DISS. ETH NO. 29341

ON BOARD MONITORING FOR RAILWAY INFRASTRUCTURE CONDITION ASSESSMENT

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

CYPRIEN HOELZL

MSc. in Civil Engineering, ETH Zurich

born on 23.05.1995

citizen of
Switzerland

accepted on the recommendation of

Prof. Dr. Eleni Chatzi (ETH Zürich, Examiner)
Dr. Vasilis Dertimanis (ETH Zürich, Co-Examiner)
Prof. Dr. Alfredo Cigada (Politecnico di Milano, Co-Examiner)
Prof. Dr. Eugene O'Brien (University College Dublin, Co-Examiner)
Dr. Lucian Ancu (Swiss Federal Railways, Co-Examiner)

2023

Acknowledgements

had a positive impact on my dissertation. I also want to thank Deepack Kumar Ravi who initially brought me into AVETH. I would especially like to thank Florentine Strudwick, who has been an inspiration as the president of AVETH while at the same time being the mother of a newborn child and successfully defending her thesis.

Recommendations and constructive and careful proofreading given by Carolin Fruehauf Hoelzl have been a great help for the final touches through my studies and research. A sincere thank you.

Finalizing my dissertation required more than academic support, and I have many people to thank for supporting me: Thanks to my friends and flatmates for all the memories and experiences we shared. You have played a key role in never helping me at letting go of work sometimes. I would like to extend my heartfelt gratitude to Milan Kovarbasic, Kyriakos Chondrogiannis, Giulia Agguzzi, and all my friends who have supported me along this journey, particularly those who shared my passion for volleyball. Your encouragement, company and team-spirit have been invaluable to me during the completion of my PhD thesis. I would like to express my sincere appreciation to Tom, for his unwavering efforts to keep the mood high with his cheerful demeanor and British wit.

My very deep gratitude goes to my family, my brother Cedric, my sister Eugenie and my parents, Anton and Carolin. I deeply thank you for your continuous support and motivation during my years of study, and through the process of researching and writing this thesis. This accomplishment would not have been possible without you and stands as a testament to your unconditional love and encouragement. Thank you.

Lastly, I express my heartfelt thanks to those whom I might have missed to mention by name, who helped directly or indirectly and supported me a lot in completion of this Ph.D. research work.

Cyprien Hölzl

14th March, 2023

Contents

Abstract

The railway track is the main component of the railway infrastructure, whose task is to carry the load of the traversing railway vehicles. Its structure is composed of numerous sub-components, namely the substructure, ballast, sleepers and rails, each of which has to fulfill different requirements in terms of load bearing capacity and durability. The recent increases in train speed, traffic density and passenger/freight loads, has urgently brought forth the need for efficient management practices, able to exploit monitoring data for ensuring optimization of availability, operation and maintenance costs and guaranteeing safety.

The determination of the condition (state) of railway assets has traditionally been based on visual on-site inspections and measurements. The nature, quality and completeness of measurements for railway tracks have been refined and significantly improved since the initial stages of track parameter measurements, conducted in the 1960s. An important catalytic agent in this respect was the introduction of automatic storage, analysis, and interpretation of digital measurement data collected by diagnostic vehicles. These are specialized vehicles, equipped with sophisticated and often costly measurement systems, which however can only infrequently traverse the network, implying limited temporal resolution for the monitoring data. Nowadays, On-Board Monitoring (OBM) vehicles, or in-service trains, can be used equipped with low cost measurement device,s such as Axle Box Accelerometers (ABAs), which regularly traverse the railway network, thus offering an almost continuous data stream that can be linked to the condition of the traversed railway infrastructure.

The primary objective of this research is to develop efficient monitoring techniques enabling the extraction of diagnostic indicators for guiding the maintenance of railway assets. To address this challenge, we rely on the principle of On Board Monitoring, i.e., the exploitation of low-cost sensors, mounted on vehicles (feasibly in service vehicles), for continuous and spatially dense monitoring of the railway network. Nevertheless, the transformation of raw measurement data, particularly accelerations, into robust and actionable condition indicators represents a significant challenge. Particularly since it is non-trivial to achieve the seamless integration of measurement data into existing asset management processes.

This thesis presents a comprehensive approach to monitoring the railway track, by primarily relying on exploitation of vibration-based data, and assessing their potency in terms of discovering long (longitudinal level) and short wavelength defects (surface defects, welds, squats). The first chapter contextualizes the problem of infrastructure monitoring and presents the state of the art in terms of conventional track monitoring techniques and vibration-based monitoring techniques. Through applications in several case studies, we show that acceleration-based approaches demonstrate high potential for complementing, or feasibly replacing in part, traditional schemes.

The second chapter proposes a robust method for examining the influence of individual parametric excitation sources. We propose adoption of the Vold-Kalman filter for identifying the periodic speed-dependent components of a measured ABA signal, such as the sleeper passage frequency and the wheel Out-Of-Roundness order. The wheel Out-Of-Roundness components are used to reconstruct the wheel profile, while the amplitude of the sleeper passage frequency is shown to form a proxy of the track stiffness. Moreover, it is demonstrated that track sections

corresponding to higher stiffness feature increased accelerations at the sleeper passage frequency, which correlate to higher forces and increased maintenance actions.

Chapter 3 zooms into the assessment of critical components of the rail in what is termed a short wave classification scheme. The rail forms an element whose structural integrity is essential for safety of operation, but which is exposed to elevated loading that leads to frequent appearance of faults such as squats, cracks or surface defects. Short wavelength rail-related effects bear a significant influence on the vehicle-track dynamics. This chapter proposes an acceleration-based automated classification scheme for classifying main rail components and surface faults (healthy rail, surface defects and squats, insulated joints and welds). Random Forests (RFs), which have been trained on engineered features, are contrasted against Convolutional Neural Networks (CNNs), which are trained on the Short Time Fourier Transform coefficients of the ABA signals. While the validation of the scheme cannot be full corroborated due to largely missing reference labels, the implementation endorses automation and motivates implementation of acceleration-based classifiers to support early detection of faults.

The influence of reliability of labels from the available condition monitoring databases and the incorporation of expert feedback into the analysis process closes the loop of our suggested scheme. The afore-mentioned short wave classification framework relies on availability of labels that are automatically extracted from rail head images. However, such labels do not adequately reflect the progressively evolving health status of critical components. The final study of this thesis focuses on welds, as salient critical components of rail infrastructure. Chapter 4 presents a Proof-of-Concept study which brings OBM-derived indicators, database information and expert feedback in the loop. The study was conducted in collaboration with asset managers and field experts of the SBB. The framework combines on-site and visual inspections, diagnostic information, and expert assessments to automate the detection of weld defects from ABA-indicator outliers. Rail head images corresponding to these outliers are reviewed by experts, leading to the identification and further labeling of suspected weld defects. This information loop is further exploited to develop robust automated defect classification schemes. Three methods are tested to this end, Binary Classification, Random Forest classifiers, and Bayesian Logistic Regression. We demonstrate that the latter additionally endows estimation with a quantification of the uncertainty involved.

In summary, this thesis contributes to the development of actionable monitoring-driven indicators and frameworks, across a range of applications; initiating from the more global identification of track geometry and stiffness to the more local identification of short wavelength faults, such as surface defects or deteriorated welds. More importantly, it delivers a holistic approach, which ties automated and continuous monitoring, sparse historic condition information, and expert opinion into one loop. This research delivers the effectuating tools and a valid proof of concept for assimilation of ABA data into the monitoring protocol of railway assets.

Zusammenfassung

Das Gleis ist der Hauptbestandteil der Eisenbahninfrastruktur, dessen Aufgabe es ist, die Last der verkehrenden Schienenfahrzeuge zu tragen. Seine Struktur setzt sich aus zahlreichen Teilkomponenten zusammen, nämlich dem Unterbau, dem Schotter, den Schwellen und den Schienen, die jeweils unterschiedliche Anforderungen an die Tragfähigkeit und die Dauerhaftigkeit erfüllen müssen. Der jüngste Anstieg der Zuggeschwindigkeiten, der Verkehrsdichte und der Passagier- und Frachtlasten hat dringend effiziente Managementverfahren erforderlich gemacht, die in der Lage sind, die Überwachungsdaten zur Optimierung der Verfügbarkeit, der Betriebs- und Wartungskosten und zur Gewährleistung der Sicherheit zu nutzen.

Die Bestimmung des Zustands von Eisenbahnanlagen basiert traditionell auf visuellen Inspektionen und Messungen vor Ort. Die Art, die Qualität und die Vollständigkeit der Messungen für Eisenbahngleise wurden seit den ersten Messungen der Gleisparameter in den 1960er Jahren verfeinert und erheblich verbessert. Ein wichtiger Katalysator in dieser Hinsicht war die Einführung der automatischen Speicherung, Analyse und Interpretation digitaler Messdaten, die von Diagnosefahrzeugen erfasst werden. Dabei handelt es sich um Spezialfahrzeuge, die mit hochentwickelten und oft kostspieligen Messsystemen ausgestattet sind, die jedoch nur selten das Netz befahren können, was eine begrenzte zeitliche Auflösung der Überwachungsdaten zur Folge hat. Heutzutage können On-Board-Monitoring-Fahrzeuge (OBM) oder in Betrieb befindliche Züge eingesetzt werden, die mit kostengünstigen Messgeräten wie Achslager-Beschleunigungsmessern (ABA) ausgestattet sind und das Schienennetz regelmässig befahren, wodurch ein nahezu kontinuierlicher Datenstrom entsteht, der mit dem Zustand der befahrenen Eisenbahninfrastruktur in Verbindung gebracht werden kann.

Das Hauptziel dieser Forschungsarbeit ist die Entwicklung effizienter Überwachungstechniken, die die Gewinnung diagnostischer Indikatoren für die Steuerung der Instandhaltung von Eisenbahnanlagen ermöglichen. Zur Bewältigung dieser Herausforderung stützen wir uns auf das Prinzip des On-Board-Monitoring, d. h. die Nutzung kostengünstiger Sensoren, die auf Fahrzeugen (möglicherweise in Dienstfahrzeugen) angebracht sind, für eine kontinuierliche und räumlich dichte Überwachung des Eisenbahnnetzes. Dennoch stellt die Umwandlung von Rohmessdaten, insbesondere Beschleunigungen, in robuste und umsetzbare Zustandsindikatoren eine grosse Herausforderung dar. Zumal es nicht trivial ist, die Messdaten nahtlos in bestehende Asset-Management-Prozesse zu integrieren.

In dieser Arbeit wird ein umfassender Ansatz für die Überwachung von Eisenbahnschienen vorgestellt, der sich in erster Linie auf die Auswertung von schwingungsbasierten Daten stützt und deren Wirksamkeit im Hinblick auf die Entdeckung von lang- (Längsebene) und kurzwelligen Defekten (Oberflächenfehler, Schweissnähte, Squats) bewertet. Im ersten Kapitel wird das Problem der Infrastrukturüberwachung kontextualisiert und der Stand der Technik in Bezug auf herkömmliche Gleisüberwachungsmethoden und vibrationsbasierte Überwachungsmethoden vorgestellt. Anhand mehrerer Fallstudien wird gezeigt, dass beschleunigungsbasierte Ansätze ein hohes Potenzial aufweisen, um herkömmliche Verfahren zu ergänzen oder teilweise zu ersetzen.

Im zweiten Kapitel wird eine robuste Methode zur Untersuchung des Einflusses einzelner parametrischer Anregungsquellen vorgeschlagen. Wir schlagen die Anwendung des Vold-Kalman-Filters vor, um die periodischen, geschwindigkeitsabhängigen Komponenten eines ge-

messen ABA-Signals zu identifizieren, wie z. B. die Schwellendurchgangsfrequenz und die Radaussenrundungsordnung. Die Komponenten der Unrundheit des Rades werden zur Rekonstruktion des Radprofils verwendet, während die Amplitude der Schwellendurchgangsfrequenz nachweislich ein Mass für die Gleissteifigkeit darstellt. Darüber hinaus wird gezeigt, dass Gleisabschnitte mit höherer Steifigkeit höhere Beschleunigungen bei der Schwellenübergangsfrequenz aufweisen, die mit höheren Kräften und verstärkten Instandhaltungsmassnahmen korrelieren.

Kapitel 3 befasst sich mit der Bewertung kritischer Komponenten der Schiene in einem so genannten Kurzwellenklassifizierungsschema. Die Schiene stellt ein Element dar, dessen strukturelle Integrität für die Betriebssicherheit wesentlich ist, das jedoch einer erhöhten Belastung ausgesetzt ist, die zu einem häufigen Auftreten von Fehlern wie Squats, Rissen oder Oberflächenfehlern führt. Kurzwellige schienenbedingte Effekte haben einen erheblichen Einfluss auf die Fahrzeug-Gleis-Dynamik. In diesem Kapitel wird ein beschleunigungsbasiertes automatisiertes Klassifizierungssystem zur Klassifizierung von Hauptschienenkomponenten und Oberflächenfehlern (gesunde Schiene, Oberflächenfehler und Squats, Isolierstösse und Schweissnähte) vorgeschlagen. Random Forests (RFs), die auf technischen Merkmalen trainiert wurden, werden Convolutional Neural Networks (CNNs) gegenübergestellt, die auf den Koeffizienten der Kurzzeit-Fourier-Transformation der ABA-Signale trainiert werden. Während die Validierung des Systems aufgrund der weitgehend fehlenden Referenzmarken nicht vollständig bestätigt werden kann, unterstützt die Implementierung die Automatisierung und motiviert die Implementierung von beschleunigungsbasierten Klassifikatoren zur Unterstützung der Früherkennung von Fehlern.

Der Einfluss der Zuverlässigkeit von Kennzeichnungen aus den verfügbaren Zustandsüberwachungsdatenbanken und die Einbeziehung von Expertenfeedback in den Analyseprozess schliessen den Kreis unseres vorgeschlagenen Systems. Der oben erwähnte Rahmen für die Kurzwellenklassifizierung stützt sich auf die Verfügbarkeit von Kennzeichnungen, die automatisch aus Bildern von Schienenköpfen extrahiert werden. Diese Kennzeichnungen spiegeln jedoch den sich ständig verändernden Gesundheitszustand der kritischen Komponenten nicht angemessen wider. Die abschliessende Studie dieser Arbeit konzentriert sich auf Schweissnähte als wichtige kritische Komponenten der Eisenbahninfrastruktur. In Kapitel 4 wird eine Proof-of-Concept-Studie vorgestellt, bei der von OBM abgeleitete Indikatoren, Datenbankinformationen und Expertenfeedback in den Kreislauf einbezogen werden. Die Studie wurde in Zusammenarbeit mit Asset Managern und Fachleuten der SBB durchgeführt. Das Framework kombiniert Vor-Ort- und Sichtprüfungen, Diagnoseinformationen und Experteneinschätzungen, um die Erkennung von Schweissnahtfehlern anhand von Ausreissern bei ABA-Indikatoren zu automatisieren. Die Bilder der Schienenköpfe, die zu diesen Ausreissern gehören, werden von Experten überprüft, was zur Identifizierung und weiteren Kennzeichnung der vermuteten Schweissfehler führt. Diese Informationsschleife wird weiter genutzt, um robuste automatische Fehlerklassifizierungsverfahren zu entwickeln. Zu diesem Zweck werden drei Methoden getestet: Binäre Klassifikation, Random-Forest-Klassifikatoren und logistische Regression nach Bayes. Wir zeigen, dass letztere die Schätzung zusätzlich mit einer Quantifizierung der Unsicherheit ausstattet.

Zusammenfassend lässt sich sagen, dass diese Arbeit einen Beitrag zur Entwicklung handlungsfähiger, überwachungsgesteuerter Indikatoren und Rahmenwerke für eine Reihe von Anwendungen leistet, angefangen von der eher globalen Identifizierung der Gleisgeometrie und -steifigkeit bis hin zur eher lokalen Identifizierung von Fehlern mit kurzer Wellenlänge, wie z. B. Oberflächenfehler oder beschädigte Schweissnähte. Noch wichtiger ist, dass sie einen ganzheitlichen Ansatz liefert, der automatisierte und kontinuierliche Überwachung, spärliche historische Zustandsinformationen und Expertenmeinungen zu einem Kreislauf verknüpft. Diese Forschung liefert die effektiven Werkzeuge und einen gültigen Konzeptnachweis für die Aufnahme von ABA-Daten in das Überwachungsprotokoll von Eisenbahnanlagen.

Résumé

La voie ferrée est l'élément principal de l'infrastructure ferroviaire, dont la tâche est de supporter la charge des véhicules ferroviaires qui la traversent. Sa structure est composée de nombreux sous-composants, à savoir la sous-structure, le ballast, les traverses et les rails, chacun d'entre eux devant répondre à des exigences différentes en termes de capacité de charge et de durabilité. L'augmentation récente de la vitesse des trains, de la densité du trafic et des charges de passagers et de marchandises a fait naître le besoin urgent de pratiques de gestion efficaces, capables d'exploiter les données de surveillance pour assurer l'optimisation de la disponibilité, des coûts d'exploitation et de maintenance et pour garantir la sécurité.

La détermination de l'état des actifs ferroviaires est traditionnellement basée sur des inspections et des mesures visuelles sur site. La nature, la qualité et l'exhaustivité des mesures effectuées sur les voies ferrées ont été affinées et considérablement améliorées depuis les premières étapes de la mesure des paramètres de la voie, réalisées dans les années 1960. L'introduction du stockage, de l'analyse et de l'interprétation automatiques des données de mesure numériques collectées par des véhicules de diagnostic a joué un rôle catalyseur important à cet égard. Il s'agit de véhicules spécialisés, équipés de systèmes de mesure sophistiqués et souvent coûteux, qui ne peuvent toutefois parcourir le réseau que rarement, ce qui implique une résolution temporelle limitée pour les données de surveillance. De nos jours, les véhicules de surveillance à bord (OBM), ou les trains en service, peuvent être équipés de dispositifs de mesure peu coûteux, tels que les accéléromètres de boîtes d'essieu (ABA), qui traversent régulièrement le réseau ferroviaire, offrant ainsi un flux de données presque continu qui peut être lié à l'état de l'infrastructure ferroviaire traversée.

L'objectif principal de cette recherche est de développer des techniques de surveillance efficaces permettant d'extraire des indicateurs de diagnostic pour guider la maintenance des actifs ferroviaires. Pour relever ce défi, nous nous appuyons sur le principe de la surveillance à bord, c'est-à-dire l'exploitation de capteurs peu coûteux, montés sur des véhicules (éventuellement dans des véhicules de service), pour une surveillance continue et spatialement dense du réseau ferroviaire. Néanmoins, la transformation des données de mesure brutes, en particulier des accélérations, en indicateurs d'état robustes et exploitables représente un défi important. D'autant plus qu'il n'est pas facile d'intégrer les données de mesure dans les processus de gestion des actifs existants.

Cette thèse présente une approche globale de la surveillance des voies ferrés, en s'appuyant principalement sur l'exploitation des données vibratoires, et en évaluant leur efficacité en termes de découverte des défauts de grande longueur d'onde (niveau longitudinal) et de petite longueur d'onde (défauts de surface, soudures, accroupissements). Le premier chapitre contextualise le problème de la surveillance des infrastructures et présente l'état de l'art en termes de techniques conventionnelles de surveillance des voies et de techniques de surveillance basées sur les vibrations. Grâce à des applications dans plusieurs études de cas, nous montrons que les approches basées sur l'accélération ont un fort potentiel pour compléter, ou remplacer en partie, les schémas traditionnels.

Le deuxième chapitre propose une méthode robuste pour examiner l'influence des sources d'excitation paramétriques individuelles. Nous proposons d'adopter le filtre de Vold-Kalman

pour identifier les composantes périodiques d'un signal ABA mesuré qui dépendent de la vitesse, telles que la fréquence de passage des traverses et l'ordre de déviation des roues. Les composantes de l'ovalisation de la roue sont utilisées pour reconstruire le profil de la roue, tandis que l'amplitude de la fréquence de passage des traverses est considérée comme une approximation de la rigidité de la voie. En outre, il est démontré que les sections de voie correspondant à une rigidité plus élevée présentent des accélérations accrues à la fréquence de passage des traverses, ce qui est corrélé à des forces plus élevées et à des actions de maintenance plus importantes.

Le chapitre 3 se concentre sur l'évaluation des composants critiques du rail dans ce que l'on appelle un schéma de classification des ondes courtes. Le rail constitue un élément dont l'intégrité structurelle est essentielle pour la sécurité de l'exploitation, mais qui est exposé à des charges élevées entraînant l'apparition fréquente de défauts tels que des déformations, des fissures ou des défauts de surface. Les effets de courte longueur d'onde liés au rail ont une influence significative sur la dynamique du véhicule et de la voie. Ce chapitre propose un système de classification automatisé basé sur l'accélération pour classer les composants principaux du rail et les défauts de surface (rail sain, défauts de surface et squats, joints isolés et soudures). Les forêts aléatoires (RF), qui ont été entraînées sur des caractéristiques techniques, sont comparées aux réseaux neuronaux convolutionnels (CNN), qui sont entraînés sur les coefficients de la transformée de Fourier à court terme des signaux ABA. Bien que la validation du système ne puisse pas être entièrement corroborée en raison de l'absence d'étiquettes de référence, la mise en œuvre approuve l'automatisation et motive la mise en œuvre de classificateurs basés sur l'accélération pour soutenir la détection précoce des défaillances.

L'influence de la fiabilité des étiquettes provenant des bases de données de surveillance de l'état disponibles et l'incorporation du retour d'information des experts dans le processus d'analyse ferment la boucle du schéma que nous proposons. Le cadre de classification des ondes courtes mentionné ci-dessus repose sur la disponibilité d'étiquettes extraites automatiquement des images du champignon ferroviaire. Cependant, ces étiquettes ne reflètent pas correctement l'évolution progressive de l'état de santé des composants critiques. La dernière étude de cette thèse se concentre sur les soudures, en tant que composants critiques importants de l'infrastructure ferroviaire. Le chapitre 4 présente une étude de validation du concept qui intègre dans la boucle les indicateurs dérivés de l'OBM, les informations de la base de données et le retour d'information des experts. L'étude a été menée en collaboration avec des gestionnaires d'actifs et des experts de terrain des CFF. Le cadre combine des inspections visuelles et sur site, des informations de diagnostic et des évaluations d'experts pour automatiser la détection des défauts de soudure à partir des valeurs aberrantes de l'indicateur ABA. Les images du champignon de rail correspondant à ces valeurs aberrantes sont examinées par des experts, ce qui permet d'identifier et d'étiqueter plus précisément les défauts de soudure suspectés. Cette boucle d'information est ensuite exploitée pour développer des systèmes robustes de classification automatique des défauts. Trois méthodes sont testées à cette fin : la classification binaire, les classificateurs forêt aléatoire et la régression logistique bayésienne. Nous démontrons que cette dernière permet en outre de quantifier l'incertitude liée à l'estimation.

En résumé, cette thèse contribue au développement d'indicateurs et de cadres de surveillance exploitables, dans une gamme d'applications allant de l'identification plus globale de la géométrie et de la rigidité de la voie à l'identification plus locale de défauts de courte longueur d'onde, tels que les défauts de surface ou les soudures détériorées. Plus important encore, il s'agit d'une approche holistique qui associe en une seule boucle la surveillance automatisée et continue, les informations historiques peu nombreuses sur l'état de la voie et l'avis d'experts. Cette recherche fournit les outils nécessaires et une preuve de concept valide pour l'assimilation des données ABA dans le protocole de surveillance des actifs ferroviaires.

Chapter 1

Introduction

1.1 Motivation and Problem Statement

Critical infrastructure systems, such as the power grid, our built environment and transport of people and goods, are essential for the functionality of modern societies and economies. However, growing requirements in developed societies place increased demands on such infrastructure. Specifically to what transportation is concerned, the ever-increasing demands for mobility result in increased complexity and interdependencies in transport networks, and railways in particular [34, 187]. In Switzerland, the railway network usage has increased by almost 40% in the last 30 years and this growth is projected to expand[112]. The amount of traffic per track kilometre in Switzerland, the Netherlands and Japan is the highest worldwide, approaching 100 vehicle-passenger-kilometre per kilometre, per day, about twice that of other European railway networks [249, 60]. In view of the growing population in urban areas and the increasing demand for climate-friendly solutions, traditional public transport (rail, urban buses...) and shared mobility solutions (car-, bike-sharing...) will dominate future developments [187, 259]. Such solutions are particularly relevant for urban areas, where about half of the world's population is concentrated today and which more than 70% of the population will inhabit by 2050 [133]. Such a booming growth and expansion, is linked with higher loads and a higher risk for disruptions due to system failures. This motivates new methods and tools for improving operation and maintenance planning.

Maintenance and renewals are necessary to ensure the safety, reliability, cost-effectiveness, compliance and efficiency of railway systems. The current approach to railway management relies on preventive maintenance schemes, which involve the scheduling of regular maintenance actions in advance in order to prevent degradation and to extend the lifespan of railway assets [200, 289]. While such a scheme reduces the risk of unexpected downtime (due to the reactive maintenance resulting from failed assets), it may also lead to unnecessary repairs when the assets are not actually in need of maintenance [88, 234]. To alleviate this disadvantage, predictive maintenance has been proposed as an approach that capitalizes on data and, often exploits models, for establishing predictive scheme for asset degradation in order to guide asset managers in optimal policy planning and scheduling of maintenance and remedial action [88, 155]. Predictive maintenance often relies on condition-based maintenance schemes, which harness sensor data and Structural Health Monitoring (SHM) approaches to determine the actual condition of the assets [7, 181, 46, 319, 149]. While Predictive Maintenance is more focused on the global state of the assets and predicting how their condition will evolve over time, condition-based maintenance is more focused on the present, monitoring the current state of the equipment [246, 20, 319, 88]. Maintenance and renewal actions result in short term costs, which can, however, result in lower costs over the life-cycle of the infrastructure.

Railway operators pursue a long-term sustainable strategy based on minimizing the costs

of the complete infrastructure over its life-cycle. Managing the infrastructure over a long-time horizon implies that high investment costs are occurring point-wise. In the case of Swiss Federal Railways (SBB), cost saving measures (Within the years 1990 - 2009) resulted in an insufficient number of renewals, which resulted in a large backlog buildup [112]. As a means to addressing such a backlog, preventive condition-based schemes enable the optimal decision of maintenance and renewal actions in order to ensure availability and safety [305, 289, 290, 10]. While preventive maintenance can alleviate these issues, condition-based maintenance can further eliminate unnecessary and costly interventions [10, 22, 136]. One of the main drivers for cost are the track assets, which account for over one third of the maintenance and renewal costs [43, 249].

The growing pressure faced by railway companies due to the previously identified challenges motivates new and innovative solutions, able to provide asset managers with precise information on the infrastructure condition. A means to this end is delivered by assimilation of intelligent and data-driven monitoring system into the management and maintenance procedures. Current options for monitoring and inspection of railway vehicles primarily include i) on site track inspections performed by field experts, such as switches or insulated joints, and ii) use of specialized diagnostic vehicles, which are equipped with high-end monitoring equipment [247], such as optical and inertial sensors [47, 14], or even more refined Non-Destructive Evaluation (NDE) tools, such as ultrasonic or Eddy current methods [124, 274].

In Switzerland, track monitoring is accomplished by means of Diagnostic Vehicles (Diagnose-fahrzeug -DFZ/gDFZ), which use optical methods to provide information on track geometry (alignment, track profile, short-pitch corrugation, rail wear) at a frequency of twice a year for regular tracks, or on a monthly basis for high-speed tracks. The diagnostic data pertaining to the health of the infrastructure collected by these vehicles is then forwarded to decision aid tools, such as the RCM Viewer [247] or SwissTAMP (Swiss Track Analysis & Maintenance Planning) [305], which support the asset managers in planning maintenance and renewal actions. As mentioned by Lederman et al. (2017) [154] monitoring of track geometry provides essential information for the safe operation of the rail infrastructure, albeit requiring special scheduling of the diagnostic vehicles, which limits their operation to relatively rare intervals in order to alleviate the associated down-time and availability restrictions [25, 71].

To overcome these hurdles and reduce Life-Cycle Costs (LCC), recent work has focused on shifting the monitoring paradigm from sophisticated monitoring vehicles to in-service passenger trains [40, 188, 167, 260]. Recent advances in sensor technology enable the simple adoption of relatively low-cost, yet sufficiently accurate sensors on revenue trains; a process that is here referred to as On-Board Monitoring (OBM). In Switzerland, a number of OBM projects, such as the OBM-ICN and the SOB-CTM [260], have been recently initiated. In practice, OBM is achieved through continuous vehicle reaction measurements of dynamic nature (vehicle position, accelerations, wheel/rail force) that are drawn from IMUs (Inertial Measurement Units) mounted on regular passenger vehicles (class "ICN"). However, accelerations in their raw form are of limited use; it is only after these measurements are processed (e.g., via feature extraction) and linked to condition that such information gains exploitable value (see Section 1.2). Aside from the technical challenge of building a complete data chain able to convey measurements from the sensors, to the processing, to the data warehouse, and eventually the end user, it is important to ensure that meaningful and repeatable indicators are established. This is commonly realized by defining and extracting condition indicators, which can inform on condition and support decisions on downstream tasks. Such indicators are often fashioned with a view to achieve detection of damage, which also forms the first level of the SHM hierarchy [238]. Such diagnostic tasks necessitates the definition of thresholds which can be used to classify the monitored condition as normal or abnormal [149, 10, 326, 52]. Reliance on this first SHM level can often prove insufficient for informing maintenance decision making, as the type and

severity of defects is ignored. The use of vibration.based monitoring, which forms an essential OBM tool, can facilitate more refined assessment. The measured dynamic parameters enable the extraction of more informative features,which enable to track the progression of the fault severity.

This thesis fuses on mobile sensing as a means to railway condition assessment. The developed tools rely on vibration-based data collected from Axle Box Accelerometers (ABA) that are mounted both on the "gDFZ" diagnostic vehicle, as well as the "OBM-ICN" vehicle [107]. In the first case, a point is made as to the potential of such sensors in complementing or possible replacing further diagnostic technologies. In the second case, we explore availability of low costs ABA solutions for ensuring a spatially and temporally denser supervision, as the in service trains regularly traverse the railway network. It is demonstrated that the indicators obtained from such a vibration-based condition monitoring approach offer a robust equivalent to optical sensors (which are often hard to maintain and preserve at a high-quality/clean state), and complement the more specialized non-destructive evaluation sensors (Ultrasonic testing, Eddy current), which are often fragile and sensitive to local features of the tracks.

1.2 State of the Art and Open Challenges

This section offers an overview of the current state of knowledge on condition-based assessment of the railway track infrastructure. The discussion first gives an overarching view on condition monitoring of railway track assets using specialized equipment, before describing dynamic based assessments using vibration measurements. Finally, the chapter concludes on the open challenges that are explored as part of this thesis.

1.2.1 Diagnosis and Prognosis

In the field of condition monitoring of railway track infrastructure, diagnosis and prognosis are two essential aspects that are closely linked. Diagnosis refers to the process of identifying the presence of a defect or a potential problem in the track, while prognosis refers to the prediction of the remaining useful life of the track components [8, 198, 85]. Railways assets typically degrade over time under repeated loading cycles, implying a slow deterioration process [167, 267, 266] that is usually modelled using empirical analysis (statistical approaches) [67]. Damage, thus, initiates earlier than it can be detected, it is eventually picked up upon its progression by the monitoring systems in place, and if no maintenance is performed failure is eventually observed. This degradation process is illustrated qualitatively in Fig. 1.1 along with the different maintenance strategies explained in the subsequent paragraphs. Maintenance and repairs that take place closer to failure occurrence result in high costs, while premature maintenance may result in unnecessary investment/expenditure. Thus, the challenge for the asset manager is to determine the optimal time point for maintenance and renewal in terms of safety, availability and costs [220].

In the absence of information regarding the state of the assets, reactive or preventive maintenance schemes form the main types of available strategies for guiding policy planning. *Reactive maintenance* follows the run-to-failure methodology, which is the repair and/or replacement work after an equipment outage has occurred [155]. This can lead to increased costs due to urgent repair work and unexpected downtime. To avoid this, *preventive maintenance* is performed based on a certain periodic interval to prevent and correct problems before breakdown without considering the actual health condition of a system [155, 246, 22], often aligned to be slightly shorter than the typical damage initiation of failure interval. In an ideal scenario, *just-in-time* (JIT) maintenance is applied just before failure occurs [155]. This cannot however be ensured via scheduled (preventive) maintenance.

A switch to data-driven maintenance paradigms allows to track and detect damage at an earlier stage, by relying on condition indicators that are used to inform maintenance and renewal decisions [319]. However, a considerable amount of degradation is often required to reach these thresholds, which hinders early damage prevention and is not optimal as scheduling maintenance and renewals is subject to time constraints [316] while the procedures when predefined defined limits are exceeded are not always clearly defined as described by Weston et al. [304].In the case of faster degradation processes, damage and failure may altogether be missed in certain cases, if the condition is not monitored at sufficiently frequent intervals. *Predictive maintenance* seeks alleviate this problem by shifting from strict thresholds to a more continuous oversight over the measured asset condition, evaluating how the asset evolves over time, in order to predict degradation trends, enabling earlier detection than *condition-based maintenance* which poses more emphasis on real-time inspections [317, 155, 132].

In the following sections, we will focus on the aspects of deriving condition indicators to serve predictive maintenance schemes, exploiting data from both state-of-the-art specialized measurement systems, as well as from lower cost ABA measurements.
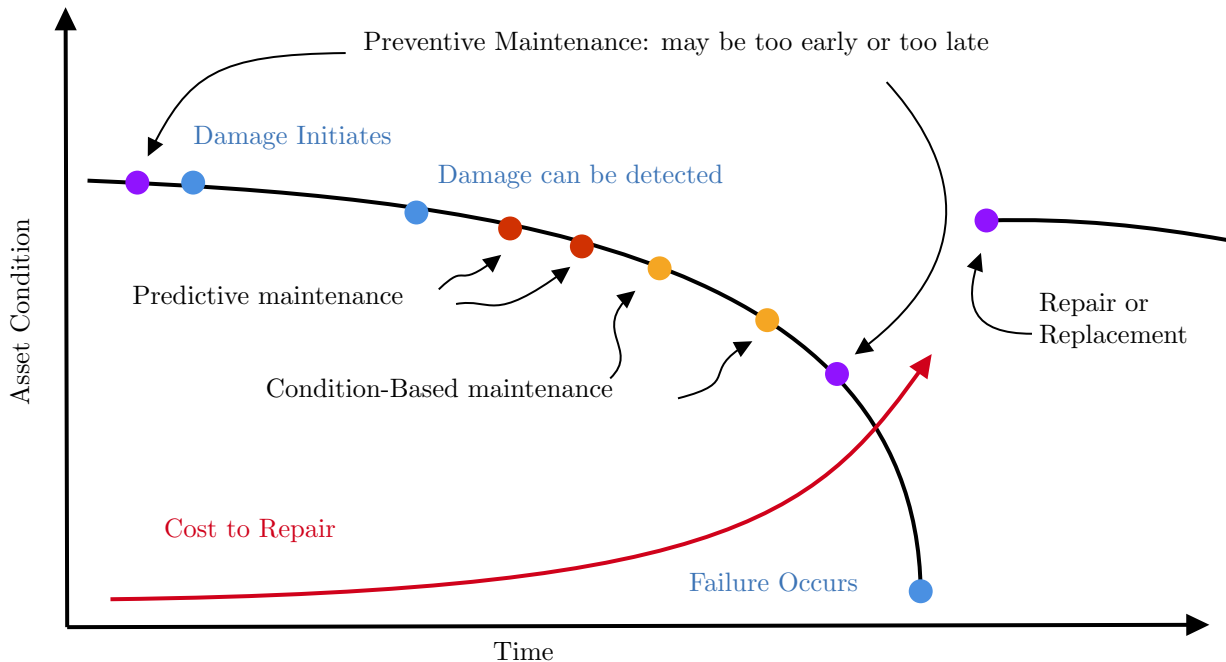
Figure 1.1: Conceptual illustration of the evolution of the condition of a railway asset and different types of maintenance. Maintenance closer to the failure of a component results in higher costs. Preventive maintenance is applied at scheduled intervals depending on the degradation rate of the asset class. Condition-based maintenance prescribed thresholds on measured condition indicators in order to signal maintenance requirements. Predictive maintenance uses the measured asset condition to predict trends in degradation, which enables earlier detection and intervention [155].

1.2.2   Monitoring Data for Informed Condition Assessment

As previously outlined, knowledge on the condition of the railway assets is of central concern to ensure efficient, reliable and safe operation. Infrastructure maintenance and renewal planning is nowadays performed in a data-rich environment, supported by four main information sources available at any position on the railway network: fixed asset information (substructure type, superstructure type, rail type, age of each component), maintenance information (number of

performed rail grinding, tamping...), loading conditions (number of axles per year, total gross tons per year) and asset condition (measurement data, inspection data and condition logs) [200, 149, 183, 27].

Fixed asset information, maintenance data and loading conditions are primarily used by railway operators to derive investment and maintenance strategies. These maintenance strategies often rely on classification schemes, such as the so-called "Standard Elements" classification, which group similar track types and loading conditions to estimate attributes such as expected lifespan, average maintenance and life-cycle cost [200, 77, 313]. Railway companies, such as the SBB and Austrian Railways (ÖBB), use the Standard Elements to improve their long-term maintenance and renewal decision making [200][77]. This approach combines data–driven approaches [77] with expert know-how to determine service life and maintenance requirements [305], in turn enabling preventive maintenance schemes.

