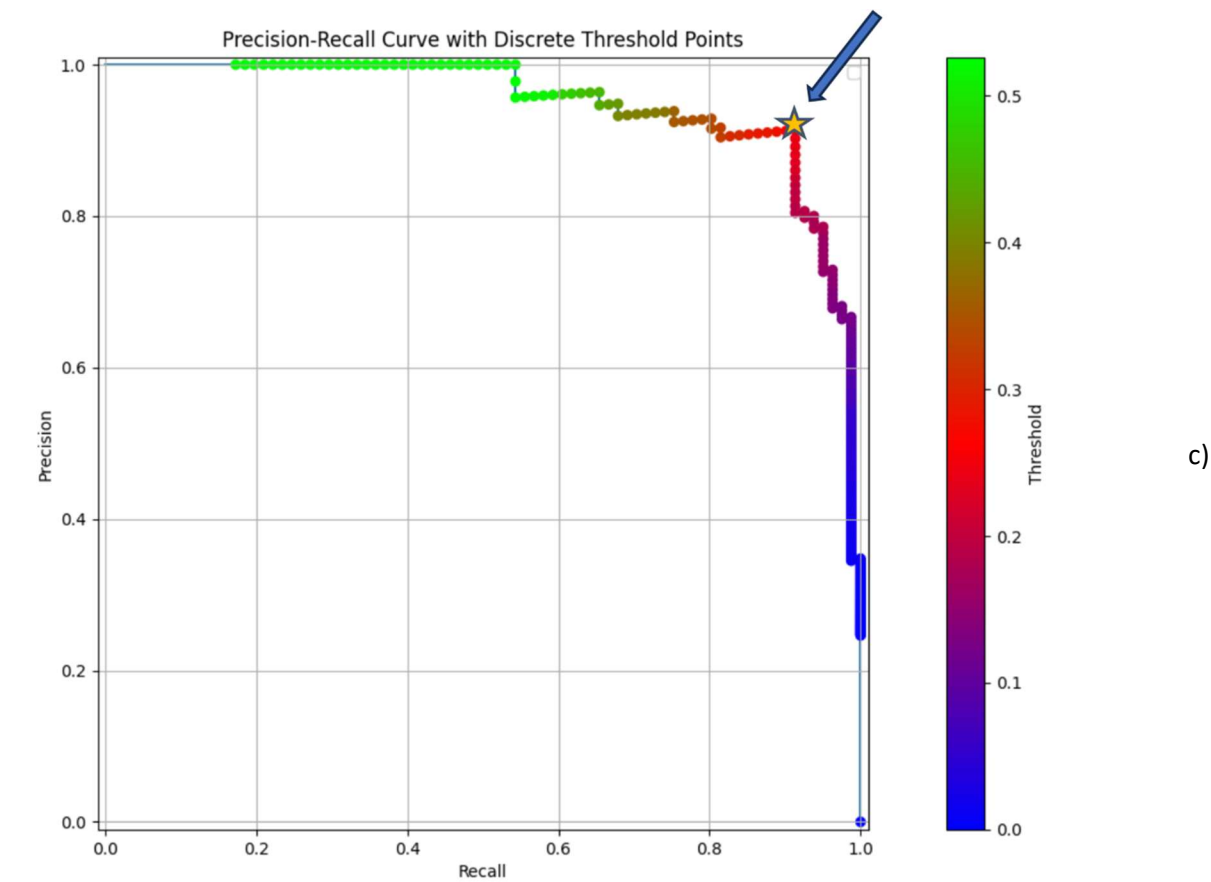
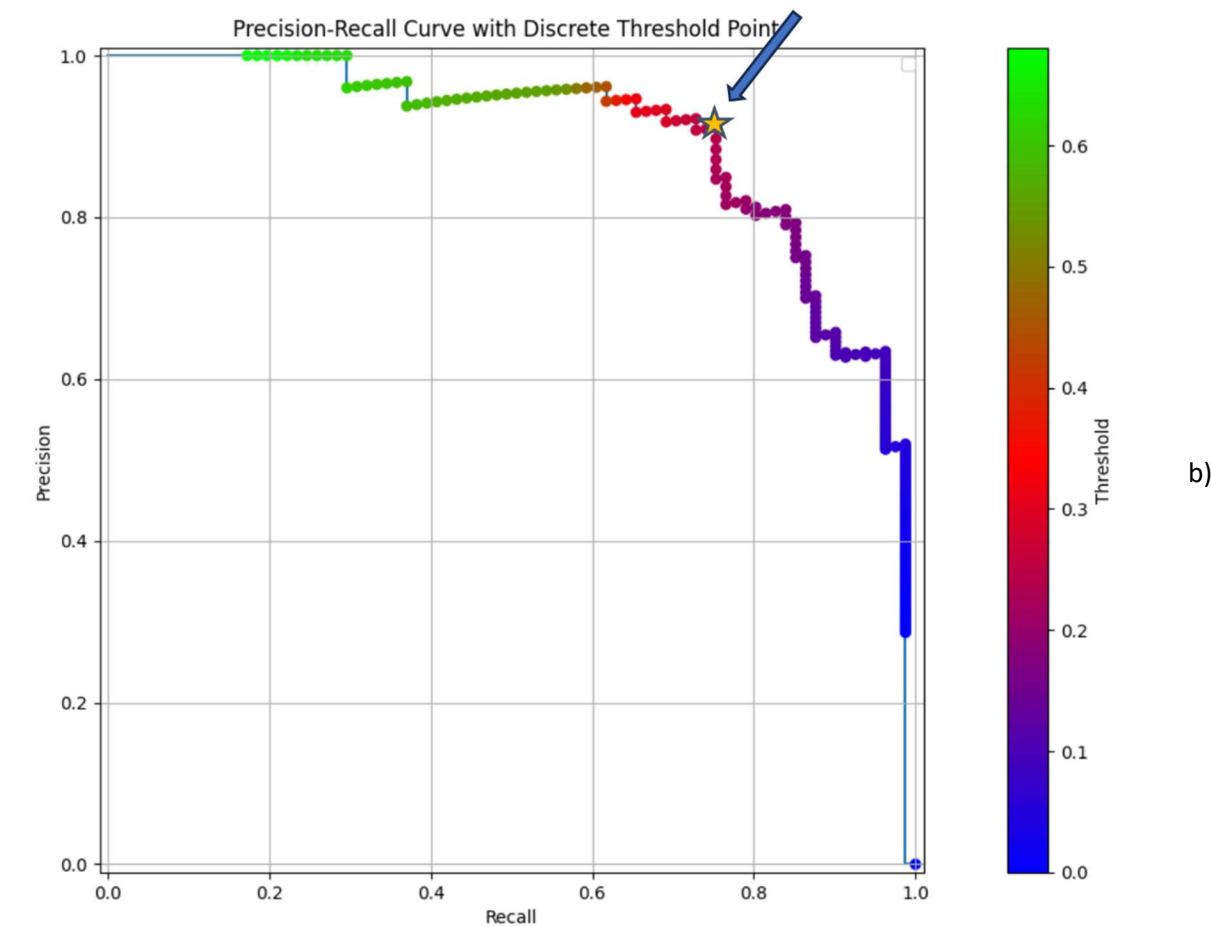
OCSVM	Retrieved instances	F1-Score	Precision	Recall	Time Performance
$L_w = 10$ (Threshold 0.1097833)	TP: 80/81 FP: 39 FN: 1 Max TP: 81	0.804	0.6723	0.987	Training: 265.20 s Fitting: 246.87 s
$L_w = 20$ (Threshold 0.199642)	TP: 74/81 FP: 13 FN: 7 Max TP: 81	0.886	0.851	0.914	Training: 293.25 s Fitting: 246.06 s
$L_w = 50$ (Threshold 0.2496439)	TP: 74/81 FP: 16 FN: 7 Max TP: 81	0.871	0.8222	0.914	Training: 204.26 s Fitting: 277.63 s
$L_w = 100$ (Threshold 0.1755193)	TP: 67/81 FP: 3 FN: 14 Max TP: 81	0.893	0.957	0.827	Training: 305.46 s Fitting: 408.21 s

OCSVM parameters: $\nu = 10^{-5}$; $\gamma = 0.01$; kernel: RBF; scaling: Robust.

In Figure 7.9 PR curves were generated by varying the set of features used to build the feature matrix. The OCSVM classifier performance was evaluated for different combinations of features, as summarized in Table 7.7. Among the tested combinations, configuration (c) yielded the best overall performance.



a)



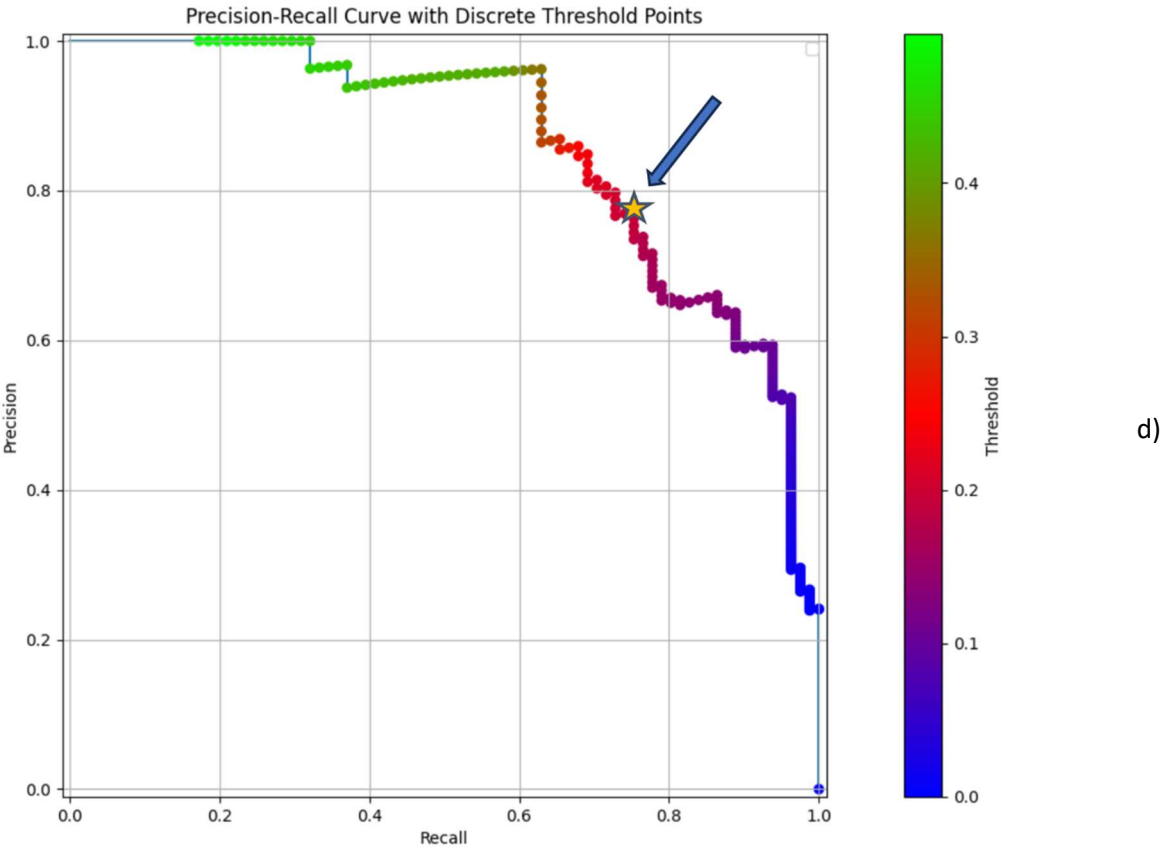


Figure 7.9: PR curves obtained varying features selection. Features are summarised in Table 7.7.

Table 7.7: OCSVM classifier comparison varying features. Detection performance summary.

OCSVM	Retrieved instances	F1-Score	Precision	Recall	Features List
a) (Threshold 0.199642)	TP: 74/81 FP: 13 FN: 7 Max TP: 81	0.886	0.851	0.914	Variance, RMS, skewness, kurtosis, peak, impulse, crest, shape, margin, clearance, 5 th order moment, 6 th order moment, mobility, complexity
b) (Threshold 0.2421137)	TP: 61/81 FP: 7 FN: 20 Max TP: 80	0.824	0.897	0.753	Variance, RMS, skewness, kurtosis, peak, impulse, crest, shape, margin, clearance
c) (Threshold 0.2408881)	TP: 74/81 FP: 8 FN: 7 Max TP: 81	0.914	0.902	0.914	Variance, RMS, skewness, kurtosis, peak, impulse, crest,

					shape, margin, clearance, 5 th order moment, 6 th order moment
					Variance, RMS, skewness, kurtosis, peak, impulse, crest, shape, margin, clearance, mobility, complexity
d) (Threshold 0.1928567)	TP: 61/81 FP: 19 FN: 20 Max TP: 81	0.762	0.763	0.753	

OCSVM parameters: $\nu = 10^{-5}$; $\gamma = 0.01$; kernel: RBF, $Lw = 20$; scaling: Robust.

Varying the scaling method revealed a consistent improvement in detection performance, evidenced by an increase in the number of retrieved instances (both TP and FP), when Robust Scaling was applied. Standard Scaling proved to be faster than Robust Scaling, albeit with slightly lower accuracy, while MaxAbs Scaling demonstrated limited detection capability but offered the shortest training and fitting times. Results of OCSVM detection obtained varying data scaling strategy are summarised in Table 7.8.

Table 7.8: OCSVM comparison varying features scaling method. Detection performance summary.

Scaler	Retrieved instances	F1-Score	Precision	Recall
<i>Standard</i> (Threshold 1.439830)	TP: 57/81 FP: 1 FN: 24 Max TP: 77	0.826	0.983	0.704
<i>Robust</i> (Threshold 1.1313796)	TP: 61/81 FP: 7 FN: 20 Max TP: 80	0.829	0.967	0.716
<i>MaxAbs</i> (Threshold 0.0)	TP: 31/81 FP: 0 FN: 50 Max TP: 31	0.554	1.7984e-05	1.0

OCSVM parameters: $\nu = 10^{-5}$; $\gamma = 0.01$; kernel: RBF; $Lw = 20$; scaling: Robust.

Both IF and LOF classifiers were also evaluated by varying their main parameters, namely, the number of estimators and contamination level for IF, and the number of neighbours for LOF, along with the scaling method and feature set.

Results confirmed that the 20-element signal window provided the best overall performance, and that Robust Scaling remained the preferred pre-processing method for both OCSVM and LOF. A different trend was observed for IF, which exhibited a significant improvement in detection performance when Standard scaling was employed.

Another noteworthy finding was that, for IF, feature selection produced different outcomes compared to OCSVM: the inclusion of Mobility and Complexity features improved detection accuracy, whereas the fifth- and sixth-order statistical moments led to a notable degradation in performance, contrary to what was observed for OCSVM. To maintain the thesis concise, only the features selection of IF case is reported in Table 7.9, as it gave a different outcome compared to the other two algorithms.

Only a limited number of tests were performed for the LOF classifier, due to its highly time-consuming and computationally intensive training and fitting procedures. The consistently poor performance obtained, compared with the other algorithms, was considered sufficient to draw conclusions and justify its inclusion in the overall comparison for completeness.

Table 7.9: IF classifier comparison varying features. Detection performance summary.

IF	Retrieved instances	F1-Score	Precision	Recall	Features List
a) (Threshold 0.0787563)	TP: 63/81 FP: 8 FN: 18 Max TP: 81	0.834	0.887	0.778	Variance, RMS, skewness, kurtosis, peak, impulse, crest, shape, margin, clearance, 5 th order moment, 6 th order moment, mobility, complexity
b) (Threshold 0.2421137)	TP: 61/81 FP: 7 FN: 20 Max TP: 80	0.824	0.897	0.753	Variance, RMS, skewness, kurtosis, peak, impulse, crest, shape, margin, clearance
c) (Threshold 0.0653092)	TP: 58/81 FP: 3 FN: 23 Max TP: 81	0.823	0.951	0.716	Variance, RMS, skewness, kurtosis, peak, impulse, crest, shape, margin, clearance, 5 th order moment, 6 th order moment
d) (Threshold 0.0595690)	TP: 73/81 FP: 10 FN: 8 Max TP: 81	0.896	0.880	0.901	Variance, RMS, skewness, kurtosis, peak, impulse, crest, shape, margin, clearance, mobility, complexity

IF parameters: $n_estim = 100$; Contamination = 0.0001; Lw = 20; Scaling: Standard.

For cross-model comparison, the optimal configuration of each classifier was selected based on the preceding analyses. A summary of these configurations and their corresponding performance metrics is reported in Table 7.10.

Table 7.10: Classifiers comparison and detection performance summary.

Classifier	Retrieved instances	Optimal F1-Score	Optimal Precision	Optimal Recall	Time Performance
OCSVM ¹ (Threshold 0.2408881)	TP: 74/81 FP: 8 FN: 7 Max TP: 81	0.914	0.902	0.914	Training: 180.612 s Fitting: 136.451 s
IF ² (Threshold 0.0510494)	TP: 74/81 FP: 10 FN: 7 Max TP: 81	0.897	0.881	0.914	Training: 249.392 s Fitting: 23.110 s
LOF ³ (Threshold 1.6000777)	TP: 44/81 FP: 8 FN: 37 Max TP: 81	0.667	0.846	0.543	Training: 6542.272 s Fitting: 13178.816 s

Classifiers parameters:

^{1.} $\nu = 10^{-5}$; $\gamma = 0.01$; kernel: RBF ; $L_w = 20$; Scaling: Robust; feature set (c)

^{2.} $n_{estim} = 1000$; Contamination = 0.001 ; $L_w = 20$; Scaling: Standard; feature set (d)

^{3.} Neighbours = 200, $L_w = 20$; Scaling: Robust; feature set (a)

In Figure 7.10 the three algorithms are compared using confusion matrices, illustrating the number of retrieved instances in relation to the ground truth to facilitate visual interpretation of the detection performance. Figure 7.11 presents the PR curves, enabling a direct comparison of the evaluation metrics of the three classifiers as the decision function threshold varies. In Figure 7.12, the detected anomalies are plotted on the contact force time series as red dots, representing the anomaly clusters, while the positions of the ground-truth defective areas are indicated by green dashed vertical lines.