Predictive or condition-based maintenance schemes, however, require more granular information on infrastructural condition [22]. Infrastructure condition is nowadays mainly assessed with diagnostic vehicles which deliver a large range of indicators pertaining to the track health such as track geometry measurements, images of the railhead and of the track bed, subsidence measurements, vibration analysis or ultrasonic testing. The systems used currently by railway authorities are summarized in Sec. 2.3.1. Aside from the automated inspection that is achieved via diagnostic measurement vehicles, at somewhat sparse time intervals, certain inspection tasks still rely on manual on-site inspection conducted by field experts. Such inspections are performed in two week intervals on critical components such as switches [121, 293] or, in some cases, are triggered by the manifestation of irregular behavior. The advantages of automated inspection via measurement vehicles are obvious from the point of view of safety, as well as in terms of optimizing cost and efficiency. Consequently, the railroad industry is currently researching approaches to automatize the tasks that still rely on on-site inspections today [25, 222].

Asset managers currently require component-based assessment of the condition, in order to optimally schedule maintenance. This is achieved, as described in Sec. 2.3.1, via a large number of state-of-the-art monitoring systems that exist to monitor the condition of railway track infrastructure. These techniques are used to detect a wide range of defects such as stiffness, track geometry faults, rail surface defects or squats, cracks in sleepers, rail corrugations, rail profile, subsurface defects and internal defects in the rails (see also Sec. 2.2). The track components monitored via diagnostic measurement vehicles are illustrated in Fig. 1.2. Certain components, such as the substructure, are more recently assessed via data-driven indicators. An example for this is found in the so-called "fractal values" indicator, which are derived from the longitudinal level, which are an indirect indicator of the substructure condition [150, 149, 148]. The rail is one of the most critical components as it is exposed to high stresses by directly carrying vehicle loads. For this reason, within the Swiss context, the rail is regularly monitored to detect both internal flaws, using ultrasonic and Eddy-current measurements, as well as surface flaws via images captured by the rail head imaging system ($VCUBE$). The latter are used to detect faults in critical rail components such as welds, insulated joints and surface defects [247]. This automated detection is achieved with the aid of an algorithm developed by SBB and CSEM (Innosuisse Project: RailCheck mit Fingerprinting).

One limitation of the monitoring via specialized vehicles is that their relatively high costs and infrequent measurements limit a continuous assessment of structural condition. In Switzerland, for instance, at a measurement frequency of only twice a year for the complete network, the delivered information is mostly exploitable for faults with a relatively low degradation rate [52]. For faults that feature fast growth such as squats, the frequency of measurements would need to be increased in order to provide a sufficient observation rate for condition-based assessment and predictive maintenance [167]. In the OBM scenario, therefore, Structural Health Monitoring

(SHM) would offer a way to continuously extract information on system performance in an almost automated manner [74]. The result would be more reliable condition status reports and, accordingly, more efficient just-in-time maintenance.
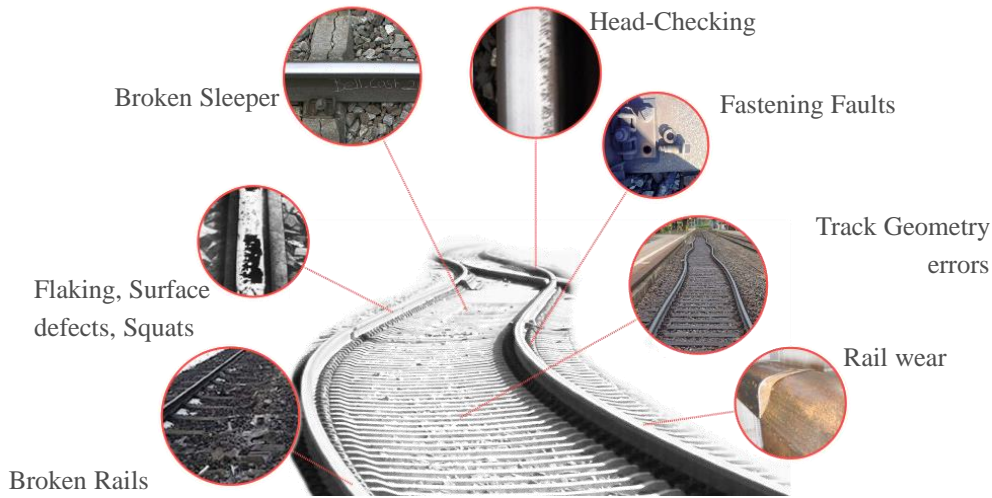
Figure 1.2: Illustration of the assets monitored on the railway track.

1.2.3   On-Board Monitoring Applications

On-Board Monitoring setups record the interaction between track and vehicle, which is influenced by the boundary conditions between vehicle and track, as well as by the specific vehicle parameters. Boundary conditions include track parameters such as mass, stiffness and rail-wheel contact parameters [140]. The vehicle's response to excitations caused by vehicle and track irregularities is a function of these parameters. Irregularities have several scales ranging from below a centimeter to over a decameter [150]. They can be periodic (short pitch corrugation, waves) or transient (turnouts, alignment, squats, welds, rail joints) [107]. The resulting excitation relates to track geometry (longitudinal level, alignment), rail defects (squats, cracks, corrugation) and wheel condition (damaged bearings, wheel Out-Of-Roundness and flat spots) [201]. An effective OBM setup needs to minimize the sources of variation (noise) in the data and, thus, maximize the information that needs to be extracted. Different approaches are followed to this end by railway companies and researchers.

Hardware implementations of OBM setups, which rely on ABA measurements, have been tested with promising results on a number of applications [201]. A complete OBM setup, illustrated in Fig. 1.3, includes accelerometers on several axles, on the bogies and the body. In some implementations, tensiometric wheelsets, which are specialized systems used to measure wheel/rail forces, are also included. The feasibility of extracting track quality indicators (TQIs), which are indicators summarizing the asset condition [316], in relation to the track geometry (waverange 3-70 m), from such measurements has been successfully demonstrated by means of both high precision sensors [280] [170] [191] as well as cheap smartphone accelerometers [314]. In several of these studies, the track geometry data is obtained via double integration of the ABA measurements and application of appropriate filters. Since 2018, for instance, such data has been reported weekly on the network of SOB [260]. In fact, a similar system is used in Germany by DB Systemtechnik, where certain tracks are monitored and maintained using indicators extracted from accelerometer measurements [196].

The identification of defects on a small scale (< 3 m) has been mainly investigated via adoption of machine learning techniques, with currently no known commercial products or applications. TU Delft has developed algorithms to detect small scale rail surface defects, such as early stage squats or damaged insulation joints, using data from their OBM vehicles [188]. Studies carried out in Delft have shown that it is possible to reduce the signal-to-noise ratio in the measurement data by using the longitudinal signal from the ABAs, which is less influenced by the track properties, and results in a better detection of small scale irregularities [189, 190, 175]. In the UK, on-board self-powered axle box sensors are used to identify wheel and bearing condition, as well as track quality, by characterizing the energy levels in frequency bands [61]. On-board accelerometers have shown high potential for monitoring high-speed lines; several studies illustrate that indications on the condition of rail welds [9] and rail pads [311] can be yielded by combining acceleration measurements with vehicle-track interaction simulation. The interpretation of acceleration-based measurements often relies on system identification techniques [107, 106] that are overviewed in the next section.
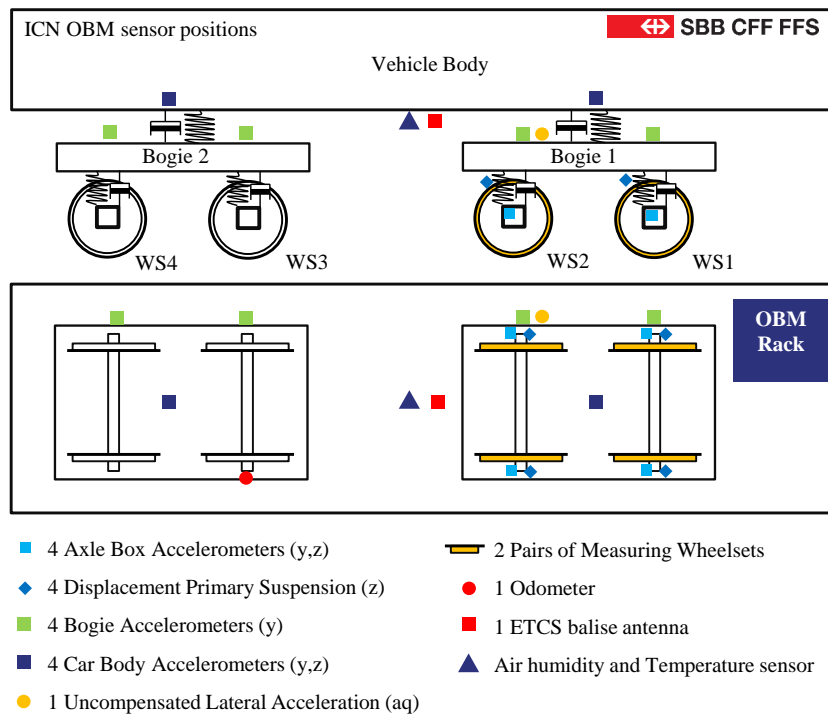
Figure 1.3: Overview of sensor positions for a complete OBM setup of an in-service train at SBB. ABAs form a salient means to the current data-driven approach to assessment of the track condition.

1.2.4 System Identification in Railways

The assessment of condition on the basis of dynamic response can be traced back to the initial work conducted, more than 200 years ago, for deepening the understanding of the vehicle/track interaction effects. Here, simulation played a crucial role. This includes running stability investigations based on analytic solutions of the railway vehicle motions [137, 41, 51]. Breakthroughs in fundamental research on stability and contact mechanics were achieved in the 1970s [186]. These linear and nonlinear stability investigations, were encoded into specialised software, such as Simpack, that enable the simulation of vehicle/track dynamics. Simpack, Medyna or Adams/rail are Multibody Body Simulation (MBS) engines, in which the vehicle

and track are assembled using substructures consisting of bodies and joints with specific properties [235]. For complex investigations MBS software, such as Simpack. allow for considerable flexibility in simulating the vehicle response [141]. Analytical solutions relying on simplified models have been commonly adopted [140] for the analysis of the vertical and lateral vehicle axle, bogie or body dynamics. To what concerns coupling of such available models, across their different granularities, with measurements, this often requires simplifying assumptions in terms of boundary conditions, contact mechanics, or abstraction/elimination of certain degrees of freedom. However, such a coupling is important to achieve in order to validate the employed models by in-situ measurements [258] or by means of hybrid-simulations [96]. Hybrid simulation is a technique which combines the use of physical testing and numerical simulations in order to analyse the behaviour of mechanical systems which have subcomponents that are highly nonlinear or safety critical which cannot be adequately modelled [257]. While such simulations enable the definition of baselines for condition indicators under varying track types and operating conditions, the variety of excitation sources and track characteristics poses a challenge in inferring the system parameters or condition [123].

One of the main excitation sources on the railway vehicles stems from the geometric alignment of the track. The characterization of the track geometry in terms of roughness and wavelength is fundamental for understanding the degradation mechanisms of the track [150]. Identification of the track alignment, short wavelength defects or the condition of the vehicle wheels or bearings from acceleration measurements can follow two main tracks, namely model-based or purely data-driven approaches [107]. While data-driven models are successful in terms of detecting changes, they are often disadvantageous for interpreting the measured variation, since they do not explicitly link to the underlying physics. On the other hand, interaction models suffer from simplifying a-priori assumptions, which often result in inaccuracies [141]. An in-depth review of parametric methods, which are often used in combination with vehicle-track models, is provided in Sec. 2.5.

In the first case, a-priori models of the vehicle (train & suspension system) are fused with measurement data, revealing an analytical connection to the underlying track geometry [231, 210, 69]. Changes in the estimated track profile across several crossings are then associated with a possible deterioration. Beyond the recovery of track geometry, model-based assessment is in many studies exploited for characterizing the effect of the track stiffness on the vehicle response [192, 29, 37, 36]. The vertical track stiffness is an important parameter since it strongly affects the wheel/rail contact forces. The stiffness is usually assessed via in-situ measurements by measuring the subsidence of the rail or sleeper under the load of a train or via specialized measurements systems. The continuous measurement of the track stiffness is obtained from so-called subsidence measurement vehicles (EMW – Einsenkungsmesswagen) [262]. Alternative track stiffness assessments exist, that exploit the dynamic response of the track to an oscillating mass, which is either connected to the axle in the case of the Portancemetre [192] or that is hydraulically actuated axle in the case of the Rolling Stiffness Measurement Vehicle (RSMV) [29]. The assessment from oscillating axles relies on the assumption of a simplified underlying spring-mass-damper systems to inversely determine the track stiffness.

In the second case, the assessment relies on the exploitation of available data and their processing in the temporal and spectral domains [153] [50]. In Sec. 2.6, we present a comprehensive overview of data-driven approaches including time-frequency analysis, numerical integration, and statistical features. The observed wavelengths of the faults on the track depend on the type of defect. For this reason, different methods of time-frequency analysis have been applied in research to extract wavelength or frequency specific components. Concerning longer wavelengths, accelerations signals are usually related to faults in the track geometry [280] [170] [191]. Short wavelength faults cause impacts on the axle, which appear as large acceleration peaks. The dynamics for such faults are quite complex as they form a combined result of the

vibrations of the axle, rail and wheel system, each of which include many unknowns. For this reason, data-driven approaches have been preferred over model-based schemes for assessing such types of faults. As an instance of such an approach, an application for squat detection, using ABAs, has been developed at TU Delft. The accelerations are filtered to remove wheel-related excitation before being transformed into a time-frequency representation using the continuous wavelet transform [188]. More recently, Yang et al. [319] have shown that both feature-based methods and Convolutional Neural Networks (CNNs) can be applied to effectively detect insulated joints from acceleration measurements. Tsunashima and Takikawa [283] propose an approach based on outliers identified from the CWT spectrograms, which were analyzed by experts who identified that ABA based detection has a higher false positive rate for small faults. In their work, Shadfar et al. [250] propose an index based on the combination of Fast Fourier Transform (FFT) coefficients, Principal Component Analysis (PCA), and a non-linear neural network for evaluating the health of rail welds. One limitation identified by Shadfar et al. is that their proposed index only serves as an indicative tool that must be further enhanced with expert judgement to build a more adequate dataset for full automation. Xiao et al. [312] combine the wavelet packet decomposition (WPD) with adaptive synchro-squeezed short time Fourier transform (ASSTFT) to locate poorly diagnosed welds on a heavy-haul railway line. Multi-frequency analysis often involves the decomposition of the signal into several modes via empirical mode decomposition (EMD) or the Hilbert-Huang transform (HHT) on the acceleration signals. The HHT has been used to characterize abnormal vibrations in damaged welds on tramway lines [126] and the EMD has been successfully applied to extract condition indicators for the catenary [115]. Such approaches, although more computationally intense are particularly well suited for non-stationary signals such as the measured ABA.

1.2.5   Uncertainty quantification

The vibration response measured on the axles, the bogie and the body is determined by the interaction dynamics between the vehicle and the track. The establishment of robust indicators from vehicle track dynamics requires the incorporation of prior knowledge relating to the track assets or the vehicle. Vehicle-track dynamics are a function of vehicle and track parameters. Vehicle parameters are generally provided by the train manufacturer during the homologation process, while track parameters are less clearly known since not all components are inventoried with the same level of detail. The vehicle parameters are non-linear and usually time-invariant. Time-variant vehicle properties such as passenger load, rail-wheel-rail contact parameters, wheel condition and dynamic are uncertain and have to be determined or estimated.

In general, existing uncertainties can be classified into the aleatoric and epistemic class. Aleatoric uncertainty pertains to the statistical uncertainty that arises due to unknowns that cannot be easily determined, eventually inducing variance in the output of an experiment [254, 197]. Noise during measurements and unknown vehicle track boundary conditions are aleatory, as no device can record every source of variance. The variance in the vehicle response for each experiment is partly due to parameters which are unknown and unmeasured. For example, the exact rail-wheel contact geometry at any point in time and the rail-wheel friction coefficient are unknowns that can only roughly be estimated. The aleatory uncertainty in a measurement system is often modeled by considering the statistical properties across multiple observations [197, 70, 38].

Epistemic uncertainty, also called systematic uncertainty, stems from factors that in principle can be inferred, but which are missing in practice. The source of this uncertainty is the incomplete knowledge of the railway system. Incorporating models and parameters can reduce the epistemic uncertainty. Epistemic can be reduced by adapting the model and its parameters. For instance, one common source of epistemic uncertainty in monitoring vehicles

is vehicle localization, whose error stems from the combination of multiple systems such as odometric data, track transponder data and in some cases GPS data [237] and from their respective measurement noise. Adding more channels via data-fusion processes can reduce the noise or errors in the estimated location of the vehicle (see also Sec. 2.3.3). Such uncertainties can be quantified via probabilistic modelling, uncertainty propagation (Monte Carlo simulation and polynomial chaos expansions) [193][17] and sensitivity analysis (correlation measures, Sobol' indices) [38][270]. Bayesian logistic regression is a statistical modeling technique that uses Bayesian inference to estimate the uncertainty associated with the predictions made by a logistic regression model via Markov Chain Monte Carlo (see also Section 7.4.2).

1.2.6 Adaptive Approaches

A common challenge when developing diagnostics or prognostics for the condition assessment of the railway track is the integration of the variability due to the previously identified uncertainties. Differences encountered in practice may stem from the following situations:

1. Variations in track type, such as different ballast conditions, sleepers, pads or rails.

2. Variations in the vehicle parameters for different monitoring vehicles.

3. Changes in operating condition such as the vehicle speed, which varies along the track position depending on the regulation.

Considering the large diversity of asset choices and operating conditions observed in railways, the resulting dynamic characteristics can vary widely, which impedes the applicability of methods beyond the tested scenarios. Finally, for the case of fully equipped vehicle fleets, the repeatability of the methodologies applied across several vehicles must be demonstrated. It is thus important to consider the varying characteristics through data processing techniques in order to ensure that the condition indicators are stable over time and space.

Several adaptive approaches and transfer learning approaches have been applied in different fields to generalize models beyond one specific set [82, 115]. In many studies this issue is circumvented by narrowing down the scope of the research, focusing on a small track sections and a limited number of defects [312, 250, 283]. The vehicle speed is often in such scenarios ignored, or is taken into account via linear correction factors [189]. The applicability of such methods decreases when increasing the monitored track sections, since more causes of variation are inevitably included in the data. Therefore, models or approaches that can adapt to different conditions are essential, since they enable the generalization to other conditions.

So far, adaptive approaches for assessing the condition of the track is still a major research issue [279]. Such approaches are particularly important for features that are dependent on other environmental or operational parameters [82]. For instance, time-frequency approaches are usually used to separate sources of vibration occurring at different scales (short wavelength rail faults to long wavelength track geometry faults). In the case of varying track properties, these sources of variation are often ignored, because the fixed asset data information is in many studies incomplete or not available. Longitudinal ABAs have been for example adopted by Molodova et al. to characterize squats, since these accelerations are less dependent on track properties [189]. Vehicle speed is often accounted by applying regression techniques to correct for its influence [189, 134]. However, such an approach only partially solves the issue, as these regression factors depends on more factors than only the speed. Adaptive approaches may further be enabled by including expert knowledge [115], since this additional information improves adaptation to different observed conditions or individual assessment.

1.2.7   Research gap and open challenges

The above mentioned research works reveal the large potential of On-Board Monitoring via ABAs for assessing the structural condition of the railway track. However, they simultaneously shed light on the challenges in extracting meaningful information to deliver reliable indications on the state of the railway assets.

Dynamic measurements intrinsically capture the response to a large number of excitations of a system that has multiple degrees of freedom and resonant modes. For instance, changes in track stiffness due to parametric variations of track parameters strongly influence the dynamic response of the axle. The resulting effects on contact forces and wear have been theoretically shown, but the practical implications have not been demonstrated [140, 286, 64].

Another aspect is that in most academic studies, the case studies remain quite limited in the length of the track or in the assessed faults, which reduces the actual problem complexity. While previous studies demonstrate the large potential for an acceleration-based assessment of short-wavelength faults, most studies do not perform component-based assessment and often neglect the assessment in term of severity due to difficulty in quantifying magnitudes of faults. Indeed, most of the published research is either restricted to small case studies with dozens of samples on short track lengths, or does not provide a sufficient level of granularity when assessing the individual components on the track [250, 326, 283, 312].

To enable large scale adoption and acceptance, the condition indicators delivered from ABA-based processes must be in tune with the needs of experts and asset managers. Expert feedback is an essential aspect to include when assessing asset condition for structural health monitoring applications, as expert feedback enables improved informed assessments. Expert feedback and labels generally stem from condition monitoring databases or alternative feedback processes. These labels generally feature an uncertainty that must also be accounted for when assessing the infrastructure using such information sources [181, 115, 279].

Overcoming the challenges in generating transparent ABA-based condition indicators and fusing them with expert feedback would ultimately lead to a comprehensive track condition assessment. Such an approach relying on low-cost sensors installed on a train fleet would result in large-scale continuous infrastructure condition assessment. In this work, we aim to address the identified challenges with the goal of achieving reliable condition monitoring using ABA measurements and infusing these in a holistic process that exploits machine learning tools and statistical methods for uncertainty quantification and robust diagnostic indicators, which are then offered to experts for further corroboration and assessment; in this way closing the loop.

1.3   Theoretical Background and Methods

As previously mentioned, acceleration-based condition monitoring is of great relevance for railway operators and poses a particular challenge when considering the more challenging feature extraction and interpretation in comparison to geometric measurements or visual inspection, whose outcome has an almost obvious connotation in terms of conditions and faults. While considerable work has been done in the field of asset monitoring via ABAs, the identification of essential parameters, such as the track stiffness or the condition of specific components such as welds, insulated joints, squats and surface defects still requires further analysis.

In this section, we present the general topics that are closely related to this dissertation. We introduce acceleration-based monitoring approaches and the relevant methods developed in the literature. A more detailed review is provided in the respective chapters.

1.3.1 Railway Infrastructure and Data Assets

Asset managers nowadays base their decisions on a large number of data sources. This expert-based approach enables more reliable decisions by integrating a multitude of factors, but such experiences and know-how is difficult to reproduce in a data-driven way. By incorporating prior knowledge about the infrastructure and the evolution of its condition over time, the uncertainty of condition indicators can be reduced. Infrastructure managers work within a data-driven environment to pursue their long-term maintenance strategy planning. Fixed Infrastructure Assets (Datenbank feste Anlagen - DfA, SBB [306]) , scheduling data (Anabel, SBB [113]) , deviation logs (Zustands-Monitoring - ZMON, SBB) , maintenance logs (Auftragsmanagement - AMGT, SBB) , and measurements (ICN-OBM, DFZ and gDFZ measurements, SwissTamp, SBB) are data streams that provide a picture of the infrastructure over its entire life-cycle [27]. This information is usually logged by most railway operators. In Switzerland, SwissTamp (SBB) forms a decision aid tool that is used by asset managers for track analysis and maintenance planning [200]. Main limitations are communication protocols, access restrictions and the processing of bulk information, which can sometimes be outdated and incorrect. Research has shown that these data streams contain valuable information that should not be neglected [27]. For example, infrastructure components such as road and railways bridges or railway tracks are still assessed based on manual inspections and other metadata. These data combined with data-driven decision support, can be used for condition assessment, remaining lifetime estimation and action recommendation [44]. Throughout this thesis, these data sources are used in order to explain and improve the proposed approaches to condition assessment.

1.3.2 Parametric and non-parametric methods for feature extraction

The condition of a structural system is often characterized by extracting indicators from measurements using parametric or non-parametric approaches as noted in Sec. 1.2.4. The traditional approach in Structural Health Monitoring requires the initial computation of characteristic indicators from time-series. More recently, the direct classification of time-series given as an input to deep learning frameworks such as Convolutional Neural Networks has been proposed [318, 146, 75]. While this approach in principle bypasses the need to calculate some features, it is more challenging to train such models since they require more input data to learn the optimal weights of the model. A comparison of feature based and time-series based approaches is given in Chapter 5 to enable the direct classification of the time-series.

The methodologies applied to extract features in this work are described in Sec. 2.6 and 2.5 for parametric methods and non-parametric methods respectively. The methods for feature extraction fall in the following categories:

- Displacement-based indicators: longitudinal level and lateral displacement can be recovered from ABA using Bayesian filters such as the Kalman Filter (see also Sec. 2.5.3) or via direct integration techniques (see also Sec. 2.6.2). They characterize the level of roughness of the track at wavelengths between 0.5 and 70 m.

- Time frequency representations: coefficients of the Short Time Fourier transform (STFT), the Discrete Wavelet Transform (DWT), or the Continuous Wavelet Transform (CWT) enable the characterization of the vehicle response in both time and frequency. These methods are especially well adapted for short wavelength defects.

- Signal filtering: Band-pass filtered time series to selected frequencies of interest. Vold-Kalman filter, which is a parametric approach used to extract periodic components from a noisy signal as a function of vehicle speed. The time series from such filters may

contain less noise by focusing on frequency ranges of interest. The Vold-Kalman filter is for example used to separate the wheel Out-Of-Roundness (OOR) and sleeper passage response from other sources of noise in Section 3.

- Essential indicators: summary statistics (mean, maximum, minimum, quantiles, kurtosis...), energy and entropy are computed on the extracted features and correspond to a reduced set of indicators for the underlying condition of the track.

These above-listed features enable an efficient and transparent application of outlier analysis and classification schemes for the assessment of the track condition. The next section introduces different methods that can be applied on these indicators for the assessment of track condition.

1.3.3 Outlier analysis and classification

Outlier analysis and classification are the subsequent stages following feature extraction with the purpose of inferring condition labels or component classes from the ABA data. The traversed (and measured) component or the related condition labels are in some cases known from inpection logs, which enables the use of supervised classification approaches. Supervised classification are used in a wide range of SHM applications to predict class labels from unseen data. Models are first trained on samples with known condition labels. Special care must be taken in order to ensure a good separation between the training data, the testing data and the validation data.

Decision Trees (DTs) and Random Forests (RFs) are commonly utilized machine learning algorithms that generate interpretable results for classification [181]. However, these algorithms perform optimally when utilizing a smaller set of parameters. RFs, in particular, are a type of ensemble model, which aggregates the results from multiple decision trees to produce a more reliable prediction compared to a single DT.

With the proliferation of time series data, Deep Neural Networks (DNNs) have gained increasing popularity for solving complex time series classification problems [75, 318, 146]. DNNs are capable of achieving results that are equal to, or better than, those produced by traditional machine learning methods such as Decision Trees (DTs) and Random Forests (RFs), without the need for feature engineering. It is important to note, however, that the injection of prior knowledge or physics into the learning system can still be advantageous in capturing the underlying dynamics of the data, as demonstrated in [146]. The ability of DNNs to automatically extract useful representations from the input data based on examples is what makes them a highly effective tool for solving complex problems. Convolutional Neural Networks (CNNs) have often been utilized in time series classification tasks. For example, a CNN was applied to raw axle box acceleration data for the detection of insulated joints in [318].

Supervised classification approaches are applied to solve two aspects of monitoring railway assets challenges in this work. The first aspect of classification between rail components such as welds, insulated joints, and squats or surface defects is described in Chapter 5. The labels automatically obtained from $VCUBE$ images (see Section 1.2.2) are used to identify acceleration time-series and features. In this study, CNNs trained on ABA time-series are compared to RF models trained on a set of essential indicators.

Such classification approaches are well suited for training models to learn the relation between the ABA features and the class labels when a large number of labels is available. However, sometimes such labels are not available. In Chapter 5, it is observed that some components such as welds do not have condition labels, which hinders the assessment of their condition.

In the absence of labels regarding the condition of the assets, outlier analysis is a statistical technique used to identify unusual or extreme data values in a dataset. The abnormal values identified may for example relate to faults. In this thesis metrics that quantify the distance

of data points to minimum acceleration levels in terms of quantiles are proposed. Outliers are samples which show abnormally high acceleration levels for their respective component class (e.g., welds). This metric is proposed for welds, which are a component whose condition is not known, since the automatic detection algorithm on the VCUBE images does not quantify the condition of welds or insulated joints. This challenge is addressed in this work by using outlier analysis techniques, namely statistical methods to characterize outliers for the welds. The identified outliers are submitted to experts for validation validation to gather feedback for training improved models. This process is described in Sec. 1.3.4.

The expert knowledge obtained from the assessment of outliers is combined with the resulting condition logs from $ZMON$ in Chapter 7, where an expert informed classification of the weld condition is proposed. This approach compares three types of models: simple Binary Choice (BC) models, Random Forest classifiers and Bayesian Logistic Regression (BLR) a applied to the ABA features. BC models are the simple and understandable models that are used as a baseddline for comparison with more complex models such as RF and BLR. BLR is a statistical method for modeling the relationship between a binary outcome variable and one or more predictor variables. BLR uses Bayesian inference to estimate the model parameters in a probabilistic framework.

For this reason, the BLR model [35], compared to BC and RF models, further delivers a probability of prediction, thus quantifying the confidence which we may attribute to the assigned labels. Its main advantages, compared to other models, are a reduced risk of over-fitting thanks to the regularization of the priors, a probabilistic estimate on the posterior under the model assumptions (i.e., the priors) and providing an indication of the predictive uncertainty.

1.3.4 Expert feedback

Expert-in-the-loop approaches in Structural Health Monitoring (SHM) involve the incorporation of human expert knowledge and expertise into the process of monitoring and analyzing the health of a structure [45]. The data obtained on the condition of assets such as the railway track is then analyzed by both automated algorithms and human experts. The human experts in such a scenario review the data, provide feedback and make decisions on how to proceed with monitoring and maintenance. Leveraging the knowledge of experts is achieved in this thesis for the identification of the condition of welds. Outlier analysis, as previously described, is used to extract ABA-based condition metrics. Samples of outliers are then given to experts after each vehicle-based automated track inspection campaign. An expert then rates these outliers in terms of binary condition labels (healthy/defect), introduces the defect samples into the condition monitoring database (ZMON). In a second stage, the expert labels are used to improve the performance of the model in a supervised classification framework developed in Chapter 7.

The knowledge gained by the ABA-based condition assessment and expert feedback also drives maintenance decisions, as each expert will depending on the observed state initiate the process leading to maintenance measures. By tracking the condition indicators over time, the effectiveness of a resulting maintenance measure can be assessed. For example, after the replacement of a damaged weld, the increased ABA-feature may return to low values, meaning the problem was solved. The analysis of the condition indicators over time gives an insight on the origin and evolution of shortcomings in the infrastructure (see also Sec. 7.4.3).

The assessment of certain components on the track is nowadays achieved via rail-head images or via on-site inspection. Both these approaches are insufficient, as they do not deliver clear metrics on the severity or urgency of a fault. For this reason, maintenance and renewal decisions are today largely in the hands of the judgement of experts, which each may have different assessment criteria. Chapter 7 identifies the challenges of biased assessment and proposes the usage of ABA, as a more quantifiable metric of the rail condition.

1.4   Thesis Objectives

The traditional approach to Life-Cycle Assessment (LCA) and infrastructure maintenance planning is based on inspections performed at regular time intervals via specialized systems or manual inspections that are triggered by the manifestation of irregular behavior. The disadvantages of this approach are obvious from both a cost and efficiency perspective. Structural Health Monitoring (SHM), for instance via accelerometers installed on passenger vehicles, offers an alternative approach that uses sensors for extracting continuous information on system performance in an almost automated manner [74]. This dissertation aims to shift the current condition assessment and monitoring paradigm for railways, based on periodic inspections by track inspectors and diagnostic vehicles, to a continuous condition assessment via on-board accelerometers mounted on specialized and/or in-service trains.

To this end, this work presents newly developed methodologies that allow for continuous monitoring and analysis of the track infrastructure condition. This is achieved on the basis of measurements from Axle Box Accelerometers (ABA) that are installed on the diagnostic measurement vehicle of the SBB (gDFZ) and the On-Board-Monitoring ICN (OBM-ICN) of the SBB. The integration of system knowledge, i.e., the fusion of locally collected data with global data sources, such as fixed asset information, physics-based vehicle models, expert feedback, condition logs or maintenance logs, forms an essential element to provide estimates with increased confidence for the estimated infrastructure condition. This thesis aims to propose component-based assessments of asset condition using axle box accelerations that are consistent with the component-based assessment pursued by the railways [247].

These aspects are further elaborated in the six detailed objectives of this thesis:

- Investigate new methods enabling robust diagnostics from ABA measurements.

- Develop an assessment methodology of the wheel Out-Of-Roundness and the track stiffness exploiting a data driven - yet dynamics aware - approach, namely the Vold-Kalman filter.

- Develop a methodology for calibrating physics based models relying on MCMC Bayesian updating, to improve the understanding of vehicle-track dynamics.

- Investigate and develop models for the automated classification of critical rail components, namely welds, insulated joints and surface defects by exploiting a Big-Data framework powered by Machine Learning tools.

- Develop new methods for automated and data-driven assessments of the condition of welds and for enabling the early detection of faults.

- Delivering a holistic framework, which combines data-driven condition indicators with robust outlier analysis and complementary expert feedback for actionable implementation in railway management.

First, a comprehensive overview of methods and applications is provided, summarizing the approaches to infrastructure condition assessment using ABA and demonstrating them using our data.

To achieve the second objective, the influence of the sleeper passage and the wheel OOR on the dynamics of the rail-axle system is investigated by extracting these components from the signal using the Vold-Kalman filter. The extracted harmonic components are then leveraged to quantify the track stiffness or wheel Out-Of-Roundness.

To achieve the third research objective of understanding vehicle-track dynamics is investigated by integrating physics-based models with data-driven approaches for model updating

using the MCMC Bayesian updating method. The resulting models enable the improved estimation of wear and fault development.

In line with the fourth goal, algorithms are developed to detect and identify deviations such as squats, insulated joints and squats or rail surface irregularities. These algorithms rely on a large database of labels that are automatically generated from an algorithm used by SBB on the Rail-Head images (VCUBE). Two approaches are explored: one that directly feeds the ABA time-series into Convolutional Neural Networks and a second one that uses engineered features as an input to Random Forests.

The last two objectives are reached by combining the welds identified from the images of VCUBE with ABA-based indications. Outlier analysis is performed on features extracted from the ABA and potentially faulty welds are given for expert feedback. More advanced classification models such as Random Forest (RF) and Bayesian Logistic Regression (BLR) which combine the expert labels with information from ZMON, are subsequently used in order to deliver improved metrics that integrate essential indicators from ABA and an essential operational parameter, namely the vehicle speed.

1.5 Thesis Outline

The thesis is structured in six chapters containing journal and conference papers produced by the author during the doctoral studies and combined in a comprehensive flow.

In line with the overall objectives of assessing the state of assets using axle box accelerations for On-Board Monitoring application, **Chapter 2** presents an in-depth analysis of the existing methodologies and approaches for railway infrastructure assessment using vehicle-based acceleration measurements. This work summarizes the approaches to assessing the condition of the complex railway infrastructure, which consists of many elements, such as ballast, sleepers rail and sleepers. Special emphasis is given to the most common defects that develop over many load cycles and affect the safety, availability and costs of the railway system. Parametric and non-parametric methods are then described and complemented by examples produced from the axle box acceleration measurements of the diagnostic vehicle of SBB, to illustrate the use of ABA for asset condition assessment. Finally, an overview of the most commonly used classification and outlier analysis methods in the railway field is given. The findings show that acceleration measurements can complement a wide range of specialized systems currently used to monitor separate components.

This work is published in *The Rise of Smart Cities – Advanced Structural Sensing and Monitoring Systems* [107].

Drawing from the first objective, **Chapter 3** shifts towards the assessment of the track stiffness, an essential parameter to quantify the track performance. The vibrations induced by the sleeper passage and the wheel Out-Of-Roundness result in periodic responses of the axle, which are extracted using the Vold-Kalman filter. After formulating the Vold-Kalman filter, we provide a novel application in the field of railway asset condition assessment by extracting periodic components. The harmonic components extracted using the VKF are then correlated with the track stiffness. Finally, increased ballast and rail maintenance can be attributed to the stiffer track sections. These findings motivate further work on improved design criteria for the asset, taking into account the dynamic nature of the vehicle/track system.

This chapter has been submitted to *Mechanical Systems and Signal Processing*.

In line with the third objective, **Chapter 4** is dedicated to the process of integrating physics-based models with data-driven approaches, specifically utilizing the MCMC Bayesian updating method, to fine-tune parameters within an ICN RABDe500 wagon model. This calibration process is based on actual OBM measurements gathered from an in-service ICN RABDe500 train. This work results in a well-calibrated model, significantly improving its ability to predict

loads and assess damage. Furthermore, the study underscores a key feature of MCMC scheme: the generation of a posterior probability distribution for parameters, which aids in quantifying uncertainty in predicted outcomes.

This work was presented at the *XII International Conference on Structural Dynamics* and will be published in the *Proceedings of the European Association for Structural Dynamics*.

In accordance with the fourth objective, **Chapter 5** presents a data-driven classification of rail irregularities from Axle Box Accelerations. Welds, insulated joints, surface defects and squats are classified from the ABA using two approaches: one, which requires the construction of features from the ABA as input to a Random Forest. The second uses the time series of accelerations, which are transformed into a time-frequency representation with using the STFT and then classified using a CNN. Both the CNN and the RF approach gave good results for the classification on unvalidated labels obtained from the Rail-Head (VCUBE) image recognition algorithm of SBB. However, welds and rail defects were not always identified reliably, because these classes are not clearly defined today and contain a mixture of healthy and defective instances. This further motivated an in-depth investigation of the condition of welds using ABA.

This work was presented at the *40th International Modal Analysis Conference* and is published in the *Proceedings of the Society for Experimental Mechanics Series* [104].

Further delving into short wavelength faults, **Chapter 6** and **Chapter 7** present a holistic framework for the automated detection of defect welds, by fusing a variety of informations, such as visual or on-site inspections with automated diagnostic information extracted from the monitoring vehicles.

Chapter 6 presents an approach relying on Extreme Value Analysis (EVA) that is performed on a set of the essential indicators obtained from the ABA measurements, to characterize outlier welds. Outliers detected via EVA are submitted to field experts, in a first of its kind Proof-of-Concept (PoC) project in collaboration with the Swiss Federal Railways (SBB) to obtain expert feedback. The expert feedback is then exploited in the subsequent chapter for expert-informed condition assessment.