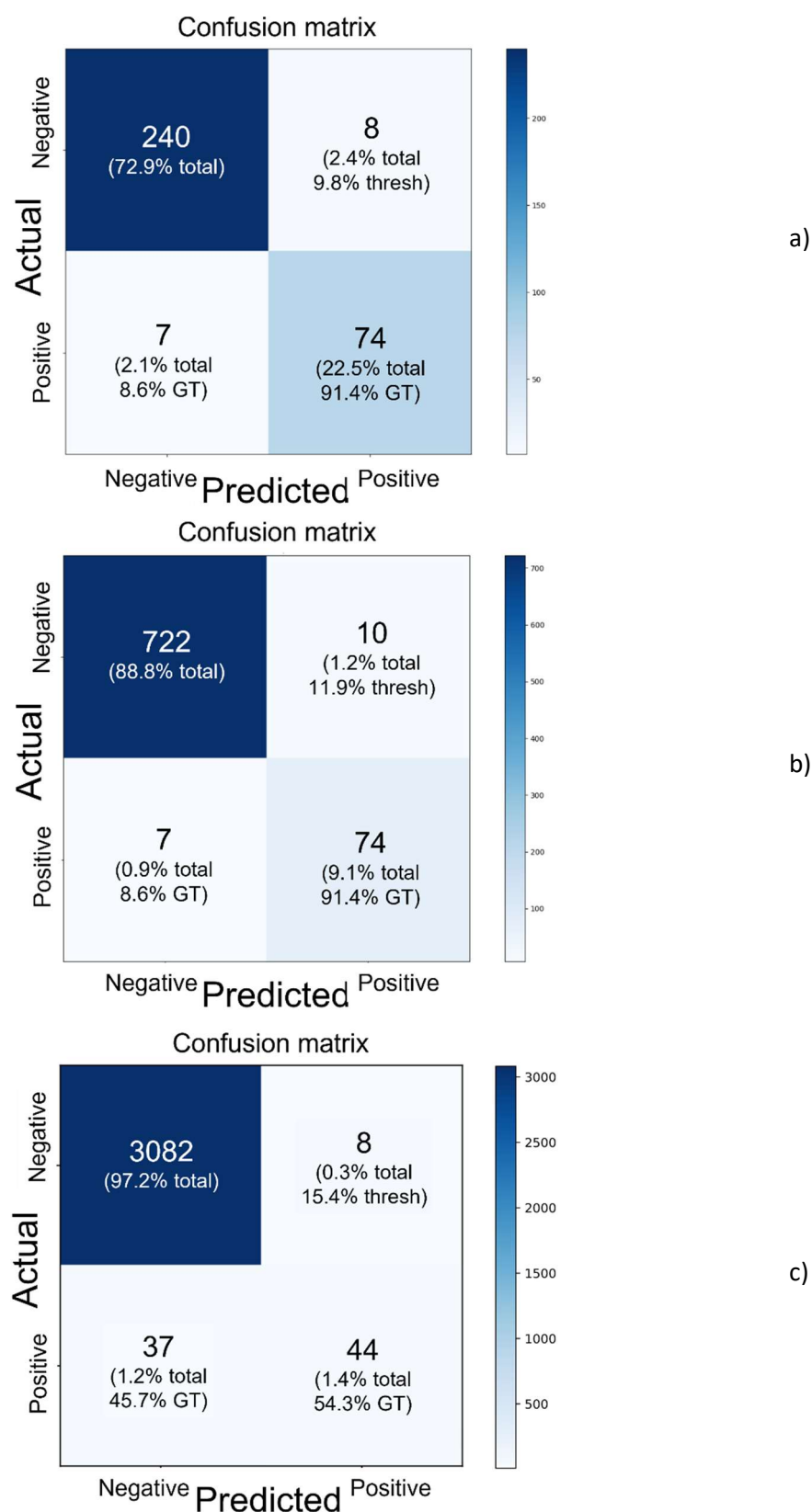
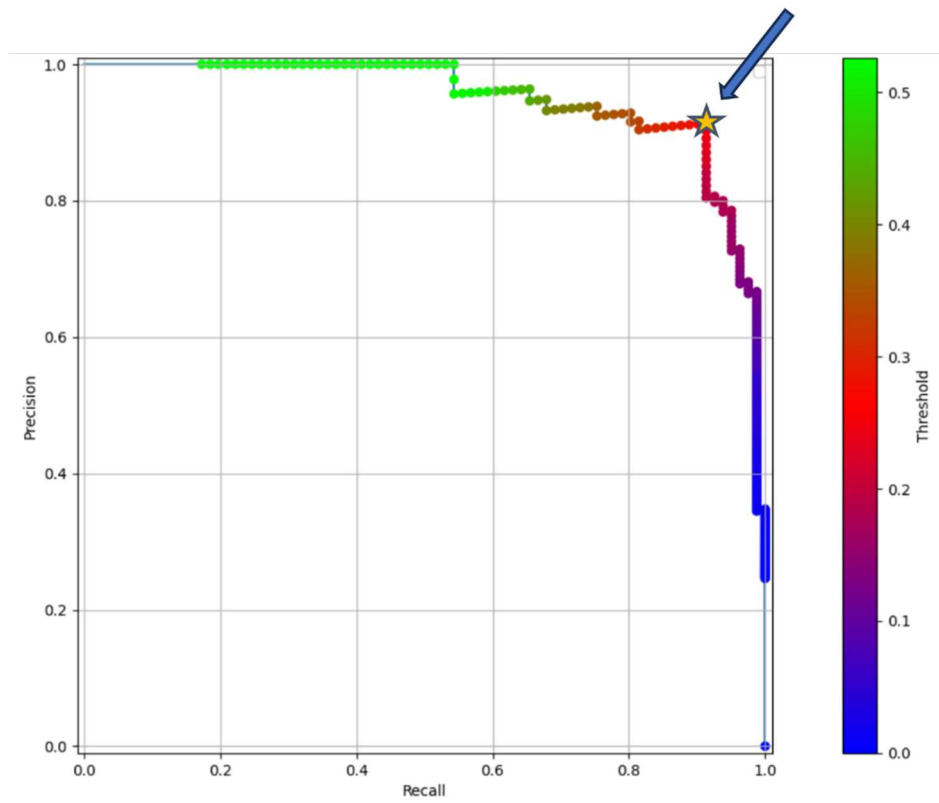
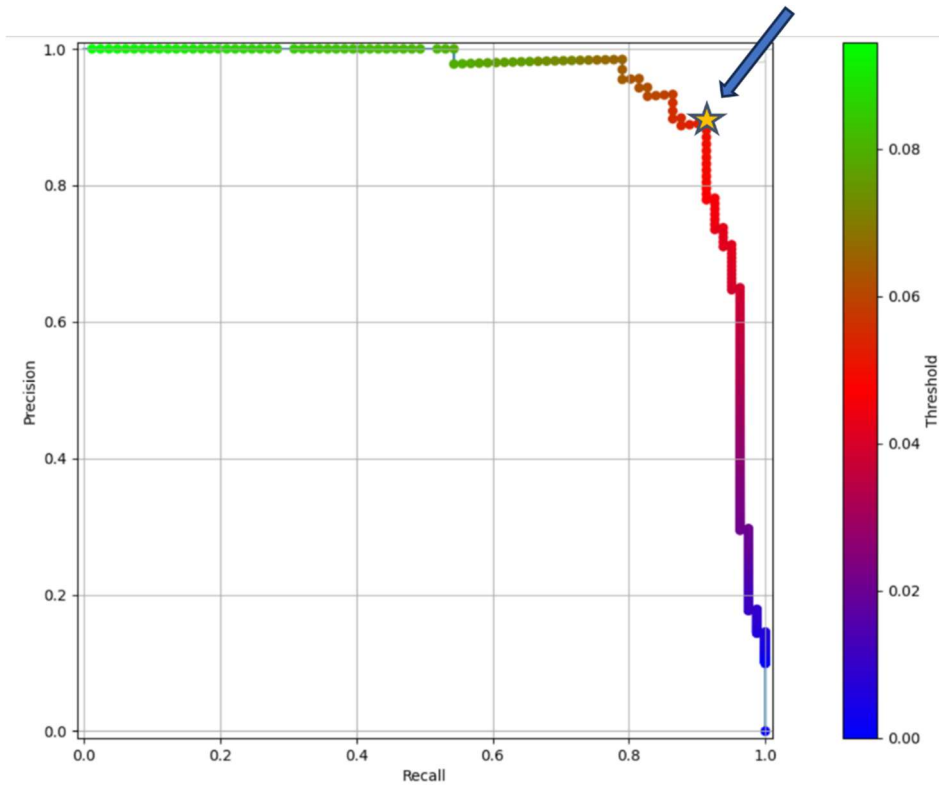


Figure 7.10: Confusion matrices of the statistical space-features method. a) OCSVM; b) IF; c) LOF classifier. Percentages are shown for total retrieved instances (“total”); percentage over the GT, and percentage over the retrieved instances after applying the optimal threshold for the FP (“thresh”).



a)



b)

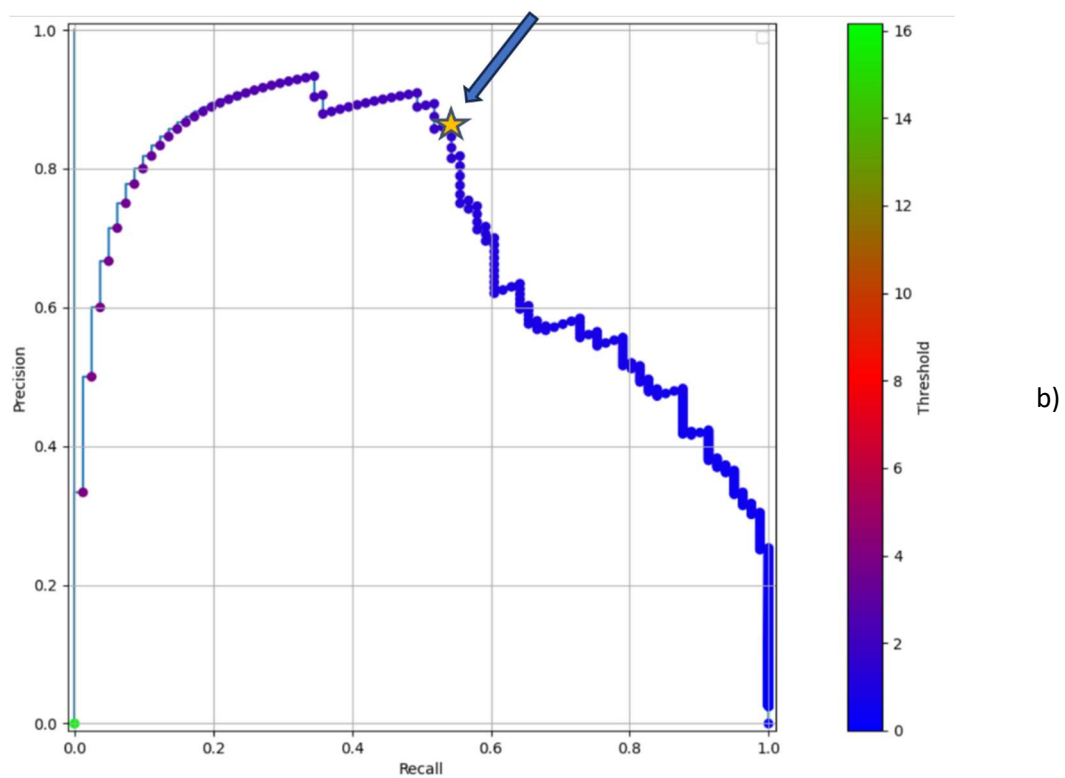
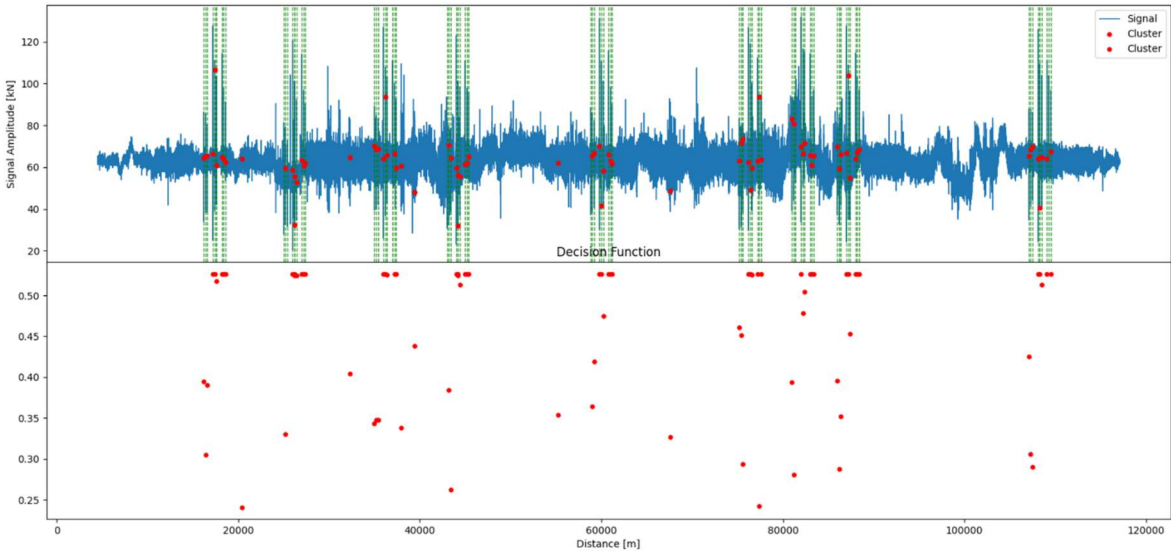
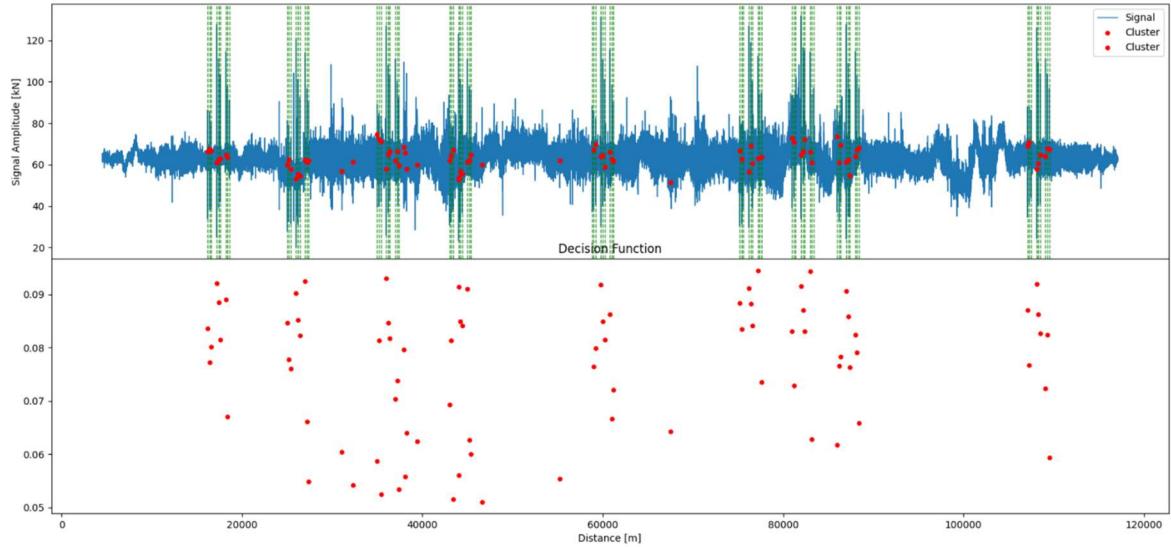


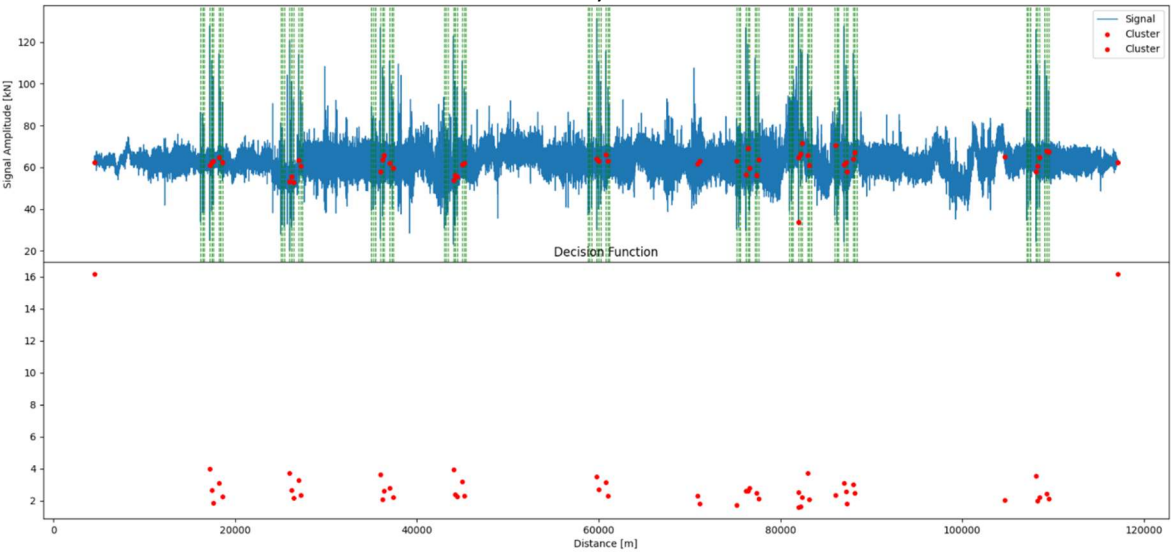
Figure 7.11: PR curves of the statistical space-features method. a) OCSVM; b) IF; c) LOF classifier.



a)



b)



c)

Figure 7.12: Classifiers outliers detection output: a) OCSVM (threshold: 0.2408881); b) IF (threshold: 0.0510494); c) LOF (threshold: 1.6000777).

Overall, OCSVM outperformed LOF, and showed a slight advantage over IF in terms of a reduced number of retrieved false positives, albeit with slightly longer training and fitting times. Compared with LOF, the advantages of OCSVM were substantial, both in detection performance and computational efficiency. Memory usage was also notably reduced. In particular, since the LOF classifier was implemented with the option “*novelty = True*”, the training dataset was not retained in memory, further decreasing resource requirements.

Considering the high percentage of true positives identified by the OCSVM model (91.4% of all instances) and the relatively low false positive rate (9.76%) and highlighted in Figure 7.13, it can be concluded that the adopted anomaly-detection strategy provides a robust foundation for developing an effective rail-head defect identification system. Such an approach shows high potential for field deployment. If validated using real track faults, and its effectiveness is confirmed beyond the use of arbitrarily modelled defects, it could lead to a significant enhancement of infrastructure maintenance strategies. In particular, it would improve the accuracy of track-faults geolocation by diagnostic trains along the network, enabling more targeted interventions while reducing logistical complexity and overall maintenance costs.

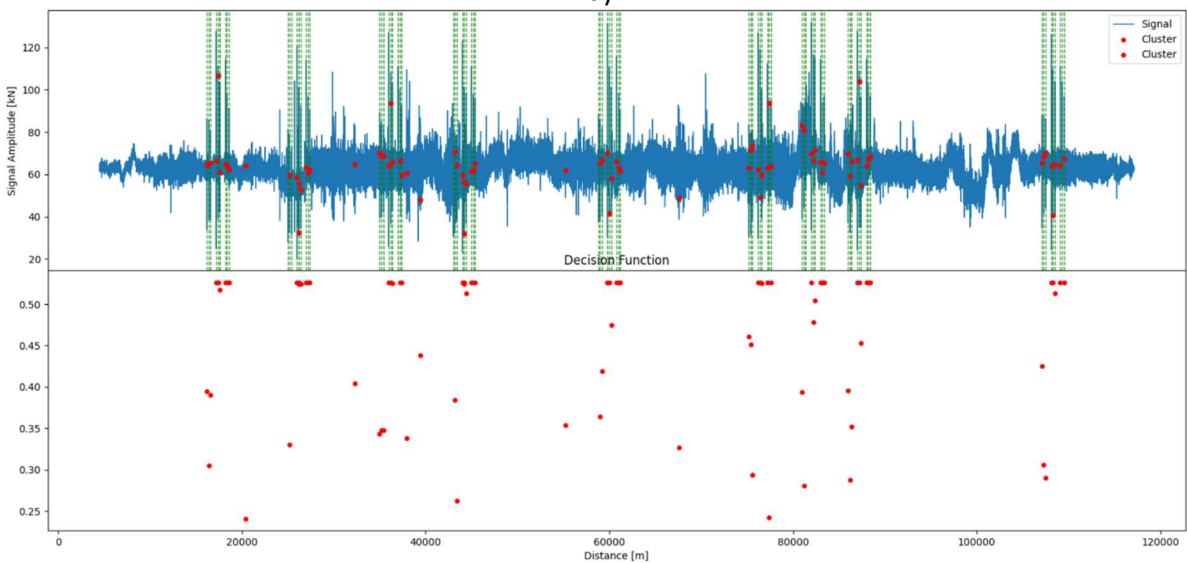
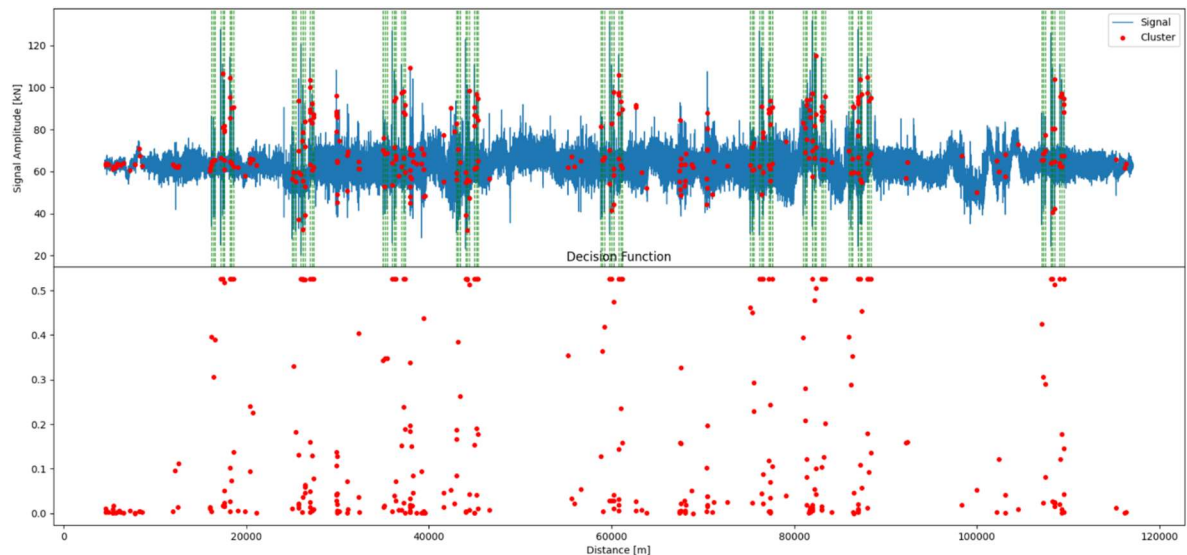


Figure 7.13: OCSVM detection output: a) default threshold (> 0); b) optimal threshold ($tr = 0.2408881$).

7.7 Graphical User Interface (GUI)

For the analysis, a graphical user interface (GUI) was developed using Python. The interface is shown in Figure 7.14, and available online in a GitHub repository.

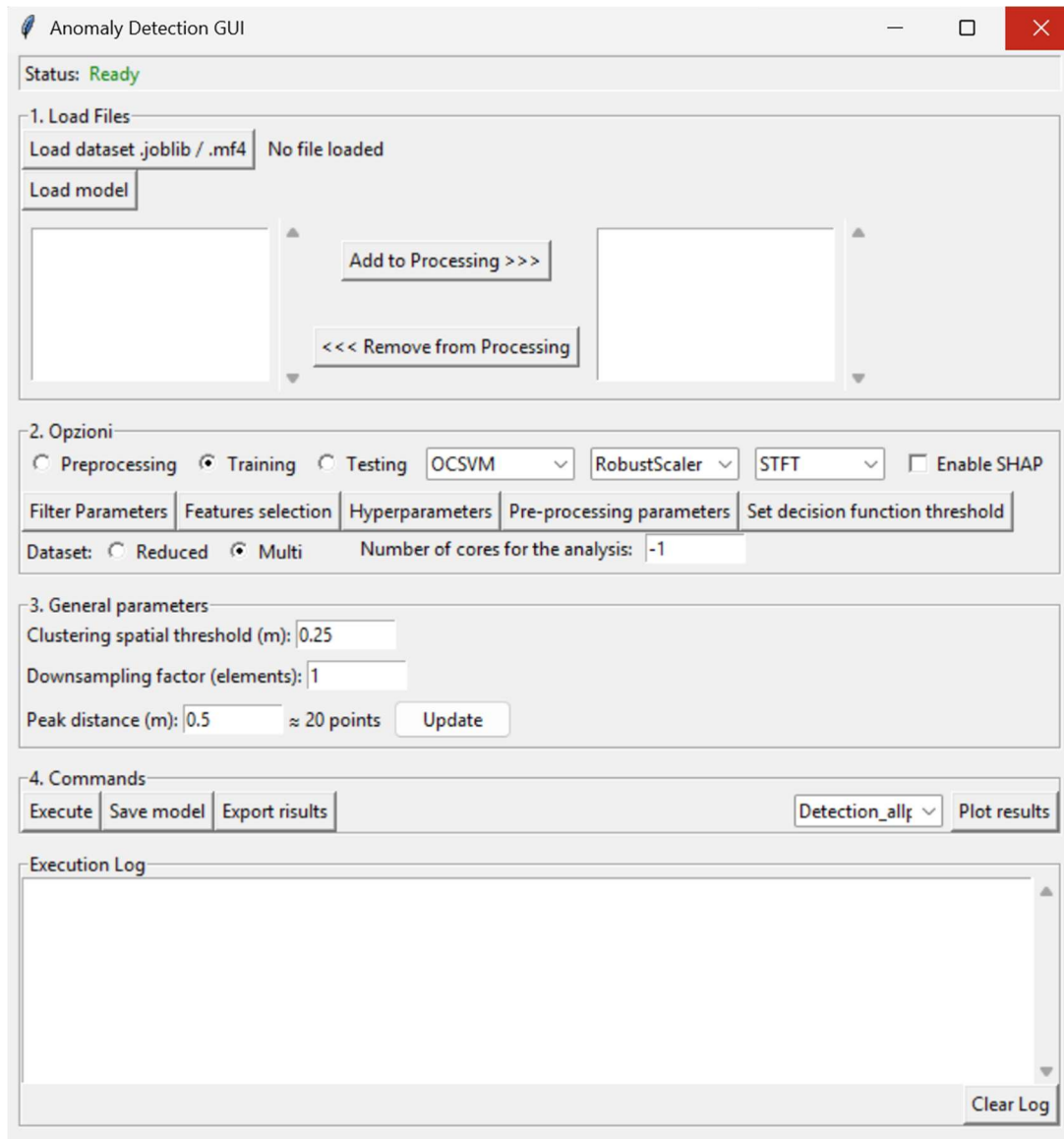


Figure 7.14: Anomaly detection GUI.

In the GUI, datasets and trained classifiers can be loaded through the “Load files” frame. Signals from the selected dataset can be chosen for inclusion in, or exclusion from, the pre-processing phase.

Within the “Options” frame, all aspects of the analysis can be configured, and three main modes are available for executing the pipeline. When the “Pre-processing” option is selected, only feature extraction is performed. This function is currently unused; its intended purpose is to allow users to plot and visually inspect feature time series against each other, or in comparison with raw force or track-geometry data. Selecting the “Training” option triggers feature extraction and

compilation of the feature matrix, which is subsequently provided to the classifier framework for model training. Once training is complete, time-performance metrics and a list of executed operations are displayed in the execution log. If the “Testing” option is selected, a pre-trained classifier can be loaded and applied to the appropriate dataset to run the novelty-detection pipeline. A trained model must be loaded to execute this process, and it is essential that all parameters used during the training phase (features, model hyperparameters, filter settings, and signal-windowing parameters) are kept identical during the fitting phase to ensure consistency of results. Upon execution, the pipeline prints a series of outputs directly in the execution log, including time-performance indicators, precision, recall, and F1-score matrices, as well as the number of retrieved instances. Using the “Commands” frame, several plots can be generated by selecting the desired option from the available list, including:

- detection results with a user-defined decision-function threshold,
- detection with the default threshold value,
- precision–recall (PR) curve,
- short-time Fourier transform (STFT) spectrogram, and
- receiver operating characteristic (ROC-AUC) curve.

Within the “Options” frame, classifiers can be selected together with the preferred data-scaling strategy. Time-frequency (or more precisely, space-wavenumber) analysis can also be enabled if frequency-domain features are used. The type of dataset can be chosen according to the analysis to be performed. The “Reduced” dataset option enables a shortened dataset and its corresponding ground-truth array, allowing rapid debugging and faster trial analyses. Conversely, the “Multi” option enables the ground truth associated with all 81 injected anomalies used for validating the classification pipeline. The number of CPU cores allocated for parallel computing is a key factor in reducing feature-computation time, particularly when the Numba library cannot be employed due to compatibility issues with other libraries (e.g., SciPy, for FFT calculations).

A series of buttons open dedicated windows in which the parameters required for analysis can be modified according to user preferences. Examples of such pop-up windows are shown in Figure 7.15. The “General parameters” frame was used in an earlier (beta) version of the GUI; it currently serves only to define the distance threshold within which an anomaly is considered sufficiently close to a ground-truth instance to qualify as a true positive (TP). At this stage of development, the other selections in this frame are not required.

Finally, in the “Commands” frame, in addition to the plot options already mentioned, three buttons allow the user to:

- execute the pipeline using the current settings,
- save the trained model (if the training mode is selected), and
- export the results.

The execution-log frame provides system messages such as errors, warnings, and results summaries printed at the end of the pipeline execution. The status bar at the top of the GUI indicates whether the system is currently processing data or available for a new analysis.

Parametri Filtro Butterworth

Sampling: 40

Ordine: 2

Cut-off min: 0.1

Cut-off max: 15.625

Tipo: bandpass

Save

a)

Select Features to Compute

Time/Spatial Features

☐ std_dev

☒ varx

☒ rms_mov

☒ f_peak

☒ f_skewness

☒ f_kurtosis

☒ f_impulse

☒ f_shape

☒ f_margin

☒ f_crest

☐ ord_mom5

☐ ord_mom6

☐ cl_f

☐ complexity

☐ mobility

Time/Space-Frequency Features

☐ s_centroid

☐ s_spread

☐ s_skewness

☐ s_kurtosis

☐ s_flat

☐ s_crestf

☐ s_rolloff_freq

☐ s_slope

☐ centroid1

☐ centroid2

☐ centroid3

☐ band_en1

☐ band_en2

☐ band_en3

☐ s_entropy

Save

b)

Iperparametri - OCSVM

nu: 1e-5

gamma: 0.01

kernel: rbf

Save

c)

Parametri di preprocessing

Sampling (samples/cycle): 40.0

Cutoff min [1/m]: 0.1

Cutoff max [1/m]: 15.625

FFT window size [elements]: 200.0

Desired Resolution [m]: 0.25

Rolloff percentage: 0.85

Save

d)



Figure 7.15: Anomaly detection GUI: “Options windows” for analysis parameters selection. a) filter parameters; b) feature selection; c) Model hyperparameters; d) pre-processing parameters; e) decision function threshold.

Since the graphical interface is still under development, some additional features have been included with the aim of performing a more comprehensive and exhaustive analysis of the ML methods. As shown in Figure 7.15(e), one of these features is labelled “LSTM”, referring to Long Short-Term Memory networks, i.e. autoencoders. Although these were preliminarily tested during the analysis, they were only evaluated in the final phase of the research, and the level of confidence and understanding achieved was not sufficient to justify their inclusion in the present work. Nevertheless, preliminary results were obtained, although they remain inconclusive. The implementation of LSTM-based architectures, and more in general auto encoders (AE), such as Convolutional Auto Encoders (CAE) and Convolutional Variational Auto Encoders (CVAE), was preliminarily tested, but they will be pursued in future developments of this research, particularly for studies focused on applying deep learning techniques to the analysis of contact-force spectrograms and the subsequent identification of anomalies based on the frequency-domain characteristics of the available data.

Some parts of the Python code were drafted or refined with the assistance of ChatGPT (GPT-5, OpenAI). The author verified and validated all outputs independently.

7.8 Conclusions

In this chapter, a data-driven methodology for rail-head fault detection was developed and evaluated. The approach relies on anomaly detection frameworks using spatial-domain statistical features as input, and it was tested across a full-scale dataset employing multiple ML algorithms and classifiers, namely OCSVM, IF, and LOF.

The performance of these one-class classifiers was compared in terms of detection effectiveness, measured by the number of TP, FP, and FN, as well as by conventional metrics such as Precision, Recall, and F1-score. To facilitate the analysis and visualization of results, a dedicated Python-based GUI was developed.

It was observed that, when statistical space-domain features (as described in Table 7.3), are employed, shorter signal windows yield better detection results. The optimal performance, corresponding to the highest number of TPs, the lowest number of FPs, and the best overall metrics, was achieved with a window length of 20 elements, closely followed by the 50-element configuration. This trend was consistently observed across all three classifiers.

Several combinations of features were also tested to assess their influence on detection performance. In particular, higher-order statistical moments (fifth and sixth order) and Hjorth parameters (Mobility and Complexity), which are generally avoided in rotating machinery

condition monitoring due to their susceptibility to noise, were examined. When the signals were high-pass filtered to remove low-frequency trends and slow-varying components, these features exhibited a certain degree of correlation with the occurrence of impulsive transients in both the original and fault-injected signals.

For OCSVM, the inclusion of fifth- and sixth-order statistical moments significantly improved detection accuracy, with all other parameters held constant. Conversely, the Hjorth parameters proved highly effective in the case of IF, resulting in a substantial reduction of FP and an increase in TP.

By comparing the classifiers under their respective optimal configurations, it was determined that OCSVM, with parameters $\nu = 10^{-5}$, $\gamma = 0.01$, RBF kernel, $Lw = 20$ elements, and feature combination (c) (see Table 7.7), achieved the best detection performance. It outperformed LOF and showed superior accuracy compared to IF. On the other hand, IF demonstrated the fastest training and inference times among the three classifiers. Regarding LOF, the excessively long training and fitting times, the large memory requirements, and the consistently lower detection performance compared with OCSVM and IF make this algorithm unsuitable for the purpose of rail-head fault detection.

Although a complete characterisation of the railway section, including the documented presence of real defects, was not available in this study, the proposed novelty detection approach based on OCSVM and IF exhibited promising potential for field deployment. In the present analysis, the ground truth was generated through artificial injection of numerical defects into the measured signals. Therefore, a full-scale experimental validation should include a reference dataset containing real defects, supported by a comprehensive characterisation of the investigated track section.

Abbreviations

The following abbreviations are used in this chapter:

AE	Auto Encoders
AUC	Area Under Curve
BP	Band-Pass
CAE	Convolutional Auto Encoders
CVAE	Convolutional Variational Auto Encoders
FFT	Fast Fourier Transform
FN	False Negatives
FP	False Positives
I/O	Input over Output
IF	Isolation Forest
LOF	Local Outlier Factor
LSTM	Long Short-Term Memory
OCSVM	One-Class Support Vector Machine
PR	Precision-Recall
ROC	Receiver Operating Characteristic

RMS	Root Mean Square
S&C	Switching and Crossings
STFT	Short Time Fourier Transform
S/N	Signal to Noise
TCA	Track Condition Analyser
TN	True Negatives
TP	True Positives

7.9 References

1. Tsunashima, H. (2019). Condition Monitoring of Railway Tracks from Car-Body Vibration Using a Machine Learning Technique. *Applied Sciences*, 9(13), 2734. <https://doi.org/10.3390/app9132734>.
2. Tsunashima, H., Ono, H., Takata, T., & Ogata, S. (2023). Development and Operation of Track Condition Monitoring System Using In-Service Train. *Applied Sciences*, 13(6), 3835. <https://doi.org/10.3390/app13063835>.
3. Gullers, P., Dreik, P., & Nielsen, J. C. O. (2011). Track condition analyser: Identification of rail rolling surface defects likely to generate fatigue damage in wheels using instrumented wheelset measurements. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 225(1), 1–13. <https://doi.org/10.1243/09544097JRRT367>.
4. Gullers, P., Andersson, L., & Lundén, R. (2008). High-frequency vertical wheel–rail contact forces—Field measurements and influence of track irregularities. *Wear*, 265(9-10), 1472-1478. <https://doi.org/10.1016/j.wear.2008.02.035>
5. La Paglia I, Di Gialleonardo E, Facchinetti A, Carnevale M, Corradi R. Acceleration-based condition monitoring of track longitudinal level using multiple regression models. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*. 2023;238(5):479-488. doi:10.1177/09544097231201513.
6. La Paglia, I., Carnevale, M., Corradi, R., Di Gialleonardo, E., Facchinetti, A., & Lisi, S. (2023). Condition monitoring of vertical track alignment by bogie acceleration measurements on commercial high-speed vehicles. *Mechanical Systems and Signal Processing*, 186, 109869. <https://doi.org/10.1016/j.ymssp.2022.109869>.
7. De Rosa, A., Kulkarni, R., Qazizadeh, A., Berg, M., Di Gialleonardo, E., Facchinetti, A., & Bruni, S. (2021). Monitoring of lateral and cross level track geometry irregularities through onboard vehicle dynamics measurements using machine learning classification algorithms. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 235(1), 107-120. <https://doi.org/10.1177/0954409720906649>.
8. Tsunashima, H., & Yagura, N. (2024). Railway Track Irregularity Estimation Using Car Body Vibration: A Data-Driven Approach for Regional Railway. *Vibration*, 7(4), 928-948. <https://doi.org/10.3390/vibration7040049>.
9. Malekjafarian, A., Sarrazebolles, C.-A., Khan, M. A., & Golpayegani, F. (2023). A machine-learning-based approach for railway track monitoring using acceleration measured on an in-service train. *Sensors*, 23(17), 7568. <https://doi.org/10.3390/s23177568>.
10. Ghiasi, R., & Malekjafarian, A. (2025). Monitoring of railway tracks maintenance needs using dynamic responses collected by an in-service train. *Railway Engineering Science*. <https://doi.org/10.1007/s40534-025-00380-w>.
11. Ghiasi, R., & Malekjafarian, A. (2025). A data-driven framework for detecting railway maintenance needs using acceleration data collected from passing trains. In C. McNally, P. Carroll, B. Martinez-Pastor, B. Ghosh, M. Efthymiou, & N. Valantasis-Kanellos (Eds.), *Transport transitions: Advancing sustainable and inclusive mobility (TRAconference 2024, Lecture Notes in Mobility)*. Springer, Cham. https://doi.org/10.1007/978-3-032-04774-8_129.
12. Ghiasi, R., Arslan, K. M., Sorrentino, D., Diaine, C., & Malekjafarian, A. (2024). An unsupervised anomaly detection framework for onboard monitoring of railway track geometrical defects using one-class support vector machine. *Engineering Applications of Artificial Intelligence*, 133, Part B. <https://doi.org/10.1016/j.engappai.2024.108167>.