This work was presented at the *European Workshop on Structural Health Monitoring (EWSHM)* and is published in *Lecture Notes in Civil Engineering [253]* [110].

Chapter 7 harnesses the expert feedback stemming from the expert assessment of outliers presented in the previous chapter. The combined information from the expert feedback and the logs in the condition monitoring database of SBB (ZMON) is used to deliver an automated classification scheme based on three different evaluated methods: Binary Classification (BC), Random Forest (RF) and Bayesian Logistic Regression (BLR). The resulting analysis shows that the relatively low classification metrics can be attributed to several sources of uncertainty, such as uncertainty in vehicle positioning and uncertainty in the expert labeling, as well as the binary labeling that does not capture the granularity of the continuous degradation in terms of severity.

This work is published in *Sensors 2023 [23(5)]* [105].

In **Chapter 8**, a summary of the main findings and conclusions of this dissertation is provided, followed by an outlook on the future developments of this research.

Chapter 2

On-Board Monitoring for Smart Assessment of Railway Infrastructure: A Systematic Review

Paper Details

Author and Co-Author Contributions

The author of this thesis conceived and developed the review and analysis developed here. Prof. E. N. Chatzi and Dr. V. Dertimanis provided supervision and guidance. Dr. M. Landgraf, Dr. L. Ancu and M. Zurkirchen provided reviews on this work.

Key Findings

- A review of track infrastructure components and monitoring techniques applied for intelligent maintenance schemes.

- Review of parametric methods and non-parametric methods for acceleration-based condition assessment, as well as application examples on the data from the diagnostic vehicles of SBB.

- Review of methodologies for classification and outlier detection techniques.

- OBM-derived indicators require further research for effective preventive maintenance. Current acceleration-based monitoring systems are customized and require human intervention and expert opinion for maintenance decisions.

- The other challenge identified relates to building efficient frameworks for processing the aggregated data and integrating these into existing processes while ensuring good model performance and reproducibility beyond the test cases.

General comments and Link to the next chapter

This study tackles the first objective of the thesis (see Section 1.4) by proposing a review of the state of the art of the assessment of track infrastructure using traditional monitoring vehicles and On-Board monitoring via accelerometers mounted on in-service vehicles.

This chapter proves the potential for assessment of railway infrastructure by mounted (OBM) sensors on railway vehicles. It also highlights the challenges in the generalization of the models beyond simple test studies and the challenges in integrating such novel approaches into existing processes. In the following chapter, one of the proposed methods for assessment, namely the Vold-Kalman filter is applied to the axle box accelerations of the railway vehicle to assess the wheel Out-Of-Roundness and the superstructure stiffness.

Abstract

The increasing demand in mobility forms a major challenge for modern cities, even more so when examined under the prism of transition from traditional to CO2-free mobility. Railway infrastructure forms a main carrier for the mobility of people and goods and a salient component of critical infrastructures. The increased traffic frequency in urban transport imposes higher capacity demands and leads to more frequent damage, more severe deterioration and associated disruptions to service and availability. Aligning with the spirit of smart cities, and data-driven decision support, infrastructure operators require timely information regarding the current (diagnosis) and future (prognosis) condition of their assets in order to sensibly decide on maintenance and renewal actions. Railway condition assessment has traditionally heavily relied on on-site visual inspections. Main measurement parameters for railway tracks are obtained since the 1960s. Quality, accuracy and precision of measurements heavily evolved since then, including aspects such as storage, analyses and interpretation of data. In recent years, specialized monitoring vehicles offer an automated means for relaying essential information on condition, obtained from diverse measurements including laser measurements, vibration, image and ultrasonic information. Powered by this information diagnostic vehicles have shifted assessment from a reactive to a predictive mode. More recently, in-service vehicles equipped with low-cost On-Board Monitoring (OBM) measuring devices, such as accelerometers, have been introduced on railroad networks, traversing the network at higher frequencies than the specialized diagnostic vehicles. The collected information includes position, acceleration and in some cases force measurements. The measured data requires interpretation into quantifiable track-quality indicators, before it can be meaningfully incorporated in asset management tools. These indicators form the basis for real-time forecasting of condition evolution and asset management, which are essential traits of a transport infrastructure that fits the vision of smart cities. This article explores the state of the art of OBM for railway infrastructure condition assessment, conducting a thorough review of data processing methodologies, which is further complemented with application examples.

2.1 Introduction

Railways form part of what is termed critical infrastructure of modern cities. The advent of digital technologies, within the grand vision of smart cities, has allowed for tremendous developments in the monitoring of railway assets and the subsequent utilization of such data for intelligent and preventive maintenance. Railway asset managers pursue a long-term infrastructure maintenance and renewal strategy, which aims at maximum safety, availability and reliability, while simultaneously reducing life-cycle costs (LCC) [200]. In order to guarantee an adequate track quality, track components are required to satisfy quality control under various maintenance measures over their entire service life [288]. Maintenance actions are either corrective or predictive and are often performed under many constraints such as budget and availability. A sound knowledge of the condition of the track infrastructure, including substructure, ballast, sleepers, rail pads and rails, is essential for applying appropriate maintenance measures in a timely manner. An estimate of the overall behaviour of the track condition is achieved using the standard elements approach [180] (both for the Swiss Federal Railways– SBB and the Austrian Federal Railways – ÖBB); a stochastic model that is used for strategic assessment of maintenance and renewal demands on a network-wide basis. The model evaluates the expected behaviour and guides on necessary interventions during the entire service life of the railway track on the basis of the most influential asset parameters, such as track type, curvature, age of components and load conditions.

The standard elements approach enables strategic planning in terms of budget-forecasting,

scheduling of required maintenance renewals, as well as setting the optimal time for reinvestment, determined by the point where the increasing maintenance costs outweigh the decreasing costs of deprecation. Such a model cannot be used for section-specific maintenance planning since deviations from the global model occur and can be attributed to different local conditions [306]. The information gathered about assets has become increasingly diverse, as it is facilitated via harnessing of network-wide measurement data from diagnostic vehicles and On-Board Monitoring solutions (OBM) [157]. This increasingly richer digital data can typically be aggregated after compression and feature extraction, in digital platforms and tools, such as the SwissTAMP in the case of SBB [200]. Such tools can be used to plan maintenance and renewal measures based on a comprehensive overview of the current and past condition of the tracks [306].

The collection of data from diagnostic and in-service vehicles offers a major boost to the process of assessment for the current (diagnosis) and future (prognosis) state of assets. Asset state has traditionally been determined through on-site visual inspections and measurements. Main measurement parameters for railway track are obtained since the 1960s in Austria [148]. Quality, accuracy and precision of measurements significantly evolved since then, including aspects such as the automatic storage, analyses and interpretation of digital measurement. Currently, data from measurement vehicles are automatically collected and transmitted, which enables an efficient assessment of the condition of the track. Diagnostic vehicles regularly measure the condition of the track with high accuracy, albeit with limited temporal resolution.

OBM vehicles are passenger vehicles equipped with measurement devices such as accelerometers that regularly traverse the network and thus provide a nearly daily data stream consisting of position and acceleration. In order to meaningfully include such information in asset management tools, the measured data must be processed into quantifiable condition indicators. Track-quality indicators (TQIs) form the basis for asset management, which is based on a real-time forecasts of development of the condition. They are powerful when it comes to describing the holistic condition of the track, as they summarize the standard deviation of several measurement values (gauge, longitudinal level, alignment, superelevation, etc) [316]. However, component-specific wear monitoring and prognosis are required for sophisticated maintenance planning (treatment of the root cause of degradation) [214]. This article first describes the most common types of damage of railway track components and the non-destructive condition measurement systems that currently exist to measure the condition. In addition, the state of the art is overviewed, primarily under the prism of acceleration-based OBM for condition assessment of railway infrastructure, while further offering a thorough review of the relevant data-processing methods and tools, which is supplemented with implementation examples. The information harnessed from a mobile sensor network, enabled via an OBM vehicle fleet, is a major step toward effectuating a vision of smarter infrastructure and smarter cities.

2.2 Track infrastructure components and condition

The railway track ensures the safe guidance of the trains on the desired path and the transfer of loads from the axles to the substructure. High-quality tracks are characterized by low maintenance and stable long-term behavior even with high train frequency, increasing travel speeds and the higher resulting loads [165]. The vehicle-track system should always be regarded at system-level, as opposed to a component-wise consideration, since all infrastructure components are interconnected. At this point, an overview of the infrastructure components and their most common deficiencies is given.

2.2.1 The track and its geometry

The two main types of track are ballast (Fig. 2.1) and, less often, slab tracks [165]. In the case of ballast tracks, the ballast material supports evenly spaced sleepers, while in the case of slab tracks, the ballast material is replaced by a concrete slab to which the sleepers are fixed. The substructure under the ballast or slab supports the superstructure and enables the drainage of rainwater. The substructure functionality is essential for track stability and durability.

Slab tracks meet higher requirements for track quality, have a longer service life and lower maintenance demands, when the substructure is completely homogeneous [66]. This track type is often used for High Speed Rail, for bridges and tunnels. The renewal is very costly and particularly time-consuming. Therefore, the planning and construction costs are higher for slab tracks [89]. Ballasted tracks, on the other hand, have a potentially long service life, albeit requiring a larger number of maintenance actions, such as tamping [102]. In contrast to slab tracks, maintenance is here easier to accomplish, which leads to an enormous advantage, since weak or deteriorating substructure conditions can be compensated by the ballast bed.
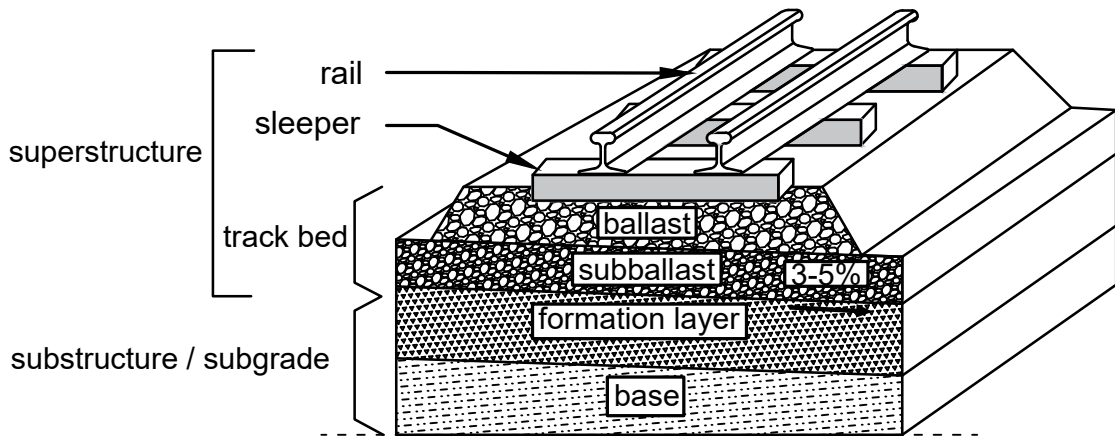

Figure 2.1: Illustration of a ballasted track.

Geometric parameters such as longitudinal level, alignment, twist, gauge and cant are regularly recorded by means of monitoring vehicles [128]. The geometric parameters are typically filtered to certain wavelength ranges that are defined as $D0$ (0.6-3 m), $D1$ (3-25 m) and $D2$ (25-70 m) in the Euronorm EN 13848 [52]. Irregularities in the geometric parameters of the track result from a combined effect of repeated load cycles on a track, with possibly decreased substructure and ballast bearing capacity, as well as insufficient drainage. Localized deviations in track geometry can occur at the location of specific components, such as switches, welds, joints, and transition zones from ballasted to slab tracks, as well as in positions of defects (squats[1]) due to changes in track parameters and repeated loading cycles [89]. Safety and comfort are guaranteed by limiting the irregularities in track geometry with maximum allowed geometric deviations. Tamping is used as a main repair (maintenance) action to correct errors in geometric parameters before these values reach critical thresholds [121].

2.2.2 Rail Connections: welds and insulated joints

Rails are mounted on the sleeper and directly bear the loads of the railway vehicles. Rails are produced as standardized components of finite length that are usually continuously welded together and, in some cases, connected using insulated joints consisting of two electrically

---

[1]A squat is type of surface defect that will be elaborated in following sections

insulated rails that are bolted together with steel plates. Welds and insulated joints form discontinuities on the rail surface and are thus picked up as such in the measured response during a vehicle run. In proximity of these components, the occurrence of stress concentrations and wear lead in irregularities in the track geometry and on the rail [179].

2.2.3 Rail corrugation

Rail corrugation is a periodic wave-like surface defect. The mechanism by which the corrugation arises is explained by a wavelength fixing mechanism and a damage mechanism [93]. The wavelength is determined by vehicle-track resonance mechanisms, such as pinned-pinned resonance (resonance of the rail), P2 resonance and rutting [277]. The P2 resonance is the vertical system resonance where vehicle unsprung mass, rails and sleepers vibrate in phase with one another. Rutting corresponds to short-pitch corrugation developing on the low rail on curves. The main damage mechanism is wear due to the resonant effects [315]. Squats sometimes arise as a result of rolling contact fatigue (RCF), caused by corrugation [205]. Corrugation is often categorized by wavelength. Asset managers typically subdivide periodic irregularities into short (2-10 cm), medium (10-30 cm) and long (30-100 cm) wavelength ones [244]. Preventive maintenance, such as grinding, can temporarily remove the effects of corrugation. The dynamic properties of selected track components during a renewal influence the growth rate of waves on the rail surface can be further reduced [272].

(a) Intact insulated joint

(b) Joint with broken joint plate
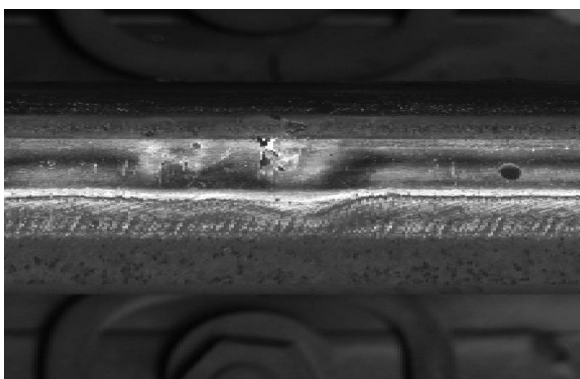
(c) Squat on rail

(d) Weld

Figure 2.2: Infrastructure components leading transient excitation (images extracted from a high speed camera, mounted on a diagnostic vehicle of the SBB)

2.2.4 Transient rail defects

Transient rail defects are surface or internal defects that are localized in space. Many mechanisms influence the appearance and growth of defects on the rail. The main types of transient defects are the following:

Cracks to the head, web, foot, weld and joint plates are events that can lead to a rail breakage. Cracks are caused by material defects that are internal or on the surface of the rail. Cracks propagate under repeated loading cycles. Fig. 2.2b shows a joint in which the joint plates cracked.

Surface defects are a broad category of defects that appear on the rail surface. These surface defects are, in some cases, due to lost goods or ballast on the rail surface and damaged wheels that lead to indentations on the rail. Surface defects and light squats share many visual characteristics and it has been observed that some surface defects grow into a squat. A rigorous definition of the rail surface defects is not an easy task as surface defects appear and develop in different locations under different conditions.

Squats are a type of rolling contact fatigue (RCF) that has been an important driver for rail replacements in the past decades due to its safety relevance [122]. Rolling contact fatigue pertains to accumulation of plastic deformations on the rail surface. Squats usually appear on tracks that are straight or have large curves. In certain cases, their appearance is related to surface corrugation, where surface defects spread under the rail surface, resulting in large breakouts [268]. The initiation of squats has been hypothesized to be associated with tracks, where the traction control employed by certain vehicles causes wheel slips. The International Union of Railways defines a squat as a "widening and localized depression of the rail/wheel contact band, accompanied by a dark spot containing cracks with a circular arc or V shape" [213]. The length of the squat is the main parameter used to measure its severity [244]. The severity classes are designated as severity 1 (L<20 mm), severity 2 (20 mm<L<40 mm) and severity 3 (L>40 mm). A squat of severity class 3 (48 mm length) is illustrated in Fig. 2.2c. Less severe squats are tentatively treated by grinding, while more severe ones require the replacement of the rail.

2.2.5 Switches

Switches are mechanical installations in which the train can be guided on different track segments. The switch forms a critical installation regulating the connection between fixed stock rails and switch rails, which are operated mechanically or hydraulically at the switch point in order to regulate the routing of the train. The frog is located at the separation point of the two tracks. This point is also termed as the crossing. The guard and wing rails, which are located next to the rails facing the frog, prevent a train derailment. The diagram in Fig. 2.3 illustrates the labeled components of a switch in the case of a switch without a movable frog. Movable frogs, which close the gap between the stock and the running rail, are sometimes chosen for high-speed tracks. These types of switches are expensive and complex, but guarantee a homogeneous rail surface, which is particularly important at high speeds. The rails on switches typically have multiple welds, while in certain cases, insulated joints can be installed in proximity of switches. The switches form an essential component for railway operations and require special attention in the life-cycle management of these components.

Currently, the monitoring of switches is performed using a combination of on-site inspections, stationary on-site measurement systems (strain gauges, accelerometers on the track, fiber optics, etc) and diagnostic monitoring vehicles. The tested properties range from the previously mentioned track geometry parameters, rail surface defects, squats, joint and weld conditions to conformity of the rail profile with the required profile geometry [293].
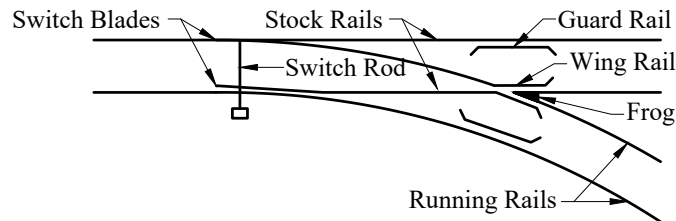
Figure 2.3: Diagram of a switch.

2.3 Condition monitoring in railways

Non-destructive monitoring aims to assess the properties or condition of an engineering asset without causing damage. In the context of railways tracks, the track condition is regularly assessed with specialized diagnostic vehicles and with On-Board-Monitoring (OBM) vehicles. Regular monitoring of the track condition ensures an optimal planning of maintenance actions, which leads to lower life-cycle costs and increased availability of railways assets.

2.3.1 Diagnostic Vehicles

Diagnostic vehicles are equipped with measurement systems, that measure quantities related to track geometry, track stiffness, rail profile, rail surface condition and the internal rail condition. Tab. 2.1 summarizes some of the systems that are commonly mounted on vehicles for monitoring the railway track. These measurement systems are specialized in detecting specific irregularities with the help of dedicated hardware, such as laser scanners and high-speed cameras. Vehicle-track interaction measurement systems (VTIMS) differ from conventional measurement systems that are specialized in recording single types of defect. VTIMS capture the effect on the vehicle-track dynamics that are caused by a large spectrum of deviations (see also Section 2.4) while using comparatively simple and inexpensive sensors, such as accelerometers.

The assessment of the infrastructure condition based on dedicated vehicles is limited by the complexity of planning and organisation of diagnostic runs on networks that are already heavily traversed by regular trains. A continuous assessment scheme using sensors on passenger vehicles can complement the periodic measurements of diagnostic vehicles with near real-time information on the state of the infrastructure.

2.3.2 On Board Monitoring

OBM vehicles typically have a simple and inexpensive hardware and software architecture that includes accelerometers mounted on the axle, bogie or body. In some cases, the vehicles are equipped with tensiometric wheel-sets [236]. These wheel-sets are modified to measure the contact forces between wheel and rail. Sometimes microphones are also installed on these vehicles. An overview of an OBM setup, such as the one implemented by SBB, is shown in Fig. 2.4. Vibration-based measurements have been tested across different railway networks to diagnose diverse pathologies, such as geometric track irregularities, corrugation, weld and joint conditions, detection of squats and monitoring of switches. Tab. 2.2 summarizes the sensor location and the approach used to monitor specific components at different railways.

Table 2.1: Non destructive track condition monitoring systems and methods

System	Method	Damage types	Limitations
Vehicle-track interaction measurement (VTIMS)	Force and acceleration	Effects and damages that impact vehicle-track dynamics	Data processing required for meaningful interpretation
Track geometry measurement system (TGMS)	Laser chord method or IMU	Geometric parameters in wavelength 1-70m [52]	Recolouring necessary to accurately represent some wavelengths for chord measurement [128][303]
Rail profile measurement system (RPMS)	Profiles from laser scanners	Rail profiles & wear of rail every 2cm	Coordinates of laser system vary with respect to rail coordinates
Track stiffness measurement system (TSMS)	Subsidence of axle	Regions of high track stiffness, and sudden stiffness variations	Maximum measurement speed 15 km/h [83]
Track surface and inspection measurement system (TSIMS)	On-Board camera	Visible damages to rail surface and track components	False positives and negatives due to light reflections or rail contamination.
Eddy current testing system (ET)	Electrical current, magnetic field	Surface cracks up to 5mm of depth	Detection sometimes inaccurate due to complex defect geometries and rail properties [124]
Ultrasonic measurement system (UT)	Ultrasound	Rail cracks under 4mm of depth	Used as a rail quality indication due to low hit rate and many false positives [274]
Ground penetrating radar (GPR)	Radar	Ballast and substructure condition	Uncertain dielectric properties of materials [28]

2.3.3 Uncertainty in vehicle localization

The localization of the vehicle on the train network is essential in order to follow the development of the infrastructure state over time. If several measurement rides are combined in an analysis in order to predict the current state and to forecast the future development, the required quality of the position depends directly on the length of influence of the observed effects. The synchronization of time–series for distributed sensor networks is a challenge [255], especially when the sensors themselves are located in a moving reference frame. The minimal setup for track-selective georeferencing is achieved by combining map information with IMU and GNSS data [237][304]. A rough position on the network is achieved through a combination of track segment information with GNSS, ETCS balise detectors [2] and odometric data. The position uncertainty obtained with these information sources is less than one meter under good conditions. On track segments where no Eurobalise data is available, where the Eurobalise position data is inaccurate or where the GPS reception is poor, the position error can increase to many meters, which in rare cases could lead to incorrect localization [200].

There are two main approaches for dealing with unsatisfying precision of the position estimation. The first lies in detecting a deterioration on the position estimation quality and

---

[2]ETCS balise (Eurobalise) are transponders used by the European Train Control System (ETCS) to give the exact location of the train on the track and to transmit signaling information to the train.

ICN OBM sensor positions   ⊕ **SBB CFF FFS**

Vehicle Body

Bogie 2

Bogie 1

WS4   WS3   WS2   WS1

OBM Rack

- ■ 4 Axle Box Accelerometers (y,z)
- ◆ 4 Displacement Primary Suspension (z)
- ■ 4 Bogie Accelerometers (y)
- ■ 4 Car Body Accelerometers (y,z)
- ● 1 Uncompensated Lateral Acceleration (aq)
- ▬ 2 Pairs of Measuring Wheelsets
- ● 1 Odometer
- ■ 1 ETCS balise antenna
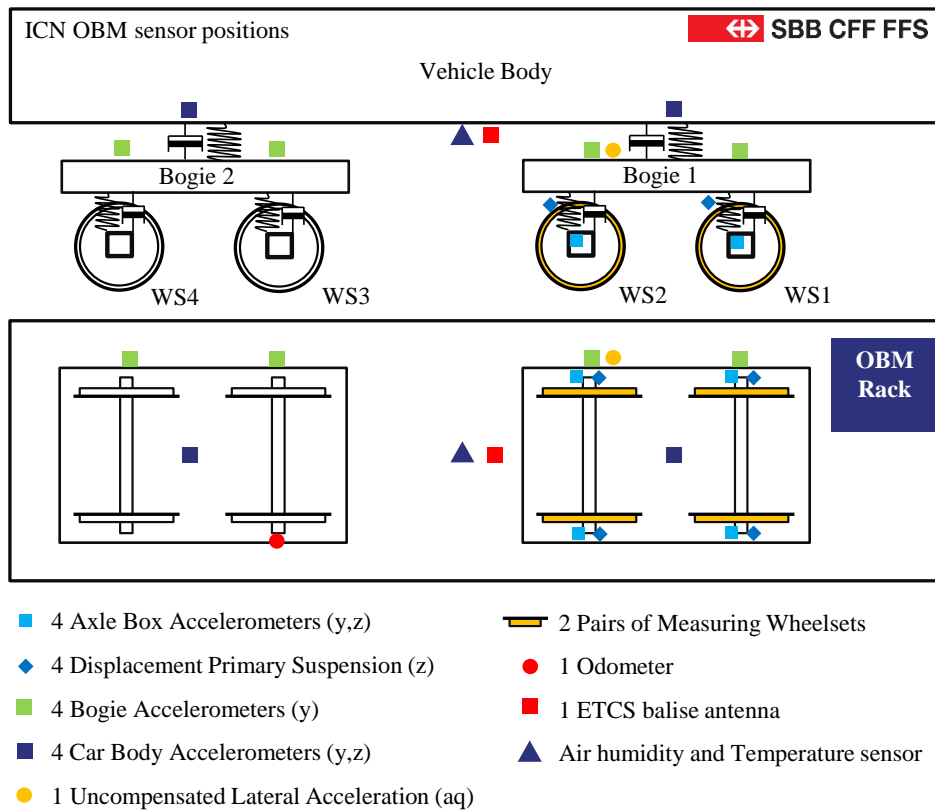- ▲ Air humidity and Temperature sensor

Figure 2.4: Overview of sensor positions for a complete OBM setup of an in-service train.

discarding these samples. This is typically succeeded via the fusion of multiple input streams such as GPS, ETCS balise, and odometric data. The second option is to synchronize the position in a post-processing step by using information extracted from previous measurement runs. Features that are most meaningful for this task are those that present reduced variability over a longer period of time. Such features can be continuous, such as the longitudinal level $D0$ and $D1$ or the uncompensated lateral acceleration. Events detected from a short wavelength response, such as crossings of insulated joints and frogs, can also be used as latching points when locating the vehicle on the network. The synchronization of time–series can be achieved by combining several methods, such as time lagged cross-correlation (or dynamic time warping) for continuous data or point set registration methods for punctual events [33][323][195].

Fig. 2.5 shows the synchronization of two measurement runs, via the indicators extracted from ABA, using the methodology described in Section 2.6.2. The reference measurement ride on 30th March 2021 is synchronized to all other measurement runs, such as the one on 27th January, using the maximization of the time-lagged cross correlation. The estimated drift between measurement rides on each track segment can be automatically corrected using such a synchronization process.

2.3.4 Digital Infrastructure

Data driven condition monitoring requires a robust digital infrastructure that allows to process, transmit and store information in a structured manner. The automated acquisition, processing and transmission of measurements from OBM systems is essential for an efficient scheme. The collected measurements are usually stored in a database, which is fed with the daily OBM measurements, thus entailing a structure that can grow rapidly with large amounts of redundant information. Data management is therefore essential in this context because, despite the obvious

Table 2.2: Recent On-Board acceleration based track condition monitoring implementations

System	Monitored components	Approach	Railway network	Ref.
Bogie vertical acc.	Longitudinal level	Longitudinal level by Extended Kalman Filter	Indian Railways	[152]
Bogie vertical acc. and gyrometer	Longitudinal level	Longitudinal level by Cross entropy optimisation	Irish Railways	[211] [212]
Body vertical acc.	Longitudinal level	Track faults using Kalman Filter and Hilbert Huang Transform	Japan Railways	[282] [281]
ABA, vertical	Longitudinal Level, TQI	Signal integration & Karhunen–Loève transformation	Polish Railways	[57]
ABA, vertical	Longitudinal level	Signal integration & filtering	Deutsche Bahn	[167]
ABA, vertical	Track irregularities	Speed normalized RMS amplitude	DLR, Germany	[21] [19]
ABA, vertical	Corrugation	Multiresolution analysis and multiple model method	Japan Railways	[99]
ABA, vertical & lateral	Corrugation	RMS value & Frequency spectrum	Subway of Milan	[40]
ABA, vertical & longitudinal	Corrugation	Scale-averaged wavelet power	ProRail, Dutch Railways	[159]
ABA, vertical & longitudinal	Squats, joints	Coefficients of continuous wavelet transform	ProRail, Dutch Railways	[188] [175] [326]
ABA, vertical & lateral	Longitudinal level, switches, welds, insulated joints and squats	Vold kalman filter, signal integration, time-frequency decomposition	Swiss Federal Railways (SBB)	ETH [108]

benefits offered by comprehensive information in refining predictive models, guaranteeing data quality through the automatic detection of sensor faults is also a challenge. The amount of data transmitted can be reduced by performing suitable pre-processing steps already on the "edge" level, i.e., the instrumented vehicle. When analyzing certain deterioration processes, such as the degradation of track alignment, a weekly measurement is sufficient, as indicated by the SOB-OBM[3] [167]. For other estimation tasks, it suffices to follow the development of indicators on certain critical components, such as insulated joints, welds, frogs or other impulses
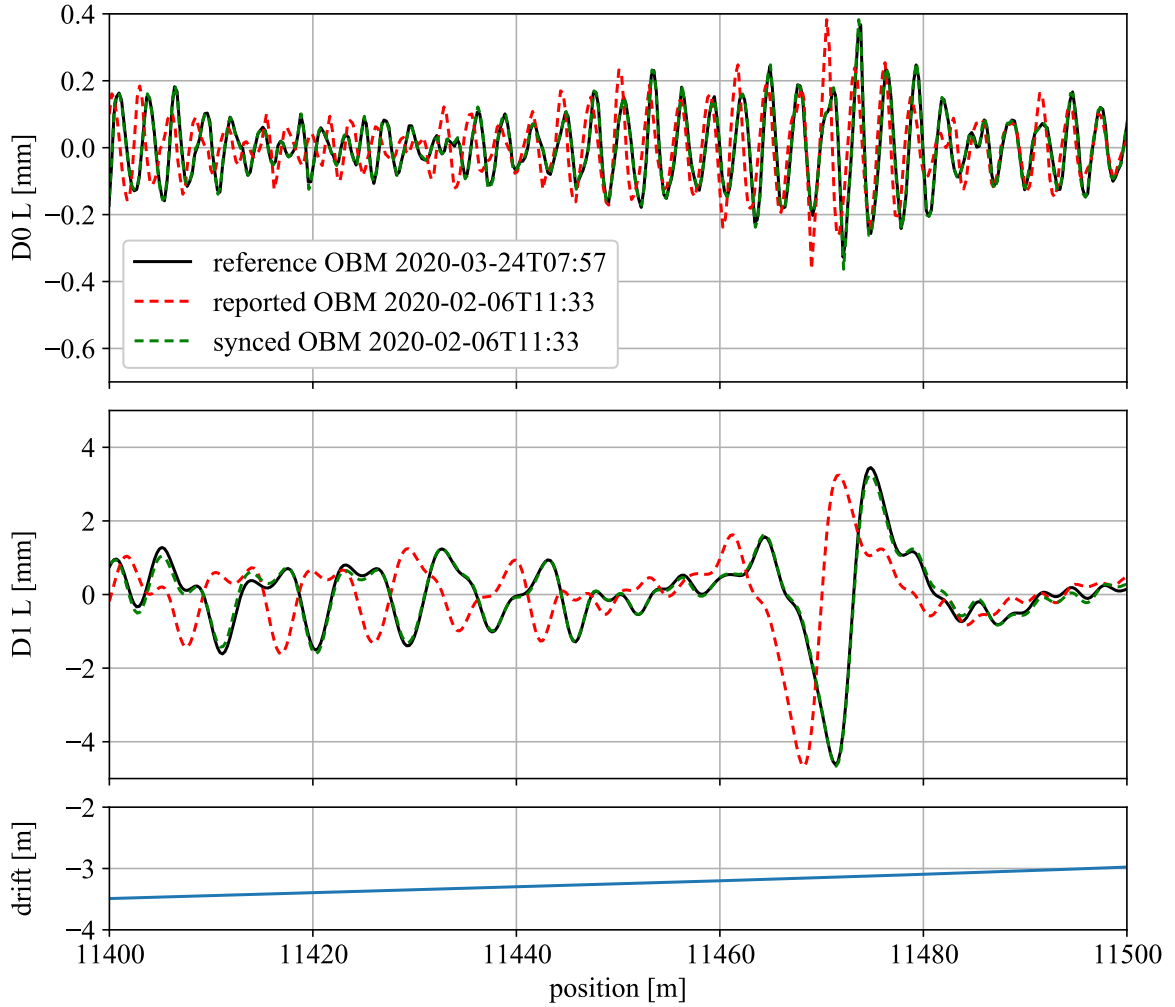
---

[3]Schweizerische Südostbahn

Figure 2.5: Illustration of fine synchronisation via time lagged cross-correlation for estimating the drift between the longitudinal level $D0$ and $D1$ from ABA (raw data from SBB).

due to surface defects or squats. The outliers generated in the measured acceleration signal from the crossings of these components are also called events. Events can be detected by means of an Outlier Detection process, shown in Fig. 2.6. The extraction of events from the processed data is then carried out in an unsupervised manner, based on baseline measurement data, complemented by ground truth labels, which can be obtained from other data sources. These events are then used to identify and classify outliers in order to generate warnings and alarms on potentially abnormal behavior. In such a scenario, a confidence statistic has to be carried out in order to avoid false alarms [207]. A measurement databases and an event database with positioned data are key to infrastructure monitoring.

2.3.5 Challenges in Condition Monitoring

One major challenge is to extract meaningful information from a large amount and variety of measurement information for decision making. Every sensing system has known strengths and weaknesses. The main causes of uncertainty in sensing data are known and quantifiable measurement errors on the one hand, and uncontrollable factors such as changes in environmental and operational conditions (EOPs) on the other hand. Noise can be decreased by adapting the measurement hardware, by accounting for EOPs or by averaging. EOPs can be accounted for by
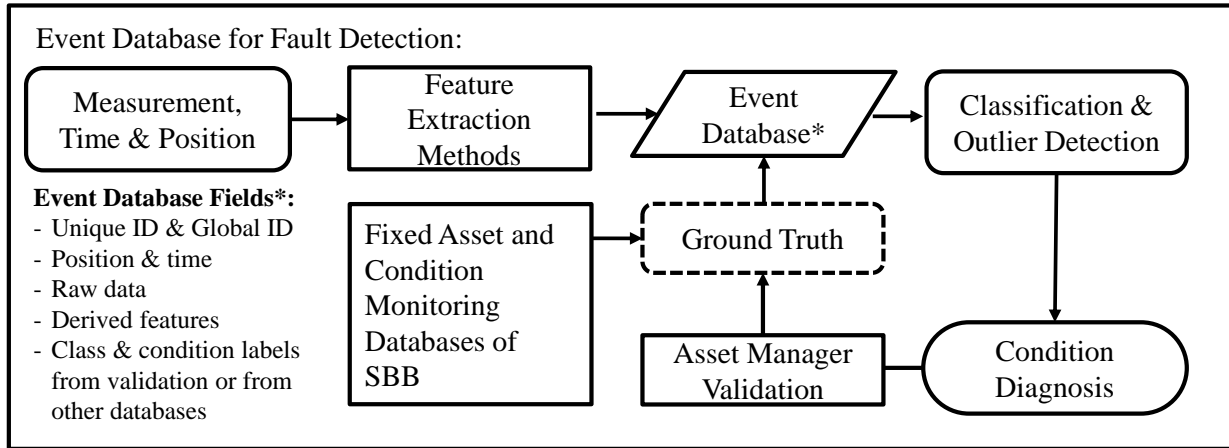
Figure 2.6: Event database generation for classification and outlier detection.

Table 2.3: Summary of excitation mechanisms [139]

Excitation source	Periodic / Nearly periodic	Transient
Track and rails	Corrugation, waves	Alignment, turnouts, squats, welds, insulated joints...
Wheel excitation	Eccentric & polygonal wheels, flat spots	-
Parameter excitation	Sleeper crossing	Hollows, crossings
Self-excitation	Sinusoidal motion	-

measuring these parameters and including them as inputs in a stochastic regression model [16] or alternatively by using projection based tools that do not require the explicit measurement of EOPs [253]. The noise can be further reduced by combining the indications or measurements from repeated rides. Such methods are particularly suitable when continuous track monitoring with OBM vehicles is carried out. In addition, the combination of several measurements from different systems enhances predictive capabilities. For example, the further exploitation of eddy current and ultrasonic measurements for assessing the rail condition can reduce the number of false positive assessments on welds [274][222]. Finally, information about the condition and recommended maintenance must be communicated to the asset managers in order to optimally plan and schedule maintenance. Decision aid tools, as the example of the SwissTAMP platform adopted in Switzerland [200], support a proactive infrastructure management by summarizing infrastructure type and condition, maintenance and planning.

2.4   Vehicle track interaction

OBM vehicles record accelerations that reflect the vehicle-track dynamics. Irregularities on the running surfaces of the wheel and on the track, as well as the vehicle dynamics cause periodic or transient excitation of the vehicle-track system. Tab. 2.3 illustrates the sources and types of excitation, which are owed to various components or dynamic processes. The category designated as "tracks and rails" pertains to excitation that can be generated both by defects or as a result of the inherent properties of the track and vehicle.

The frequency range of interest depends on the phenomena considered. Frequencies up to

20 Hz have the largest impact on vehicle dynamics. Forces that are transferred from the wheel to the axle and the bogie are generally considered to occur up to 500 Hz. The contact surface between the wheel and rail has a radius of approximately 1 cm and thus when analyzing rail–wheel contact dynamics, in terms of short-wave irregularities, the upper limit is between 1.5 and 2 kHz. Higher frequencies up to 20kHz relate to acoustics and noise emissions. Acceleration amplitudes beyond 100g can be observed during impacts of the wheel on switch components or on severe surface defects.

Fig. 2.7 illustrates the relationship between vehicle speed, frequency and wavelength. The transverse lines are lines with a constant wavelength. The colored areas indicate the main excitation wavelength ranges, which range from short-wave corrugation to long-wave geometric deviations ($D2$). The lowest bound of the frequency range of commercially available MEMS[4] based sensors is often around 0.5 Hz. At low speeds, this leads to certain wavelengths being cut off at the minimum frequency of the sensor, which results in missing low frequency vibration information. For shorter wavelength defects at high speeds, a sufficiently high sampling frequency is required to detect an excitation with a higher frequency (i.e. 3.2 kHz).

From the point of view of monitoring the structural condition, the consideration of the wavelength is essential, since different defects occur at different scales and lead to diversified dynamics at these corresponding scales. For example, it has been shown that ballast and substructure conditions relate to the lower wavelength ranges $D1$ and $D2$ [150]. Issues related to the sleeper such as hollows, are proposed to affect the wavelength $D0$ and the wavelength of the sleeper passage [166]. Wheel out-of-roundness and wheel flats affect the frequencies which correspond to integer multiples of the wheel circumference [164]. Rail corrugation induces pronounced vibration in the wavelength range that corresponds to the wavelength of the corrugation [93]. For this reason, acceleration–based condition indicators must take into consideration the variation of the scale of resulting vibrations depending on the frequency and wavelength characteristics of the excitation.

2.5 Parametric methods

For analyzing time–series data, parametric time–series models are often chosen because of their transparency and low computing power requirements. These models feature parameters that define the dynamics of the time–series and determine the relationship between the potentially unknown (or unmeasured) excitation source and the observed response. The residual difference between the measurement and the predicted data indicators can serve as an indicator (or damage sensitive feature) for detecting faults that affect vehicle-track dynamics. The next sections offer an overview of parametric methods that can be applied in the context of railway infrastructure condition assessment, including the AutoRegressive Moving Average model (ARMA), the Linear Parameter Varying AutoRegressive model (LPV-AR), the Vold Kalman filter (VKF), the linear and Extended Kalman Filter (KF/EKF), as well as the Unscented Kalman Filter (UKF).

2.5.1 Autoregressive Moving Average Models

Autoregressive Moving Average Models (ARMA) occur via coupling of the AutoRegressive (AR) and the Moving Average (MA) model classes. ARMA models can be used in a k-step ahead prediction mode, in order to predict future response based on past observations. The moving average part in particular is often used as beneficial due to its smoothing effect on
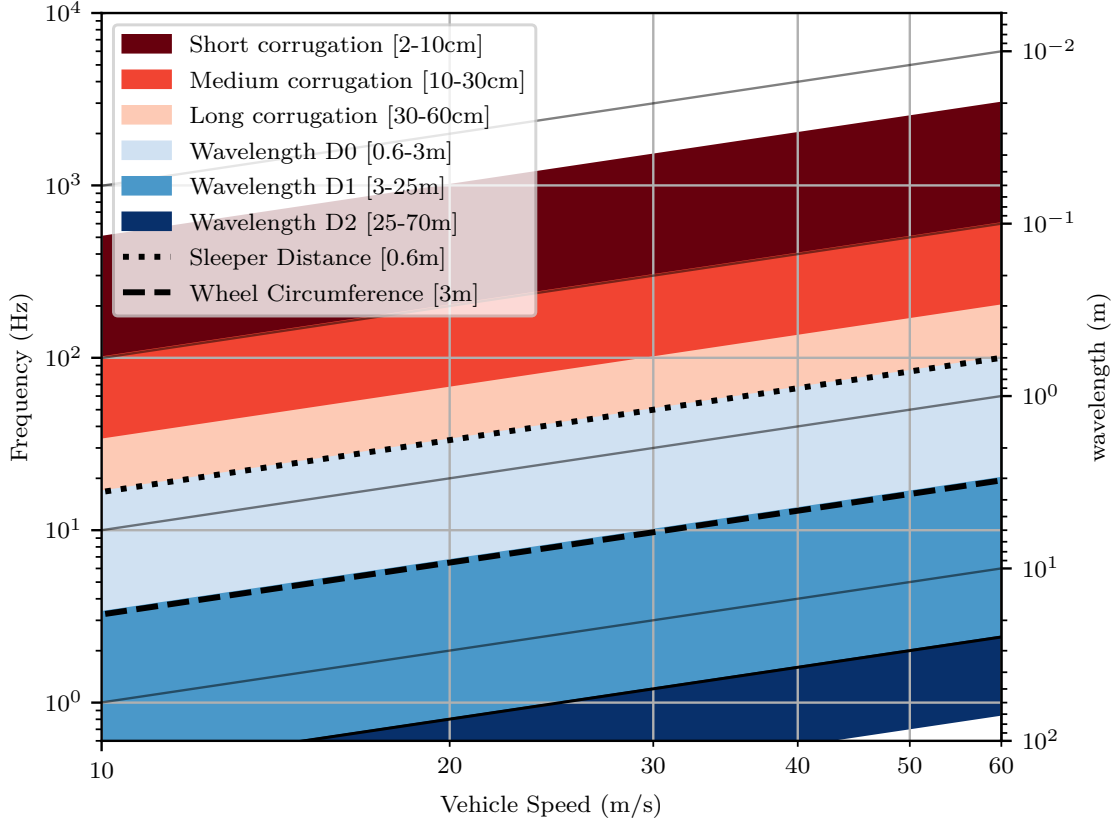
---

[4]Microelectromechanical system

Figure 2.7: Relation between vehicle speed $[m/s]$, frequency $[Hz]$ and wavelength $[m]$.

time–series data [86].

$$y[t] = \alpha_0 + \sum_{i=1}^{n} a_i \cdot y[t-i] + w[t] + \sum_{i=1}^{q} b_i \cdot w[t-i] \qquad (2.1)$$

where $y[t]$ reflect the 1-step ahead prediction at time instance t, $w[t]$ is a white noise source of zero mean and variance σ_w^2; a_i, b_i determine the AR and MA coefficients respectively, while n and q designate the AR and MA orders, which define the number of previous observations $y[t-i]$ and previous innovation terms $w[t-i]$ to be used, respectively [173]. A fitted ARMA model has been used to identify railway track irregularities for forecasting maintenance actions [127][156].

2.5.2 Linear Parameter Varying Autoregressive models

Linear Parameter Varying AutoRegressive models (LPV-AR) form an extension of the classic AutoRegressive (AR) models, which are used to model non-stationary vibration response when the dynamics are controlled by an external scheduling variable. In this sense, LPV-AR models are defined as time-dependent AR models, where the values of the AR parameters $a_i(\beta[t])$ are expressed as a function of a scheduling variable $\beta[t]$. In vehicle-track dynamics, the scheduling variable is the vehicle speed. The regression vector of LPV-AR models is defined as [15]:

$$y[t] = -\sum_{i=1}^{n_a} a_i(\beta[t]) \cdot y[t-i] + w[t] \qquad w[t] \sim NID(0, \sigma^2(\beta[t])) \qquad (2.2)$$

41

$$a_i(\beta[t]) \triangleq \sum_{j=1}^{p_a} a_{i,j} \cdot G_{b_{a(j)}}(\beta[t]) \qquad \mathcal{F}_{AR} = \{G_{b_{a(1)}}(\beta[t]), \ldots, G_{b_{a(p_a)}}(\beta[t])\} \qquad (2.3)$$

$$\sigma_w^2(\beta[t]) \triangleq \sum_{j=1}^{p_s} s_j \cdot G_{b_{s(j)}}(\beta[t]) \qquad \mathcal{F}_{sigma_w^2} = \{G_{b_{s(1)}}(\beta[t]), \ldots, G_{b_{s(p_s)}}(\beta[t])\} \qquad (2.4)$$

where $w[t]$ reflects the sequence of zero mean NID innovations with variance σ_w^2, $\mathcal{F}$ stands for a functional subspace of the respective quantity, $b_{a(j)}$ and $b_{s(j)}$ are the indices of the specific basis functions that are included in each functional subspace, while $a_{i,j}$ and s_j stand for the projection coefficients of the AR parameters and the innovations variance. LPV-AR models have been used to characterize the time-varying dynamics of wind turbine structures, where the scheduling variable $\beta[t]$ reflects the azimuth (position) of the rotor, while the additional dependence on varying Environmental and Operational Parameters (EOPs) was captured by additional use of Gaussian Process regression [15]. For OBM, LPV-AR enables the modeling of the response behavior of the vehicle axle for different types of track superstructure depending on the vehicle speed [108]. Fig. 2.8 shows the LPV-AR based Power Spectral Density (PSD) of the axle response in function of the vehicle speed and the type of track superstructure.
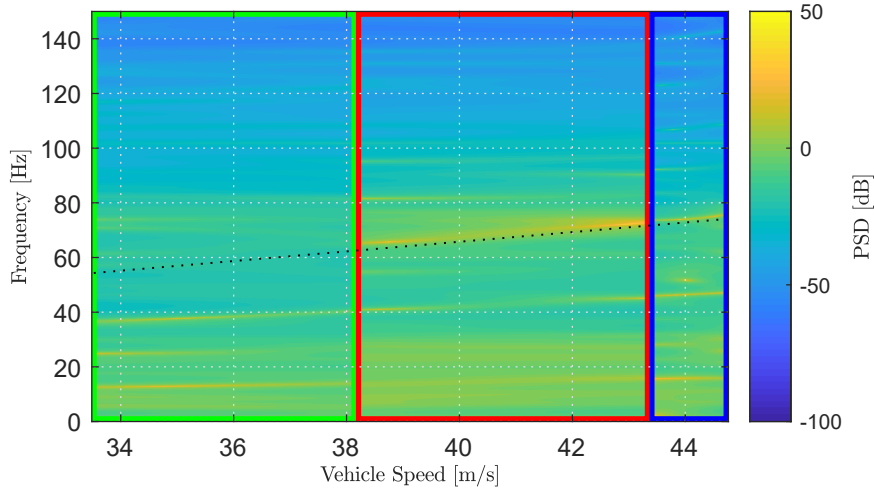

Figure 2.8: LPV-AR based PSD of the response for different track superstructures as a function of speed (adapted from [108]).

2.5.3 Bayesian Filtering

Bayesian filtering methods, such as the Kalman filter (KF), offer an efficient means for monitoring the state of linear dynamical systems. This approach has found broad application in real-time response estimation for the purpose of diagnostics and control. The KF requires coupling with a dynamic model of the system, which is used to establish the filter's state (process) equation. Such a model assumes that the state at time k is formulated as:

$$x_k = A_k x_{k-1} + B_k u_k + w_k \qquad (2.5)$$

where A_k and B_k denote the state and input matrices, correspondingly, of the linearized vehicle–track interaction model, x_k is the vector of dynamic states of the system, in discrete time, which comprises a concatenation of displacement and velocities, u_k is the vector of the inputs (excitation) to the system and w_k is the process noise, which is assumed to be drawn from a zero

mean multivariate normal distribution, $\mathcal{N}$, with covariance, $Q_k : w_k\, \mathcal{N}(0, Q_k)$.

The second equation necessary for the filter formulation is the measurement (observation) equation. The vector of sparse observations of the system response (e.g. accelerations of the axle), y_k, is given as:

$$y_k = C_k x_k + D_k u_k + v_k \qquad (2.6)$$

where C_k and D_k denote the observation and feed-through matrices and v_k is the observation noise, which is assumed to be zero mean Gaussian white noise with covariance, $R_k : v_k\, \mathcal{N}(0, R_k)$. The observation vector, or rather the discrepancy between the estimated and actually observed quantities (innovation vector), is used to update the model-based estimate of the system's response at time t_k. The interested reader is referred to [18] for more information on the KF formulation.

The KF has been used to detect faults on vehicle suspension systems [171] [125] [300]. Lederman et al. [154] have applied the adaptive KF, which weighs the data according to their estimated reliability, in order to combine data from several inexpensive sensors for monitoring the railway track geometry. Tsunashima et al. have shown that by applying KFs to the body motions of Shinkansen trains, irregularities in the track geometry can be reconstructed with acceptable accuracy [282]. Fig. 2.9 shows the comparison of the time–series and the PSD of the longitudinal level estimate obtained via application of a Dual Kalman Filter (DKF) on acceleration data (estimate) versus the true (measured) level. The DKF is a variant of the KF, which achieves dual estimation of the dynamic system's response (state) and the (unmeasured) input, which in this case corresponds to the rail level. The procedure proposed by Dertimanis et al. [69] can be used to accurately reconstruct the longitudinal level from accelerations.
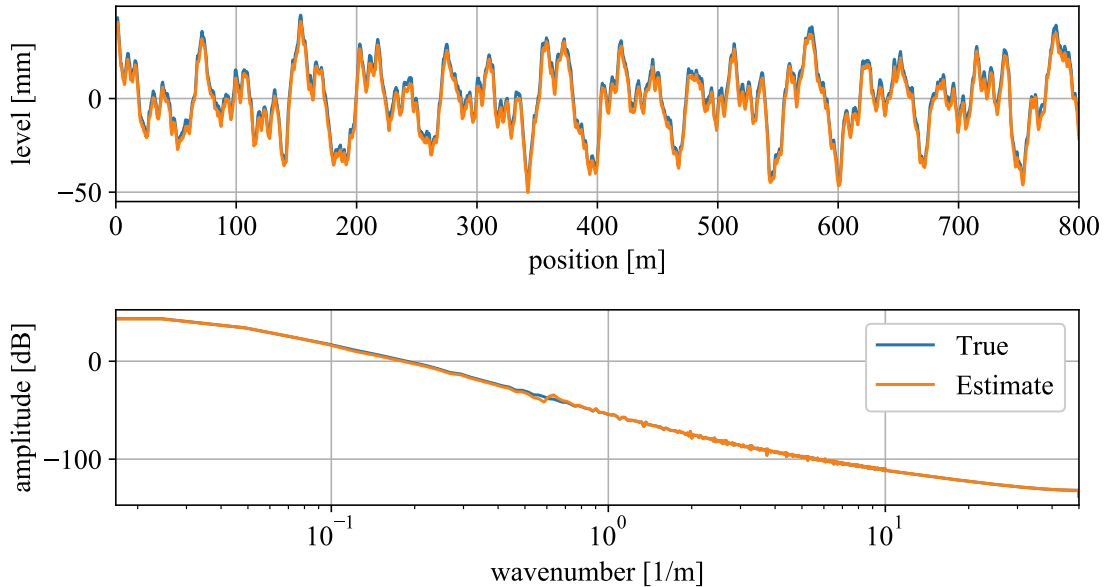

Figure 2.9: Time–series and PSD of KF–based estimate of longitudinal level (adapted from [69]).

2.5.4   The Vold Kalman Filter

KFs require knowledge of a-priori information on system dynamics in order to recursively estimate unknown states. This is succeeded via use of a proper state-space model, i.e., a model of the dynamics, as the employed process equation. The Vold Kalman Filter (VKF) operates on a different premise, whereby the known system dynamics is reflected in the measurable harmonic components within a response signal. The VKF is used to decompose a stationary stochastic process into the uncorrelated sum of a purely deterministic and a purely stochastic process [308].

The deterministic process is driven by the instantaneous frequencies of the system's periodic components [292][297]. The dynamics generated by a rotating structure, such as the axle of a train, forms a process $y(t)$ that can be described as the sum of a deterministic harmonic signal $x(t)$, whose parameters are function of the vehicle speed $v(t)$, and a stochastic process $\eta(t)$.

$$y(t) = x(t) + \eta(t) \tag{2.7}$$

Within the context of rail condition assessment, the process, $y(t)$, corresponds to the measured axle box accelerations, while $\eta(t)$ reflects noise contamination in the measurements. The harmonic signal, $x(t)$, is formulated as the product of a complex envelope $A_k(t)$ with a phasor $\rho_k(t)$.

$$x(t) = \sum_{s \in \mathcal{S}} \sum_{k \in \mathcal{K}_s} A_{sk}(t) \rho_{sk}(t) \tag{2.8}$$

where the phasor $\rho_{sk}(t) = e^{ik \int_0^t \omega_s(u) du}$ for each tracked order k of each measured instantaneous shaft speed s. Each component in the harmonic signal X corresponds to a process generated by a tracked order k and a shaft speed x. The VKF has been applied used to diagnose rotating equipment components, such as bearings and shafts of machines, vehicles and wind turbines [321][160][296][79][80] that are subjected to non-stationary conditions. While the VKF has been used extensively for order tracking in axles of vehicles and machinery, its use to diagnose the condition of the axle, wheel or track of OBM systems has not yet been reported. Here we show the potential of using the sleeper passage frequency, extracted using the VKF, for the estimation of the track stiffness. A high track stiffness results in increased contact forces and accelerations, which may result in premature failure of the infrastructure [165]. The track stiffness is commonly measured using special subsidence measurement vehicles[5]. Fig. 2.10 indicates that locations with increased track stiffness exhibit increased VKF sleeper passage components (corresponding to a wavelength of 0.6 m).

2.5.5   Extended-Kalman Filter

The extended Kalman filter (EKF) is an approximate filter for nonlinear systems, based on first-order linearization of the process and measurement functions. It is frequently used in joint parameter and state estimation problems for linear systems with unknown parameters [173]. In this case, the inclusion of the unknown parameters in the state vector of the system renders the estimation problem a nonlinear one. In the context that relates to rail condition estimation, the EKF has been combined with a weighted global integration procedure to identify the parameters of a beam under a moving mass load [116]. Lathe and Gautam [152] demonstrate the use of the EKF for reconstruction of vertical profile irregularities from OBM vibration measurements.

---

[5]For instance, the vertical track subsidence under an axle load of 20 tons is used by the SBB to measure the track stiffness[262]
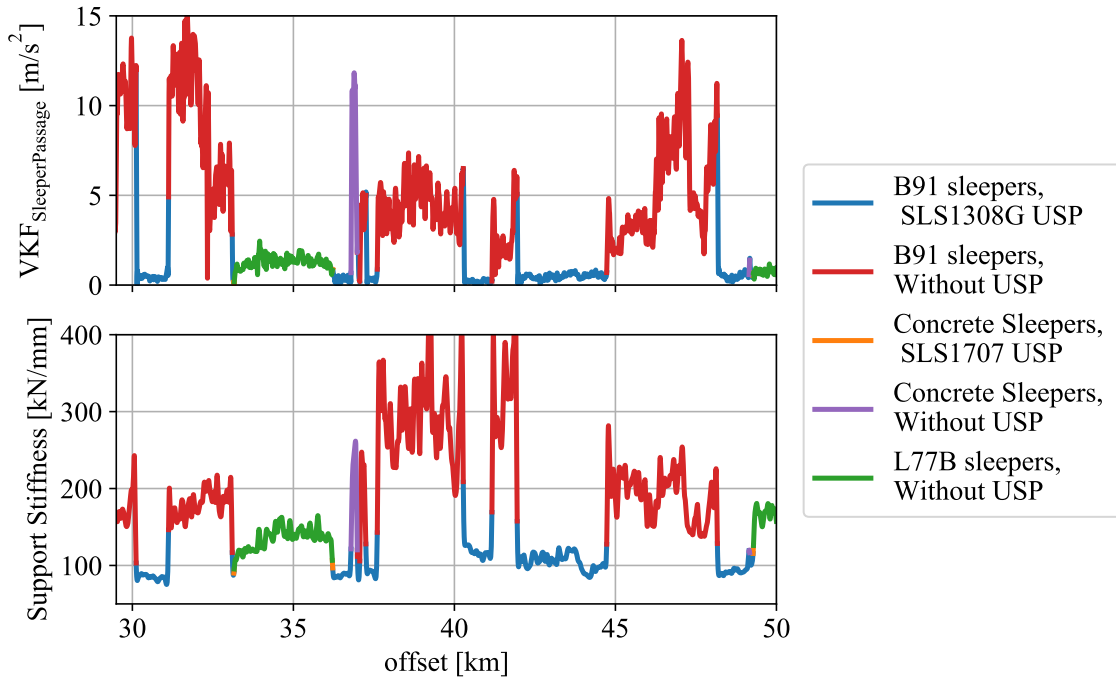
Figure 2.10: VKF Sleeper Passage acceleration amplitude compared to the measured track stiffness for varying sleeper and Under Sleeper Pad (USP) types (raw data from SBB).

2.5.6 Unscented-Kalman Filter

The Unscented-Kalman Filter (UKF) is a nonlinear variant of the KF that employs a deterministic sampling approach to estimate the state of nonlinear systems [130], i.e., systems where the state and/or observation equations are governed by nonlinear functions. Instead of linearizing the transformation function, as is the approach followed by the EKF, the UKF propagates the deterministic sample points (sigma points) through the nonlinear system functions in order to yield a posterior estimate of the system response (state). The friction coefficient is essential for estimating the lateral and creep forces exerted on the rail. The UKF has been used in railway condition assessment applications to indirectly estimate the rail-wheel friction coefficient, relying on measurements of the observed traction motor behavior [324], or of the dynamic response of a wheelset [217]. The rail-wheel friction coefficient has been indirectly estimated using the UKF applied on measurements of the observed traction motor behavior [324], or from the dynamic response of a wheelset [217].

2.6 Non-parametric methods

Non-parametric methods have gained popularity in dynamic signal analysis, since they alleviate the requirement for fitting a specific (parametric) model structure. Parametric and non-parametric methods are not mutually exclusive and their combination can prove beneficial for certain tasks, such as correlation and outlier analysis from signal statistics. Non-parametric methods include time and frequency domain analysis methods via time- or space-frequency decomposition.

2.6.1 Time-frequency analysis

Time–series of non-stationary signals, such as accelerations due to vehicle-track interactions, can be analyzed using a time-frequency or time-scale decomposition. Transformation methods used for non-stationary signals are the short-time Fourier transform (STFT), discrete wavelet transform (DWT),continuous wavelet transform (CWT) and fractal dimension analysis (FA) [150]. Further techniques include adaptive time-space methods, such as the empirical mode decomposition and the Hilbert-Huang transform, which are used to decompose a signal into intrinsic modes [39].

For the frequently employed fractal analysis indicator, the fractal dimension corresponds to the degree of roughness of a time–series for varying wavelengths. The fractal dimension is calculated from the slope of the logarithmic representation of the polygonal length to the logarithmic wavelength [150]. The fractal values, first applied to railway condition assessment by Landgraf [148], are commonly used to detect ballast and substructure damage. In the case of the Fourier and Wavelet transforms, the convolution of a functional basis with the time–series results in a time-frequency representation [92]. The choice of the optimal time–frequency analysis method to select, according to a specific assessment goal, is guided by the properties of each algorithm. In contrast to the STFT, which uses a fixed window size, wavelet transforms use an adaptive window size that is scaled proportionally to the length of the functional basis. Adaptive techniques, such as the wavelet transforms, result in an optimal resolution in the time and frequency domain. The continuous wavelet transform results in a significant redundancy at the expense of a large set of coefficients, while on the opposite spectrum, the discrete wavelet transform offers an efficient representation, where each level corresponds to a frequency band [92]. Invertibility is a desirable property that enables the recovery of the original signal from the set of coefficients. Both the STFT and the DWT satisfy the property of invertibility.

The most common time-frequency representations are shown in Fig. 2.11 for the time–series of vertical axle box accelerations during the crossing of a surface defect. The limitation in the resolution of the STFT is apparent, as compromises are made on both the time and the frequency resolution for a given window size. The CWT results in a redundant representation, while the DWT allows a condensed representation of the signal in both time and frequency.

Spectral analysis based on the Fourier transform has been applied to ABA signals to inversely determine the parameters of the vehicle from the measured response spectrum. The comparison between response spectra of different vehicles at different track locations has been used to identify damage locations on the vehicle suspension of Shinkansen trains [209]. Zili et al. [326] used the wavelet coefficients from a continuous wavelet transform (CWT) to classify squats and welds from axle box acceleration signals. Schenkendorf and Dutschk [246] used a similar approach as Molodova et al. [190] for identification of insulation joints. The irregularities on rail welds are characterized by identifying frequency and amplitude characteristics in the time-frequency response at joints and augmenting the identification process with a FE-Model [9]. In [55] axle box accelerations have been analyzed by means of the empirical mode decomposition and the Hilbert-Huang Transform in order to detect characteristic response signatures due to short-wave rail defects. The empirical mode decomposition and the Wigner-Ville distribution are used to characterize the amplitude of the wheel out-of-roundness [263].

Time-frequency analysis has been used to match acceleration patterns to the excitation source. However the limited generalizability due to variations of track parameters, vehicle parameters, environmental parameters and vehicle speed remains a major challenge. The numerical integration procedure described in Section 2.6.2 can be used to increase the interpretability of time-frequency analysis.
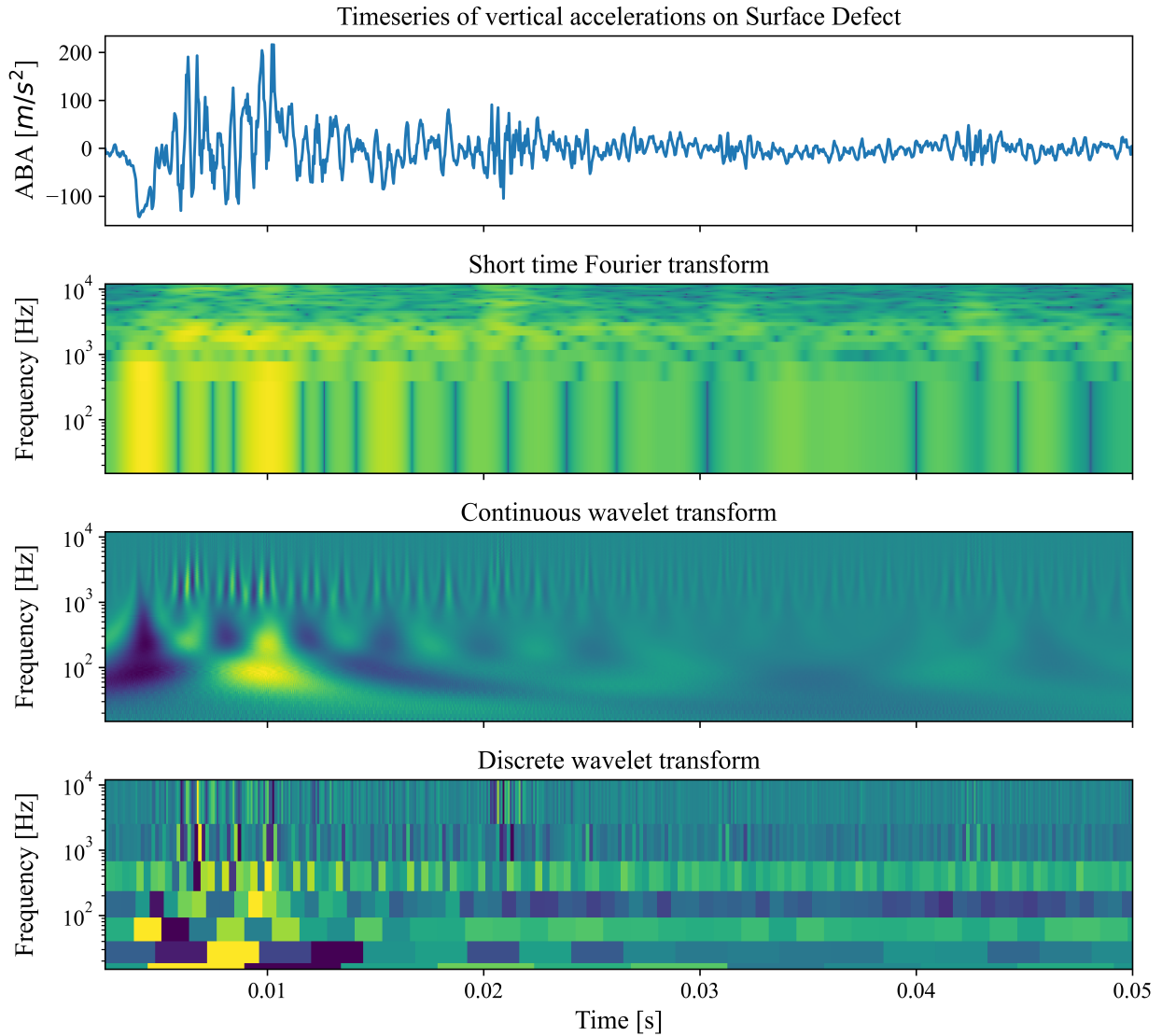
Figure 2.11: Time frequency representations of acceleration time–series on surface defect crossing.

2.6.2 Numerical Integration

The double integration of a discretely sampled acceleration signal at an observed degree of freedom makes it possible to recover the time–series of its displacement. Studies have shown that axle and track vibration are mostly independent of other vehicle degrees of freedom [69][6]. Therefore, the vertical displacement of the axle, obtained by double integration of the axle box accelerations corresponds to the combination of the longitudinal level and the displacement due to the wheel out-of-roundness. Due to its periodic nature, wheel out-of-roundness can be separated from the vertical displacement with suitable filters or with model based approaches such as the afore-mentioned Vold-Kalman Filter. The acceleration based longitudinal level has been proposed as a robust indicator in several studies [57][167]. The lateral axle displacement and the longitudinal level are band-pass filtered to the wavelength bands $D0$, $D1$ and $D2$ that are specified in the standards [52] in order to obtain a space-wavelength decomposition with three wavelength levels. Using the geometric properties of the track, the twist can be derived from the longitudinal level and the rail gauge. From the longitudinal level $D0$, the dip angle corresponding to the slope of the signal, is an indication for wheel rail impact loads that occur

on switches and joints [278][23]. Fig. 2.12 shows the comparison between the longitudinal level $D1$ obtained by TGMS (a laser based system) and by OBM (ABA based measurement). The difference between the longitudinal level $D1$ obtained from both systems is generally under 0.5[mm]. The indicators obtained from integration offer numerous advantages. Such indica-
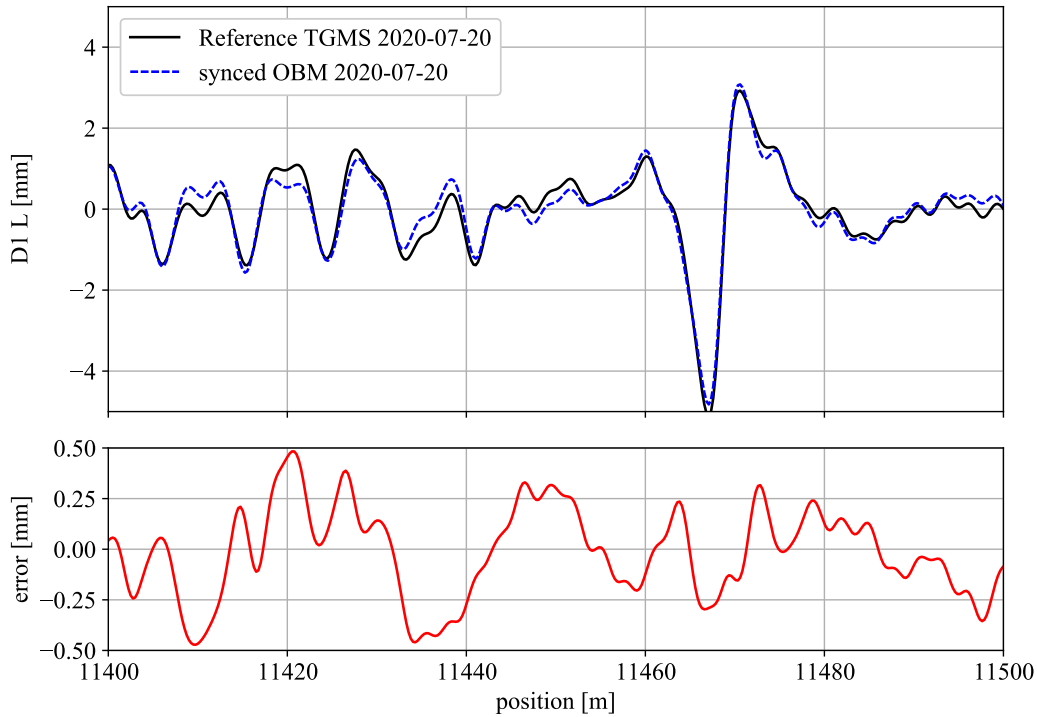

Figure 2.12: Comparison between TGMS (optical system) and OBM (ABA) based longitudinal level $D1$ (raw data from SBB).

tors are easy to interpret, largely independent of the vehicle speed and are straightforward to assimilate with existing measurements (TGMS Tab. 2.1); a trait which allows for the direct integration into existing processes and norms [52].

2.6.3 Statistical Features

Statistical Features are numerical measures, which allow the condensation of the signal into a set of indicators, resulting in a sparse representation of essential attributes [233]. These features include minimum, maximum and mean values, as well as standard deviation and higher statistical moments (skewness, kurtosis) and entropy. Depending on the scale of the phenomenon under consideration, these statistics can be calculated on the raw (original time–series or on processed data over different lengths of influence. The influence length (or window size) is usually chosen to be greater than the length of signal that is influenced by the effect of interest. For OBM data, the comparison of statistics between measurement rides is sensitive to the positioning precision, as well as to the influence length. The process of generating an event database containing statistical features is illustrated in Fig. 2.6. Regression, classification and outlier detection based on statistical features that are contained in such an event database can be used to define thresholds and related alerts for triggering decisions on remedial actions.

2.6.4 Parametric versus non-parametric methods

In general, parametric representations of times-series offer important benefits in comparison to their non-parametric counterparts. Starting with *parsimony*, parametric methods succeed in modelling time-series with a limited set of parameters. Indicatively, the spectral description of a signal of length N requires the same amount of data, whereas an ARMA model may require 20-40 parameters (depending on the complexity of the underlying dynamics), leading to significant *data compression*.

Parametric methods are further attributed with improved *accuracy* while using considerably less data. This is due to the fact that the associated finite parametrizations (difference equations, modal models, and state-space models) offer *closed form expressions* for engineering quantities of interest (e.g., frequencies, damping ratios, mode shapes, impulse responses, and spectra) at a considerably higher *resolution*, as well as reduced, theoretically sound, and quantified *uncertainty* (in most cases).

Arguably, parametric methods stand as the only available solution, when further analysis is required, as for example in applications that involve signal prediction, condition monitoring and control.

The power of non-parametric methods, on the contrary, lies in the immediate accessibility of the visual information they offer. Indeed, it is a matter of microseconds to derive the power spectrum, or the wavelet transform of a digital signal. This is further supported by the very low requirements for user-expertise, as opposed to parametric modelling, which requires a certain degree of familiarity with the associated methods. These methodologies are not mutually exclusive. For instance, statistical analysis or time-frequency analysis is often applied on the outputs of parametric or non parametric schemes [108][15].

2.7 Classification and outlier analysis

Non-parametric and parametric methods provide indicators that are related to the condition of the asset. The approximate state of the asset components at the time of an OBM measurement can be known a-priori from previous assessments. The resulting decision making is based on discrete (i.e. inspection labels) and continuous (i.e. exploitation percentage) condition quantities that are defined in railway norms and regulations. OBM provides additional indicators that should be linked to condition definitions based on a physics–based model or data driven model. Identification and classification methods presented in this section are often employed to achieve this goal.

2.7.1 Classification

Classification methods are machine learning algorithms that enable the prediction of a discrete outcome variable based on the value of one or multiple predictor variables. The outcome variable in monitoring railway tracks is often a continuous fault indicator or a discrete label. The predictor variables correspond to the measurements that are used to diagnose the infrastructure. The most common classification algorithms in supervised and unsupervised analysis are summarized in Tab. 2.5 [5]. For predictive maintenance of railway tracks, data driven methods are preferred, in which machine learning makes up three quarters of available literature [313]. Support Vector Machines (SVMs) (33%), Neural Networks (26%) and tree-based models (21%) are the most frequently adopted algorithms. These algorithms can be applied to raw time–series data or to features extracted from that data using the methods exposed in Section 2.6 and 2.6 [35].

Table 2.4: Comparison of parametric versus non-parametric methods

Method	Features	Limitations
Parametric Methods		
AR, ARMA	Stationary time-series modelling. Does not require input data (output-only)	Assumes linearity. Requires moderate user-expertise.
LPV-AR	Non-stationary time-series modelling. Does not require input data (output-only)	Assumes the observation of scheduling variables. May reach high model orders. Requires high user-expertise.
KF	Linear state, or state-input observer. Optimal for Gaussian process and measurement noise	Requires knowledge of the system dynamics
EKF	Nonlinear state, state-input, and state-input-parameter observer. Linearizes a model around an operating point.	Requires knowledge of the system dynamics. Not optimal, even for Gaussian process and measurement noise. Stability and convergence problems.
UKF	Nonlinear state, state-input, and state-input-parameter observer. Approximates the state probability density by a deterministic cloud sampling.	Requires knowledge of the system dynamics. Stability and convergence problems.
VKF	Separation of time-series of periodic system processes	Requires the observation of the system phasor and the assumption of the filter bandwidth
Non-Parametric Methods		
STFT, DWT, CWT	Compact time-frequency representations	Dependence on track and vehicle parameters or speed are not explicitly considered
Numerical Integration	Recovery of the time-history of directly observed states	Requires the direct observation of each degree of freedom that is to be analyzed
Statistical Measures	Can be applied for outlier detection on the parameters and signals of all previous methods	Best used in combination with other parametric or non-Parametric Methods

Table 2.5: Classification methods in supervised and unsupervised analysis [5]

Supervised Model	Unsupervised Analogs
Support-vector machines	One-class SVM
Neural Networks	Replicator neural networks
Decision Trees	Isolation Trees
Random Forests	Isolation Forests
Naive Bayes	Expectation-maximization
Rocchio	Mahalanobis method, Clustering
k-nearest neighbor	k-NN distance, LOF, LOCI
Rule-based	FP-Outlier
Linear Regression	Principal Component Analysis

A SVM has been applied to the continuous wavelet transform coefficients of acceleration signals to identify rail defects in [246]. The detection was improved by applying model inversion to the raw data. The model inputs are obtained by applying the inverse system derived from a quarter-car model to reconstruct the inputs that cause the dynamic response. In [318] a Convolution Neural Network (CNN) has been applied to raw axle box acceleration data to detect insulated joints. The CNN yields similar performance to a decision tree model trained on features that are computed on wavelet coefficients, but at the cost of higher complexity and elevated computational cost.