13. Rosano, G., Massini, D., Bocciolini, L., Zappacosta, C., Di Gialleonardo, E., Somaschini, C., et al. (2024). Diagnostics of the railway track - Possibility of development through the measurement of accelerations and contact forces [JB]. *Ingegneria Ferroviaria*, 81–102.
14. Zhu, T., Wang, X., Wu, J., Zhang, J., Xiao, S., Lu, L., Yang, B., & Yang, G. (2025). Comprehensive identification of wheel–rail forces for rail vehicles based on the time domain and machine learning methods. *Mechanical Systems and Signal Processing*, 222, 110667. <https://doi.org/10.1016/j.ymssp.2025.110667>.
15. Walther, S., Müller, S., Renggli, R., Ünlü, F., & Fuerst, A. (2022). Prediction of wheel–rail contact forces using simple onboard monitoring system and machine learning. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 237(5), 553–562. <https://doi.org/10.1177/09544097221118872>.
16. Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PLOS ONE*, 11(4), e0152173. <https://doi.org/10.1371/journal.pone.0152173>.
17. Moya, M. M., & Hush, D. R. (1996). Network constraints and multi-objective optimization for one-class classification. *Neural Networks*, 9(3), 463–474. [https://doi.org/10.1016/0893-6080\(95\)00120-4](https://doi.org/10.1016/0893-6080(95)00120-4).
18. Esveld, C. (2017). *Modern railway track* (3rd ed.). MRT Productions.
19. Iwnicki, S., Spiryagin, M., Cole, C., & McSweeney, T. (Eds.). (2019). *Handbook of railway vehicle dynamics* (2nd ed.). CRC Press. <https://doi.org/10.1201/9780429465089>.
20. Zhao, X., Li, Z., & Dollevoet, R. (2013). The vertical and the longitudinal dynamic responses of the vehicle–track system to squat-type short wavelength irregularity. *Vehicle System Dynamics*, 51(12), 1918–1937. <https://doi.org/10.1080/00423114.2013.844315>.
21. Dukkipati, R. V., & Dong, R. (1999). The dynamic effects of conventional freight car running over a dipped joint. *Vehicle System Dynamics*, 31(2), 95–111. <https://doi.org/10.1076/vesd.31.2.95.2081>.
22. Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7), 1443–1471. <https://doi.org/10.1162/089976601750264965>.
23. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining (ICDM)* (pp. 413–422). IEEE. <https://doi.org/10.1109/ICDM.2008.17>.
24. Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data* (pp. 93–104). ACM. <https://doi.org/10.1145/342009.335388>.
25. Scikit-learn developers. (2025, October 25). `sklearn.svm.OneClassSVM`. Scikit-learn documentation. <https://scikit-learn.org/stable/modules/generated/sklearn.svm.OneClassSVM.html>.
26. Scikit-learn developers. (2025, October 25). `sklearn.ensemble.IsolationForest`. Scikit-learn documentation. <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html>.
27. Scikit-learn developers. (2025, October 25). `sklearn.neighbors.LocalOutlierFactor`. Scikit-learn documentation. <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.LocalOutlierFactor.html>.

28. Caesarendra, W., & Tjahjowidodo, T. (2017). A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing. *Machines*, 5(4), 21. <https://doi.org/10.3390/machines5040021>
29. Lee, J., & Liao, L. (2009). Methods for prognosing mechanical systems (U.S. Patent No. US 8,301,406 B2). U.S. Patent and Trademark Office. <https://patents.google.com/patent/US8301406B2/en>.
30. Shrivastava, A., & Wadhwani, S. (2013, May). Development of fault detection system for ball bearing of induction motor using vibration signal. *International Journal of Scientific Research*, 2(5). Retrieved from http://theglobaljournals.com/ijsr/file.php?val=May_2013_1367588548_a9f79_85.pdf.
31. Sreejith, B., Verma, A. K., & Srividya, A. (2008, August). Fault diagnosis of rolling element bearing using time-domain features and neural networks. In *Proceedings of the IEEE Region 10 and the Third International Conference on Industrial and Information Systems (ICIIS 2008)* pp. 1-6. IEEE. <https://doi.org/10.1109/ICIINF5.2008.4798444>.
32. Caesarendra, W., Tjahjowidodo, T., Kosasih, B., & Tieu, A. K. (2017). Integrated Condition Monitoring and Prognosis Method for Incipient Defect Detection and Remaining Life Prediction of Low Speed Slew Bearings. *Machines*, 5(2), 11. <https://doi.org/10.3390/machines5020011>.
33. Hjorth, B. (1970). EEG analysis based on time-domain properties. *Electroencephalography and Clinical Neurophysiology*, 29(3), 306-310. [https://doi.org/10.1016/0013-4694\(70\)90143-4](https://doi.org/10.1016/0013-4694(70)90143-4).

8 DATA-DRIVEN METHODS: NOVELTY DETECTION USING CONTACT FORCE SPECTRA FEATURES AND MACHINE LEARNING

8.1 Introduction

As discussed in Chapter 7, anomaly detection algorithms have gained increasing attention within the scientific community due to their effectiveness and reliability in fault detection across a wide range of domains and applications. The railway sector is no exception: anomaly detection frameworks are now widely applied for monitoring vehicle and infrastructure conditions and for detecting faults in subsystems such as vehicle components, bridges, catenaries, and tracks.

In train and track condition monitoring, accelerometers and rate gyros are commonly used to analyse motion and estimate displacements for vibration analysis. These measured or estimated quantities provide valuable diagnostic information, supporting maintenance planning, asset management, and operational reliability.

At present, most monitoring data are collected using Track Recording Vehicles (TRVs), which are diagnostic trains equipped with an extensive suite of sensors designed for infrastructure monitoring, covering rails, bridges, catenaries, and data transmission systems.

Some studies have explored the use of in-service vehicles for monitoring purposes. While this approach is feasible with well-established technologies such as accelerometers and gyroscopes, it is less suitable for testing new instrumentation whose performance has not yet been fully validated, such as the system investigated in this research.

Following the structure of this doctoral work, this study focused on the use of instrumented wheelsets equipped with strain gauges for track condition monitoring. These systems measure the contact forces at the wheel-rail interface. Although this technology is well established for applications such as homologation, vehicle and component certification [1], derailment risk assessment, and static load measurements [2, 3], it has not been extensively used for continuous

condition monitoring, as already noted in the previous chapter. Unlike accelerations, which can be directly measured on components such as axle-boxes, bogies, and the carbody, contact forces at the wheel–rail interface must be estimated through inverse identification, following a procedure similar to that described in Chapter 3. Strain gauges are arranged in a Wheatstone bridge configuration and mounted on the wheel web and axle to enable this estimation. Furthermore, the interaction between vertical and lateral force components introduces cross-sensitivity, which complicates direct measurements; this issue is typically mitigated through a calibration matrix [4].

Despite these challenges, contact forces are highly sensitive to short-wavelength railhead faults such as squats, dipped joints, peaked or dipped welds, soft spots, and hollows. This sensitivity makes instrumented wheelsets a promising tool for gaining deeper insights into railhead condition.

The large volume of data recorded during a single run poses significant challenges for traditional analysis methods such as purely statistical or time–frequency analyses. Although time-frequency techniques (e.g., [5–7]) can identify the position of anomalous force peaks, distinguishing between normal structural features (such as joints or frogs in switch and crossing areas) and genuine defects (such as worn or damaged components approaching the end of their service life) remains challenging. Early detection of critical faults is essential to prevent hazardous failures, including derailments. Ensuring the safety of railway operations—whether for passenger or freight transport, and on both conventional and high-speed lines—therefore remains a fundamental objective of condition monitoring.

8.2 State of the art and literature review

Only a limited number of studies in the literature have investigated the use of contact force measurements for condition monitoring. As mentioned earlier, most research focuses on acceleration-based measurements from axle-boxes (ABA), bogies, and carbodies. Methods for monitoring the track can broadly be classified into three categories: time–frequency methods, model-based methods, and data-driven approaches.

During this thesis, the first two approaches were explored and are extensively described in Chapters 5 and 6. Model-based methods rely on mathematical formulations describing the relationships among system components, making use of established physical knowledge of the phenomena under investigation. These models, whether simplified or highly detailed, are always approximations of the real system and are typically solved through numerical simulations.

Their complexity ranges from simplified systems with a reduced number of degrees of freedom [8] to sophisticated multibody and finite element (FE) models. Developing such models requires in-depth knowledge of the mechanical properties and dynamic behaviour of both the vehicle and the infrastructure to ensure accurate system representation. This task is inherently challenging, as it demands detailed information about the physical characteristics of all involved components. On the one hand, traditional data analysis techniques, such as statistical analysis and frequency or time, frequency analysis [1, 6, 7], can provide valuable insights into system behaviour but are limited in their ability to capture complex fault signatures. On the other hand, data-driven methods offer an alternative approach by exploiting mathematical relationships within the recorded data itself, uncovering hidden patterns and correlations without the need for detailed

physical models. These methods typically require large datasets to train models capable of generalising to unseen data effectively.

In this chapter, we focus again on data-driven ML methods for analysing contact force data to detect faulty track sections. In the context of unsupervised and semi-supervised anomaly or novelty detection, data-driven approaches are particularly suitable for handling large unlabelled datasets [1, 9-11]. For these learning paradigms [12] the training datasets should contain mostly (or exclusively) normal data, with very few or no anomalies. During the testing phase, new unseen data are fed into a trained model, also known as a classifier or novelty detector, which identifies and isolates anomalous data using a decision function that assigns an anomaly score. The higher the anomaly score, the greater the likelihood that the data point corresponds to abnormal behaviour.

8.3 Time-Frequency analysis

Spectral analysis is a fundamental approach for signal investigation, with the Short-Time Fourier Transform (STFT) being widely used for analysing non-stationary signals whose frequency content changes over time. Unlike the traditional Fourier Transform, which gives only global frequency information, the STFT divides a signal into short, overlapping segments using a sliding window (e.g., Hann or Hamming). A Fourier Transform is computed for each segment, producing a time-localized frequency spectrum. Stacking these spectra generates a spectrogram, where time (or space), frequency (or wavenumber), and spectral intensity are plotted on two axes and colour, respectively. This representation allows tracking spectra evolution over time (or space), which is particularly useful in applications such as vibration analysis and condition monitoring. The STFT, however, involves a trade-off between time and frequency resolution: shorter windows improve time resolution but reduce frequency resolution, and vice versa. Despite this, its simplicity and clarity make it a key tool for analysing transient or time-varying phenomena such as contact forces. Referring to equation 7.3 (Chapter 7), the one-sided Discrete Fourier Transform (DFT) of a finite-length sequence $x_w[n]$ (where $n = 0, 1, \dots, N-1$) is given by:

$$X[k] = F_d(x_w[n]) = \sum_{n=0}^{N-1} x_w[n] e^{-\frac{j2\pi kn}{N}}, \quad \left(k = 0, \dots, \left\lfloor \frac{N}{2} \right\rfloor\right). \quad 8.1$$

Where $X[k]$ is the complex DFT coefficient at frequency bin, $x[n]$ is the input signal at time (space) index, N is the total number of samples, j is the imaginary unit ($j^2 = -1$).

Other so-called time-frequency methods involve the use of wavelets or Empirical Mode Decomposition and Hilbert transform. These methods represent an alternative to the widely known STFT. They depict the same phenomena but under different assumptions and the signal is parsed through different procedures, obtaining not only different representations, but also other new or complementary information on the signal.

We will focus mainly on the STFT method which has been extensively used for the analysis. The wavenumber (in rad/m) and spatial frequency (in cycles/m) for bin k are:

$$k_k = \frac{2\pi k}{N \Delta s}, \quad 8.2$$

$$f_k = \frac{k}{N \Delta s} \quad 8.3$$

and the corresponding wavelength is:

$$\lambda_k = \frac{1}{f_k} = \frac{N \Delta s}{k}, \quad (k \geq 1). \quad 8.4$$

The one-sided power spectrum is defined as

$$P[k] = \frac{2}{N U_w} |X[k]|^2, \quad \left(k = 1, \dots, \left\lfloor \frac{N}{2} \right\rfloor\right). \quad 8.5$$

where $U_w = \frac{1}{N} \sum_{n=0}^{N-1} w^2[n]$ accounts for the window energy.

8.3.1 Definition of wavelength bands

Following the scheme of Figure 8.1, based on typical railway irregularities, three wavelength ranges are considered, as reported on Table 8.1.

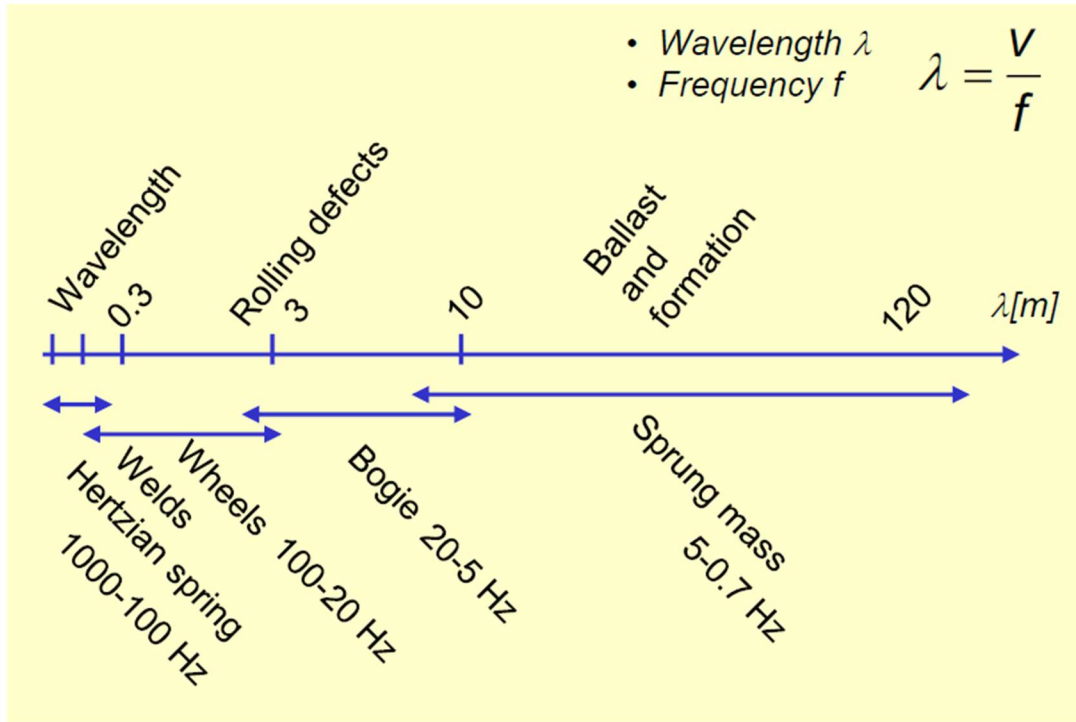


Figure 8.1: Typical wavelength and frequencies for common vehicle and track components, defects and faults [14].

Table 8.1: Wavelength bands definition.

Band	Wavelength range (m)	Frequency range (cycles/m)
Long (L)	$\lambda \geq 5$	$f \leq 0.2$
Medium (M)	$0.25 \leq \lambda < 5$	$0.2 < f \leq 4$
Short (S)	$\lambda < 0.25$	$f \geq 4$

In terms of spectral bins:

$$K_L = \{k: 0 < f_k \leq f_{L,max}\}, \quad (f_{L,max} = 0.2), \quad 8.6$$

$$K_M = \{k: f_{M,min} \leq f_k < f_{M,max}\}, \quad (f_{M,min} = 0.2, f_{M,max} = 4), \quad 8.7$$

$$K_S = \{k: f_k \geq f_{S,min}\}, \quad (f_{S,min} = 4), \quad 8.8$$

(Units of frequency are cycles/m; to express them in rad/m, multiply by 2π).

The spectral energy in each band is computed as:

$$E_B = \sum_{k \in K_B} P[k], \quad B \in \{L, M, S\}. \quad 8.9$$

The total energy (excluding DC) is:

$$E_{tot} = \sum_{k=1}^{N/2} P[k]. \quad 8.10$$

The normalized band ratios, which serve as input features, are:

$$R_B = \frac{E_B}{E_{tot} + \varepsilon'}, \quad B \in \{L, M, S\}. \quad 8.11$$

where ε is a small positive constant (e.g. 10^{-12}) to avoid division by zero.

Optionally, the log-ratio form can be used for small energy bands:

$$\tilde{R}_B = 10 \log_{10} \left(\frac{E_B}{E_{tot} - E_B + \varepsilon'} \right), \quad 8.12$$

8.4 Features extraction

Rather than analysing the entire waveform directly, each windowed segment was represented by a set of descriptive statistics.

While traditional studies have focused largely on time- or space-domain features such as RMS, crest factor, skewness and kurtosis, the present work emphasises frequency-domain features and Hjorth parameters, motivated by their ability to reveal shifts in energy distribution and irregularities associated with local defects.

For each spatial window $x[n]$ of length N , the magnitude spectrum $X[k]$ was computed using the Fast Fourier Transform (FFT). Only the positive-frequency components were retained. The features were then derived as described in Table 8.2:

Table 8.2: Extracted spectral features.

Parameter	Symbol	Formula
Spectral Centroid	C_s	$C_s = \frac{\sum_k f_k X[k] ^2}{\sum_k X[k] ^2}$
Spectral Spread	S_s	$S_s = \sqrt{\frac{\sum_k (f_k - C_s)^2 X[k] ^2}{\sum_k X[k] ^2}}$
Spectral Skewness	Sk_s	$Sk_s = \frac{\sum_k (f_k - C_s)^3 X[k] ^2}{S_s^3 \sum_k X[k] ^2}$
Spectral Kurtosis	K_s	$K_s = \frac{\sum_k (f_k - C_s)^4 X[k] ^2}{S_s^4 \sum_k X[k] ^2}$
Spectral Crest Factor	CF_s	$CF_s = \frac{\max(X[k])}{\sqrt{\frac{1}{N} \sum_k X[k] ^2}}$
Spectral Flatness	SF_s	$SF_s = \frac{(\prod_k X[k])^{1/N}}{\sqrt{\frac{1}{N} \sum_k X[k] ^2}}$
Band-Energy Ratios	E_b	$E_b = \sum_k X[k] ^2$

Spectral Centroid represents the energy-weighted mean frequency (or wavelength). A shift in C_s indicates that the dominant excitation moves towards higher or lower wavelength bands. Spectral Spread quantifies the dispersion of spectral energy around the centroid. Spectral Skewness measures the asymmetry of the spectral distribution. Changes in SK_s may indicate localised peaks linked to defect-induced resonances. Spectral Kurtosis indicates the “peakedness” of the spectrum and is sensitive to impulsive events. Spectral Crest Factor highlights the presence of narrow-band components or spikes in the spectrum. Spectral Flatness

distinguishes between tone-like (low flatness) and noise-like (high flatness) spectra. Band-Energy Ratios the integrated spectral energy in three wavelength bands - long (>5m), medium (0.25–5 m), and short (< 0.25m) - was computed and normalised by the total energy. These bands are motivated by known wavelength ranges of different rail irregularities.