The continuous equivalent of classification is regression analysis. Regression methods are machine learning methods that allow to predict a continuous outcome variable based on the value of one or more predictor variables. Continuous outcome variables are in some cases reference measurements from an established system, such as track geometry or rail profile measurements. Sometimes continuous variables relate to operational conditions such as the vehicle speed. In [189] it is demonstrated how a polynomial regression implemented on a parametric analysis model has allowed to predict the expected peak acceleration from the crossing of a light squat at a given vehicle speed.

2.7.2 Outlier analysis

Outlier Analysis involves the identification of anomalous, deviant or discordant observations in a dataset. Outliers are observations of unusual behaviour in a system or process. There are two main causes of abnormal observations: Outliers can be caused by errors in the collected data itself due to hardware or software issues. Outliers can also reflect an abnormality in the system itself due to irregularities. All outlier detection schemes require a model that classifies normal behavior and anomalies. The model can be based on a physical vehicle track interaction model (see Section 2.5) or on machine learning classifiers [35].

A fully supervised scenario, in which both normal and abnormal samples are available and clearly labeled, is often preferred. However, the number of labels for a large amount of generated samples or measurements is in practice limited. Most often, examples of outliers and normal data are available from the existing track inspection processes. These labels can have an unknown proportion of biases, errors and noise. When a limited number of most informative observations are available, semi-supervised classification methods can be applied [46]. Unsupervised classification methods in Tab. 2.5 are suitable for identifying outliers from unlabeled data. Outliers can be qualified with binary labels or in terms of an outlier score by assessing each sample based on its deviation from a regular data model [4].

In real-world applications, such as OBM measurement scenarios, the data (and labels) can contain a significant amount of noise that blurs the line between normal samples and significant

anomalies. Each sample lies in a continuous spectrum from low to high outlier scores with the score uncertainty depending on the amount of noise. Fig. 2.13 illustrates the spectrum of outlier scores for three classes: normal data, noise and anomalies. The separation between these three classes is usually continuous. Some authors qualify noise and anomalies as weak outliers and strong outliers [5] [138].
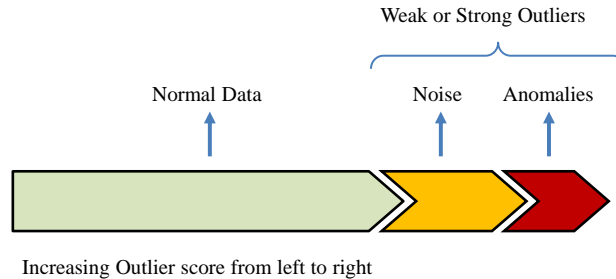

Figure 2.13: Spectrum from normal data to outliers [4].

With regard to OBM, outlier analysis aims to identify outliers that are due to irregularities on the rolling contact surface or to the dynamics caused by irregularities in the track geometry. Fig. 2.14 shows an application of outlier analysis to the insulated joint crossings stored in the event database, which is presented in Fig. 2.6. While typical insulated joints result in a longitudinal level $D0$ of in average -0.4 mm and a longitudinal level $D1$ around -2.5 mm, joints with loose screws - or even broken joint plate - result in severely increased values. Fig. 2.14 illustrates this effect for a joint with a broken joint plate. The longitudinal level at the location of the joint presents a dip in $D0$ of over 2.5 mm and a dip in $D1$ of around 13 mm after the damage occurrence. The evolution of outlier scores over measurement runs can be used to detect potential degradation of the condition of an asset and carry out interventions in time.

2.8 Conclusion

This work examines the potential rendered by exploitation of On Board Monitoring (OBM) data, that have been extracted from diagnostic or in service vehicles for monitoring the condition of railway infrastructure. Railways form a main pillar of modern transport infrastructure and a principal enabler of mobility. Following the idea of smart cities and the principles of digitization that underlie such a vision, it is imperative to build a mobility system of the future, i.e., one that exploits current monitoring potential for ensuring, safe, sustainable and resilient transport of people and goods. OBM measurements can be carried out at regular intervals in order to identify early damage, or deterioration of the condition of the tracks and further critical rail components, and to prompt associated maintenance measures with optimal planning. The use of such a diagnostic vehicle comes with operation costs, while further requiring trained personnel and careful planning, so as not to interfere with regular rail traffic. To overcome these obstacles, OBM via in service trains was recently proposed for supporting preventive and reactive maintenance tasks with a view toward automation.

This work offers an extensive overview of state-of-the-art research approaches for monitoring the condition of the rail and the track using data from simple, inexpensive and robust sensors installed in regular rolling stock. This offers an efficient and cost-effective complement to the calibrated measurements of the specialized diagnostic vehicles, which can yield significant benefits in preventive maintenance planning. A suite of parametric and non–parametric methods are presented in this article, along with exemplary applications on data from the Swiss federal
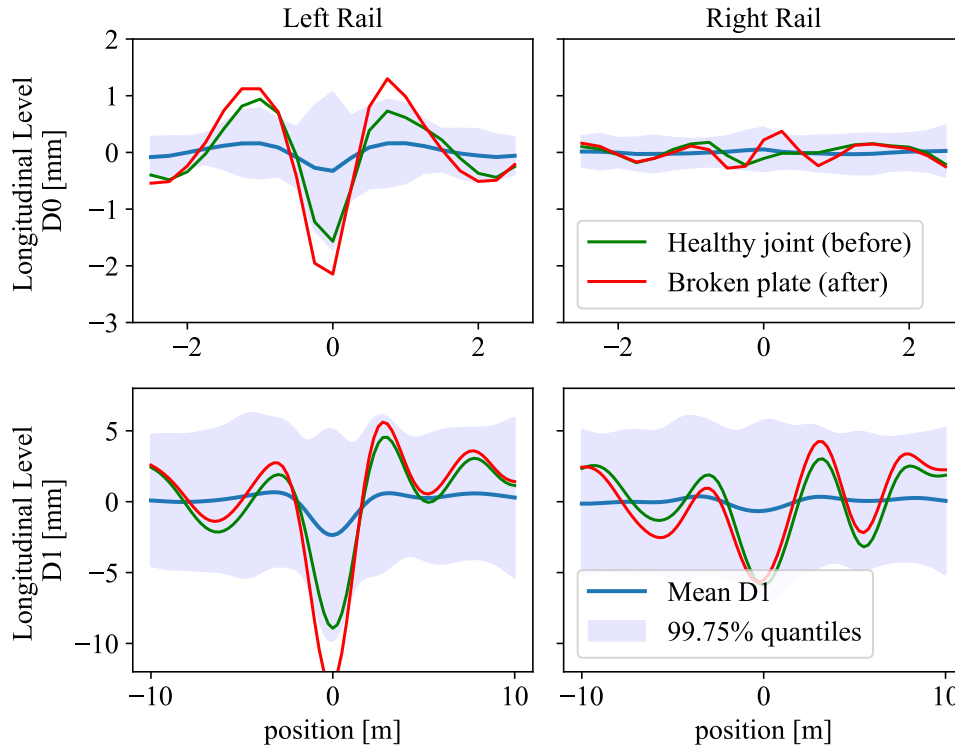
Figure 2.14: Outliers in Longitudinal level $D0$ (0.6-3 m) and $D1$ (3-25 m) from ABA (raw data from SBB) for left and right rail during crossings of insulated joints on the left side of the rail. Before (Fig. 2.2a) and after damage (Fig. 2.2b).

Railways (SBB) database. These methods are applied for extracting condition indicators which in a second step are coupled with classification techniques for damage characterization.

Owing to its potential for broad network coverage at frequent intervals, OBM via in-service vehicles offers a valuable tool for the timely detection of faults in railways. This approach may further be coupled with less frequent but higher precision measurements from specialized diagnostic vehicles in an effort to increase the predictive capabilities of the demonstrated diagnostic algorithms. Specialized diagnostic vehicles remain important for measuring the parameters related to the condition of the track that cannot be inversely determined via in-service OBM accelerometers (such as rail profile wear). The intelligent and meaningful evaluation of sensor data mounted on regular passenger trains represents a challenge, both in terms of the deployment aspects and the necessary investment in the required technology. Further research is necessary for direct exploitation of OBM-derived indicators for preventive maintenance. Existing acceleration-based monitoring systems are often customized to individual vehicles and require human intervention and expert opinion, for supporting further decision-making processes, including maintenance and repair actions. Challenges lie in building efficient frameworks for processing the aggregated data and integrating these into existing processes, reducing manual workload, and simultaneously dealing with potential false alarms or errors. A fully automated implementation is not a "low hanging fruit" - but bears tremendous potential for an intelligent and integrated railroad infrastructure of future smart cities.

Chapter 3

Vold-Kalman Filter Order tracking of Axle Box Accelerations for Rail Stiffness Assessment

Paper Details

Author and Co-Author Contributions

Key Findings

- The proposed multi-order extraction scheme separates deterministic (periodic) components from non-deterministic contributions to the measured axle response.

- The Vold-Kalman Filter (VKF) is a suitable approach to extract condition indicators characterizing the railway track stiffness and the condition of the vehicle wheels.

- The VKF can estimate amplitude components corresponding to the wheels and sleeper passage, providing information on wheel condition, potential flaws, and changes in track stiffness.

- The VKF-derived stiffness indicator is proportional to the rail-wheel forces and can improve the planning of optimal maintenance measures. High stiffness leads to increased damage accumulation on the rails and ballast. Coupled with existing substructure condition indicators such as the fractal values, this information could further support the Root Cause Analysis of the origins of increased stiffness.

General comments and Link to the next chapter

This study tackles the second objective of the thesis (see Section 1.4) by proposing a novel approach, exploiting a Vold-Kalman filter scheme, to assess the wheel Out-Of-Roundness and the track stiffness.

This chapter concludes the analysis of continuous excitation effects due to parametric excitation. It demonstrates the significant effect of track parameters on the vehicle axle response and the resulting degradation of infrastructure. In the following chapter, we shift the assessment focus to the classification of short-wavelength components and defects.

Abstract

Intelligent data-driven monitoring procedures hold enormous potential for ensuring safe operation and optimal management of railway infrastructure in the face of increasing demands on cost and operations efficiency. Numerous studies have highlighted track stiffness as a main parameter influencing the evolution of degradation that drives maintenance processes. As such, the measurement of track stiffness is fundamental for characterizing the performance of the track in terms of deterioration rate and noise emission. In this work, we propose a rail stiffness assessment scheme relying on low-cost On Board Monitoring (OBM) sensors, namely axle-box accelerometers, that are mounted on in-service trains and enable frequent, real-time monitoring of the railway infrastructure network. A Vold-Kalman filter is proposed for decomposing the signal into periodic wheel and track related excitation–response pair functions. We demonstrate that these components are in turn correlated to operational conditions, such as wheel out-of-roundness and the rail type. We further illustrate the relationship between the track stiffness, the measured wheel-rail forces and the sleeper passage amplitude, which can ultimately serve as an indicator for predictive track maintenance and prediction of track durability.

3.1 Introduction

The measurement of track stiffness is considered a fundamental driver for the continuous development of railway engineering, holding both theoretical and practical significance [298]. Track stiffness forms a primary feature, which can define the performance of railway infrastructure in terms of degradation progression [204] and acoustic noise emission to the surrounding environment [215].

The degradation of railway tracks is typically modeled by an analytical degradation model (e.g. a settlement equation), which relates cycles of dynamic forcing to incremental damage accumulation (i.e., ballast settlement or component wear) [63]. In such models, track stiffness is one of the critical parameters influencing loads. Real *et al.* [232] demonstrate that higher track stiffness, or equivalently low subsidence, generates axle loads that are distributed between fewer sleepers. As a result, the rail–wheel and sleeper-ballast contact forces increase, which may lead to premature rail and ballast wear [286]. A sudden decrease in track stiffness is often related to local damage of the ballast, substructure, and subgrade (changes in track type, fouled ballast, mud pumping or hanging ties) [106]. Given the defining nature of track stiffness for rail condition assessment, reliable measurements of this property can substantially support maintenance decisions [64].

Track stiffness can be estimated by (i) controlled experiments, where a measured external excitation is applied, e.g. using impact hammers, oscillating masses, or hydraulic cylinders [252, 147, 218, 30, 81]; (ii) standstill measurements of rail deflections at distinct track locations, due to the axle load of crossing vehicles [298]; or (iii) various types of vehicle–based measurement systems [262, 299, 48, 184, 117, 192, 229, 29, 30]. The latter category, which often uses the subsidence of the track under an axle load as an indicator of track stiffness [262], is increasingly preferred over the first two, since it delivers a high-accuracy assessment along extended track lengths. A common conclusion of all approaches is that the track stiffness is strongly dependent to the excitation frequency [142, 78, 141].

In terms of proposed methods, the stiffness may be directly calculated from measured force-displacement hysteresis curves [117, 192, 229], or by applying an explicit model (e.g. the Winkler model), which relates the subsidence to the load [184]. A wide class of studies applies some form of model updating [30], by fitting features of identified frequency response functions, either manually [81, 158], or by solving an associated optimization problem [62, 131, 310, 252]. The effects of varying track stiffness, resulting from the discretely supported rails, is also investigated

in several studies, primarily in a forward approach where the response of the track is simulated with finite element models [252, 239, 37, 36, 103]. In such scenarios, the rails are most commonly modeled as 4 DOF Timoshenko beam elements linked to the sleepers via discrete [239, 37] or distributed springs and dampers [36, 103]. The vertical wheel/rail contact stiffness is modeled via a Herzian spring. Vehicle models are most commonly composed of an axle and a sprung mass when considering only the track or axle response. From such studies, one can observe that peak responses occur at specific frequencies, such as the P2 resonance (50-100 Hz) or the rutting and roaring effects (250-1000 Hz). Such resonances are shown to lead to higher fatigue and damage [239].

Recently, on-board monitoring (OBM) vehicles have been introduced for conducting sparser measurements [107]. OBM vehicles are conventional passenger vehicles equipped with sensors, such as axle box accelerometers (ABA), which enable a nearly daily data collection and can be be processed into quantifiable track quality indicators [316].

Song *et al.* [263] characterize the wheel out-of-roundness (OOR) by applying the Empirical Mode Decomposition and the Wigner-Ville distribution to ABA. The Hilbert Huang transform spectrum characteristics have been used to study the effect of wheel flats on ABA while proposing different the frequency bands at different running speeds [164]. Li *et al.* [159] propose a short pitch corrugation detection method based on signatures in the wavelet power spectrum. Huang *et al.* [119] identify stiffness using convolutional neural networks from ABA response of the vehicle axle in a FE model. Quirke et al. [230] estimate both stiffness and damping from bogie accelerations.

Most of the aforementioned studies rely on the availability of a physical model. Purely data-driven schemes may also be applied, however such approaches are inherently challenging, especially when considering the particular nature of the excitation and disturbance sources for this problem, as well as the many parametric unknowns that occur in practice. It is noted that several methods have been applied to relate ABAs with geometric flaws, including the use of Kalman filters [69], double integration and filtering techniques [107] and linear parameter varying autoregressive models [108]. Non-parametric methods, such as time frequency analysis via Discrete [104] or Continuous Wavelet Transforms [326], have been also applied to identify rail flaws such as squats. Other fault effects, such as wheel OOR, are characterized using the Empirical Mode Decomposition (EMD) and the Hilbert-Huang transform (HHT) [263]. Infusing such non-parametric methods with physics–based models is proposed by Quirke *et al.* [230], who infer the track stiffness from the vertical bogie accelerations using a methodology that combines a vehicle-track interaction model with cross–entropy optimization.

In suggesting a scheme that is fit for an on–board, automated setting, we propose a stiffness indicator based on use of a Vold-Kalman Filter (VKF) on Axle Box Acceleration (ABA) measurements, which exploits the physical parameters of the wheel and the track to deliver wheel condition and track condition indicators. Unlike standstill stiffness and subsidence measurements, the proposed indicator can be applied in an in-service monitoring scheme, which attempts to deliver continuous assessment. The VKF is a parametric identification method that can be used to decompose the signal into harmonic responses, or orders, of periodic loads [100, 42]. It has been extensively used for monitoring the condition of rotating machinery equipment, such as wind turbine gearboxes [114, 160].

In the current setting, the VKF is used to extract the response functions corresponding to the sleeper passage frequency and to the wheel OOR frequencies from ABA measurements. Few previous studies have identified the large influence of periodic excitation such as the sleeper passage [239] and the wheel OOR excitation [26, 263] on the degradation of infrastructure. We demonstrate that these response functions are related to operational conditions, such as subsidence, wheel OOR and vehicle speed. We further show the relation between track stiffness, rail–wheel forces and the sleeper passage amplitude, suggesting that a VKF-derived stiffness

indicator can indeed support maintenance actions in a straightforward manner.

The structure of this paper is organized as follows: In Sec. 3.2 we describe the inverse problem of estimating track stiffness from ABA measurements. The fundamental aspects of the Vold-Kalman Filter are described in Sec. 3.3, whereas Sec. 3.4 contains the measurement data description and the results of the application of the proposed stiffness and wheel OOR identification. Finally, Sec. 3.5 outlines the core concept and results of the proposed methodology.

3.2 Problem formulation

3.2.1 Railway track stiffness

A stable track with generally consistent stiffness that can absorb the cyclic vehicle loads is a essential for an efficient low-maintenance rail system [271]. The characteristics of the substructure and the superstructure, including the drainage system, the ballast, and the sleeper, affect the degradation of the geometry and the rail. In particular, a homogeneous substructure improves the track geometry durability [106]. The vertical track stiffness is an essential parameter, which strongly affects the dynamic behaviour of the track and the resulting rail-wheel contact forces [140]. Vertical stiffness is often used to detect substructure flaws [262, 285], but it has in some cases also been related to rail faults such as squats [163].

The static track stiffness, or in other words, the overall stiffness of the track k_s, corresponds to the ratio of vertical load to the vertical deflection of the track,

$$k_s = \frac{Q}{u} \tag{3.1}$$

where u is the subsidence or deflection under the applied load Q. Variations in the static track stiffness can be attributed to the condition of the track superstructure and substructure, especially due to changes in selected components (rail, fasteners, sleepers, ballast and subgrade). The transition between superstructure or substructure types can cause sudden variations in stiffness resulting in increased wear and maintenance.

The dynamic track stiffness is defined as the inverse of the track receptance $\alpha(f)$ [298],

$$k_{dyn}(f) = \frac{1}{\alpha(f)} = \frac{Q(f)}{u(f)} \tag{3.2}$$

where u is the subsidence, or deflection, under the load Q and f. is the is the frequency of excitation. Dynamic stiffness expresses the ability of the railway track to resist dynamic loads such as wheel-rail interaction forces caused by the motion of trains. It is a measure of the track's ability to absorb and distribute these dynamic loads, reducing track deformation and preventing excessive track wear [94]. The dynamic track stiffness is influenced by factors such as track geometry, rail fastening systems, ballast condition, and subgrade stiffness. Resonance results in high deflections since the railway system is force-driven [276].

Table 3.1: Meaning of the $D_{A,S}$ naming convention of the sensors.

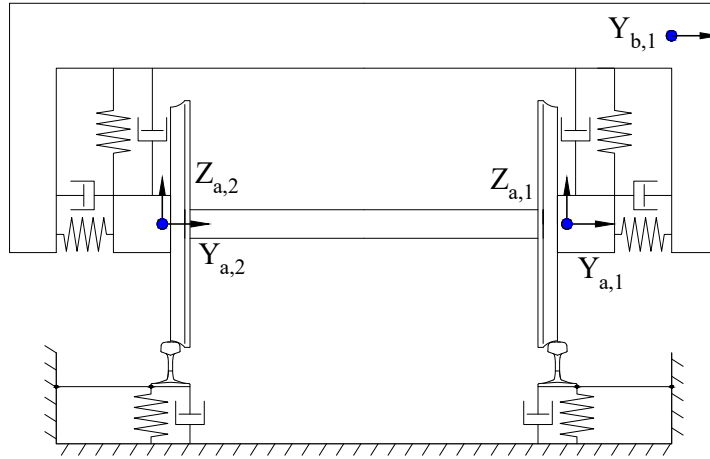| Letter	Explanation	Possible entries
D	direction	Y for lateral, Z/Q for vertical
A	axle number	1 to 4, starting from the front (leading) axle
S	vehicle side	1 for right, 2 for left (w.r.t vehicle's top view)

Figure 3.1: Schematic 2D primary suspension representation of wheelset, bogie and track with corresponding measured degrees of freedom.

3.2.2 Objective of this work

We consider the case of a diagnostic vehicle travelling along a railway track. The vehicle is instrumented with ABAs, mounted on its leading and trailing axles. Figure 3.1 displays a simplified roll-bounce dynamics representation of a vehicle axle connected to track and bogie, and shows the position of the sensors. The naming convention of the ABAs, complying with railway–specific standards, is $D_{A,S}$, with the related letter-entries being explained in Tab. 3.1.

The vibration response of the axle is composed of several periodic processes driven by the wheel OOR and the periodicity of sleepers. The wheelset is excited by the geometric irregularities of the track [267], OOR of the wheel [266] and parametric excitation (sleeper passage frequency, track parameters) [141]. Each of these excitation sources comprises a specific wavelength range (e.g. sleeper spacing and factors of the wheel diameter).

The vehicle-track interaction is dominated by four main classes of resonant effects: vehicle modes, the P2 effect and roaring or rutting [141]. Roaring or rutting result from the rail corrugation that occurs through a wavelength "fixing mechanism", which represents the dynamic behaviour of the vehicle–track system [93]. The P2 effect corresponds to the vertical resonance of the coupled train–track system [129]. Figure 3.2 illustrates the relation between vehicle speed, frequency and wavelength, as well as the frequency ranges for the main vehicle-track resonances that stem from vehicle-track interaction [107]. The vehicle speed is the main factor modulating the excitation frequencies relating to these effects [141].

Many studies demonstrate that increased dynamic response occurring at resonant frequencies, such as P2 resonance (50-100 Hz), or the rutting and roaring resonance (250-1000 Hz), leads to higher fatigue and damage on both on the railway vehicle (wheel OOR) and the railway track (corrugation) [129, 93]. The track stiffness has a major influence on these degradation processes, as contact forces directly relate to the track stiffness [141]. The periodic excitation stemming from the discrete rail supports (sleeper passage) can interact with these resonant modes, such as the P2 resonance and the track stiffness, resulting in high-amplitude dynamics [239].

The wheel OOR and the sleeper passage excitation are the two periodic excitation sources with known wavelength. The ABA response to these two periodic excitation sources is the result of their interaction with resonant modes, such as the P2 one. The response from the sleeper passage excitation is largely dependent on the stiffness and the damping of the track [239], but is also affected by the P2 resonance. Evidently, for tracks with comparatively higher superstructure stiffness, the amplitude of the acceleration response is increased in the frequency
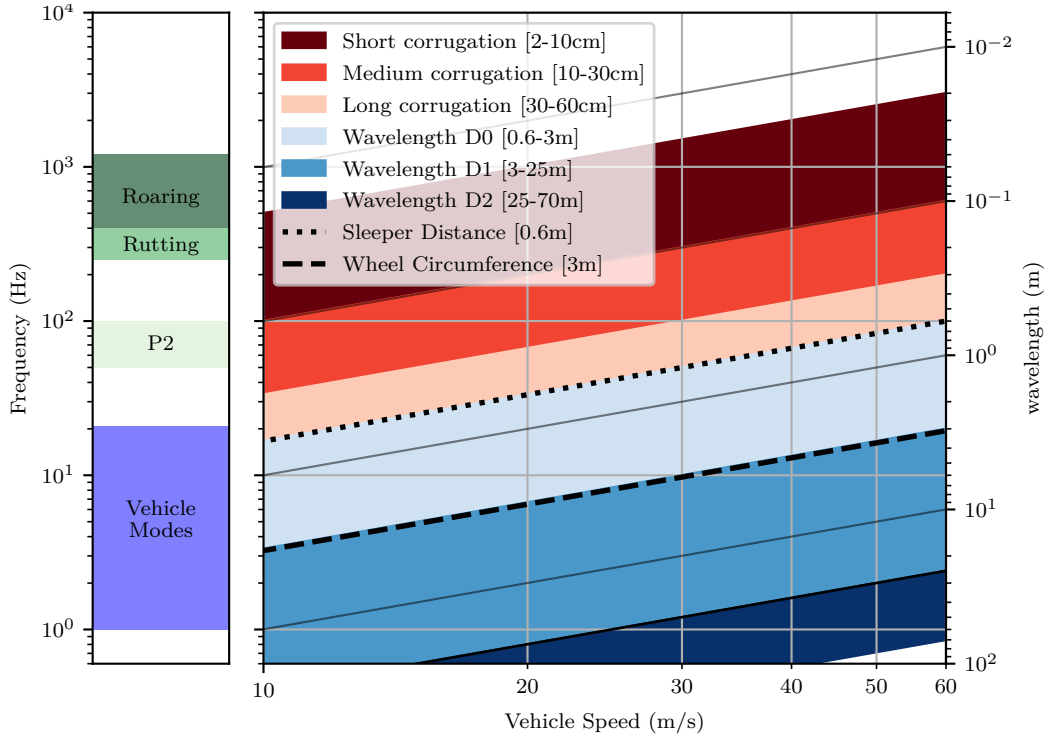
Figure 3.2: Relation between vehicle speed, frequency and wavelength, as well as frequency ranges for the main vehicle-track resonances. Excitation due to geometric irregularities (D0, D1, D2), corrugation, sleeper spacing and wheel OOR occur in specific wavelength ranges. The frequencies of excitation corresponding to these wavelengths are proportional to the vehicle speed. The main resonance levels (vehicle modes, P2, rutting and roaring) are highlighted on the left part of the figure

range up to 200 Hz [37, 36, 103].

The *angular velocity of the wheel* depends on the measured vehicle speed $v(t)$ and the known diameter of the wheel (herein $d_w = 0.92$ m) by

$$f_w[\ell] = \frac{v(t)\ell}{\pi d_w} \tag{3.3}$$

and it is used for extracting indicators of the wheel condition (OOR and flat spots). Each order ℓ of wheel OOR causes an instantaneous frequency of $f_w[\ell = 1, 2, ...n]$. The *track stiffness* is related to the instantaneous frequency caused by the sleeper passage, defined as

$$f_s = f_w[1]\frac{\pi d_w}{d_s} \tag{3.4}$$

with d_s denoting the sleepers' spacing (herein $d_s = 0.6$ m). The remaining ABA signal components comprise non-deterministic processes and measurement noise and may be used to characterize all non-periodic effects.

Under this setting, given estimates of f_w and f_s, the aim is to apply a data–driven scheme and identify the deterministic periodic components related to wheel OOR and sleeper passage, and correlate these to the dynamic stiffness of the track. This is succeeded by applying the VKF, outlined in the next section, to the measured ABAs of the vehicle.

3.3 The Vold-Kalman Filter

3.3.1 The process and measurement equations

Given a realization of a discrete-time, noise-corrupted ABA stochastic process, the Wold decomposition reads [307],

$$y[k] = x[k] + \eta[k] \tag{3.5}$$

where $k \in \mathbb{Z}^+$, $t = kT_s$ (s) are the associated discrete time indices for a given equidistant sampling period T_s, and $x[k]$, $\eta[k]$ comprise the mutually uncorrelated deterministic and stochastic components, respectively, with the latter corresponding to broadband random noise.

For the purpose of the current study, $x[k]$ is assumed to form a superposition of harmonic components, which are represented as modulated carrier waves of varying frequency [284], parameterized over the vehicle speed $v(t)$. That is

$$x[k] = \sum_{n=1}^{s} A_n[k] e^{j\phi_n[k]} \tag{3.6}$$

where $A_n[k]$ is a complex envelope and

$$\phi_n[k] = \sum_{\ell=1}^{k} 2\pi \frac{f_n[\ell]}{F_s} \tag{3.7}$$

is the phase of the complex phasor, with $f_n[\ell]$ and F_s denoting the instantaneous frequency and the sampling rate in Hz, respectively.

The Vold-Kalman filter aims at identifying the complex envelopes $A_n[k]$, given noise-corrupted observations $y[k]$, for $k = 1, \ldots, N$, and estimates of the instantaneous frequencies $f_n[\ell]$. Assuming that the complex envelopes comprise *smooth, slowly-varying* modulations of their associated phasors, they may be approximated by

$$\nabla^q A_n[k] = \epsilon_{n,q}[k] \tag{3.8}$$

where ∇^q is the finite difference operator of order q and $\epsilon_{n,q}[k]$ may be considered as the q-th gradient of $A_n[k]$. Under the adopted assumptions, Eq. 3.8 implies that the finite differences among successive time instants of the complex envelopes are "small". Moreover, it introduces a filtering effect. Indicatively, for $q = 1, 2, 3$ we get,

$$A_n[k] - A_n[k-1] = \epsilon_{n,1}[k] \tag{3.9a}$$
$$A_n[k] - 2A_n[k-1] + A_n[k-2] = \epsilon_{n,2}[k] \tag{3.9b}$$
$$A_n[k] - 3A_n[k-1] + 3A_n[k-2] - A_n[k-3] = \epsilon_{n,3}[k] \tag{3.9c}$$

which pertain to standard, non-homogeneous difference equations, represented by associated fixed-pole digital transfer functions in the $\mathcal{Z}$ domain.

By considering Eqs. 3.8 and 3.5 as *process* and *measurement* equations, respectively, a state-space system of the form,

$$\nabla^q \mathbf{a}[k] = \boldsymbol{\epsilon}[k] \tag{3.10a}$$
$$y[k] = \mathbf{c}[k]\mathbf{a}[k] + \eta[k] \tag{3.10b}$$

is generated, where

$$\mathbf{a}[k] = \begin{bmatrix} A_1[k] & A_2[k] & \ldots & A_s[k] \end{bmatrix}^T \qquad [s \times 1] \tag{3.11}$$

is the unknown state vector,

$$\boldsymbol{\epsilon}[k] = \begin{bmatrix} \epsilon_{1,q}[k] & \epsilon_{2,q}[k] & \cdots & \epsilon_{s,q}[k] \end{bmatrix}^T \qquad [s \times 1] \tag{3.12}$$

$$\mathbf{c}[k] = \begin{bmatrix} e^{j\phi_1[k]} & e^{j\phi_2[k]} & \cdots & e^{j\phi_s[k]} \end{bmatrix} \qquad [1 \times s] \tag{3.13}$$

and $\phi_n[k]$ is given by Eq. 3.7.

3.3.2 A least-squares solution

We cannot apply the original Kalman filter to the set of Eqs. 3.10, since both $\boldsymbol{\epsilon}[k]$ and $\eta[k]$ are unknown and the entries of $\mathbf{c}[k]$ are usually estimated from data. Instead, a Kalman *smoothing* solution can be established on the basis of available measurements [76]. To demonstrate this solution, define the $[N \times 1]$ vectors

$$\mathbf{A}_n = \begin{bmatrix} A_n[1] \\ A_n[2] \\ \vdots \\ A_n[N] \end{bmatrix}, \qquad \mathbf{E}_n = \begin{bmatrix} \epsilon_{n,q}[1] \\ \epsilon_{n,q}[2] \\ \vdots \\ \epsilon_{n,q}[N] \end{bmatrix} \tag{3.14}$$

for $n = 1, 2, \ldots, s$ and

$$\mathbf{y} = \begin{bmatrix} y[1] \\ y[2] \\ \vdots \\ y[N] \end{bmatrix}, \qquad \boldsymbol{\eta} = \begin{bmatrix} \eta[1] \\ \eta[2] \\ \vdots \\ \eta[N] \end{bmatrix} \tag{3.15}$$

Then, Eq. 3.8 can be written over the available data indices as

$$\mathbf{S}_q \mathbf{A}_n = \mathbf{E}_n \tag{3.16}$$

for $n = 1, 2, \ldots, s$, where $\mathbf{S}_q$ is a *known* $[N \times N]$ matrix that depends only on the filter's order q (i.e., it is n-independent, as demonstrated in Eq. 3.9). By further defining the $[Ns \times 1]$ vectors

$$\mathbf{a} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_s \end{bmatrix}, \qquad \mathbf{e} = \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \vdots \\ \mathbf{E}_s \end{bmatrix} \tag{3.17}$$

allows expanding Eq. 3.16 over the index n as

$$\mathbf{S}\mathbf{a} = \mathbf{e} \tag{3.18}$$

for $\mathbf{S} = \mathbf{I}_s \otimes \mathbf{S}_q$ ($[Ns \times Ns]$).

Accordingly, Eq. 3.10b may be written as

$$y[k] = \sum_{n=1}^{s} c_n[k] A_n[k] + \eta[k] \tag{3.19}$$

with $c_n[k]$ denoting the n-th entry of $\mathbf{c}[k]$ in Eq. 3.13. Writing Eq. 3.19 for $k = 1, 2, \ldots, N$ gives

$$y[1] = c_1[1]A_1[1] + c_2[1]A_2[1] + \cdots + c_s[1]A_s[1] + \eta[1]$$
$$y[2] = c_1[2]A_1[2] + c_2[2]A_2[2] + \cdots + c_s[2]A_s[2] + \eta[2]$$
$$\vdots \qquad\qquad \vdots$$
$$y[N] = c_1[N]A_1[N] + c_2[N]A_2[N] + \cdots + c_s[N]A_s[N] + \eta[N]$$

or, using the definitions of Eqs. 3.14–3.15

$$\mathbf{y} = \sum_{n=1}^{s} \mathbf{C}_n \mathbf{A}_n + \boldsymbol{\eta} \tag{3.20}$$

with $\mathbf{C}_n = \mathrm{diag}\{c_n[1], c_n[2], \ldots, c_n[N]\}$. Then, for

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 & \ldots & \mathbf{C}_s \end{bmatrix} \quad [N \times Ns] \tag{3.21}$$

Eq. 3.20 becomes

$$\mathbf{y} = \mathbf{C}\mathbf{a} + \boldsymbol{\eta} \tag{3.22}$$

A least-squares problem can be now formulated by considering the objective function

$$V(\mathbf{a}) = \left(\mathbf{R}\,\mathbf{e}\right)^T \left(\mathbf{R}\,\mathbf{e}\right) + \boldsymbol{\eta}^T \boldsymbol{\eta} \tag{3.23}$$

where $\mathbf{R} = \mathrm{diag}\{\mathbf{I}_N r_1, \mathbf{I}_N r_2, \ldots, \mathbf{I}_N r_s\}$ ($[Ns \times Ns]$) is a diagonal matrix of weighting factors, which balance the influence of the process equation to the objective function [76]. Substituting Eqs. 3.18,3.22 and setting the gradient of $V(\mathbf{a})$ equal to zero finally yields

$$\left[\mathbf{C}^H \mathbf{C} + \left(\mathbf{R}\mathbf{S}\right)^T \left(\mathbf{R}\mathbf{S}\right) \right] \mathbf{a} = \mathbf{C}^H \mathbf{y} \tag{3.24}$$

with H denoting Hermitian transpose. Due to the special structure of the involved matrices, Eq. 3.24 constitutes an ill-conditioned problem, the solution of which requires iterative methods, such as the preconditioned conjugate gradient one. UMFPACK [65] is a library which was used to efficiently solve this sparse inverse problem. The reader is referred to Feldbauer and Holdrich [76] for further details.

3.3.3 Implementation and solution

The resulting linear system of coupled equations is formulated as a sparse matrix product and is solved using a sparse direct solver [65]. The execution time of this solver depends linearly on the matrix order. For a large time series, the direct solution is obtained by dividing the time history into overlapping bins. The decomposed signals are reassembled with a Hanning window taper on the overlapping parts. The source code for the Python 3 implementation of the second order VKF, developed as a part of this paper, has been made openly available [101] for reuse by interested readers.

3.3.4 Validation on a simulated signal

In demonstrating the workflow of the VKF, the filter is first applied on a synthetic multi-harmonic acceleration signal of the form,

$$y[k] = x[k] + \eta[k] \qquad (3.25)$$

where

$$x[k] = \sum_{n=1}^{3} a_n \cos\left(2\pi\frac{f_n}{F_s}k + \theta_n\right) = x_1[k] + x_2[k] + x_3[k] \qquad (3.26)$$

with $k = 1,\ldots,N$, $N = 6 \times 10^5$ and $F_s = 12$ kHz. The amplitudes a_n, frequencies f_n and phases θ_n are given by

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} 600 + 120\cos\left(2\pi\frac{0.03}{F_s}k\right) \\ 240 - 120\cos\left(2\pi\frac{0.01}{F_s}k\right) \\ 252 - 126\cos\left(2\pi\frac{0.01}{F_s}k\right) \end{bmatrix} \qquad (3.27a)$$

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} \frac{0.02}{F_s}k + 1 \\ 0.7 + 0.4\sin\left(2\pi\frac{0.02}{F_s}k\right)\cos\left(2\pi\frac{0.04}{F_s}k\right) \\ 0.9 + 0.4\sin\left(2\pi\frac{0.02}{F_s}k\right)\cos\left(2\pi\frac{0.04}{F_s}k\right) \end{bmatrix} \qquad (3.27b)$$

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} \frac{0.06}{F_s}k - 2 \\ 0 \\ -1 \end{bmatrix} \qquad (3.27c)$$

respectively, while $\eta \in \mathcal{N}(0, \sigma_\eta^2)$, with $\sigma_\eta = 0.75$ m/s^2. Equations 3.26–3.27 indicate that $x_1[k]$ features a sinusoidal frequency variation and a linearly varying amplitude variation and phase, $x_2[k]$ attains a sinusoidal frequency and amplitude variation, whereas $x_3[k]$ corresponds to $x_2[k]$ with a small shift in frequency, amplitude and phase to demonstrate the approach on two closely spaced phasors. The spectrogram (Welch's method with $N_{FFT} = 256$ and 50% overlap) of $y[k]$ is displayed in Figure 3.3a.

The implementation of the VKF to the synthetic, noise-corrupted data of Eq. 3.25 proceeds by estimating first the complex envelopes A_n of Eq. 3.6 via Eq. 3.24 and accordingly recovering the amplitudes a_n and phases θ_n from

$$\hat{a}_n = ||A_n|| \qquad (3.28a)$$

$$\hat{\theta}_n = \arg\{A_n\} \qquad (3.28b)$$

The results are illustrated in Figs. 3.3b–3.3c and show excellent match to Eqs. 3.27b–3.27c, allowing a proper reconstruction of the simulated signal. Indeed, as Fig. 3.4 demonstrates, the VKF-estimated noise–free signal

$$\hat{x}[k] = \sum_{n=1}^{3} \hat{a}_n \cos\left(2\pi\frac{f_n}{F_s}k + \hat{\theta}_n\right) \qquad (3.29)$$

is in very good agreement to the original one.

In demonstrating the robustness of the filter against noise–corrupted observations, the multi-harmonic synthetic signal $x[k]$ defined in Eq. 3.26 is corrupted with white noise $\eta[k]$ of varying
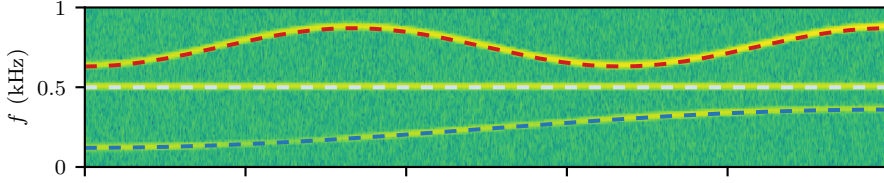
(a) Spectrogram of $y[k]$ and associated time-varying frequencies $f_n[k]$ of the constitutive components.
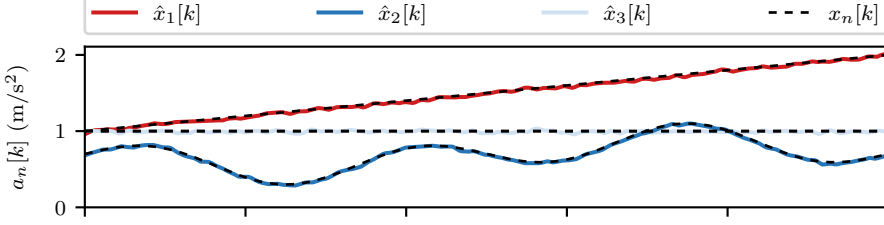
(b) VKF-extracted amplitudes (coloured curves) over their theoretical counterparts (black dashed curves).
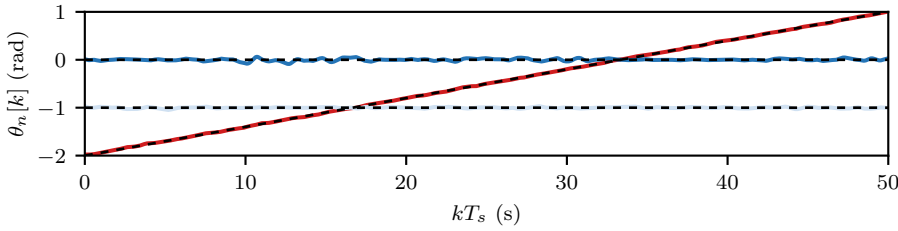
(c) VKF-extracted phases (coloured curves) over their theoretical counterparts (black dashed curves).

Figure 3.3: The synthetic, noise–corrupted signal of Eq. 3.25 and the VKF estimates of order $q = 2$. The extracted amplitudes and phases agree to the original ones.

Noise-to-Signal Ratios (NSRs), herein defined as

$$\text{NSR} = 100\frac{\sum_k \eta[k]^2}{\sum_k x[k]^2} \quad (\%) \tag{3.30}$$

while the error on the extracted VKF components is quantified in terms of Normalized Mean Square Error (NMSE)

$$\text{NMSE} = 100\frac{||x_n - \hat{x}_n||}{||x_n - \bar{x}_n)||} \quad (\%) \tag{3.31}$$

where $\bar{x}_n$ are the mean values of the theoretical signal components x_n and $\hat{x}_n$ are the estimated VKF components. The NMSE of the estimated VKF components is shown in Fig. 3.5, from where the robustness of the filter is confirmed: even for noise levels of around 100%, the error remains constrained below 0.5%.

3.4 Case study on real-world ABA measurements

3.4.1 Data description

To demonstrate the use of the VKF for the estimation of the wheel OOR and the dynamic stiffness of the railway track, data from different sources is analyzed, covering more than 100 km of track on the network of the Swiss Federal Railways (SBB). This data, which is generally used by railway infrastructure asset managers for decision-making purposes [200], is composed of periodic inspection data, infrastructure inventory information and maintenance data.
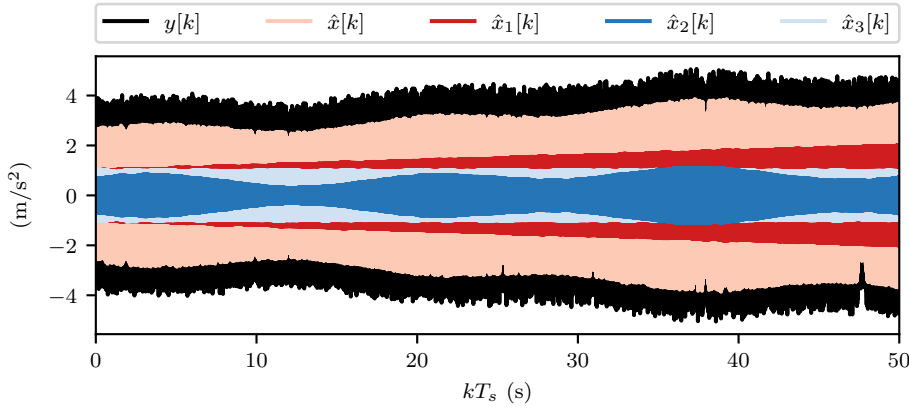
(a) Full time series of the noisy signal $y[k]$, the VKF-filtered noise free signal $\hat{x}[k]$, and its constitutive components $\hat{x}_1[k]$, $\hat{x}_2[k]$, $\hat{x}_3[k]$.
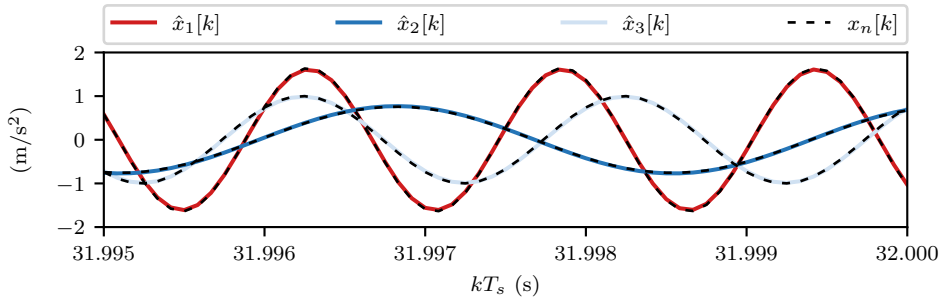
(b) 5 ms zoom showing the VKF-filtered components $\hat{x}_n[k]$ against the theoretical ones $x_n[k]$.
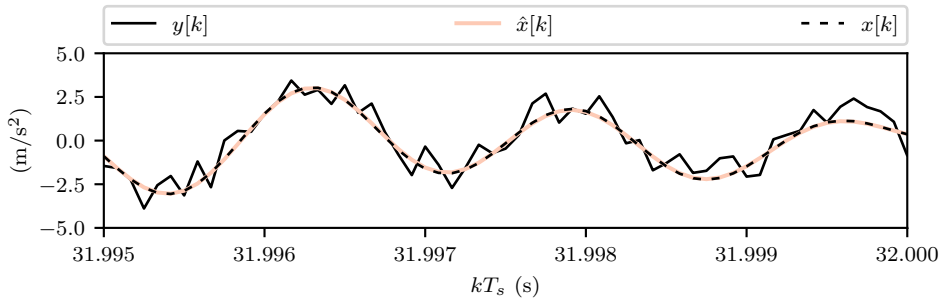
(c) 5 ms zoom showing $y[k]$, $\hat{x}[k]$, and $x[k]$.

Figure 3.4: Time-series of $y[k]$, $\hat{x}[k]$, and its constituent components, $\hat{x}_1[k]$, $\hat{x}_2[k]$ and $\hat{x}_3[k]$, estimated with the VKF.
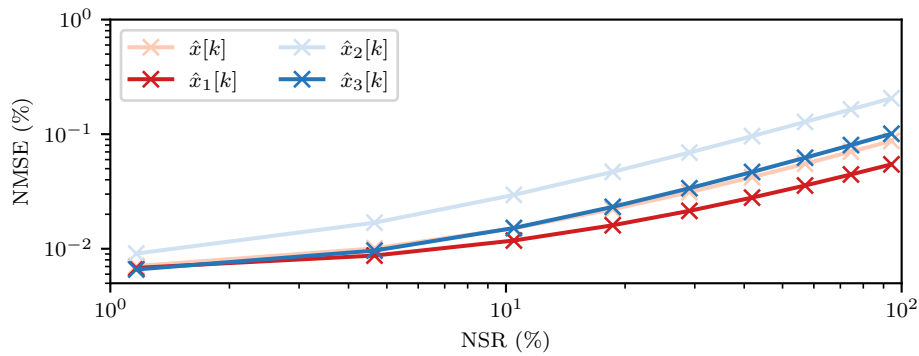

Figure 3.5: Normalized mean square error (NMSE) of the estimated VKF components $\hat{x}[k]$, $\hat{x}_1[k]$, $\hat{x}_2[k]$, $\hat{x}_3[k]$ for varying NSRs.

66

Table 3.2: Summary of VKF frequencies for wheel OOR and sleeper passage signal extraction.

VKF parameters	Wheel OOR	Sleeper passage
f (Hz)	$f_w(l = 1, ..., 11)$ (Eq. 3.3)	f_s (Eq. 3.4)

In more detail, inspection data acquired from a specialized diagnostic vehicle (gDFZ) and a subsidence measurement vehicle (EMW) are herein assessed. The gDFZ, which is a specialized measurement vehicle equipped with a variety of instrumentation systems, offers dynamic measurements of acceleration and force, using ABA and tensiometric wheelsets (TWS), respectively. TWS are special wheelsets measuring wheel-rail contact forces [248]. The ABA and the TWS are mounted to the leading and the trailing axle of the vehicle, as shown in Fig. 3.1. The subsidence measured by the EMW is inversely proportional the track static stiffness k_s

$$k_s = \frac{Q_{EMW}}{u} \tag{3.32}$$

where u is the subsidence and Q_{EMW} is the static axle load of the EMW.

Dynamic measurements from the gDFZ vehicle, conducted since Spring 2019, are processed using the VKF and subsequently correlated to the subsidence. Measurement of the latter was carried out on the same track section in 2016 by the EMW (InfraMT & SBB) [262]. Moreover, information on the track superstructure components and their maintenance history is extracted from the fixed asset database (DfA of SBB) and used to correlate the static and the dynamic track stiffness characteristics to the track type.

3.4.2 Order extraction with Vold-Kalman filter

Harmonic signals are extracted from the ABA signals to account for the sleeper passage frequency and 11 orders of wheel OOR, as defined in Tab. 3.2. Figure 3.6 illustrates the spectrograms of the original noisy signal and the VKF-based estimated harmonics, as a function of the vehicle speed. The VKF clearly succeeds in separating the deterministic components of the noise–corrupted observations. These components are composed of the ones relating to the wheel geometry (OOR) and excitation linked to passage over the sleepers, which are spaced usually at a distance of 60 cm. The response associated to sleeper-passage is non-stationary, since the track properties (stiffness, inertia, damping) are varying.

Figure 3.7 illustrates the time series of the raw ABA signal Z_{12}, the filtered components $\hat{Z}_{12}$ and $\hat{Z}_{12,i}$ and the residuals η. The VKF filtered signal separates the excitation stemming from the wheel OOR and the sleepers spacing from the remaining noise. The VKF specifically filters the frequencies corresponding to the periodic excitations, which in some cases may overlap further resonant effects described in the problem formulation. While the specification of the target frequency does not allow distinguishing between specific sources of excitation that may be jointly contributing in that frequency, the VKF captures the complete response corresponding to that frequency; in this case, the sleeper passage and wheel OOR wavelengths.

3.4.3 Wheel OOR

The wheel OOR causes a harmonic excitation, whose components are apparent in the acceleration spectrogram of sensor Z_{12} in Fig. 3.6b. The profile of the wheel can be reconstructed by summing up the components of the 11 VKF–orders extracted in Sec. 3.4.2 that correspond to

(a) Spectrogram of the unfiltered ABA signal Z_{12}. The signal reflects the large diversity of several excitation sources and operating conditions.

(b) Spectrogram of the extracted VKF components for sensor channel Z_{12}. The noise free VKF components corresponding to the frequencies defined in Tab. 3.2
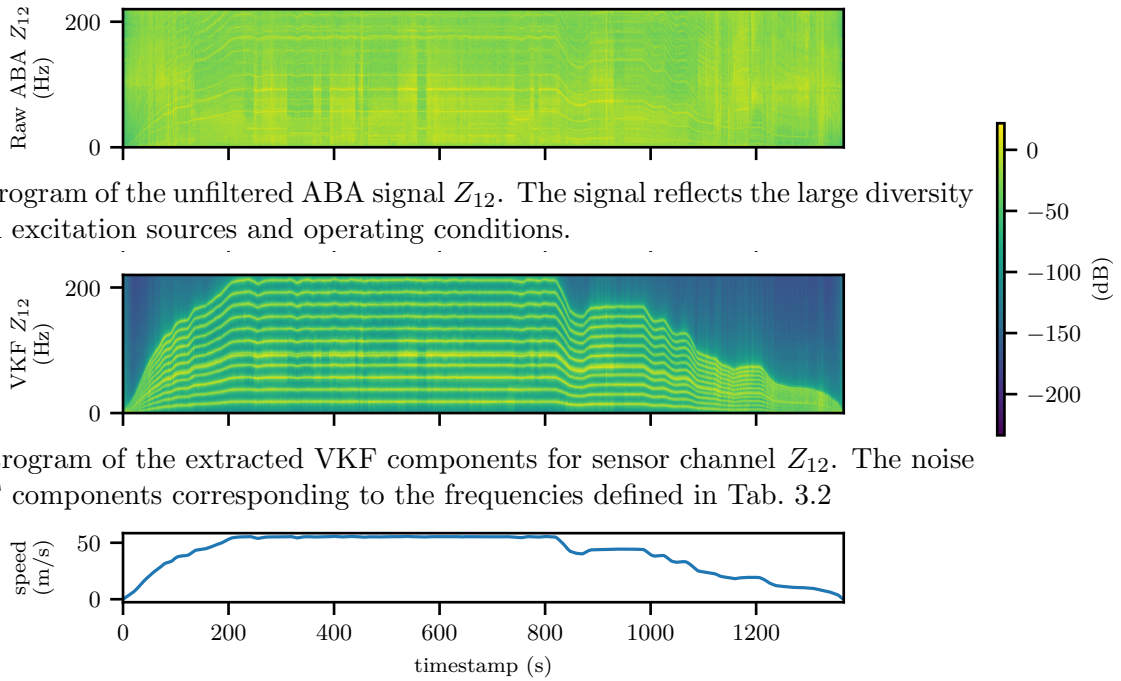
(c) The non-stationary vehicle speed $v(t)$ can reach up to 200 km/h.

Figure 3.6: Spectrogram of the noisy signal and of the corresponding harmonics whose frequency is proportional to the vehicle speed. The VKF filtered signal separates the excitation stemming from the wheel OOR and the sleepers spacing from the remaining noise.

the wheel's OOR harmonic frequencies, i.e.,

$$r(x) = 0.5d_w + \sum_{k=1}^{11} A_k(x) \tag{3.33}$$

where the radius r of the wheel is the sum of the mean wheel radius $r_w = 0.5d_w = 0.46$ m and the amplitude of the envelope function A_k for each wheel OOR order k along a position x of the wheel circumference. Figure 3.8a shows a polar plot with the wheel OOR for the four wheels that are equipped with vertical acceleration sensors. The maximum OOR amplitude shown in Fig. 3.8b lies below 300 μm, which is equivalent to the OOR observed in the field tests by Nielsen and Johansson [206]. The left wheel (Z_{42}) and the right wheel (Z_{41}) on axle 4 have a similar OOR to the left wheel (Z_{12}) and right wheel (Z_{11}) on axle 1. While an exact profile is not available for these wheelsets, it is known that they are in good condition at the time of measurement. This is corroborated by the low amplitude of the VKF-based wheel profile reconstruction.

3.4.4 Track stiffness

The sleeper passage excitation stems from the discrete support condition of the rail which is fixed to the sleepers at intervals on 60 cm. The amplitude of the sleeper passage component of the axle vibration varies in function of the vehicle speed and track stiffness. As the vehicle speed varies, the parametric excitation due to the vehicle or the track may overlap with the frequency of the vibration modes, resulting in increased accelerations. The amplitude of the sleeper passage acceleration component, obtained in Sec. 3.4.2 is henceforth notated as VKF Z_{12} (60 cm wavelength).

Figure 3.9b displays the VKF–estimated Z_{12} amplitude, extracted for a measurement per-
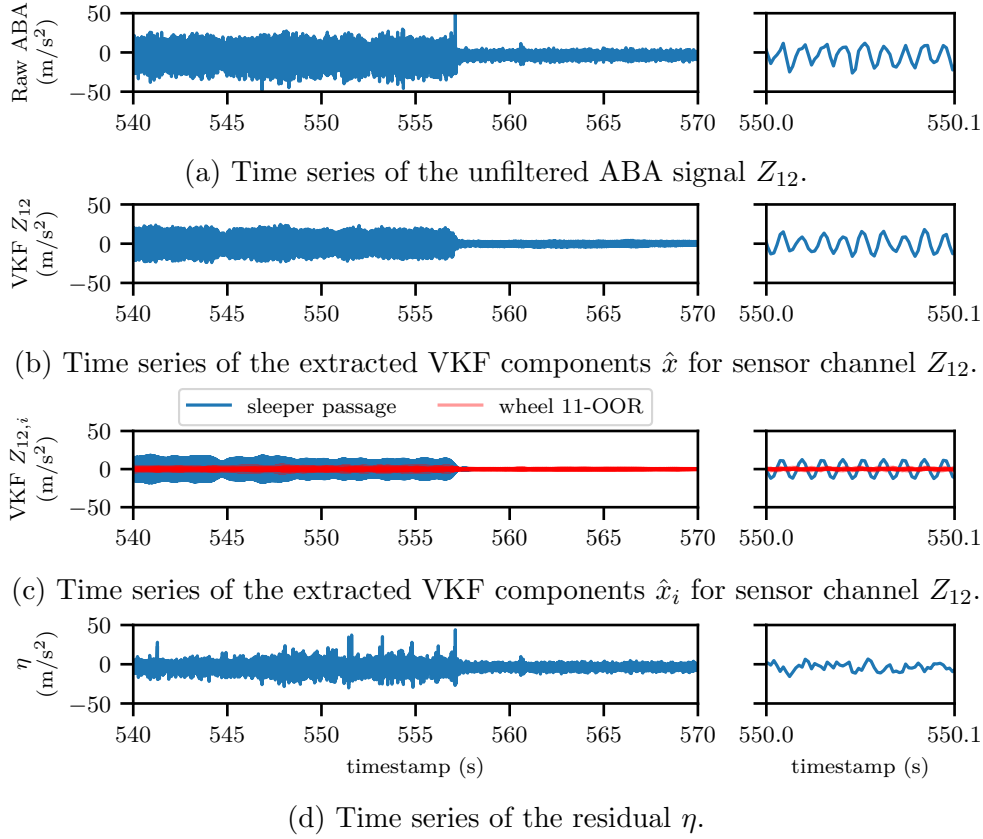
(a) Time series of the unfiltered ABA signal Z_{12}.

(b) Time series of the extracted VKF components $\hat{x}$ for sensor channel Z_{12}.

(c) Time series of the extracted VKF components $\hat{x}_i$ for sensor channel Z_{12}.

(d) Time series of the residual η.

Figure 3.7: Time series of the raw ABA signal Z_{12}, the filtered components $\hat{Z}_{12}$ and $\hat{Z}_{12,i}$ and the residuals η. The VKF filtered signal separates the excitation stemming from the wheel OOR and the sleepers spacing from the remaining noise.

formed on a 30 km section. The corresponding subsidence measurement, obtained from the EMW, is shown in Fig. 3.9a. One can observe that high sleeper passage acceleration amplitudes are located in areas of low subsidence. This is further supported by Fig. 3.10a, which illustrates that the subsidence closely matches the VKF Z_{12} amplitude. The three main clusters observed in Fig. 3.10a correspond to concrete sleepers with padding (B-91 sleepers with Sylomer SLS 1308G under sleeper pads), slab tracks (L77-B sleeper) and concrete sleepers with under-sleeper pads (B-91 sleepers without under sleeper pads).

Figure 3.10a demonstrates that when the subsidence approaches -0.5 mm, the acceleration amplitude of the in-phase sleeper passage mode Z_{12} increases significantly. Indeed, when the track stiffness is high (or equivalently when the subsidence is low), the energy of the crossing vehicle is mostly dissipated via the dynamics of the rail and the wheel, resulting in increased VKF–estimated Z_{12} amplitude. In fact, the track subsidence u is proportional to the logarithm of the acceleration amplitude Z_{12}, i.e.,

$$u = a \log\{Z_{12}\} + b \tag{3.34}$$

The Pearson correlation of the VKF–estimated Z_{12} amplitude and the subsidence is high ($r^2 = 0.74$). The slopes a and the intercepts b are estimated via conventional regression analysis and are summarized in Tab. 3.3, along with the Root Mean Square Error (RMSE), which remains below 0.35 mm for all channels.

The error between subsidence and the VKF–estimated Z_{12} amplitude has two main origins: On one hand, differences in track stiffness can be due to changes in stiffness over time. Indeed, the EMW subsidence measurement was performed in 2015, while the earliest ABA measure-
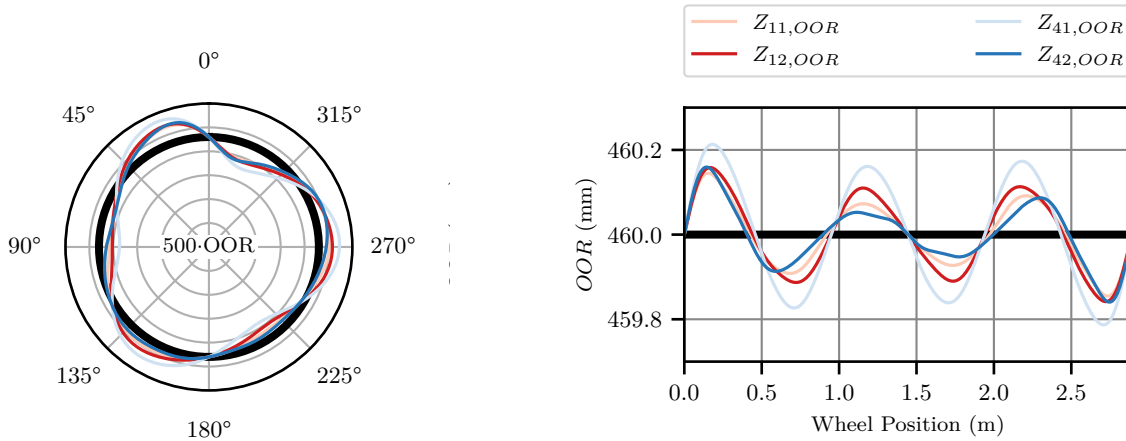
(a) Polar plot of the wheel OOR, illustrating the shape of the OOR–profile for the four wheels equipped with ABA. A four wheels monitored have a similar OOR.

(b) Wheel OOR for the position along the circumference. The radius of the wheels varies between 459.8 mm and 460.2 mm. The 3rd OOR order is prevalent, revealing that the wheels have a slightly tri-oval shape.

Figure 3.8: Wheel Out–Of–Roundness profile obtained from the sum of the first 11 OOR orders for wheel Z_{11}, Z_{12}, Z_{41} and Z_{42} as defined in Eq. 3.33.

Table 3.3: Parameters of the subsidence prediction model defined in Eq. 3.34. The RMSE achieved by the predictor is around 0.33 mm.

Regressor	Wheel 11	Wheel 12	Wheel 41	Wheel 42
b	-1.624	-1.617	-1.701	-1.760
a	0.304	0.285	0.356	0.343
RMSE	0.324	0.311	0.350	0.333
Samples	18605	18605	18605	18605

ments are available from 2019. On the other hand, the discrepancy can also be partly attributed to the different measurement processes: the VKF–estimated Z_{12} amplitude reflects the dynamic stiffness occurring at the regular line speed, whereas the subsidence measurements reflects the static stiffness that occurs at low speeds. Both the subsidence and the VKF–estimated Z_{12} amplitude are albeit not perfectly correlated, well in agreement with each other, illustrating the significant effect of the subsidence on the vehicle axle dynamics.

3.4.5 Longitudinal Level

Sources of excitation, such as the longitudinal level measurement obtained by the optical systems of the gDFZ, may also affect the sleeper passage response. The relation between stiffness, VKF–estimated sleeper passage acceleration and the longitudinal level must be assessed. The standard deviation over a length of 50 m is computed for the longitudinal level that was filtered to the wavelength range 1 to 70 m. This is a commonly used Track Quality Indicator (TQI) [316] indicating the roughness of the track. However this TQI does not contain the wavelength of 0.6 m, which was previously identified as critical for indicating effects due to variation in stiffness. The standard deviation of the longitudinal level is then compared to the subsidence measured by the EMW and to the VKF–estimated sleeper passage acceleration Z_{12} obtained from the gDFZ in ABA. Figure 3.10 illustrates the weak correlation between the VKF Z_{12} amplitude and the longitudinal level LL $Z_{\sigma_{200m}}$. The Pearson correlation coefficient between

(a) Subsidence along track position

(b) VKF Z_{12} sleeper passage component along track position

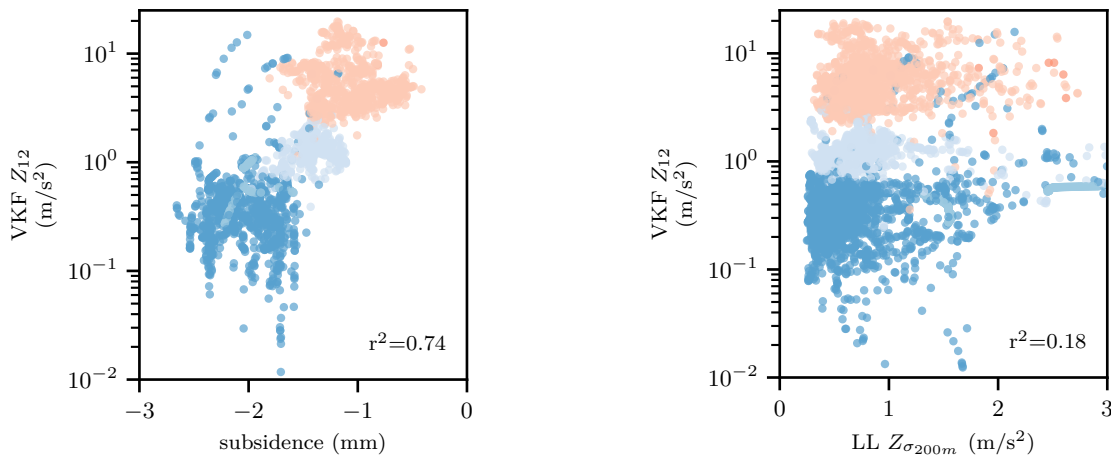(c) Standard deviation of longitudinal level 1-70 m along track position
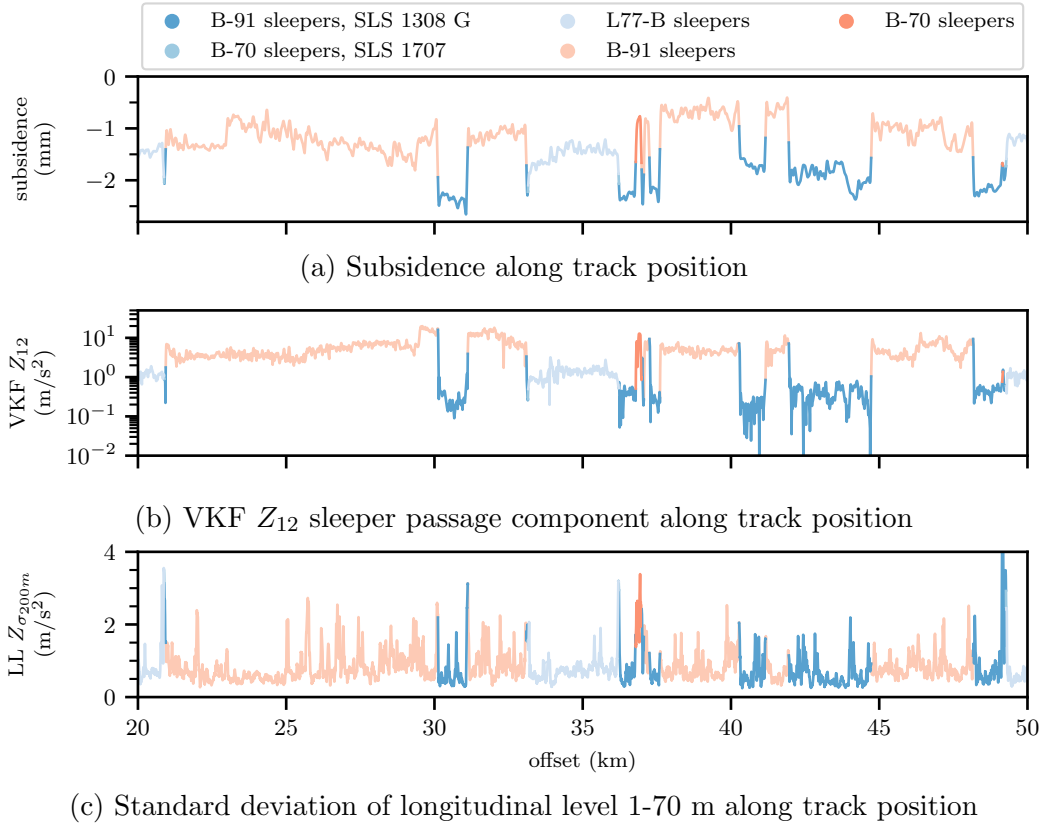
Figure 3.9: The subsidence and VKF Z_{12} acceleration amplitude are well correlated to each other. The increased VKF acceleration on stiff sections is primarily caused by the superstructure type. A low correlation is observed between the standard deviation of the longitudinal level 1-70 m (LL $Z_{\sigma_{200m}}$) and the VKF Z_{12} indicator.

these two parameters is low (r^2 = 0.18). Figure 3.9a-c shows the relation between changes in subsidence or the VKF–estimated sleeper passage acceleration Z_{12} and the standard deviation of the longitudinal level. The relative changes in subsidence in Fig.3.9a are also observed by the VKF in Fig.3.9b . The effect of changes in stiffness results in the appearance of geometric irregularities corresponding to the peaks in Fig.3.9a at kilometer offset 29.4, 30.1 or 31.1, but stiffness changes is not the only origin of geometric errors.The longitudinal level has overall a low correlation to the sections with low subsidence and high VKF–estimated sleeper passage acceleration Z_{12} signifying that this indicator is not highly sensitive to the geometric irregularities as shown in Fig.3.10b. Nevertheless, the sudden changes in stiffness result in measurable geometric irregularities that require regular maintenance.

3.4.6 Relation between acceleration, forces and damage accumulation

The VKF order extraction process, carried out in Sec. 3.4.2, is now applied to the force measurements from the TWS, in order to separate the sleeper passage force amplitude. The VKF–estimated sleeper passage force Q_{12} and acceleration Z_{12} amplitude, extracted for a 30 km track section, are shown in Fig. 3.11a and Fig. 3.11b respectively. Figure 3.11c illustrates the direct relation between the amplitude of rail–wheel force and ABA amplitudes. The resulting VKF–estimated forces Q_{12} are proportional to the VKF–estimated Z_{12} amplitudes and, expectedly, high forces and accelerations are observed for stiffer superstructures. In contrast, at flexible superstructures, the VKF Z_{12} amplitude is lower, since the load is uniformly distributed across multiple sleepers.

71

(a) Scatter of VKF–estimated Z_{12} amplitude and subsidence.

(b) Scatter of VKF Z_{12} amplitude and standard deviation of longitudinal level

Figure 3.10: The VKF–estimated sleeper passage acceleration amplitude Z_{12} is compared to the subsidence and the standard deviation of the longitudinal level 1-70 m (LL $Z_{\sigma_{200m}}$). A strong Pearson correlation ($r^2 = 0.74$) is observed between the subsidence and VKF Z_{12} indicator while a weak Pearson correlation ($r^2 = 0.18$) is observed between the longitudinal level indicator and the VKF Z_{12} indicator.

(a) Forces along track position

(b) Accelerations along track position

(c) Scatter of forces and acceleration
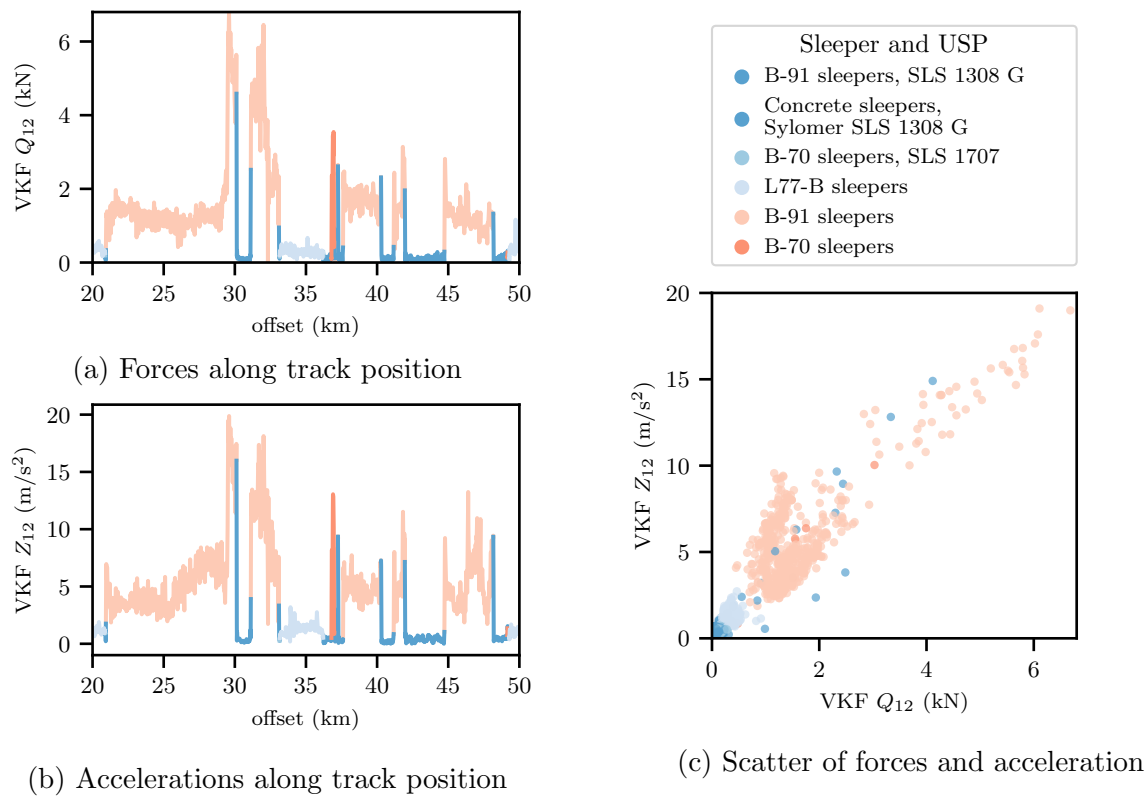
Figure 3.11: The VKF–estimated sleeper passage force Q_{12} is strongly correlated to the acceleration amplitude Z_{12}. Both quantities exhibit similar responses for varying superstructure types.

The direct relation between forces and accelerations is useful in quantifying track deterioration, since the degradation of the alignment of ballasted tracks is a time-dependent process that originates from the dynamic excitation of vehicles. Repeated loading cycles, during which
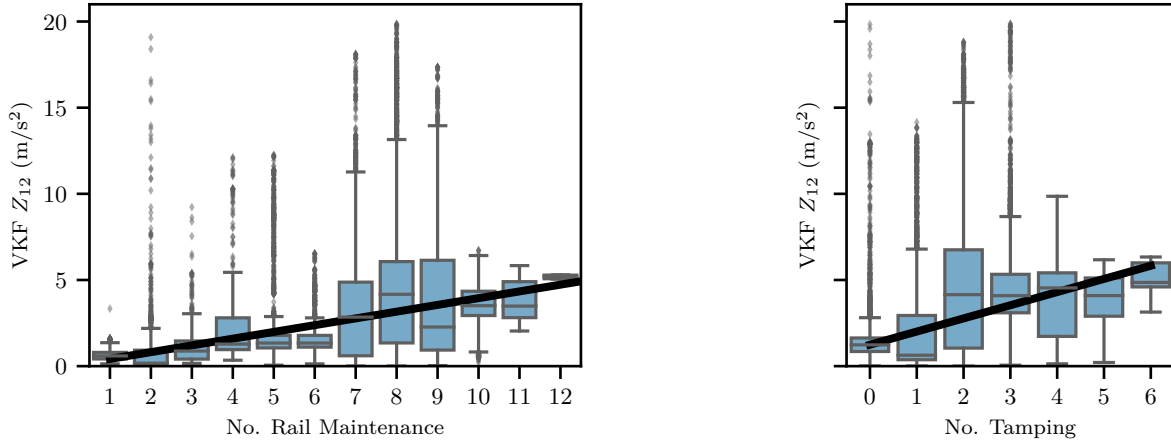
(a) The VKF sleeper passage acceleration amplitude Z_{12} corre- lates to the number of rail maintenance actions.