8.5 Data Preparation

The data were processed using a band-pass filter within the wavenumber range [0.1 – 15.625] cycles/m. The lower cut-off was selected to ensure the interpretability of the spectrogram, following a preliminary Fourier analysis. It was observed that, below a certain threshold, it becomes difficult to effectively distinguish relevant spectral features, as illustrated in Figure 8.2. The upper cut-off was determined according to the Shannon-Nyquist sampling theorem, by applying the following relation:

$$f_{max} = \frac{f_{samp}}{2.56} \quad 8.13$$

e.g. in actual case:

$$f_{max} = \frac{40}{2.56} = 15.62 \text{ cycl} \cdot \text{m}^{-1} \quad 8.14$$

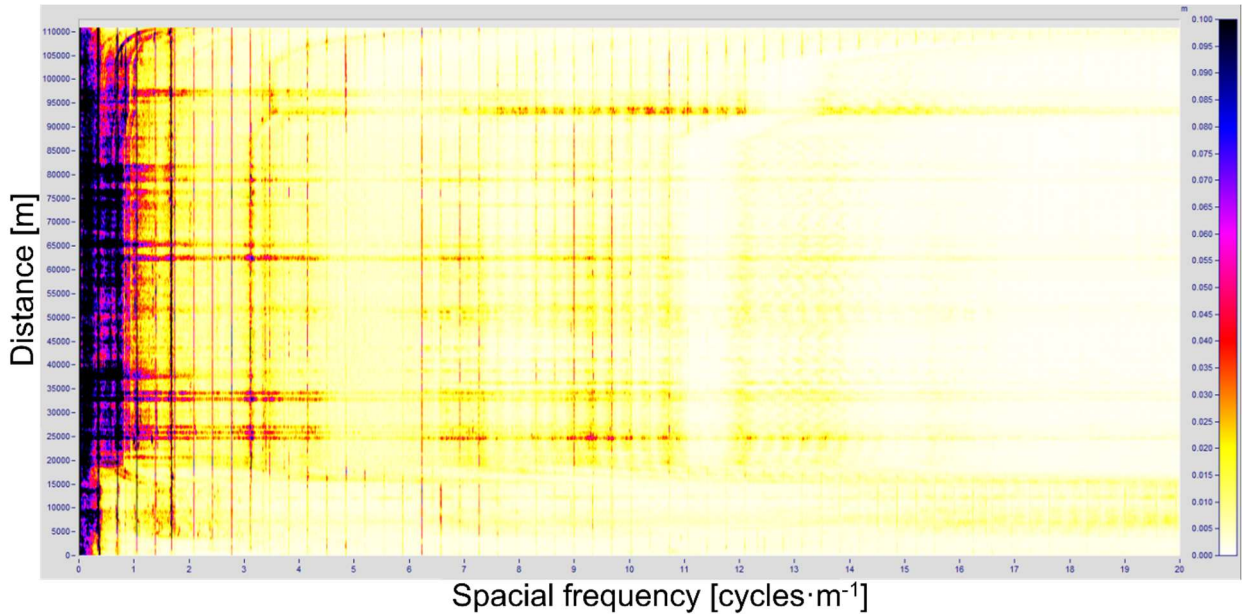


Figure 8.2: STFT of the C14 recording.

8.6 Machine learning and novelty detection pipeline

The proposed ML pipeline uses one-class classifiers trained on force-derived statistical spectral features to distinguish between normal and abnormal (anomalous) data. The models tested include the One-Class Support Vector Machine (OCSVM), Isolation Forest (IF), and Local Outlier Factor (LOF). These classifiers were selected in continuity with the previous space-features analysis for their ease of implementation, fast execution, and scalability, given the large

dataset size (over four million data points in the force array). As in the previous chapter, models were trained on normal data (C14), and tested on a modified version of the C12 dataset, which contained 81 artificially generated defects simulating literature-based rail-head faults. These defective segments served as ground truth for evaluation. These defective sections served as ground truth for evaluation. The input data consisted of spectral features extracted from the contact force STFT. For each signal window, the FFT was computed, and features such as spectral centroid, spread, skewness, kurtosis, crest factor, flatness, and roll-off (at 85% of the cumulative spectral energy) were calculated. Additional features captured the energy distribution and centroid shifts within three wavelength bands: short (0.15–1 m), medium (1–5 m), and long (5–25 m). These features were compiled into a feature matrix, which was then used as input for the classifiers. Model performance was evaluated using Precision, Recall, F1-score, and the number of true positives (TP) and false positives (FP). Computational performance, including training and inference time, was also assessed.

8.7 Results and discussion

As in the previous chapter, three classifiers (OCSVM, IF and LOF) were tested using different configurations of hyperparameters, window lengths (L_w) for features computation, feature sets, and data-scaling strategies. Since the results were consistent with those obtained in the preceding analysis, they are omitted here for brevity. Once the optimal configuration was identified, the models were compared to select the most effective approach for detecting railhead faults through anomaly detection pipelines based on statistical frequency-domain features. This approach represents an alternative to the statistical time- or space-domain features commonly adopted in rotating machinery condition monitoring and in OCSVM-based classification of axle-box acceleration (ABA) data recorded from in-service vehicles [9].

The detection results for the three classifiers are summarized in Table 8.3. Dataset C12, with the injected artificial defects, was used for classifier evaluation, enabling novelty detection. Ground-truth anomalies are shown as vertical green dashed lines for visual reference, corresponding to 81 injected anomalies (see Figure 8.3).

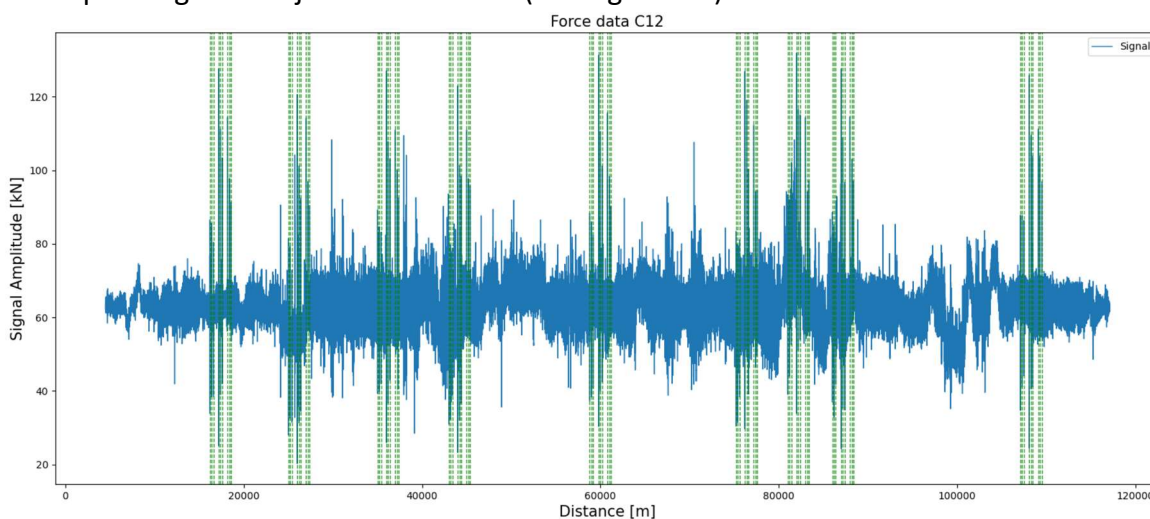


Figure 8.3: C12 contact force recording with ground truth anomalies.

As indicated in Table 8.3, IF achieved the fastest training and fitting times. In contrast, OCSVM and LOF were significantly slower, with the latter consuming large amounts of memory (several gigabytes per model), whereas OCSVM and IF were considerably lighter. Even under the tested setup, LOF exhibited poor detection performance. IF underestimated the number of positive instances, suggesting that relying solely on statistical frequency-domain features is not recommended when using this type of classifier.

OCSVM outperformed both IF and LOF, achieving the highest Precision, Recall, and F1-score. However, detection was not perfect, with two missed faults and nine false positives. Compared with earlier studies using space-domain features [9] and the tests in the previous chapter, the results suggest that relying solely on spectral features may be insufficient for accurate fault detection when using IF and LOF. Combining space-domain and time–frequency (or space–wavenumber) features could improve both detection accuracy and reliability. Although expanding the feature set would increase computational demands for feature computation, training, and fitting, these drawbacks could be offset by higher F1-scores and a more robust fault detection system, ultimately providing maintenance teams with a more trustworthy diagnostic tool.

Table 8.3: Classifiers comparison and detection performance summary.

Classifier	Retrieved instances	Optimal F1-Score	Optimal Precision	Optimal Recall	Time Performance
OCSVM ¹ (Threshold 0.0457308)	TP: 79/81 FP: 9 FN: 2 Max TP: 81	0.940	0.898	0.975	Training: 877.15 s Fitting: 1145.77 s
IF ² (Threshold 0.0382393)	TP: 51/81 FP: 8 FN: 30 Max TP: 81	0.734	0.864	0.630	Training: 127.42 s Fitting: 194.58 s
LOF ³ (Threshold 0.6029069)	TP: 59/81 FP: 30 FN: 22 Max TP: 81	0.698	0.663	0.728	Training: 5835.86 s Fitting: 8907.15 s

Classifiers parameters:

^{1.} $\nu = 10^{-5}$; $\gamma = 0.01$; kernel: RBF; Lw = 50; Scaling: Robust.

^{2.} $n_estim = 1000$; Contamination = 0.001; Lw = 50; Scaling: Standard.

^{3.} Neighbours = 200 Lw = 50; Scaling: Standard.

In Figure 8.4 the three algorithms are compared using confusion matrices, illustrating the number of retrieved instances in relation to the ground truth to facilitate visual interpretation of the detection performance. Figure 8.5 presents the PR curves, enabling a direct comparison of the evaluation metrics of the three classifiers as the decision function threshold varies. In Figure 8.6, the detected anomalies are plotted on the contact force time series as red dots, representing the anomaly clusters, while the positions of the ground-truth defective areas are indicated by green dashed vertical lines. Figure 8.6 shows the outliers detected by the three algorithms along the entire dataset when the optimal threshold was applied (value that maximises the F1-score).

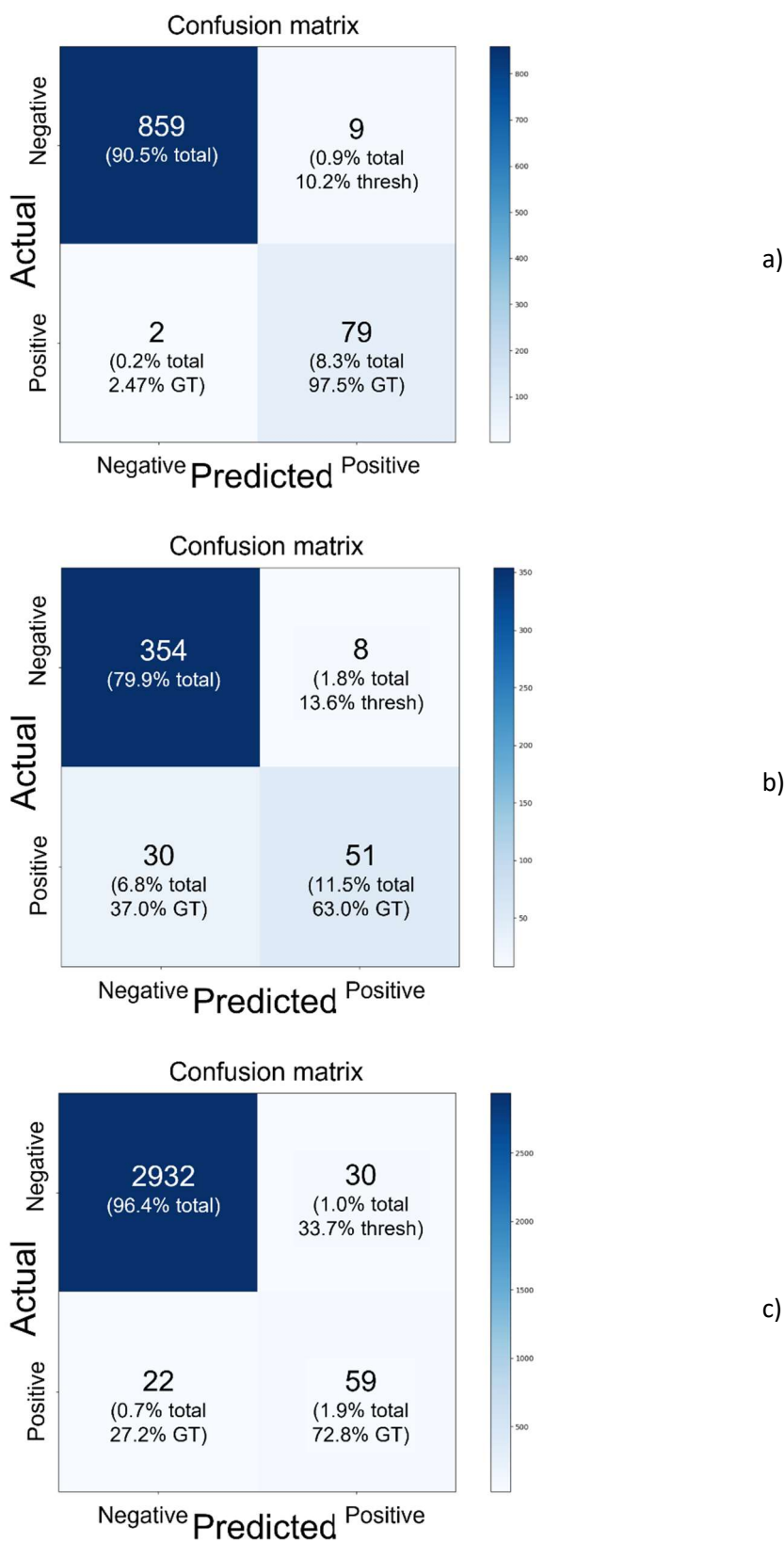


Figure 8.4: Confusion matrices of the statistical frequency-features method. a) OCSVM; b) IF; c) LOF classifier. Percentages are shown for total retrieved instances (“total”); percentage over the GT, and percentage over the retrieved instances after applying the optimal threshold for the FP (“thresh”).

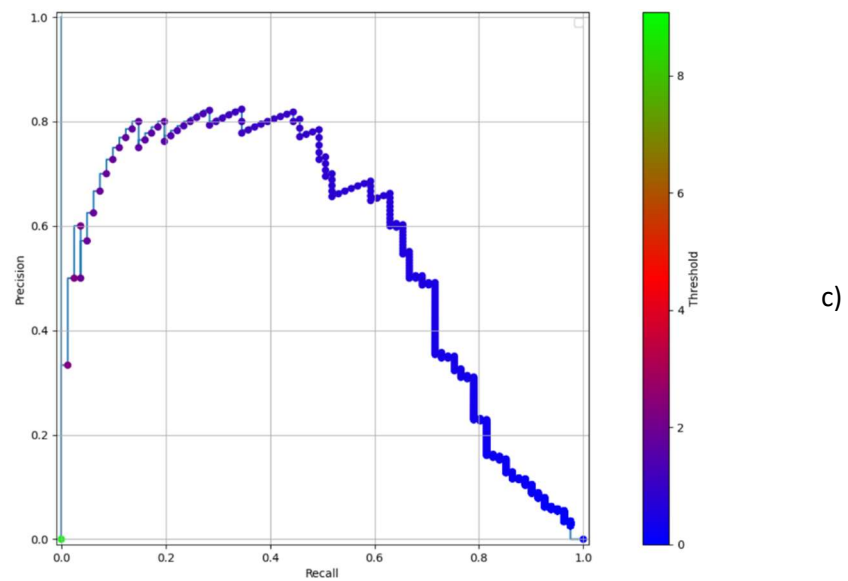
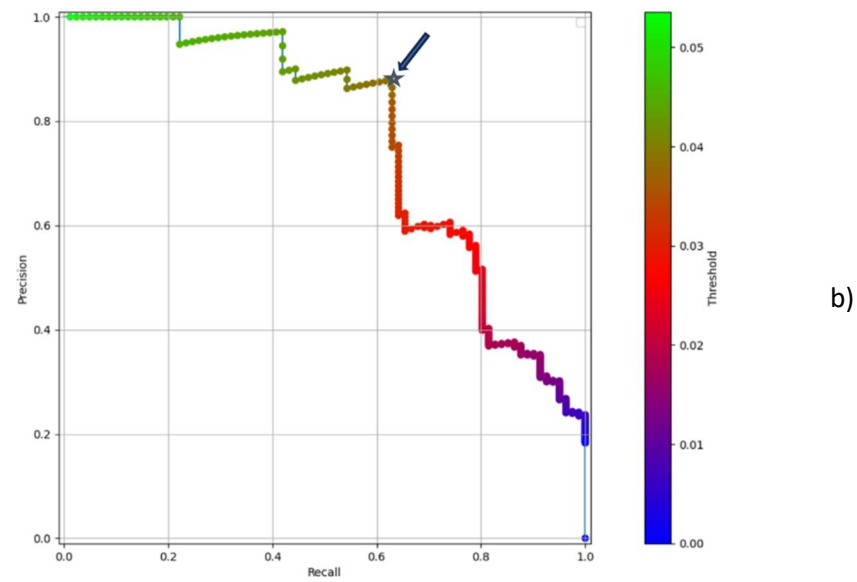
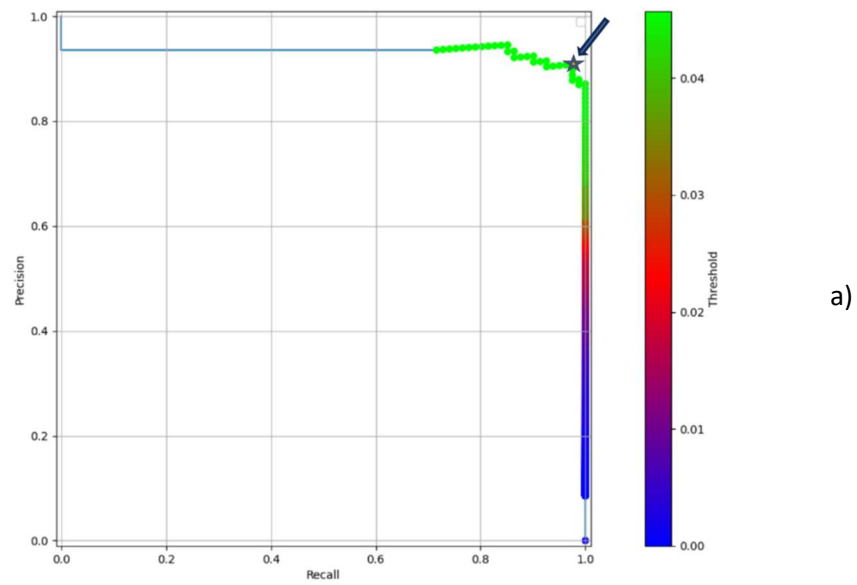
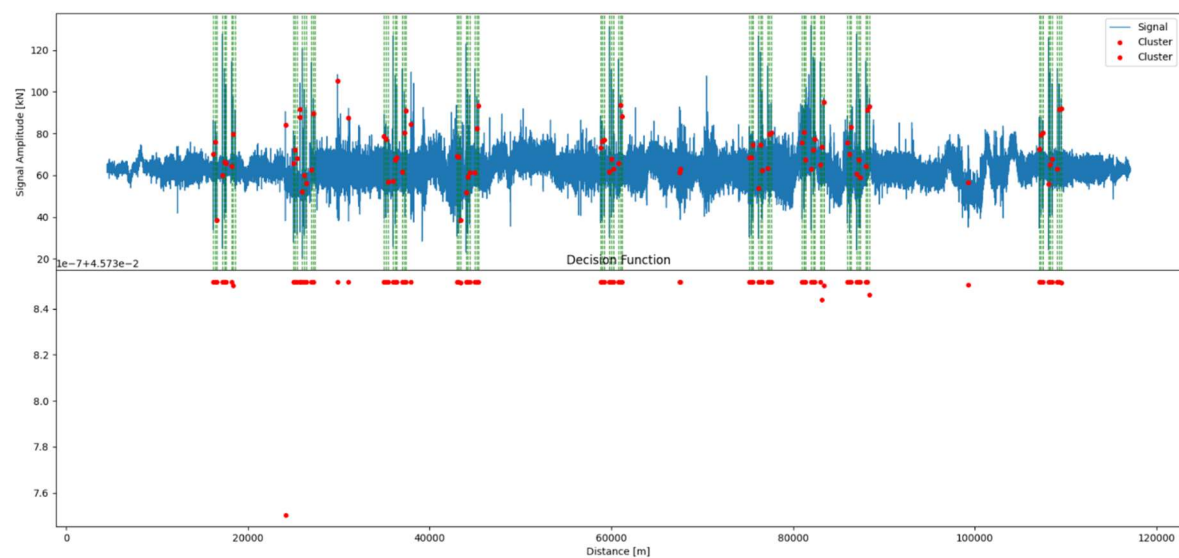
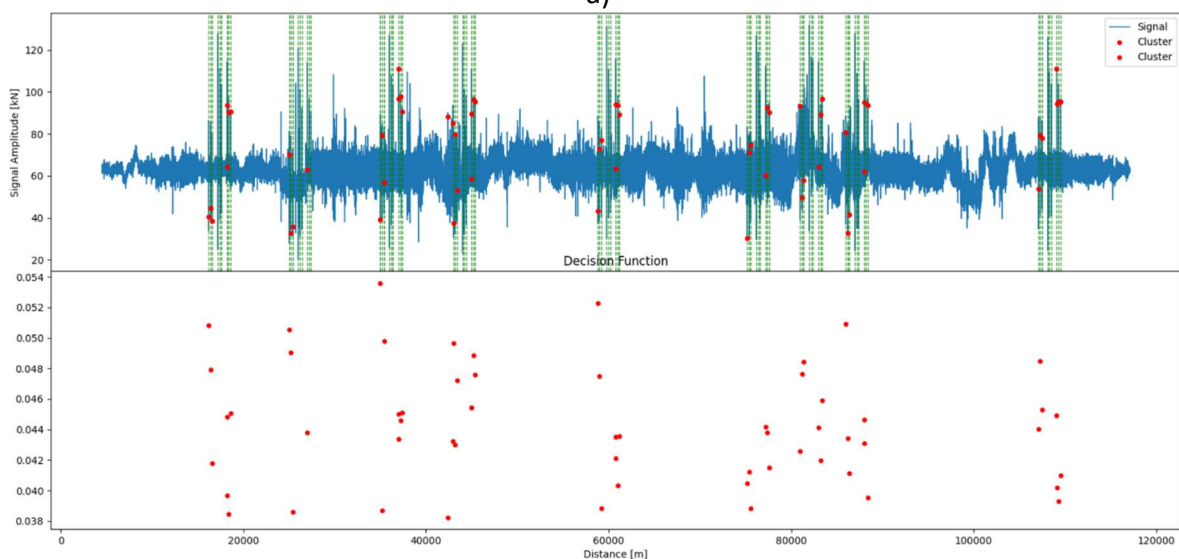


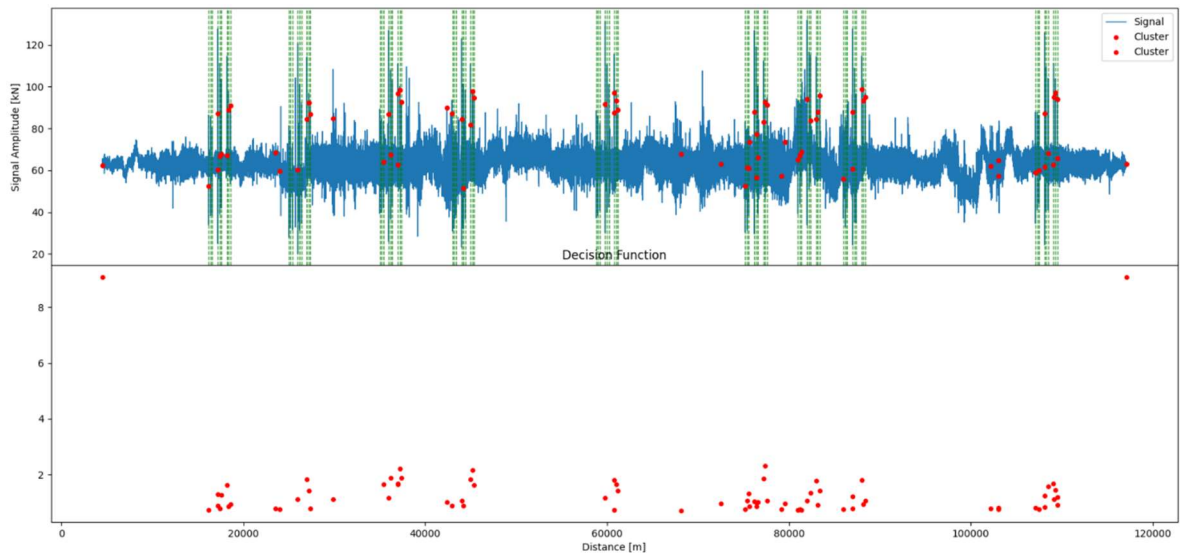
Figure 8.5: PR curves of the statistical frequency-domain features method. a) OCSVM; b) IF; c) LOF classifier.



a)



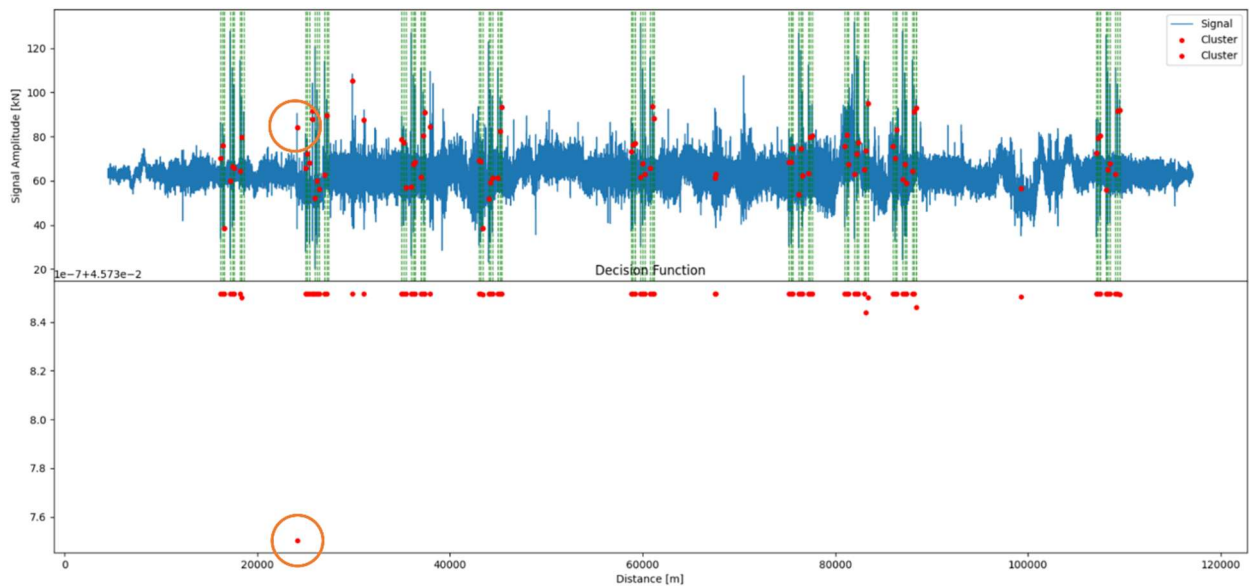
b)



c)

Figure 8.6: Classifiers outliers detection output: a) OCSVM (threshold: 0.045730750); b) IF (threshold: 0.038239348); c) LOF (threshold: 0.709274613).

As illustrated in Figure 8.6 (a), the OCSVM decision function for the filtered anomalies is more consistent compared with the other classifiers, which exhibit greater scatter. An important observation emerges from Figure 8.6 (a): since the decision function threshold is varied discretely rather than continuously, the chosen value that maximises the F1-score may not necessarily correspond to the absolute best detection performance. For example, increasing the threshold slightly, from 0.0457307504553889 to an interpolated value of 0.0457307504553900, reduces the FP count (from 9 to 8) and results in higher Precision and F1-score metrics (see Figure 8.7).



a)

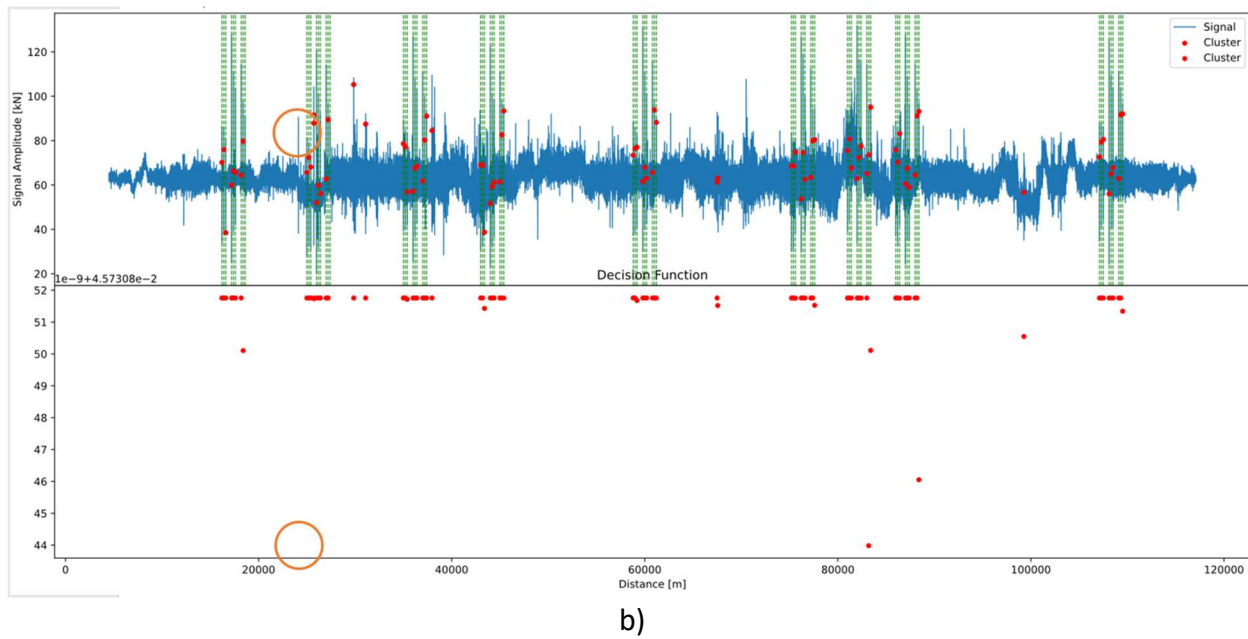


Figure 8.7: OCSVM detection. a) threshold: 0.0457307504553889, proposed discrete value; b) threshold: 0.0457307504553900, interpolated value.

8.7.1 Influence of signal windowing

The length L_w of the moving window used for feature calculation did not exhibit a consistent influence on detection performance, unlike in the case of statistical space-domain features. For all classifiers, the general trend observed for OCSVM indicated an increase in the maximum number of detected TPs with decreasing L_w (using the default threshold > 0). However, when using the optimal threshold, selected to maximise the F1-score, the number of detected TPs decreased steadily. This suggests that the algorithm retrieves a larger number of positive instances (outliers) as L_w decreases, but most of them correspond to FPs with higher decision-function values than true anomalies.

Contrary to initial expectations, shorter signal windows did not compromise the information captured by the FFT. No significant loss of spectral characteristics was observed when smaller L_w values were used.

8.8 Conclusions and future developments

This chapter proposed a novel semi-supervised anomaly detection approach for rail-head fault identification based on statistical frequency-domain features. Unlike previously explored methods that rely on statistical time- or space-domain indicators, the present approach employs spectral analysis and the STFT to extract frequency-domain information from overlapping sliding windows of the contact force signals measured with instrumented wheelsets (IWS). The extracted features were assembled into a feature matrix and fed into three classifiers, OCSVM, IF, and LOF. The performance of each anomaly detection algorithm was compared in terms of Precision, Recall, F1-score, inference time, and the number of retrieved instances, with the optimal decision threshold (the one maximising F1-score) applied to filter outliers.

Based on the results, OCSVM remains the most promising classifier for novelty detection of rail-head faults, followed by IF and LOF. The training and inference time of all tested algorithms was longer compared to the statistical space-domain feature method. The increase can be attributed not only to the STFT process, requiring FFT computation for each signal window, which is inherently time-consuming, but also to limitations in parallelisation: the STFT step cannot be executed within the Numba parallel computing framework, as it is incompatible with the FFT modules in SciPy and NumPy. Consequently, the pipeline was divided into two stages:

- FFT computation using Joblib.Parallel, which represents the computational bottleneck, and
- Parallelised statistical feature computation with Numba, which significantly accelerates numerical processing.

Comparing these results with previous trials conducted on smaller datasets (omitted here for brevity), it can be concluded that training dataset length plays a fundamental role in the success of novelty detection. Using a reduced training set (one quarter of dataset C14, containing over 4.5 million elements) led to a significant performance drop across all classifiers, especially for OCSVM, which otherwise demonstrated superior detection accuracy. Given these findings, OCSVM is highly recommended for anomaly detection pipelines based on statistical frequency-domain features. Although the TPs obtained using the optimal threshold do not fully match the ground-truth length, and FPs still represent about 10% of the total retrieved instances, the achieved performance is remarkable and confirms the potential of this method for large-scale field applications.

Future developments will include the Wavelet Transform as an alternative to the traditional STFT, with the goal of achieving improved frequency resolution and more precise characterisation of transient events caused by rail defects. These enhancements aim to further improve the overall performance of the detection pipeline, reducing both false positives and missed faults.

Abbreviations

The following abbreviations are used in this chapter:

AUC	Area Under Curve
BP	Band-Pass
FFT	Fast Fourier Transform
FN	False Negatives
FP	False Positives
GT	Ground Truth
IF	Isolation Forest
IWS	Instrumented Wheelsets
LOF	Local Outlier Factor
OCSVM	One-Class Support Vector Machine
PR	Precision-Recall
RMS	Root Mean Square

STFT	Short Time Fourier Transform
TCA	Track Condition Analyser
TN	True Negatives
TP	True Positives

8.9 References

1. Rosano, G., Massini, D., Bocciolini, L., Zappacosta, C., Di Gialleonardo, E., Somaschini, C., et al. (2024). Diagnostics of the railway track - Possibility of development through the measurement of accelerations and contact forces [JB]. *Ingegneria Ferroviaria*, 81–102.
2. Bernal, E., Spiryagin, M., and Cole, C.: "Onboard Condition Monitoring Sensors, Systems and Techniques for Freight Railway Vehicles: A Review," in *IEEE Sensors Journal*, vol. 19, no. 1, pp. 4-24, 1 Jan.1, 2019, doi: 10.1109/JSEN.2018.2875160.
3. Zhu, T., Wang, X., Wu, J., Zhang, J., Xiao, S., Lu, L., Yang, B., Yang, G.: Comprehensive identification of wheel-rail forces for rail vehicles based on the time domain and machine learning methods. *Mechanical Systems and Signal Processing* (222), (2025).
4. Bellacci, G., Pugi, L., Baldanzini, N. (2024). Validation of a Wheelset Finite Element Model for Static Structural Analysis and Inverse Force Identification. In: Concli, F., Maccioni, L., Vidoni, R., Matt, D.T. (eds) *Latest Advancements in Mechanical Engineering*. ISIEA 2024. *Lecture Notes in Networks and Systems*, vol 1124. Springer, Cham.
5. Bellacci, G., Baldanzini, N., Pugi, L. (2025). On the Use of Wavelet Analysis for the Detection and Position Estimation of Short Wavelength Rail Head Faults. In: Carbone, G., Quaglia, G. (eds) *Proceedings of I4SDG Workshop 2025 - IFToMM for Sustainable Development Goals*. I4SDG 2025. *Mechanisms and Machine Science*, vol 180. Springer, Cham.
6. M. Molodova, Z. Li, A. Núñez and R. Dollevoet, "Automatic Detection of Squats in Railway Infrastructure," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 1980-1990, Oct. 2014
7. Z. Li, M. Molodova, A. Núñez and R. Dollevoet, "Improvements in Axle Box Acceleration Measurements for the Detection of Light Squats in Railway Infrastructure," in *IEEE Transactions on Industrial Electronics*, vol. 62, no. 7, pp. 4385-4397, July 2015
8. Bellacci, G., Entezami, M., Weston, P., Baldanzini, N., Pugi, L. (2025). Evaluating the Feasibility of Estimating Rail Longitudinal Level Through Double Integration of Vertical Contact Forces. In: Carbone, G., Quaglia, G. (eds) *Proceedings of I4SDG Workshop 2025 - IFToMM for Sustainable Development Goals*. I4SDG 2025. *Mechanisms and Machine Science*, vol 180. Springer, Cham.
9. Ghiasi, R., Arslan, K.M., Sorrentino, D., Diaine, C., Malekjafarian, A.: An unsupervised anomaly detection framework for onboard monitoring of railway track geometrical defects using one-class support vector machine. *Engineering Applications of Artificial Intelligence*, Volume 133, Part B, 2024.
10. Walther, S., Müller, S., Renggli, R., Ünlü, F., Fuerst, A.: Prediction of wheel-rail contact forces using simple onboard monitoring system and machine learning. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*. 2022;237(5):553-562.
11. S. Kanarachos, J. Mathew, A. Chroneos and M. Fitzpatrick, "Anomaly detection in time series data using a combination of wavelets, neural networks and Hilbert transform," *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, Corfu, Greece, 2015, pp. 1-6.
12. Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PLOS ONE*, 11(4), e0152173. <https://doi.org/10.1371/journal.pone.0152173>.
13. Gullers, P., Dreik, P., Nielsen, JCO., et al. Track Condition Analyser: Identification of Rail Rolling Surface Defects, Likely to Generate Fatigue Damage in Wheels, Using Instrumented Wheelset

Measurements. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit. 2011;225(1):1-13.

14. Esvelde, C. (2001). Modern Railway Track (3^a ed.). MRT Productions. ISBN 978-9080032439.

9 DATA-DRIVEN METHODS: COMBINING ONE-CLASS NOVELTY DETECTION FRAMEWORKS

9.1 Introduction

To achieve a reliable and precise track fault-detection pipeline, the previously developed approaches were combined. Both space-domain features and space–frequency-domain (or, more precisely, space-wavenumber-domain) analyses offer specific advantages and limitations. By combining them, it is possible to mitigate their individual weaknesses and enhance overall detection capability. This brief chapter presents and discusses the results obtained from the combined methodologies.

9.2 Limitations

As already observed in Chapters 7 and 8, signal windowing exerts a strong influence on the results of the analysis, so much so that it can radically affect anomaly detection. Shortening the window length, up to 20 elements, for feature extraction improves the performance of the space-feature pipeline, leading to a reduction in FP and FN and an increase in TP across all three tested algorithms. A similar trend can be noted in the space-frequency pipeline ($L_W = 50$ elements), although no clear or consistent relationship could be established. In general, the windowed signal must retain sufficient spectral content to allow meaningful frequency analysis. Consequently, the window length cannot be too short. Therefore, when combining the two methods, a compromise in window length must be reached, which inevitably limits the performance of the space-feature

approach. Taking the smallest faults size as a minimum window dimension may not be the wisest strategy. Although technically valid, choosing a window length comparable with the size of real head defects should imply that faults can be generalised, meaning that all faults can be detected with the shortest window length. Considering a squat with 30 mm of extension (such as an early-stage squat), even if we admit that this can be identified, the sampling frequency of the force signal (and all the related signal filtering), does not allow the observation of the dynamic trace of the contact force deriving from the interaction between the wheel and the defect. A minimum of ten times this length should be used to identify most of the injected faults with poor detection performance, as results from the analysis. Raising the value up to 0.5 m or 1 m (20 or 50 elements), depending on the adopted technique (space, frequency or mixed) demonstrated an improvement for the detection.