(b) VKF Z_{12} correlates to the number of tamping actions.

Figure 3.12: Boxplots and regression line with amplitude of the VKF acceleration Z_{12} in function of the number of rail maintenance actions and tamping actions between 2006 and 2020 for a sampling interval of 25 cm on the 100 km track sections. The VKF-derived sleeper passage amplitude correlates to the number of rail maintenance actions and tamping actions.

energy is dissipated via mechanical processes (damping of pads, friction of ballast grains,etc) cause material degradation. The consequence of the degradation of the ballast and the substructure is a reduced bearing capacity, leading to geometric deviations and settlement of the track structure [267, 266]. Aside from the number of loading cycles, the most important parameter of mechanistic degradation models is the dynamic excitation forces on the ballast [243]. The VKF–estimated dynamic excitation forces are indications of increased dynamic response of the vehicle axle under certain superstructure configurations. Hence, the damage accumulation to the infrastructure components is closely related to the load; larger load variations induce a higher rate of damage accumulation.

In verifying the relationship between the extracted indicators and the prompting for suitable maintenance actions, the track maintenance actions recorded in the Database of Fixed Assets (DfA), between 2006 and 2020, are extracted for the analyzed track segments for a total length of more than 100 km. To estimate the number of maintenance actions at any location along the track, we use the type of maintenance, start and end position on the track and the maintenance date extracted obtained from the DfA. The number of track maintenance actions that are related to the ballast and the rail are subsequently derived along the track at spatial intervals of 25 cm, by summing up the number of interventions that were performed at any position along the track section over a period of 14 years (2007–2020). The distribution of the VKF-derived sleeper passage acceleration amplitude Z_{12} given a number of maintenance actions can be visualized with boxplots. Figure 3.12a and Fig. 3.12b show that the number of maintenance actions to the ballast (tamping) and the rail (grinding) correlates with the VKF-derived sleeper passage amplitude. Thus, we demonstrate that decreased VKF sleeper passage acceleration and force amplitude due to lower superstructure stiffness results in a more durable track [286].

The large variance in Fig. 3.12 of the VKF-derived sleeper passage acceleration amplitude is explained by two main factors. Firstly, maintenance is usually performed on a large scale and therefore these actions overlap between degraded and healthy sections. Secondly, preventive maintenance actions are not considered and may also influence the evolution of the overall track condition. Figure 3.12 demonstrates that the use of under sleeper pads (effectively resulting in track lower stiffness) has a beneficial effect by significantly lowering rail–wheel force amplitudes. This is confirmed by the work of Dahlberg [64], who suggests that the wheel/rail contact forces

are considerably reduced by decreasing the local stiffness variation via use of grouting or under-sleeper pads.

3.5 Conclusion

The efficient and continuous monitoring of vehicle-track dynamics is essential for maximizing the lifespan of the railway infrastructure, while minimizing the required maintenance actions. The Vold-Kalman Filter (VKF), a time-domain filtering technique, is here suggested as a fitting approach to extract condition indicators that characterize the condition of railway tracks and vehicle wheels. The proposed multi-order extraction scheme separates the deterministic (periodic) components from non-deterministic contributions to the measured axle response.

The VKF–estimated amplitude components corresponding to the wheels are used to reconstruct the wheel profile, which in turn provides information on the condition of the wheel and the potential existence of flaws, such as wheel OOR, thus serving for condition assessment of wheel assets. The VKF–estimated component related to the sleeper passage, on the other hand, contains information that is indicative of changes in the track stiffness. It is shown that the VKF sleeper passage acceleration amplitude is proportional to the rail–wheel forces. The VKF-derived indicator, thus, delivers an indirect, yet reliable, means to assess the underlying track stiffness, which serves as a proxy of the track condition and implies that OBM-based ABA measurements could potentially supplement such subsidence recordings, delivering temporally denser measurements. The VKF-based indicator can be exploited to improve the planning of optimal maintenance measures, since, railway track sections with higher stiffness were demonstrated in this and prior works to lead to increased forces at the rail–wheel interface, resulting in damage accumulation on the rails and ballast. Future work will investigate the effects of stiffness changes stemming from fouled ballast and hanging sleepers by performing additional stiffness measurements with the EMW. Moreover the coupling of the VKF-derived stiffness indicator with existing substructure condition indicators, such as fractal values [149] will be explored to further support he characterization of the specific source of increased stiffness.

Chapter 4

Data-Driven Railway Vehicle Parameter Tuning using Markov-Chain Monte Carlo Bayesian updating

Paper Details

Author and Co-Author Contributions

Key Findings

- Literature review focusing on the development of models and approaches used to capture coupled vehicle-track system dynamic with an emphasis on their role in facilitating the interpretation of dynamic measurements.

- This chapter employs an MCMC Bayesian updating method to optimize the parameters of an ICN RABDe500 wagon model system based on actual on-board measurement (OBM) data from an in-service train. This approach results in a well-calibrated model capable of predicting wheel rail forces and accelerations.

- The proposed Bayesian model updating approach method proves to be an efficient and suitable tool for optimizing wagon model parameters by minimizing the estimation error between simulation model contact forces and OBM-measured forces. It particularly enhances the prediction of lateral forces by optimizing the friction coefficient between rails and wheels.

- Several avenues for further research and improvement of the MCMC approach are proposed, including defining an automatic termination criterion, exploring alternative evaluation criteria, considering additional monitored parameters, and incorporating the probabilistic nature of parameter estimates. Additionally, the proposed approach is extensible to other vehicle types and track sections, indicating its broader applicability.

General comments and Link to the next chapter

In pursuit of the fourth objective outlined in the thesis (see Section 1.4), this chapter delves into the critical task of understanding the dynamic interplay between railway vehicles and tracks. This understanding is essential for accurately predicting the condition of both the vehicles and the tracks, ultimately ensuring the safety and longevity of railway infrastructure. Physics-based models are commonly used to simulate the dynamic response of railway vehicles, however such models contain numerous parameters requiring validation and potential tuning.

MCMC is a suitable tool for vehicle modelling updating, which can increase simulation accuracy and decrease model building workload. We employ a Bayesian model updating approach with Markov-Chain Monte Carlo (MCMC) to estimate optimal vehicle parameters and their uncertainties, minimizing the mean squared error between simulated and measured contact forces. The process described in the following chapter, results in an improved 'vehicle twin,' providing valuable insights into vehicle-track dynamics.

Abstract

Understanding the dynamics of the interaction between railway vehicles and tracks is essential for forecasting vehicle and track conditions and performing maintenance actions to preserve the safety of railway infrastructure. In this work, physics-based models are deployed to predict the dynamic response of railway vehicles to track alignment and irregularities. Such models comprise a large number of parameters that need to be validated and possibly tuned, a task often accomplished on the basis of expert knowledge or trial-and-error without necessarily reflecting the condition of the operating system as-is. Therefore, model uncertainties persist due to assumptions on the actual system behavior, as well as due to varying operational parameters (e.g., wheel profile, rail profile, rail moisture). To improve modelling accuracy, this study adopts a multi-body vehicle model, realized in SIMPACK software, and performs parameter tuning based on on-board measurement data (accelerations and forces) from an instrumented tilting train, which regularly traverses the Swiss Federal Railways (SBB) network. The pertinent vehicle model is optimized in terms of the interaction forces at the wheel-rail contact using a Bayesian model updating approach, relying on Markov-Chain Monte Carlo (MCMC). Specifically, the MCMC method is applied to estimate both optimal vehicle parameters and their uncertainty by minimizing the mean squared error between simulated and measured contact forces, leading to an improved vehicle twin.

4.1 Introduction

Ensuring the quality and safety of rail transport requires monitoring and evaluating the condition of rail infrastructure. Specifically, for the track system, continuous monitoring can identify faults at an early stage [200, 149, 107, 106], preventing severe damage to track and vehicle components, reducing rolling noise, minimizing passengers' discomfort due to excessive vibrations, and, ultimately, enhancing the overall safety of rail transport [23, 317]. This degradation of tracks mainly owes to vehicle load cycles and vehicle-track interaction [93, 267]. For this reason, the modelling of the dynamic interaction between railway vehicles and tracks has been a topic of significant research over the past decades [301, 141], with some studies also proposing operational criteria to avoid adverse effects due to vehicle-track interaction. For example, Knothe and Stichel [141] proposed a set of short- and long-term criteria regarding the dynamics a vehicle must meet for operational approval [1]. Short-term criteria are those related to the vehicle's operation (e.g., excitation of the vehicle to the rails). In contrast, long-term criteria relate to consequences on the infrastructure after hundreds or thousands of cycles (e.g., damage to the infrastructure due to fatigue). To define these criteria, diverse factors are considered, mainly associated with three essential aspects: safety, comfort, and cost.

However, in order to develop such safety operational criteria, accurate models that reflect the actual dynamic behavior of coupled railway vehicle-track systems are required. Simplified or reduced-order models are often adopted for simplicity or to reduce computational complexity when studying specific components or processes, such as the vehicle axle response [273]. For example, Dertimanis et al. [69] used a two-dimensional (2D) half-car model approximating the vehicle dynamics to estimate the track roughness and vehicle-track forces. Although computationally more efficient, reduced-order models may not fully represent the actual dynamics of the vehicle. To this end, full-order vehicle models (Fig. 4.1) enable more realistic vehicle-track dynamic simulations since they can capture more complex dynamic phenomena than the simplified models [301, 141]. However, such models are composed of a large number of bodies and connections whose properties are inherently uncertain and may sometimes vary with environmental conditions [240]. The calibration of full-order models nowadays often relies on trial-and-error approaches or requires specific dynamic tests performed in the lab [240].
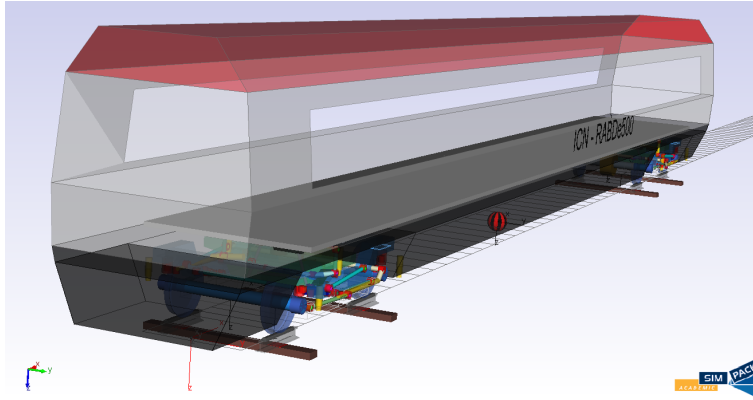
Figure 4.1: 3D multi-body simulation model of the ICN RABDe500 wagon in SIMPACK software [143].

To avoid such cumbersome approaches, more comprehensive parameter optimization methods relying on genetic algorithms are often used to tune the vehicle parameters [120].

Towards this direction, in this study, we use a Markov-Chain Monte Carlo (MCMC) algorithm to optimize the parameters of the vehicle model. MCMC-based approaches are commonly applied to estimate the parameters of a system which are inherently uncertain. For example, Arcieri et al. [12] modelled the problem of optimal railway maintenance scheduling policies as a partially observable Markov decision process and inferred the underlying transition dynamics and observation-generating process via MCMC sampling. MCMC sampling has also been used to calibrate finite element models for structural dynamics problems [221]. Thus. in the same premise, this work employs an MCMC method to optimize vehicle parameters by minimizing the error between simulated and measured wheel-rail contact forces via iterations. Simulations are performed with the aid of the ICN ("Intercity Neigezug") wagon (RABDe 500) of SBB [143] in the multi-body simulation software SIMPACK [256]. The on-board monitoring (OBM) system installed on the same train measures contact forces and accelerations, which are then used for calibrating the model.

4.2 Methodology

This section presents a methodology to calibrate the ICN RABDe500 model. Fig. 4.2 graphically illustrates the proposed calibration method step-by-step. The goal is to estimate optimal vehicle parameters that minimise the mean squared error (MSE) between the wheel-rail forces as predicted by the simulations in SIMPACK software and those measured on-board. Minimising the MSE between estimated and measured contact forces results in a well-calibrated wagon model which can be used in subsequent simulations.

4.2.1 Simpack Simulation Model

Fig. 4.1 demonstrates the three-dimensional (3D) ICN RABDe500 wagon model [143], implemented in SIMPACK software, which comprises a first-class carriage modelled as a rigid body. The track on which the train runs is modelled as a continuously supported two-layer model (the rail and sleeper layers) composed of two rails elastically connected to a sleeper. The track gauge is set to the Swiss standard of 1435 mm. Both rails and sleepers are modelled as rigid bodies. The rail-sleeper and sleeper-ground contacts are realised via force elements of linear stiffness and linear damping. The mm-accurate track geometric layout and excitation, used as
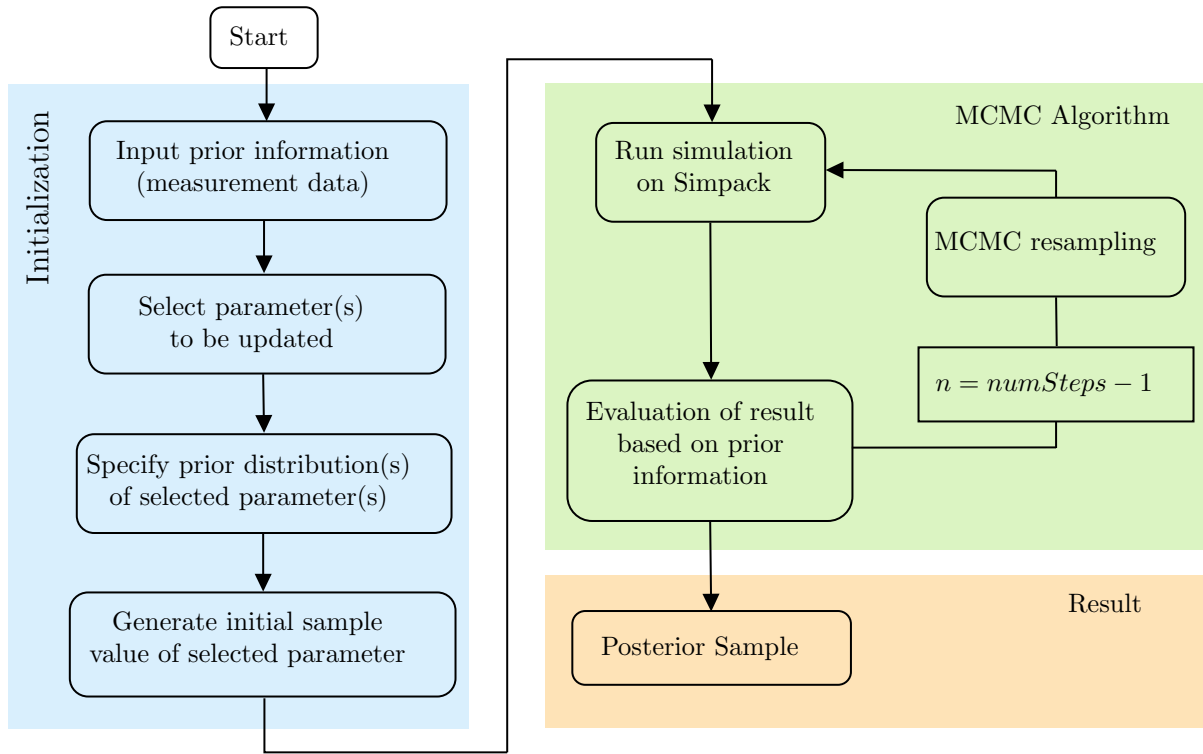
Figure 4.2: Proposed methodology for vehicle parameter optimization.

an input for the simulation, are obtained from measurements of a diagnostic vehicle [107].

The interaction surface between the wheels of the wagon and the rails is modelled as discrete elastic, i.e., it considers the actual wheel-rail contact area for the estimation of the normal and tangential forces. The normal (vertical) force is composed of an elastic normal force and a damping normal force evaluated at each slice of the contact patch. For the tangential and lateral interaction, the FASTSIM method of J.J. Kalker is used [256]. The Kalker weighting factor is set to 1, and the friction coefficient is initially set to 0.35. The input variables for initiating the simulation are a geometry profile (track layout), an excitation profile (geometric irregularities), a velocity profile and a tilting profile and, optionally, a measured wheel and/or a measured rail surface profile. Fig. 4.3 illustrates the geometry and excitation profiles for the selected section. The velocity profile is taken directly from a measurement run, while the tilting profile is calculated from several measurement properties of the corresponding measurement run. For each of the measurement runs, both wheel-rail force and accelerations are measured.

The initial simulations, using the default parameters, have a reasonably good performance in approximating the measured wheel-rail forces. However, their values can further be improved by incorporating changes that may be attributed, for example, to environmental parameters, which can vary in day-to-day operations. For instance, the vertical forces are influenced by the train's body mass and centre of gravity, which changes under different passenger loads. On the other hand, lateral forces are strongly affected by the wheel-rail friction coefficient, which is unknown and dependent on environmental conditions [141]. Therefore, in the next section, an MCMC methodology for parameter optimization is proposed which provides probabilistic estimates of these parameters in an automated manner.

4.2.2 MCMC-supported Parameter Optimization

The parameters of the model presented in Fig. 4.2.1 are optimized using the MCMC algorithm. A thorough treatment of the MCMC algorithm can be found in [174], while this Section will also
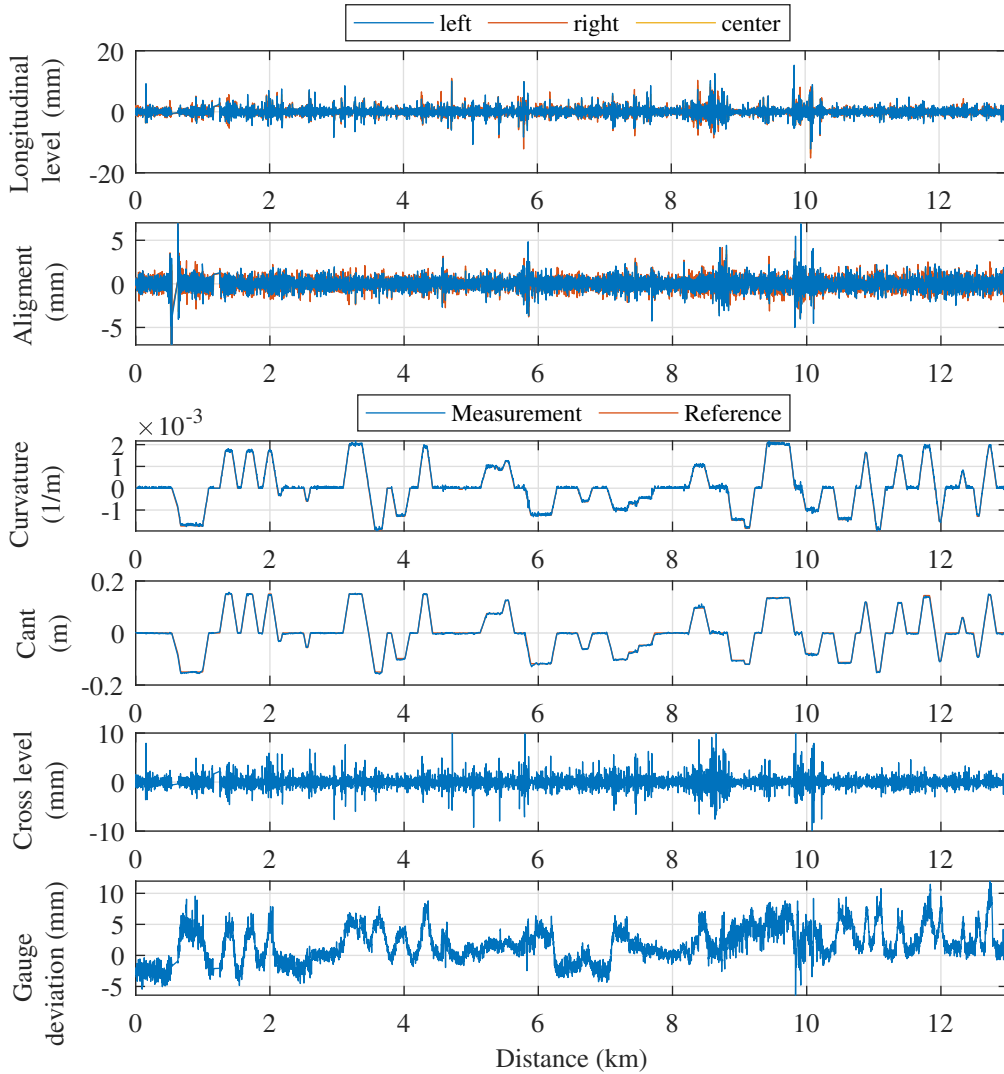
Figure 4.3: Geometry profile (curvature and cant) and excitation profile (longitudinal level, alignment, cross level, and gauge deviations) measured by the diagnostic vehicle of SBB [106].

provide a brief overview. Fig. 4.2 also gives a general insight into the procedure for optimising the parameters.

First, in the initialization phase, the parameters to be identified are chosen, and suitable prior distributions are assumed for their initial values. The MCMC algorithm then updates these prior distributions to find the posterior distributions on these parameters, or, in other words, the likely distribution of these parameters given the measured forces and assumed model. This process works through sampling; samples are drawn from the prior distributions, and simulations are run using these samples as the model parameters. The sampled parameters are then evaluated using a defined likelihood function which determines how well the simulation outputs correspond to the measurements. Given this likelihood value, the sampled values are accepted or rejected by the Metropolis-Hastings algorithm [185], thus updating the distribution. This process is repeated until convergence is achieved, and the final output consists of the samples from the posterior distribution. The MCMC updating process is implemented in Matlab software using the UQLab toolbox [294]. As previously mentioned, the MCMC algorithm requires a likelihood function that evaluates a certain parameter vector to be defined. In this case, the likelihood is chosen as the MSE between the estimated and measured forces in both lateral (Y21 and Y22) and vertical (Q21 and Q22) directions for a 1 km long section of the 13 km long track that the wagon runs on.

Table 4.1: Overall posterior parameter optimization results and comparison with individual MCMC given the prior estimates.

Parameters	Prior estimate, $\mathcal{N}(\mu, \sigma^2)$, $\mathcal{U}(a,b)$	Individual result	Overall result	Difference $(|\mu_{\mathbf{overall}} - \mu_{\mathbf{indiv}}|)$
Car body mass	μ_{prior}=35487 kg, σ_{prior}=3000 kg	μ_{indiv}=37027 kg, σ_{indiv}=325.7 kg	μ_{overall}=36891 kg, $\sigma_{\mathrm{overall}}$=92.3 kg	136 kg
Centre of gravity sideways	μ_{prior}=-0.0267 m, σ_{prior}= 0.5 m	μ_{indiv}=-0.0611 m, σ_{indiv}=0.01 m	μ_{overall}=-0.0405 m, $\sigma_{\mathrm{overall}}$=0.01 m	0.02 m
Friction coefficient	a=0, b=0.7	μ_{indiv}=0.1993, σ_{indiv}=0.02	μ_{overall}=0.2137, $\sigma_{\mathrm{overall}}$=0.02	0.01
Flange friction coefficient	a=0.6, b=0.18	μ_{indiv}=0.1393, σ=0.02	μ_{overall}=0.1687, $\sigma_{\mathrm{overall}}$=0.005	0.03
Navigator factor	μ_{prior}=1, σ_{prior}=0.25	μ_{indiv}=0.8403, σ_{indiv}=0.04	μ_{overall}=0.8399, $\sigma_{\mathrm{overall}}$=0.03	0.00
Primary spring *cx* factor	μ_{prior}=1, σ_{prior}=0.25	μ_{indiv}=6.2835, σ_{indiv}=1.07	μ_{overall}=5.1403, $\sigma_{\mathrm{overall}}$=0.11	1.14

Tab. 4.1 demonstrates the results of the optimization procedure. The parameters update assumes a single measurement run with a specific speed and tilting profile. The first column of Tab. 4.1 presents the optimized parameters using the MCMC-based scheme. Tab. 4.1 also shows the posteriors resulting from the MCMC optimization, which are discussed in Section 4.3. The selection of the initial values of the vehicle parameters relies on prior knowledge of the vehicle dynamics. In Tab. 4.1, cx denotes the factor which is multiplied with the translational stiffness constant of the primary suspension in the x-direction (longitudinal direction of the rail). The navigator factor indicates how much the axes can be adjusted radially. A factor of one means that both axes (leading and trailing) are perfectly perpendicular to the rails. Note that a perfect coupling is usually not observed in practice due to friction and parasitic stiffness. The body mass and center of gravity can vary under passenger load and thus are modelled with Gaussian priors $\mathcal{N}(\mu, \sigma^2)$. The wheel-rail friction coefficient is modelled with a uniform prior $\mathcal{U}(a,b)$. In the next section, the results of the MCMC-based optimization are presented.

4.3   Results & Discussion

The force data measured by the OBM system installed on the in-service ICN vehicle are compared with the simulations in SIMPACK software for a case study that considers a 13 km track section. First, the simulation is carried out with the initial parameter values, and the predicted wheel-rail forces are compared with the measured values recorded by the recording system of the ICN vehicle. Then, MCMC is used to optimize the initial parameters, as those are likely to strongly influence the response of the vehicle. It is worth noting that the parameter optimization uses a shorter track section (of 1 km in length), while the performance of the simulation is assessed across the entire length of 13 km. This 1 km track segment, highlighted in Fig. 4.6, was used for the MCMC identification as it includes curves in both directions. The good results
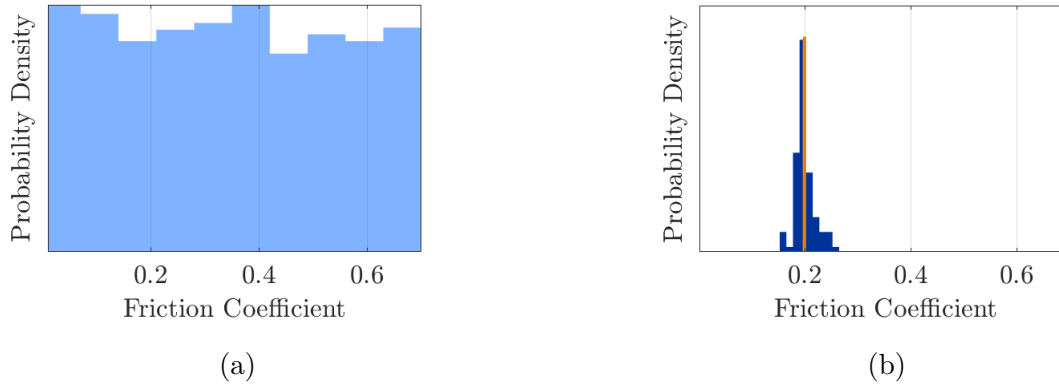
Figure 4.4: Prior (a) and posterior (b) distribution of the friction coefficient according to the MCMC algorithm.
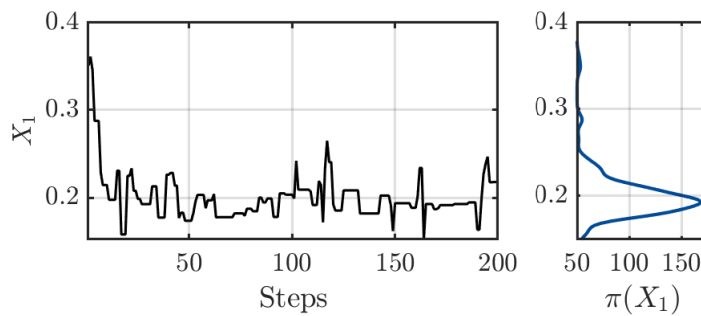

Figure 4.5: Illustration of the convergence of a single chain for the friction coefficient.

along the entire 13 km track section indicate that the parameters identified by the MCMC are generalizable, performing well also on track sections not included in the training data.

This Section further elaborates on the MCMC identification process for a single parameter, the friction coefficient. Fig. 4.4a shows a sample drawn from the assumed prior distribution for the friction coefficient, which assumes a uniform distribution between 0 and 0.7. Fig. 4.4b shows the identified posterior distribution of the friction coefficient following the MCMC process. The distribution has significantly narrowed, and both shape and mean value have considerably changed. Fig. 4.5 shows the process of parameter identification through the convergence of one of the Markov chains used in the MCMC algorithm, where only the first 200 steps are shown for visualization purposes. The chain quickly converges to the mean value of the identified posterior distribution. The kernel density estimation of the posterior distribution is also shown, which is estimated based on the samples from this Markov chain after the first 50 % of values have been discarded as the *burn-in period* [174]. All identified parameters converge, and their posterior estimates demonstrate low variance, as shown in Tab. 4.1.

Fig. 4.6 shows the time series for the vertical (Q21) and lateral (Y21) forces of wheel 21 (right wheel of the second axle in rolling direction) along the track. The lateral forces (Fig. 4.6b) resulting from the simulation with the initial parameters do not perfectly match the measured accelerations, especially in the curved track sections, which are identified as the regions of high amplitude lateral forces. However, after optimization of the parameters via MCMC, we observe that the simulated lateral forces show better agreement with the measured lateral forces. On the other hand, according to Fig. 4.6a, the vertical forces after parameter optimization improve less than the lateral ones.

Fig. 4.7 further supports this finding showing that the MSE shows significant improvement only in the case of lateral forces (Y21 and Y22). In contrast, in the case of vertical forces (Q21 and Q22) the MSE is practically the same before and after optimisation. Tab. 4.1 summarizes
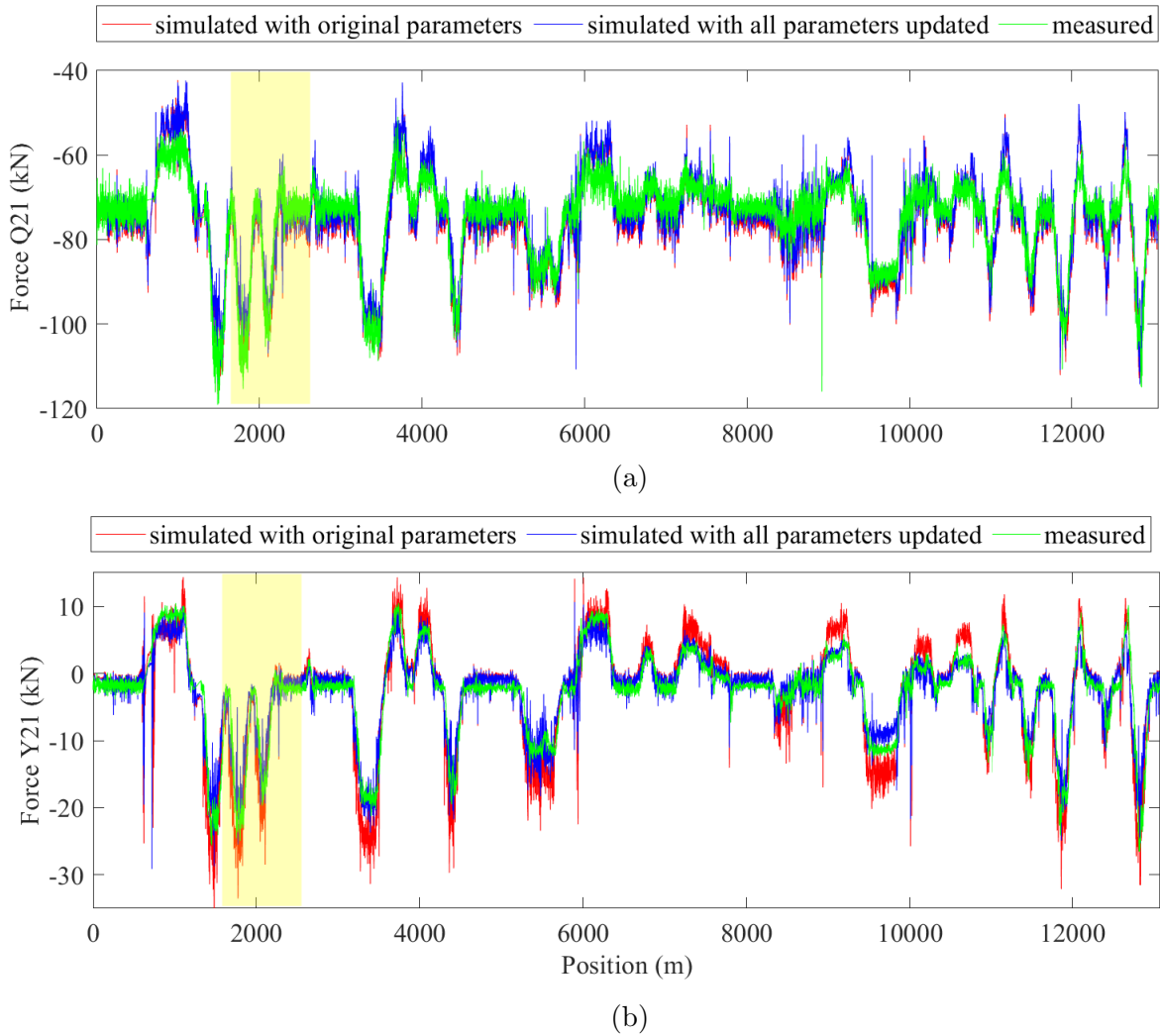
Figure 4.6: Force-position histories of the (a) vertical Q21 and (b) lateral Y21 contact forces of the right wheel of the second wheelset of the wagon, for simulations with the initial (red) and updated (blue) vehicle parameters, along with the measured contact force values (green). The highlighted yellow section indicates the section used for the MCMC algorithm.

the mean and standard deviation (or a, b of the uniform distributions) of the prior estimates. The statistics on the posteriors of the individual and overall optimizations deviate only slightly. The larger standard deviation of the lateral forces compared with vertical forces stems from differences in the friction coefficient, which is strongly influenced by environmental conditions (weather, humidity, etc.), among other parameters. Thus, lateral forces are more sensitive to variations in the rail surface moisture [141] and other rail surface contaminants, such as oil and dirt [227], that can affect the contact between rails and wheels. However, specific conclusions about the interdependence of weather and lateral forces are difficult to draw; thus, further analysis is required.

The MCMC algorithm is programmed in MATLAB R2022a software using the UQLab toolbox[294]. The simulations invoked by the algorithm are performed using SIMPACK Version 2021x.6 on a device with "Intel(R) Xeon(R) CPU E3-1246 v3 @ 3.50GHz" processor. This results in simulation times listed in Tab. 4.2. MCMC is computationally expensive; however, this offline process does not require engineering judgment and, thus, direct supervision. Moreover, parallelization, which was not implemented herein, can further reduce computation time.
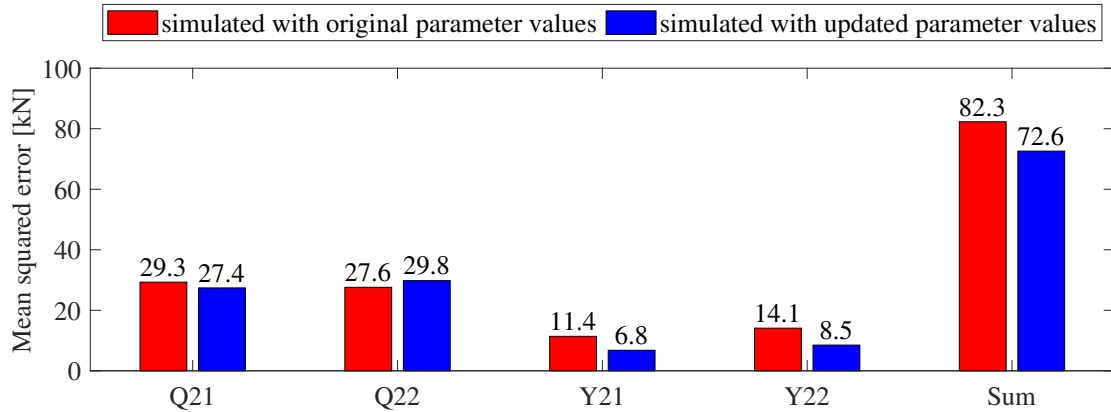
Figure 4.7: MSE between simulated and measured contact forces in the vertical and lateral directions for simulations with the initial (red) and updated (blue) vehicle parameters.

Table 4.2: Simulation times

Simulation specification	duration [hh:mm:ss]
Single run $\approx$ 13 km	$\approx$ 00:35:00
Single run $\approx$ 1 km	$\approx$ 00:02:30
MCMC simulation, 6 parameters updated, 400 steps, $\approx$ 1 km	$\approx$ 15:30:00

4.4 Conclusion

To preserve and further develop railway infrastructure, it is crucial to understand vehicle-track dynamics and develop simulation models that can capture the actual dynamics of coupled vehicle-track systems. One way to achieve that is by fusing physical models with data-driven approaches for model updating. Therefore, in this study, we employ an MCMC Bayesian updating method to tune/optimize the parameters of an ICN RABDe500 wagon model system based on actual OBM measurements from an ICN RABDe500 in-service train. This leads to a well-calibrated model that can be used for predicting loads and damage, and improving maintenance periods.

To realize this scheme, we first perform simulations based on the initial wagon model values. These initial simulations already show a good agreement with the measured data in terms of contact forces. However, there is still space for improvement by updating the simulation parameters of the wagon. To this end, MCMC proves to be a very efficient and suitable tool to accomplish this task. MCMC optimizes the wagon model parameters by minimizing the MSE between the contact forces of the simulation model and the forces measured by the OBM system of the ICN train. The friction coefficient between the rails and the wheels appears to be a critical optimization parameter, the tuning of which significantly improves the prediction of lateral forces. On the other hand, the vehicle's body mass and centre of gravity mainly influence the vertical contact forces, but the effect of those parameters is not as strong. Opposite to current approaches that calculate optimal values using a trial-and-error approach, MCMC determines these values automatically. However, this process requires a prior selection of a set of parameters and the definition of their prior distribution. A key feature of the MCMC approach is that the resulting output offers a posterior probability distribution of the parameter, enabling uncertainty quantification of the predicted results. Eventually, the calibrated model can be used for further simulations.