9.3 Results and Discussion

The results obtained from combining space-feature-based and frequency-feature-based novelty detection frameworks for the investigated one-class classifiers are summarised in Table 9.1. As a general trend, when comparable window lengths are used in both the space and space-frequency methods, generating a combined feature matrix that merges statistical space- and frequency-domain features produces a positive effect, particularly in reducing FP retrieved by the algorithms (notably for IF and LOF). At the same time, the number of TP increases, and FN instances decrease, compared with the use of the single separated methods. When compared with the space-feature method alone (using very short signal windows, $Lw = 50$), the improvement in detection is evident, showing a higher number of retrieved TP and a lower FP count for all classifiers. When compared to the statistical frequency-feature approach, the advantage of the combined strategy remains evident for IF and LOF, although it results in a slight drop in performance for OCSVM. Although unexpected and contrary to the trend observed for the other two classifiers, the behaviour of OCSVM suggests that greater attention should be paid to the feature extraction and selection strategies. A more refined selection strategy should be implemented to identify the most informative features, for example through feature importance ranking methods such as LIME, SHAP, or ANOVA-based algorithms.

It is important to remark that all the tests performed on the C12 dataset relied on artificially injected numerical faults. Although these were inspired by literature-based wheel-rail contact force time series representing typical defects and rail-head faults, such synthetic anomalies may differ significantly from real-world irregularities.

Moreover, only a limited set of representative faults was recreated and investigated with the three classifiers, while other fault types, such as corrugation (with varying wavelength and depth), spalling, and ballast voids, remain unexplored. For instance, a frequency- or wavenumber-feature-based approach might provide greater benefits in the case of corrugation, or in scenarios where the spectral content of the analysed track segment changes significantly between fault-free and faulty sections. This could include impulsive events exciting a broad frequency bandwidth, or defects whose wavelength varies across a given track portion.

9.3.1 Considerations over inference time

Given the time-consuming nature of the training and fitting processes, real-time embedded implementation of these algorithms appears currently impractical. Furthermore, they require

large quantities of data to train the classifiers properly, data that must be entirely free from anomalies. Therefore, model training should be performed offline, using recorded datasets spanning several kilometres, which should then be stored locally on diagnostic vehicles and subsequently transferred to a central data repository for secure storage and further processing. When comparing inference times, IF consistently exhibited the fastest performance, followed by OCSVM, whereas LOF remained the slowest. Despite the slight improvement observed from combining the two feature domains, LOF's poor detection performance does not justify its considerably higher computational cost.

Table 9.1: Classifiers comparison and detection performance summary.

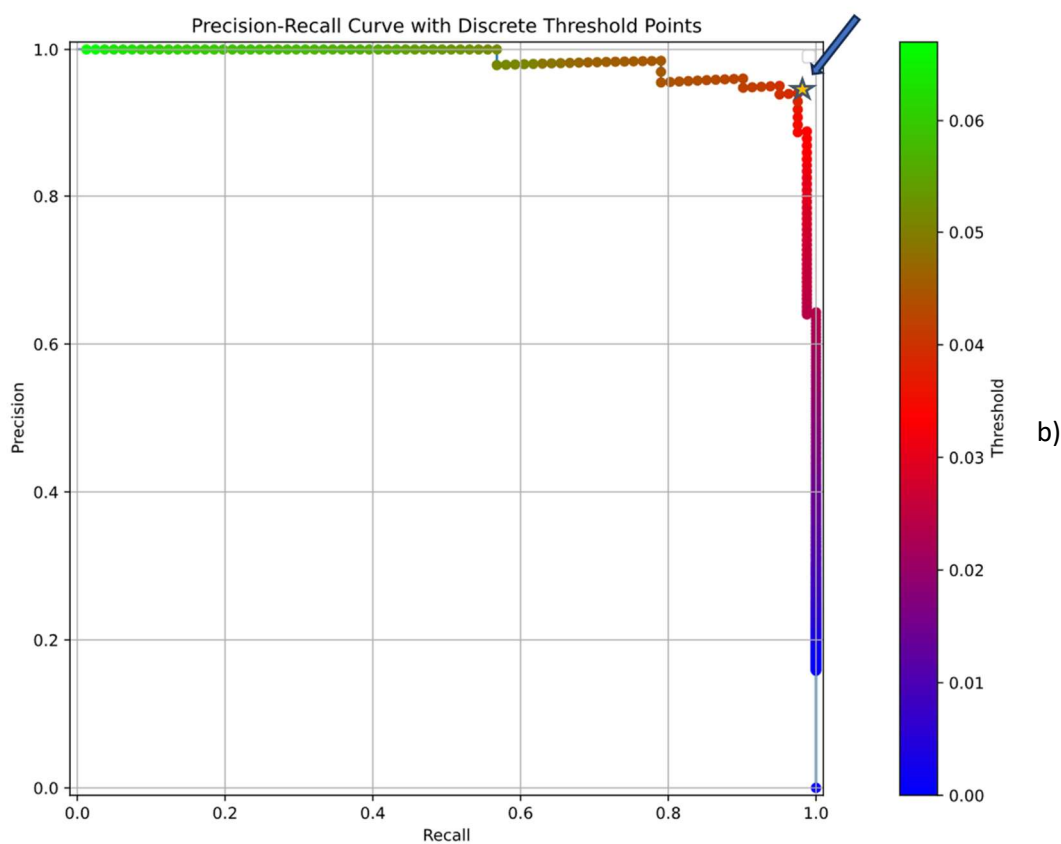
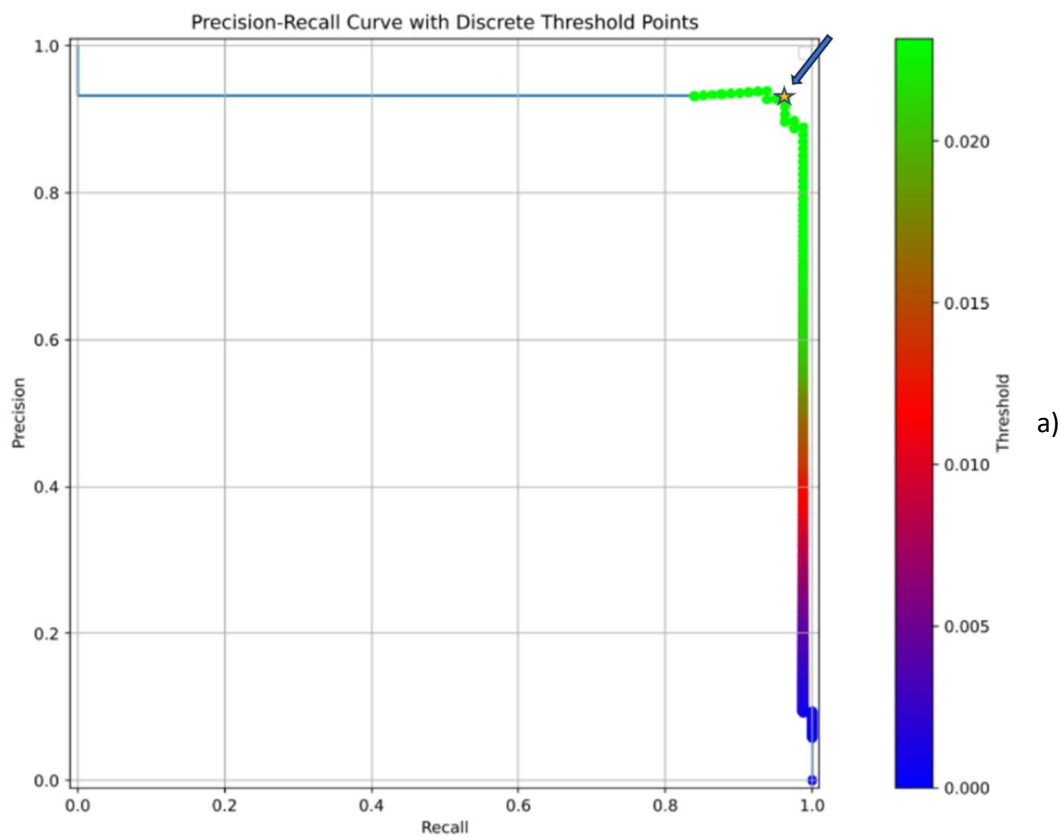
Classifier	Retrieved instances	Optimal F1-Score	Optimal Precision	Optimal Recall	Training Time Performance	Fitting Time Performance
OCSVM ¹ (Threshold 0.0231815)	TP: 78/81 FP: 6 FN: 3 Max TP: 81	0.945	0.929	0.963	Training: 2586.90 s Parallel-feat.: 700.2952 s Numba-feat.: 3.90 s Tot: 3298.81 s	Fitting: 2443.63 s Parallel-feat.: 682.39 s Numba-feat.: 3.93 s Tot: 3143.59 s
IF ² (Threshold 0.0436101)	TP: 79/81 FP: 6 FN: 2 Max TP: 81	0.952	0.930	0.975	Training: 136.63 s Parallel-feat.: 698.38 s Numba-feat.: 5.99 s Tot: 851.33 s	Fitting: 219.33 s Parallel-feat.: 703.60 s Numba-feat.: 3.85 s Tot: 896.20 s
LOF ³ (Threshold 1.6995327)	TP: 69/81 FP: 40 FN: 12 Max TP: 81	0.730	0.633	0.851	Training: 7759.93 s Parallel-feat.: 694.42 s Numba-feat.: 3.83 s Tot: 8468.56 s	Fitting: 16060.22 s Parallel-feat.: 711.16 s Numba-feat.: 3.94 s Tot: 16788.85 s

Classifier parameters:

^{1.} $\nu = 10^{-5}$; $\gamma = 0.01$; kernel: RBF; $Lw = 50$; Scaling: Robust.

^{2.} $n_{estim} = 1000$; Contamination = 0.001; $Lw = 50$; Scaling: Robust.

^{3.} Neighbours = 200; $Lw = 50$; Scaling: Standard.



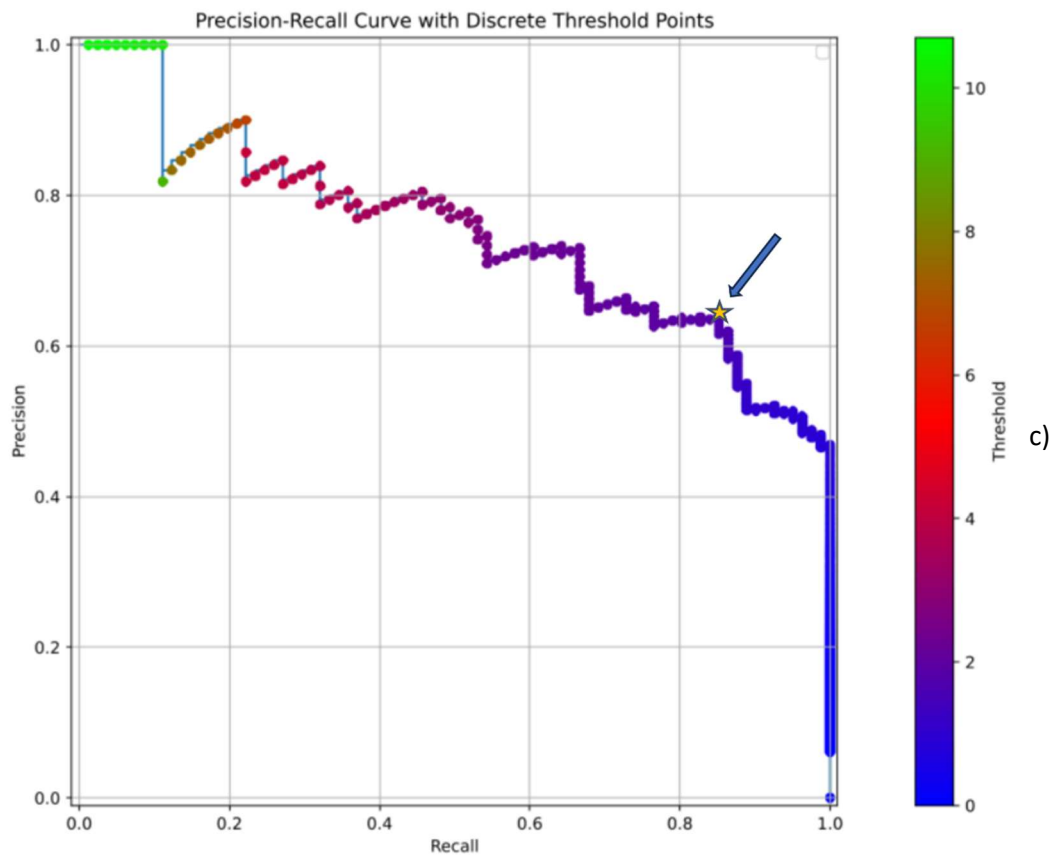


Figure 9.1: PR curves of a) OCSVM; b) IF; c) LOF classifiers with combined set of features.

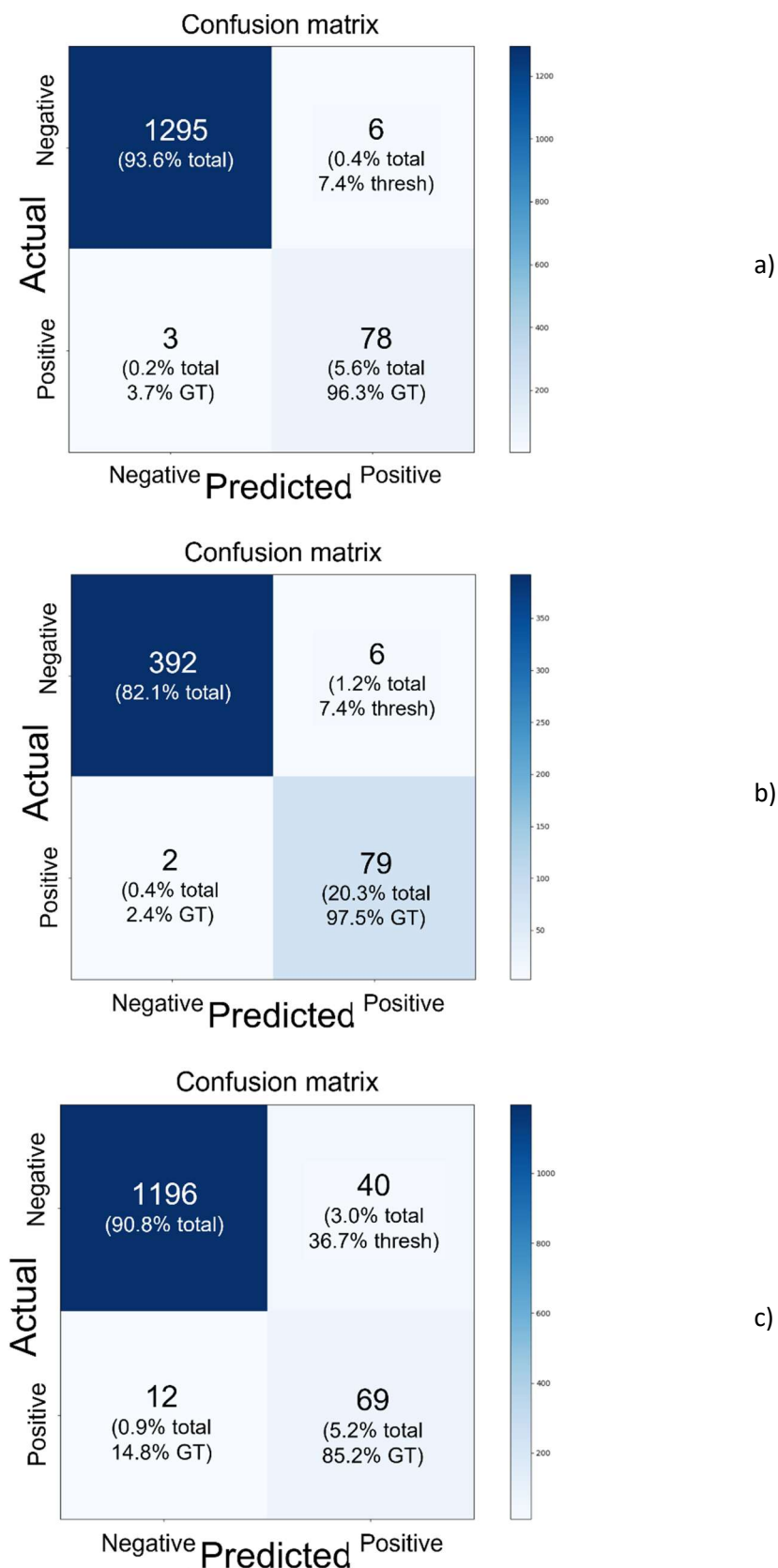
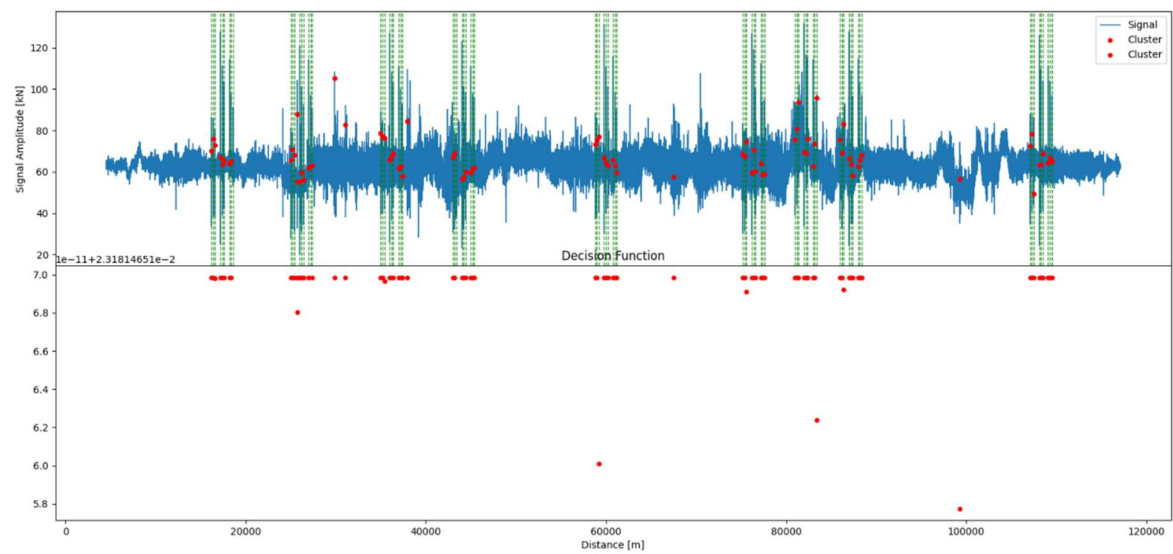
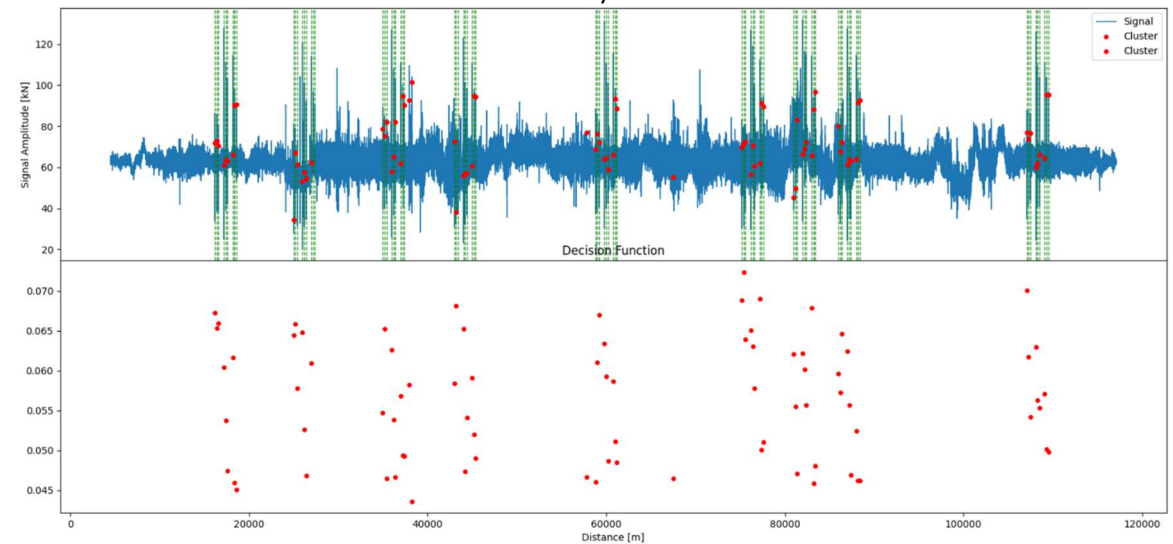


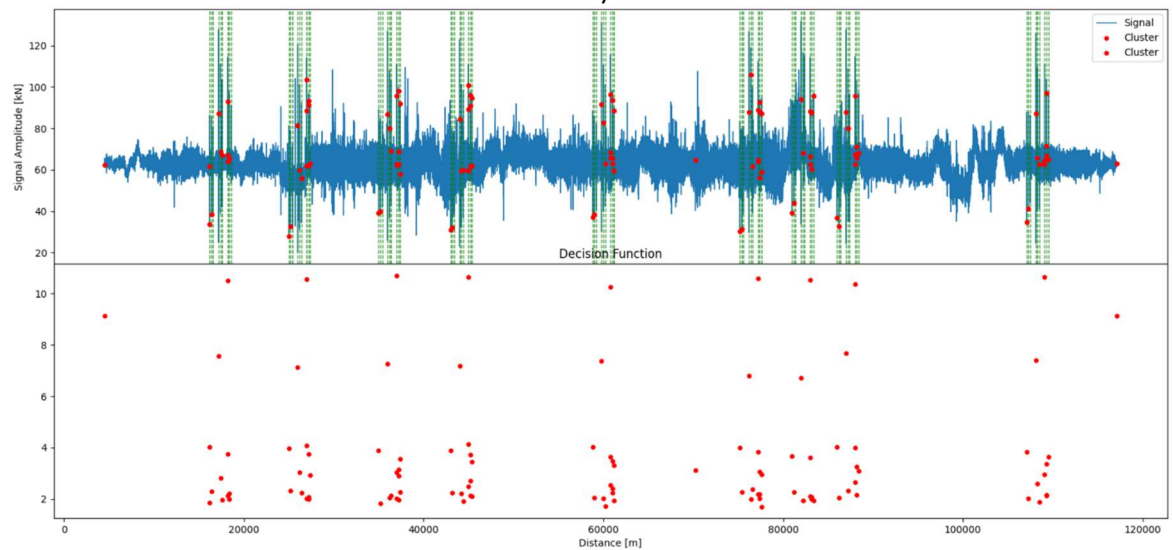
Figure 9.2: Confusion matrices of a) OCSVM; b) IF; c) LOF classifiers. Statistical space-domain-feature method combined with statistical frequency-domain-feature method. Percentages are shown for total retrieved instances (“total”); percentage over the GT, and percentage over the retrieved instances after applying the optimal threshold for the FP (“thresh”).



a)



b)



c)

Figure 9.3: Classifiers outliers detection output: a) OCSVM (threshold: 0.0231815); b) IF (threshold: 0.0436101); c) LOF (threshold: 1.6995327).

9.4 Conclusions

This chapter presented the combination of space-domain and space-frequency-domain novelty detection frameworks applied to one-class classifiers. The results confirmed that integrating the two feature domains enhances fault detection performance, especially by reducing FP and FN while increasing TP for IF and LOF classifiers. Although OCSVM showed a slight decrease in accuracy under the combined setup, its overall reliability remains high, suggesting that the interaction between feature domains deserves further optimisation, particularly in the feature selection phase.

Despite the promising results, computational limitations, mainly related to training duration and data volume, currently restrict real-time applications. Nevertheless, the combined approach represents a meaningful step toward a more comprehensive and robust rail condition monitoring framework, bridging the gap between statistical, spectral, and hybrid feature-based anomaly detection strategies.

10 ORIGINAL CONTRIBUTION

10.1 List of Publications

- P1. Bellacci, G., Neri, F., & Pugi, L. (2024). Preliminary frequency response analysis of a contact force measurement system for rail applications. In F. Bellotti et al. (Eds.), *Applications in electronics pervading industry, environment and society (ApplePies 2023) (Lecture Notes in Electrical Engineering, Vol. 1110, pp. 240–247)*. Springer. https://doi.org/10.1007/978-3-031-48121-5_20.
- P2. Bellacci, G., Pugi, L., & Baldanzini, N. (2024). Validation of a wheelset finite element model for static structural analysis and inverse force identification. In F. Concli, L. Maccioni, R. Vidoni, & D. T. Matt (Eds.), *Latest advancements in mechanical engineering (ISIEA 2024) (Lecture Notes in Networks and Systems, Vol. 1124, pp. 236–244)*. Springer. https://doi.org/10.1007/978-3-031-70462-8_23.
- P3. Bellacci, G., Entezami, M., Weston, P., Baldanzini, N., & Pugi, L. (2025). Evaluating the feasibility of estimating rail longitudinal level through double integration of vertical contact forces. In G. Carbone & G. Quaglia (Eds.), *Proceedings of I4SDG Workshop 2025 – IFToMM for Sustainable Development Goals (I4SDG 2025) (Mechanisms and Machine Science, Vol. 180, pp. 737–744)*. Springer. https://doi.org/10.1007/978-3-031-91179-8_63.
- P4. Bellacci, G., Baldanzini, N., & Pugi, L. (2025). On the use of wavelet analysis for the detection and position estimation of short wavelength rail head faults. In G. Carbone & G. Quaglia (Eds.), *Proceedings of I4SDG Workshop 2025 – IFToMM for Sustainable Development Goals (I4SDG 2025) (Mechanisms and Machine Science, Vol. 180)*. Springer. https://doi.org/10.1007/978-3-031-91179-8_61.
- P5. Bellacci, G., Entezami, M., Weston, P. F., & Pugi, L. (2025). Estimation of Railway Track Vertical Alignment Using Instrumented Wheelsets and Contact Force Recordings. *Machines*, 13(10), 963. <https://doi.org/10.3390/machines13100963>.

- P6. Bellacci, G.; Baldanzini, N.; Pugi, L. Anomaly Detection in Railway Tracks Using Contact Force Spectra and Machine Learning. In Applications in Electronics Pervading Industry, Environment and Society. ApplePies 2025; Ruo Roch, M., et al., Eds.; Lecture Notes in Electrical Engineering, Vol. 1553; Springer: Cham, 2026. https://doi.org/10.1007/978-3-032-17174-0_49.
- P7. Bellacci, G.; Di Carlo, L.; Fiaschi, M.; Bocciolini, L.; Zappacosta, C.; Pugi, L. Online Track Anomaly Detection: Comparison of Different Machine Learning Techniques through Injection of Synthetic Defects on Experimental Datasets. MDPI Machines (Submitted for publication)

11 CONCLUSIONS

This doctoral thesis explored the use of instrumented wheelsets (IWS) as diagnostic tools for railway track condition monitoring, combining model-based, traditional, and data-driven methodologies. The overarching goal was to assess whether contact force measurements could be used not only for acceptance and certification purposes, but also as a reliable source of information for detecting and classifying track irregularities. Through a progressive and multidisciplinary approach, this work established the methodological and technological foundations for transforming IWS from experimental tools into continuous monitoring devices capable of supporting predictive maintenance strategies.

11.1 Main findings

11.1.1 Finite Element Modelling and Frequency Response Characterisation

A full finite element (FE) model of the instrumented wheelset was developed and validated under static load conditions to support inverse identification of wheel-rail contact forces. The model successfully reproduced the strain and stress distribution within the wheel web, confirming the reliability of the adopted sensor configuration. A frequency response analysis further characterised the system dynamics, identifying a usable measurement bandwidth of approximately 200 Hz. Although this range was limited by a not-rotating object, subjected to a single point excitation with the restricted excitation energy of the instrumented hammer, the study provided a clear understanding of the dynamic sensitivity and transmissibility of the IWS, laying the foundation for the subsequent signal analyses.

11.1.2 Finite Element Modelling and Frequency Response Characterisation

Using Wavelet Transform and time–frequency methods, the study demonstrated that the temporal and spatial localisation of short-wavelength rail defects can be achieved directly from the vertical contact force signals recorded by IWS. Irregularities within the 87–282 Hz frequency band, corresponding to wavelengths of 9.5–31 cm, were effectively detected, including

developed squats, dipped joints, welds, ballast voids, and long-pitch corrugation. Despite the absence of a validated ground truth, visual comparison confirmed that most of the major force impulses corresponded to relevant railhead irregularities.

This work highlighted the diagnostic potential of IWS compared to traditional optical systems, thanks to their higher sampling rate and superior sensitivity to dynamic phenomena. The results also suggested a natural integration with machine learning frameworks, where wavelet-derived scalograms could be used for automated classification and localisation of short-wavelength defects through image-recognition algorithms.

11.1.3 Model-Based Approach

A simplified 3-DOF vertical dynamic model representing half a railway coach was formulated to estimate the mean vertical alignment of the track using only force recordings as input. The model successfully reconstructed the D1 wavelength range (3–25 m) of track irregularities with good agreement against reference TRV measurements, particularly at train speeds between 160 km/h and 200 km/h. High correlation coefficients ($R > 0.8$ for most datasets) and consistent spectral content confirmed the validity of the approach despite its reduced order. Residual errors were attributed to neglected degrees of freedom and filtering approximations; however, these simplifications provided substantial advantages in terms of computational efficiency and ease of deployment. The proposed estimator can thus serve as a low-cost complementary tool for real-time or post-processing correction of optical data, enhancing robustness in mixed diagnostic architectures.

11.1.4 Data-Driven Approaches

In the final part of the research, several machine learning frameworks for unsupervised and semi-supervised anomaly detection were developed. Three classifiers, One-Class Support Vector Machine (OCSVM), Isolation Forest (IF), and Local Outlier Factor (LOF), were trained on force-derived statistical features extracted in both the space and space-frequency domains. OCSVM consistently achieved the best detection metrics (Precision, Recall, F1-score), while IF provided faster execution times. Although LOF exhibited lower performance and higher memory demand, its inclusion enriched the comparative understanding of one-class classifiers in this context. A further improvement was achieved by combining space- and frequency-domain features, which reduced FP and FN for IF and LOF, though it slightly decreased OCSVM accuracy. A summary of the comparison among the methods is shown in Figure 11.1 and Figure 11.2. As can be evinced from these graphs, the combination of space and spectral features held an extremely positive effect on IF, which almost reached a complete detection with the smallest number of FN and FP, which is a remarkable result. The highest number of TP was achieved by OCSVM too, in case of Spectral features, but with a higher number of FP. Just by observing the performance metrics, it can be said that IF and combined features method shows the most promising result, with great advantage in training and inference time compared with the other two algorithms. From Figure 11.2, it is evident that this method reaches the best detection in terms of number of TP. When FP are accounted, the reduced number of such instances makes IF extremely appealing and trustworthy for a field application with known ground truth defects.

These findings demonstrate that hybrid feature domains can enhance the robustness of data-driven pipelines, and that feature selection and interpretability (e.g., using SHAP or LIME) are key directions for future refinement.

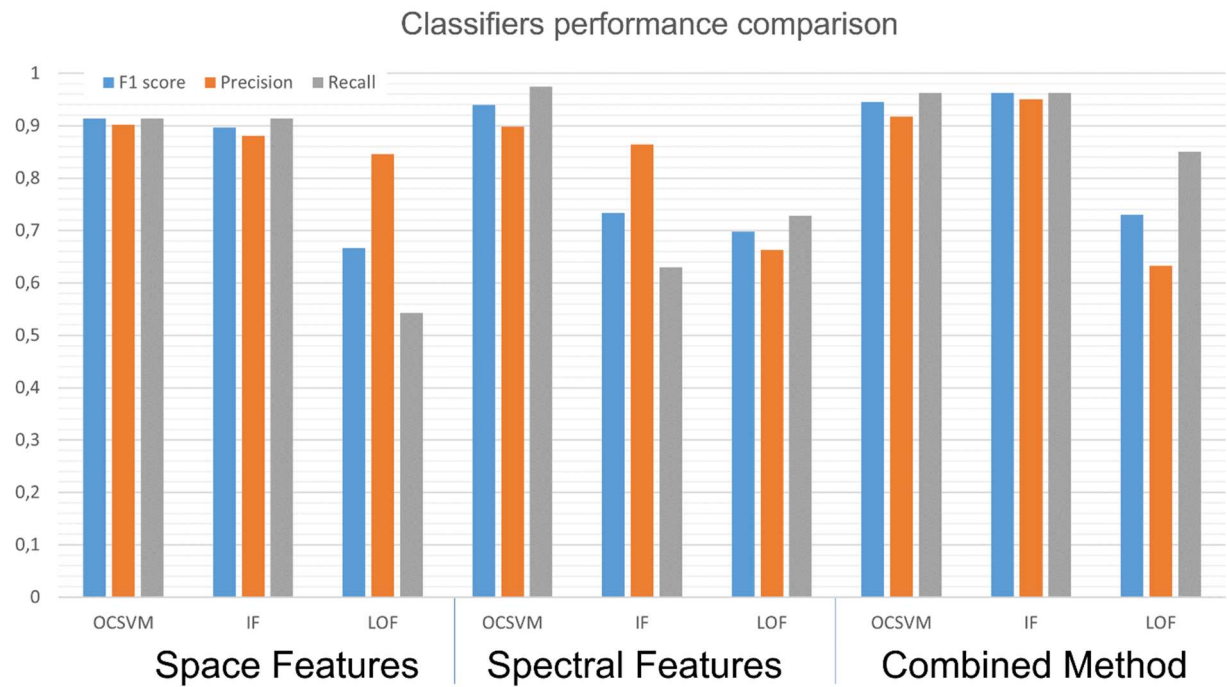


Figure 11.1: Summary of classifiers’ performance. Comparison among the tested methods.

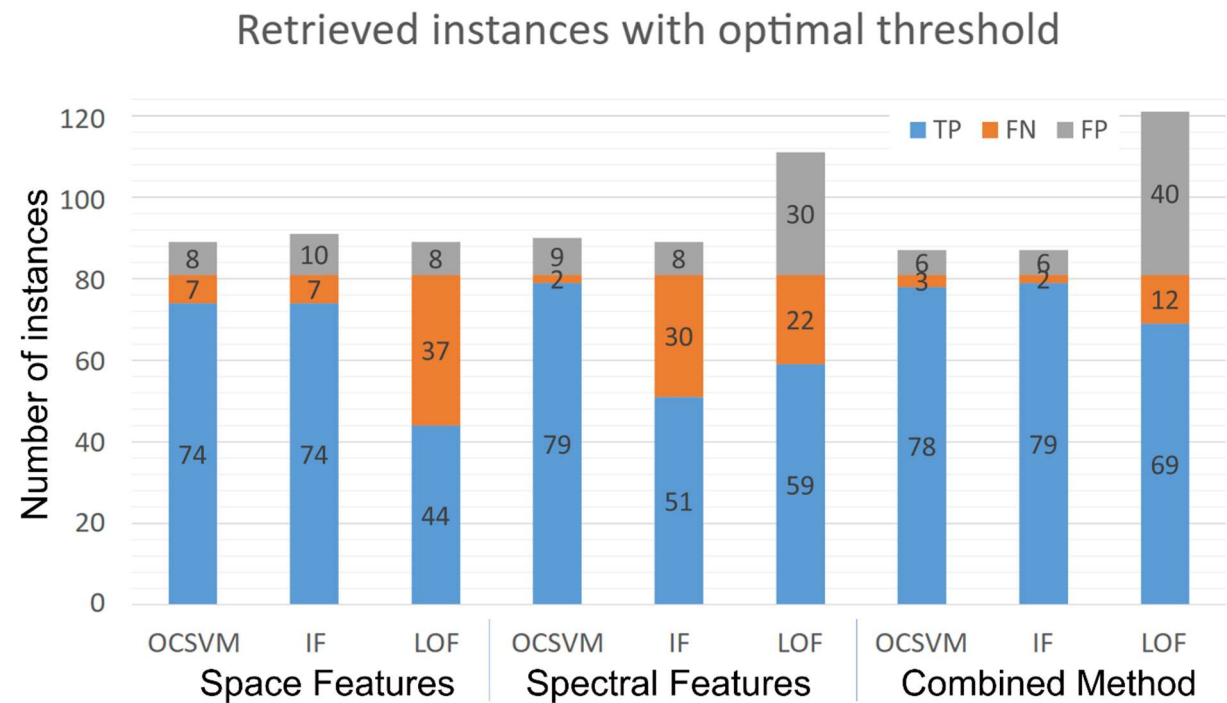


Figure 11.2: Retrieved instances with optimal decision function threshold applied for each classifier and case. Comparison among the tested methods.

11.2 Overall Contributions

The principal contributions of this thesis can be summarised as follows:

- Validated FE model of an instrumented wheelset enabling inverse identification of wheel-rail contact forces and dynamic characterisation.
- Experimental determination of the IWS frequency response and identification of the effective measurement bandwidth (~ 200 Hz).
- Application of traditional time–frequency methods (Wavelet Transform, STFT) for identifying localised railhead defects from force recordings.
- Development of a simplified 3-DOF model for estimating track longitudinal level using force-based inputs.
- Design and comparison of ML novelty detection pipelines (OCSVM, IF, LOF) based on space- and frequency-domain features.
- Integration of hybrid approaches combining physical models, spectral analysis, and data-driven learning for a comprehensive diagnostic framework.
- Demonstration of the feasibility of IWS-based monitoring as a viable complement, or potential alternative, to conventional optical inspection systems.

11.3 Future Developments

Building on the outcomes of this research, several directions are proposed for future work:

- Model refinement: Extend the current 3-DOF system to a 6-DOF or full-vehicle model, incorporating pitch and roll dynamics and employing Kalman filters for parameter calibration, disturbance estimation, and sensor fusion.
- Data fusion and hybrid modelling: Combine force-derived information with inertial, optical, or vibration data to improve fault classification accuracy and reliability.
- Field validation: Conduct extensive measurement campaigns to correlate detected anomalies with verified track defects, providing robust ground truth for model and ML validation.
- Advanced spectral analysis: Integrate Wavelet or Empirical Mode Decomposition (EMD) with Hilbert-Huang transform to improve transient resolution and complement the limitations of the STFT.
- Feature selection and explainability: Implement interpretability-driven approaches (e.g., SHAP, LIME, ANOVA) to identify the most informative features contributing to fault detection.
- Real-time deployment: Explore edge-computing or cloud-assisted implementations to enable on-board real-time monitoring and decision support for maintenance operations.

11.4 Final Remarks

This thesis demonstrated that contact force measurements obtained through instrumented wheelsets can serve as a powerful foundation for multi-level diagnostic frameworks. By integrating physics-based models, traditional signal analysis, and data-driven intelligence, the work bridges the gap between experimental measurement and predictive maintenance. The

methodologies developed herein provide both interpretability and scalability, enabling their application across a wide spectrum of railway monitoring scenarios. The results contribute to the ongoing evolution of smart railway infrastructure, where digital twins, self-calibrating models, and AI-driven diagnostics cooperate to enhance safety, reduce maintenance costs, and ensure sustainable railway operation.