MCMC is a suitable tool for vehicle modelling updating, which can increase simulation accuracy and decrease model building workload. However, further investigation is required to enhance its performance and ease of application. An extension could define an automatic termination criterion instead of implementing a fixed number of iterations at the beginning of the algorithm, which could decrease the computation time. Such a termination criterion would be based on the convergence of the likelihood evaluation. Furthermore, a different evaluation criterion, other than the MSE between contact forces, could be chosen in the log-likelihood function, such as frequency differences estimation, to improve accuracy and computational efficiency. In addition, since more than 50 vehicle parameters are monitored, a similar approach combining acceleration measurements could be implemented. Finally, incorporating the probabilistic nature of parameter estimates provided by MCMC could help capture uncertainty in the simulation predictions. Lastly, the proposed approach is extensible to other vehicle types and track sections.

Chapter 5

Classification of Rail Irregularities from Axle Box Accelerations using Random Forests and Convolutional Neural Networks

Paper Details

* First authors.

Author and Co-Author Contributions

The author of this thesis prepared the data, performed the initial conception, validation and description of the classification schemes presented here. L. Ancu conceived the CNN-based framework. M. Grossmann and D. Ferrari supported the development and testing of the RF-based and CNN-based frameworks respectively. Prof. E. N. Chatzi and Dr. V. Dertimanis helped to conceive the framework, provided supervision and guidance.

Key Findings

- We demonstrate the feasibility of detecting and classifying rail components (welds, insulated joints, squats) in a Big-Data framework using ABA only.

- Two approaches were compared: Random Forest trained on features from the ABA and CNN trained on the time-frequency representation of the ABA using the STFT.

- Insulated joints and defect-free rails can be distinguished quite well. Welds, surface defects and squats, however, feature more uncertainty in classification which can partly be traced back to the ground truth labels.

General comments and Link to the next chapter

This study tackles the fourth objective of the thesis (see Section 1.4) by extracting automatically labeled accelerations obtained from the image recognition on the rail-head images of $VCUBE$ in a data-driven classification framework which can accurately predict the rail components using accelerations only.

This chapter proves the potential of machine learning tools applied on ABA measurements for distinguishing between critical rail components and faults. This chapter further demonstrates the limitations in the precise identification of surface defects, since these typically reflect damage of various intensities, which in itself requires a more targeted scheme for classification of damage intensity. Likewise, it is noticeable that welds can feature various levels of damage. Considering that there are far more welds than insulation joints or surface defects in the network, the next chapter investigates the potential for refining detection of defective welds using ABAs by fusing information from ABA, rail-head images ($VCUBE$), and expert-informed labeling.

Abstract

The continuously increasing demand for mobility results in increased loading of the Swiss railway network, which is further associated with higher wear and deterioration of the rail infrastructure. Safety relevant surface defects on railway tracks, such as squats, have acted as an important driver of rail replacements in Europe. The early detection of such defects can support the planning of appropriate maintenance measures, such as grinding, which prolong the remaining life of the rails. On-board monitoring has redefined the paradigm of railway infrastructure monitoring, via use of in-service vehicles as mobile sensing systems. Such vehicles are equipped with sensors, e.g. axle box accelerometers in order to continuously collect information on the track and vehicle condition, and support the monitoring of railway assets and infrastructure. Acceleration–based monitoring has been shown to bear tremendous potential for offering temporally and spatially dense diagnostics of railway infrastructure. While the potential of such a monitoring scheme has been proven, the generalization has been limited due to the small sample sizes in existing studies.

We propose a methodology to recognize and classify between the most common rail irregularities, namely surface defects, insulated joints and welds, by exclusively relying on availability of on board acceleration measurements. We combine labeled information, stemming from railhead image based detection, with acceleration measurements. Two classification approaches are compared in this work. The first methodology exploits Convolutional Neural Networks (CNNs) that are applied to the Fourier coefficients, which are computed from acceleration time-series data. The second methodology relies on a more classical machine learning approach, applied on features that are extracted from the acceleration time series, which are then classified using Random Forests. Finally, the uncertainty of the acceleration metrics and of the ground-truth labels is analyzed, motivating the application of acceleration based detection for improvement of rail condition monitoring. The resulting classifiers can be deployed on regular passenger trains for enabling the continuous and automated monitoring of the rail condition.

5.1 Introduction

Frequent monitoring to assess the condition of the rails is indispensable for applying predictive maintenance. Most railway tracks consist of continuously welded rails that are discretely supported onto sleepers. In some cases, instead of *welds*, the rails are connected using *insulated joints*, which consist of two electrically insulated rails that are bolted together with steel plates. The appearance of a fault usually initiates with small material imperfections that under accumulated load grow to more severe faults. *Surface defects* are a broad category of defects that have many origins. They can be caused by ballast on the rail surface, lost goods or damaged wheels leading to indentations on the rail. It has been observed that some surface defects grow into a squat. *Squats* are defined by the International Union of Railways as a "widening and localized depression of the rail/wheel contact band, accompanied by a dark spot containing cracks with a circular arc or V shape" [213]. *Cracks* to the head, web, foot, weld and joint plates are events that can lead to a broken rail. This type of flaw is a more rare occurrence that is further elaborated in the UIC Code 712 [213].

Diagnostic vehicles and their low cost counterpart - OBM vehicles - have been introduced to measure and collect data relating to the track state. Diagnostic vehicles are equipped with sensitive and highly accurate measurement systems, but require specially organized measurement rides and periodic system maintenance. OBM vehicles on the other hand are in-service vehicle equipped with simple sensors such as accelerometers, to continuously gather data related to track and vehicle condition.

(a) Intact insulated joint

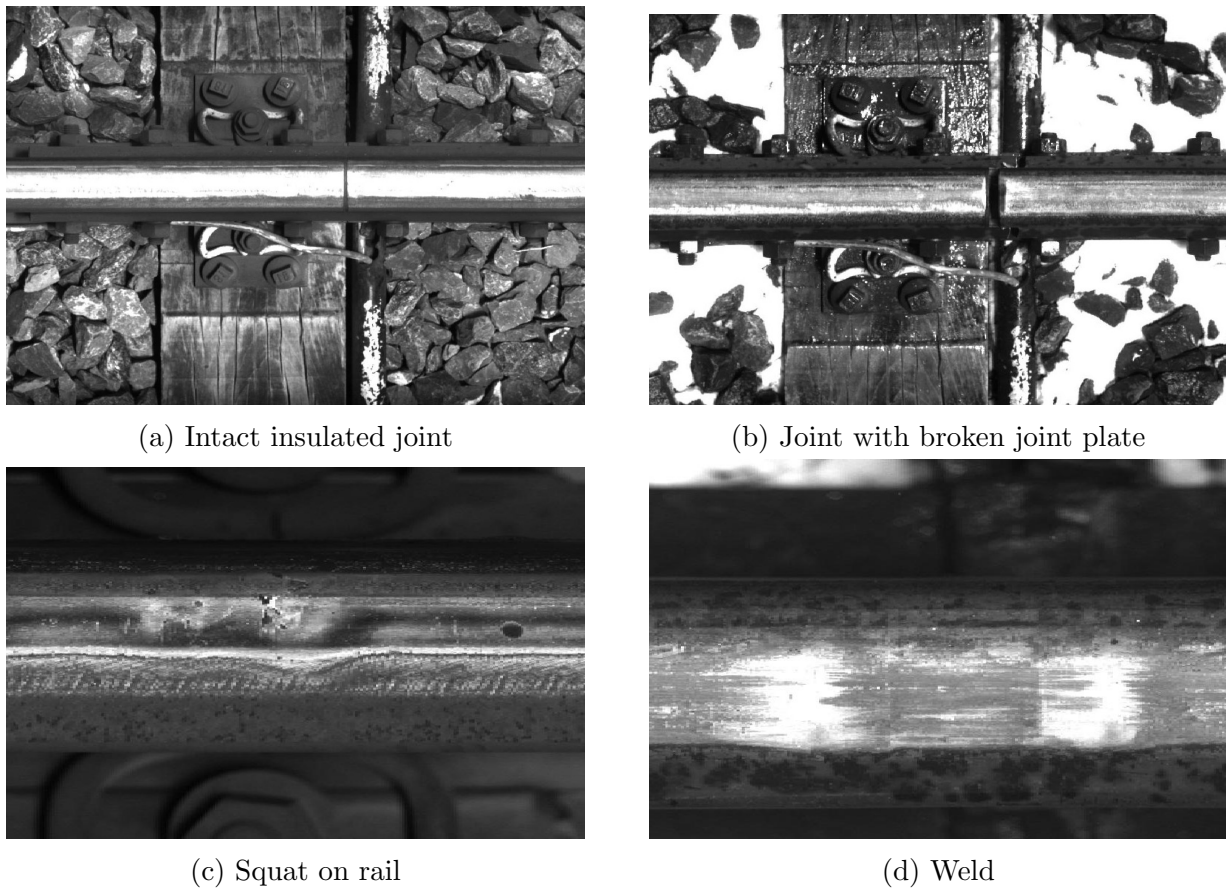(b) Joint with broken joint plate

(c) Squat on rail

(d) Weld

Figure 5.1: Characteristic defect and component classes (images extracted from a high speed camera, mounted on a diagnostic vehicle of the SBB)

Two approaches are generally followed when dealing with identifying and classifying accelerations time–series: parametric and non parametric methods. Parametric methods, such as the Kalman filter can be used to reconstruct the longitudinal level profile [69]. The longitudinal level is obtained via double integration of axle box accelerations (ABA) [6]. Non-parametric methods include time frequency domain analysis methods such as Fast Fourier Transform (FFT), Short Time Fourier Transform (STFT), Discrete Wavelet Transform (DWT) or Continuous Wavelet Transform (CWT). Molodova [326] used the wavelet coefficients from the CWT to classify squats and welds from axle box acceleration signals. This generalization of a classifier requires a dataset of sufficient size containing various types of track, welds, defects and operational conditions.

The methodology we propose to recognize and classify between the most common rail irregularities, namely surface defects, insulated joints and welds exclusively relies on axle box acceleration measurements from the vehicle track interaction measurement system of the diagnostic vehicle. Labeled information stemming from rail-head image based detection are combined with regular acceleration measurements to allow the capture of variations in measurement conditions such as measurement speed and weather. Two classification approaches are compared in this work. The first classifier exploits Convolutional Neural Networks (CNNs) that are applied to FFT coefficients, which are computed from acceleration time-series data. The second classifier relies on a more classical machine learning approach, applied on features that are extracted from the acceleration time series, which are then classified using random forests (RFs). RFs are compared to CNNs, because RFs allow for root cause analysis while CNNs perform well accross many time–series classification tasks. Finally, the uncertainty of the acceleration metrics and of the ground-truth labels is analyzed, motivating the application

of acceleration based detection for improvement of rail condition monitoring.

5.2 Methodology

Machine learning algorithms have had numerous successful applications for classifying signals from accelerations. The traditional approach requires the extraction of representative features from the acceleration time–series. These features are then used for classification. Kubera et al. [145] compare Support Vector Machines (SVMs) and Random Forests (RFs) to detect the change of vehicle speed using basic features computed from the time–frequency representation of audio data. Operational parameters, such as the vehicle speed, directly relate to the response of the axle and thus these parameters are important to be taken into account. The longitudinal level D0 (wavelength 1-3m), D1 (wavelength 3-25m) and D2 (wavelength 25-70m) are geometric parameters defined the railway norm EN13848-1 [52]. The longitudinal levels extracted from Axle Box Accelerations using integration and filtering techniques have been shown to be robust, speed independent and repeatable indicators [69][107]. The high frequency effects such as the vibrational modes of rail, wheel and axle effect the wavelengths of under one meter. The representation of these high frequency effects in the time–frequency domain, is obtained by decomposing the signals using DWT or STFT. Both the STFT and the DWT satisfy the property of invertibility, but the DWT unlike the STFT offers a high resolution in time and in frequency [92]. The condensation of a signal into a sparse representation of essential attributes is generally achieved by computing statistical features [233]. These features include minimum, maximum and mean values, as well as standard deviation and higher statistical moments (skewness, kurtosis).

Decision Trees (DTs) and Random Forests (RFs) are classic machine learning algorithms, that yield interpretable classification results, but similarly to SVMs work best on a reduced set of parameters. RFs are ensemble models, that aggregate several decision trees to achieve a more robust prediction than an individual DT. A semi-supervised interpretable machine learning framework for Sensor Fault detection, based on SVMs trained on numerous features extracted from acceleration time series, combined with the SHAP algorithm [275] was proposed by Martakis et al. [181]. The methodology proposed here employs RFs to classify the class label based on the previously described features (see also Fig. 5.2).

Deep Neural Networks have increasingly become popular in solving complex time series classification tasks with the increase of time series data availability [75]. Deep learning can achieve similar or better results, while avoiding to perform the feature engineering required for machine learning methods such as DTs and RFs. This is not to say that feature engineering has no place within deep learning; as the injection of some form of prior knowledge or physics in a learning system is often beneficial in better capturing the governing dynamics [146]. However, the ability of a neural network to ingest data and extract useful representations on the basis of examples is what makes deep learning so powerful. CNNs have often been used in tasks relating the classification of time series. In [318] a CNN has been applied to raw axle box acceleration data for the detection insulated joints. The methodologies for classifying axle box acceleration signals using CNNs & RFs are illustrated in Fig. 5.2.

5.3 Case Study

5.3.1 Data Description

The acceleration data are labelled under three main categories: *welds*, *insulated joints (Insul)* and *surface defects (SurfDef)*. The model is trained to differentiate between these three cat-
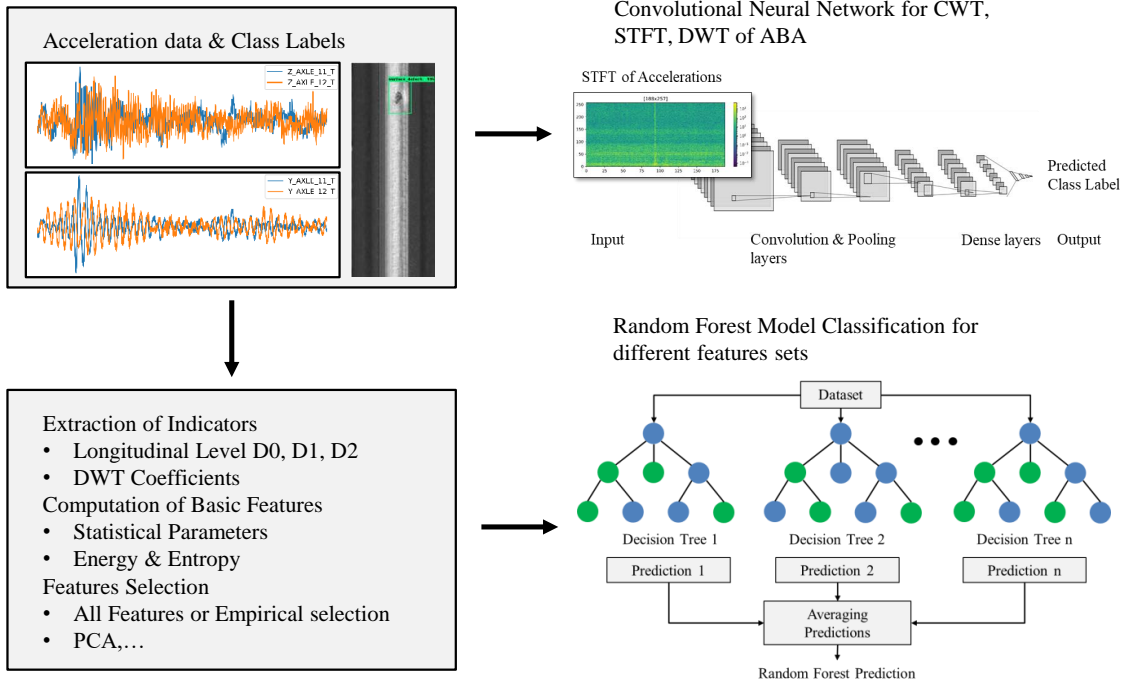
Figure 5.2: Proposed methodologies using CNN & RF for classifying labeled acceleration time series.

egories and a fourth one, named *no event (NoEvent)*. This category corresponds to baseline samples where no label was identified for the signal. The labeled acceleration events are created by merging the peaks in the acceleration signal to the labels stemming from deep learning models applied on rail head images. Defects recognized in the image are not always crossed by the wheel. Such occurrences are filtered out during the matching of acceleration peaks to component labels. The features used for training include accelerations, longitudinal level $D0$ and $D1$ obtained from accelerations, vehicle speed, DWT wavelet coefficients (bior 2.2) and their stochastic analysis (mean, percentiles, number of 0 crossings,..). The RFs are trained on this reduced features set. The CNN models are trained on the STFT representation of the time series.

The available data are highly imbalanced as the two classes *welds* and *no event* have a number of observations that is a magnitude higher than the one of *insul* and *surface defects*. Indeed, when a track is split into equal segments, the number of segments presenting a defect is much smaller than the number of healthy segments. Insulated joints are less common than welds, as they are only chosen over welds when necessary. The samples count is summarized in Tab. 5.1. The data was subdivided into a training and a testing set. A balanced dataset is required in order to train the model for each category and not just for the most represented classes. Balanced datasets are obtained either by decreasing the class weights of the over-represented category, or by oversampling the minority class via use of a Synthetic Minority Oversampling Technique (SMOTE) [54].

Table 5.1: Number of Samples in the database.

	No Events	Welds	Insulated Joints	Surface Defects
Sample count	180'042	179'097	11'808	9'657

5.3.2 Classification Results

The RF and the CNN are trained using the data pre-processing steps illustrated in Fig. 5.2. The model performance is compared for different data preprocessing steps in Tab. 5.2. The confusion matrices are illustrated for the classifiers with the best scores for the RF classifier in Fig. 5.3 and for the CNN model in Fig. 5.4. RF and CNN based models result in similar metrics.

Table 5.2: Average Accuracy scores of the most important models based on CNN and RF.

Classifier	Preprocessing	Avg. Acc.[1]
CNN	Speed normalized FFT	0.65
CNN	Speed normalized CWT	0.62
RF	Statistics on D0, D1 and DWT Bior 2.2 of ABA, speed	0.71
RF	PCA of statistics on D0, D1 and DWT Bior 2.2 of ABA, speed	0.64

[1] Average Accuracy.

For both RF and CNN models, a test score of around 62-71 % is achieved for the classification of component and damage labels using only preprocessed accelerations and vehicle speed as main inputs. The best RF model(71% average accuracy) was achieved by taking only the data from the accelerometers on the first axle raw data on which the following features were computed: Stochastic analysis of the accelerations, bandpass filtered signals D0 and D1, DWT of ABA (bior 2.2) and speed. Using Principal Component Analysis (PCA) to combine correlated features did not increase the model accuracy. STFT delivered the best classification results out of the different preprocessing approaches tested (STFT and CWT). The CWT and its subcategories scored similar results, although always slightly worse than the Fourier approach. Normalizing the data with the speed of the carriage, balancing the input data and increasing the number of train events were the steps that brought the major improvements to the results. Using only one axle instead of two, swapping the ABA data in order to have all defects on the same side and changing the number of layers in the TensorFlow model only had a minor impact on the results.

For all models, it can be observed that there is always a trade-off between accuracy and precision: It is possible to get over 80 % of the welds recognised correctly, however in a such model many surface defects are also classified as welds. On the other hand, when the model has a good detection rate of the surface defects, the detection rate of the welds drops to around 60 %. Insulated joints do not show such a clear trade-off as the acceleration response over this component can be separated quite easily from the other classes. In the conclusion, the sources of these trade-offs are explained and process and model improvements are proposed which may yield improved classification results.

5.4 Conclusions

The rail surface irregularities can be classified with a mean accuracy between 62 % and 72 %. Certain classes, such as insulated joints, are clearly separated from the other fault classes. The classification of surface defects, welds and no events is more challenging, as they lie in a continuum of axle box acceleration response dynamics. A perfectly executed weld is for instance hard to distinguish from a flat rail surface (no event). Certain welds may assimilate to surface defects due to a surface or internal material irregularity. Smaller surface defects do not cause a significant acceleration response, while large surface defects cause a stronger response. A finer subdivision of categories, such as welds and insulated joints, into *healthy* and *damaged*

	Confusion matrix, without normalization Train score: 0.94 Test score: 0.71			
True label	Insul	NoEvent	SurfDef	Welds
Insul	2313	27	185	367
NoEvent	61	23315	166	5315
SurfDef	116	557	1384	1179
Welds	1042	1e+04	3224	28331

(a) Absolute numbers

	Normalized confusion matrix Train score: 0.94 Test score: 0.71			
True label	Insul	NoEvent	SurfDef	Welds
Insul	0.8	0.0093	0.064	0.13
NoEvent	0.0021	0.81	0.0058	0.18
SurfDef	0.036	0.17	0.43	0.36
Welds	0.024	0.24	0.075	0.66

(b) Normalized

Figure 5.3: Confusion matrix after applying optimized hyperparameters for RF.

	Welds	Insulated joints	Surface defects	No event
Welds	687 (0.48)	126 (0.09)	267 (0.19)	354 (0.25)
Insulated joints	85 (0.07)	965 (0.78)	183 (0.15)	11 (0.01)
Surface defects	252 (0.20)	173 (0.14)	582 (0.47)	236 (0.19)
No event	270 (0.19)	53 (0.04)	38 (0.03)	1047 (0.74)
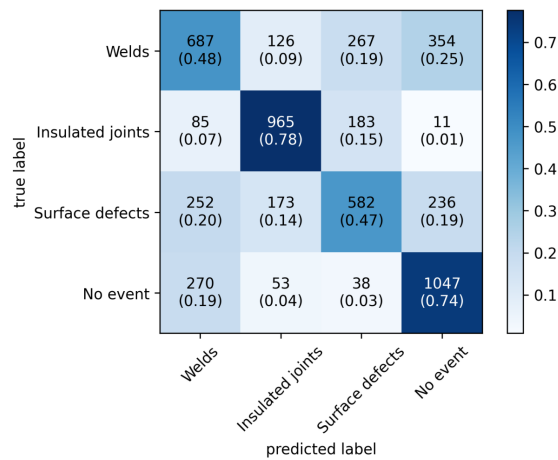
Figure 5.4: Confusion matrix of CNN-model on speed normalized STFT.

component categories, is being explored as a refinement of the scheme. The labels used in the classification are first automatically assigned from images. The assigned labels may in some cases be erroneous (false positive or misclassified samples). The categorization of a defect is not always 100 % clear and even when manually evaluating the image, there is some room for interpretation. The use of only human validated events may, at the cost of a lower sample size, help to filter out such unclear or erroneous instances. While the proposed models are successful at classifying accelerations, further improvements can be obtained by some extensions of the current framework. The combination of the RF model with the SHapley Additive exPlanations (SHAP) algorithm could be used to achieve a more interpretable classification framework [181]. The slightly lower performance of the CNN compared to RF may be explained by the smaller dataset used for training the model. Aside from training the model on a more powerful machine, directly feeding the untransformed measurement data into the CNN may result in better classification results. In such an instance, the CNNs would essentially act as feature detectors, that may be combined with an LSTM for the classification task [194]. The application of the classifiers to tracks with no prior knowledge is an essential step to validate the performance of data driven machine learning models. These models may in future enable rail condition monitoring using low–cost accelerometers on OBM vehicles.

Chapter 6

Weld condition monitoring using expert informed Extreme Value Analysis

Paper Details

Author and Co-Author Contributions

The author of this thesis prepared the data, performed the initial conception, validation and description of the outlier detection scheme presented here. A. Kollros and L. Ancu supported the conception of the expert validation framework including the setup of the processes that enabled an actionable feedback loop. Prof. E. N. Chatzi and Dr. V. Dertimanis supported the conception, provided valuable supervision and guidance.

Key Findings

- The study proposes a process to obtain labeled information from experts, by identifying potentially damaged welds using extreme value analysis performed on data-driven condition indicators.

- Extreme value analysis (EVA) models calibrated on various ABA metrics are used to identify outlier welds in an unsupervised fashion.

- Field experts provide feedback on rail-head images corresponding to the EVA-identified outliers, resulting in characterization of these outliers which are suspected weld defects.

- Uncertainty in the expert assessment is observed, as we notice that more severe faults are more commonly perceived as defect, while earlier stage faults are often not as clearly discerned by experts.

- Finally, we initiate the exploration of continuous monitoring approaches on the condition, by characterizing the weld condition in terms of outlier scores and their growth rates over time.

General comments and Link to the next chapter

This study addresses the last two objectives of the thesis (see Section 1.4) by presenting a salient approach to generate condition labels harnessing expert assessment. Extreme Value Analysis is performed to identify acceleration outliers that are then submitted for expert validation. The outlier detection framework combines is part of a first of its kind Proof-Of-Concept performed in collaboration with track experts at SBB to improve the condition assessment of welds.

In the following chapter, the expert feedback is exploited to propose actionable weld condition indicators.

Abstract

On-board acceleration measurements bear significant potential for the early detection of damage to railway infrastructure components. Condition assessment forms a complex problem in this case, due to the mobile nature of the On Board monitoring (OBM) solution, and the uncertainties associated with the rail-wheel contact dynamics, a lack of knowledge on the excitation sources (track, rail and wheel irregularities, parameter and self excitation) and the variability of the environmental and operational conditions. Welds are, amongst critical railway components, an essential element where high response amplitudes can occur. The monitoring of welds is still largely based on human assessment via typically visual means. We propose an automated approach to weld condition diagnostics via use of Extreme Value Analysis; an outlier detection schema which allows the early detection of flaws. In a second assessment phase, these potentially damaged welds are then assessed by experts during in-office and on-site inspections. The evolution of these OBM-based weld condition indicators is then tracked over time, leading to early detection of damaged welds.

6.1   Introduction

Railway tracks are composed of continuously welded rails that are discretely supported onto sleepers (e.g. Fig. 6.1a). Faults on welds often initiate due to material imperfections, as for instance the resulting discontinuity in strength of consecutive sections [68]. The increased loads that are generated on welds that feature a geometric irregularity (flaw) favor the growth of defects on welds [161]. *Squats* are a class of geometric irregularity (e.g. Fig .6.1b and 6.1c), which is defined by the International Union of Railways as a "widening and localized depression of the rail/wheel contact band, accompanied by a dark spot containing cracks with a circular arc or V shape" [213]. *Cracks*, on the other hand, form a defect which occurs more rarely on welds, as further elaborated in the UIC Code 712 [213].

The assessment of the rail condition via frequent, or feasibly continuous, monitoring is indispensable for effectuating efficient predictive maintenance schemes. Diagnostic vehicles and their lower cost counterparts - in-service On Board Monitoring (OBM) solutions - have been introduced to measure and collect data relating to the track condition. Diagnostic vehicles are equipped with highly sensitive and accurate measurement systems, but require specially organized measurement rides. In-service OBM vehicles are, on the other hand, typically equipped with lower cost, and thus lower precision, sensors such as accelerometers, to continuously gather data related to track and vehicle condition. Welds are most often visually inspected using images from diagnostic vehicles [107] or by performing field inspections. Track inspections often rely on use of handheld rail geometry measurement devices (laser devices such as Calipri C40 [202]) to assess the quality and diligence of the connection [2, 11]. Eddy current or Ultrasound-based measurements are often additionally used to identify welds with internal material flaws or cracks [274].

When seeking to execute data-driven damage detection on the basis of acceleration data, two are the prevailing approaches, namely supervised or unsupervised schemes [35]. Supervised techniques, such as regression or classification, are applicable when the target variable is known. While in our scenario, we know the existence and location of a weld, its condition is unknown. Unsupervised learning methodologies, such as Density Estimation [309], Data Clustering [24] and Extreme Value Analysis (EVA) [261], are well suited in such a scenario. We investigate the detection of potentially faulty welds by applying outlier detection, via use of EVA. The identified welds are, in a second - validation - stage, assessed by experts during in-office or on-site visual inspections. These validated samples are then used as ground truth-basis to further improve the automated algorithm for weld-defects identification.

(a) Healthy weld (camera mounted on a diagnostic vehicle of the SBB).

(b) Squat on weld

(c) Squat on weld

(d) Weld 6.1c after manual grinding

Figure 6.1: Characteristic defect and component classes (images from track inspections)

6.2 Methodology

Supervised machine learning approaches, such as Random Forests and Convolutional Neural Networks, have been shown to successfully allow the classification of surface defects, welds and insulated joints [104]. Supervised machine learning approaches are often limited by the quality of the available labels. The automatic image labeling algorithm of the SBB currently does not assess the condition of welds or insulated joints. Unsupervised machine learning has to navigate an unlabeled space and accomplishes this either via direct application to raw time history signals, such as accelerations, or to processed features that are derived from acceleration measurements during the crossing of welds. A common approach to outlier detection on acceleration data extracted from multiple channels is to first reduce the dimensionality of the acceleration data (e.g. via use of a Principal Component Analysis [253] or Autoencoder operator [242]), and then apply an outlier metric to the reduced manifold. For instance, Ulriksen, Tcherniak and Damkilde use Principal Component Analysis (PCA) to combine the acceleration data from 12 accelerometers, whose principal scores are then employed as damage features in the Mahalanobis metric in order to detect damage-induced anomalies [287]. A similar approach is also used by Avedano, Chatzi and Tcherniak for the assessment of wind turbine blades [16]

Following a similar logic, in this work, representative features that are extracted from the acceleration time–series are used for outlier detection. The longitudinal level D0 (wavelength 1-3m), D1 (wavelength 3-25m) and D2 (wavelength 25-70m) are geometric parameters defined in the railway norm EN13848-1 [52]. The longitudinal levels extracted from Axle Box Accelerations (ABA) using integration and filtering techniques are robust, speed independent and repeatable indicators [69][107]. Esveld and Steenbergen [72] developed geometrical standards for rail welds, by relating the slope of the longitudinal level D0 to the dynamic wheel-rail contact force and, thus, the condition of the welds. High frequency effects correspond to the vibrational modes of rail, wheel and axle. These effect can be represented in the time–frequency domain,
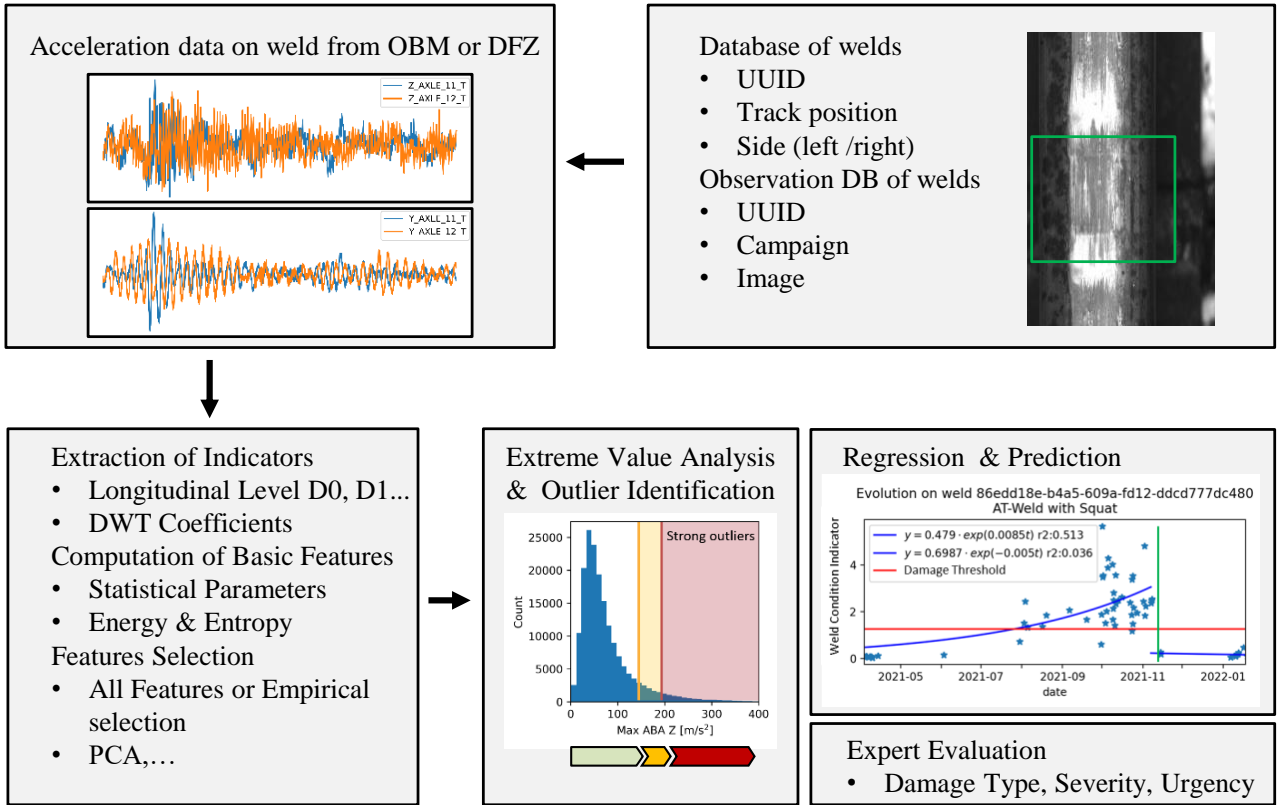
Figure 6.2: Event database generation for outlier detection, validation and condition prediction

by decomposing the signals using suitable transformations, such as the Discrete Wavelet Transform (DWT) or the Short Time Fourier Transform (STFT) [49]. Statistical features are often computed to condensate the signal into a sparse representation of essential attributes[233]. These features include minimum, maximum and mean values, energies, as well higher statistical moments (skewness, kurtosis).

To obtain information a first estimate on the damage state of welds, a feedback loop with experts was developed. We use EVA to identify outlier welds that are subsequently labeled by experts using railhead and track inspection system images from the diagnostic vehicle. The experts assess the condition of welds based on criteria defined in the deviation catalog of SBB [244]. Deviations in this catalog are for instance welds featuring a squat or a surface defect. The time history of damaged welds is subsequently analyzed to model the evolution of the defect. The flow of the proposed framework, from data preparation and feature extraction, to EVA and expert evaluation is illustrated in Fig. 6.2.

6.3 Case Study

6.3.1 Data Description

The railway tracks specifically selected for this study are regularly inspected on–site. They are located in the west, south, center and east regions of Switzerland. The data used for the expert validation in the feedback loops stems from the diagnostic vehicle (gDFZ) of SBB, which is equipped with 24kHz ABAs on the left and the right side of the first and last axle. The images from the rail head imaging system (Fig. 6.1a) is used to generate a database of automatically labeled welds. The acceleration samples related to these welds are processed to extract an array of features, such as the longitudinal level $D0$ and $D1$, DWT wavelet coefficients and their

stochastic analysis (mean, maximum, minimum, percentiles,...).

Table 6.1: Number of welds in the database.

Condition:	Healthy	Deviation	Unlabeled condition
Low Outliers	0	0	57659
Medium Outlier	276	24	17860
Strong Outlier	318	38	4416

The OBM–ICN of the Swiss Federal Railways (SBB) is an in-service train, equipped with 1 kHz ABAs that allows nearly daily collection of measurement data. The operating region of the in-service OBM train includes the tracks crossed by the diagnostic vehicle during the validation routes mainly conducted in the eastern region. This data source, while of lower frequency and resolution compared against the 24kHz ABA sensors of the diagnostic vehicle, is nevertheless useful for assessing the variance in vehicle response on a rail with a defect weld. The frequent measurements of in-service OBM trains allow to more frequently monitor the evolution of the track parameters over time. Tab. 6.1 summarizes the number of individual welds on all crossed tracks and the number of welds signaled as faulty or healthy during the feedback loop. Not all strong and medium outliers are assessed by experts, as not all track segments are part of the case study.

6.3.2   Application of Extreme Value Analysis

The first rounds of validation use prior-knowledge to identify outlier metrics. The longitudinal level D0 is an indication on the quality of the weld geometry [72]. Molodova has shown that squats induce a vibration of the rail–axle system in the frequency range of 200-1200Hz, which corresponds to wavelengths of less than 1 m [189]. Strong outliers, such as the ones caused by large squats, are identified as sharp impacts in the vertical ABA channels. Thus, maximum acceleration or energy metrics yield meaningful indicators for welds that feature short-wave irregularities such as squats. The 98% quantile for the maximum vertical ABA was selected based on previously observed welds with breakouts, as well as the capacity of review of the experts:

$$O_{score,strong} = \frac{ABA_{Z,max}}{q_{98\%}(ABA_{Z,max})} > 1 \tag{6.1}$$

Medium outliers are defined as the combination of outliers in the longitudinal level D0 and the maximum acceleration. These samples feature either high accelerations due to short wave irregularities or due to low geometric quality:

$$O_{score,medium} = \left(\frac{(D0)^2_{Z,min}}{q_{95\%}((D0)_{Z,min})^2} + \frac{ABA^2_{Z,max}}{q_{95\%}(ABA_{Z,max})^2}\right)^{0.5} > 1 \tag{6.2}$$

The corresponding outlier regions, along with the distribution of the vertical D0 dip and maximum acceleration are illustrated in Fig. 6.3. One can see that most welds have a dip of less than the 0.3 mm measured on good quality welds [72]. The high outliers (red region) represent 2 % of the welds, while the medium outliers (orange region), composed of the combination of two parameters, correspond to 10% of the samples.

6.3.3   Expert Assessment

The experts are asked to assess all the 356 samples with a high outlier score, while in the case of welds with a medium outlier score, only a random selection of 300 samples is validated. The
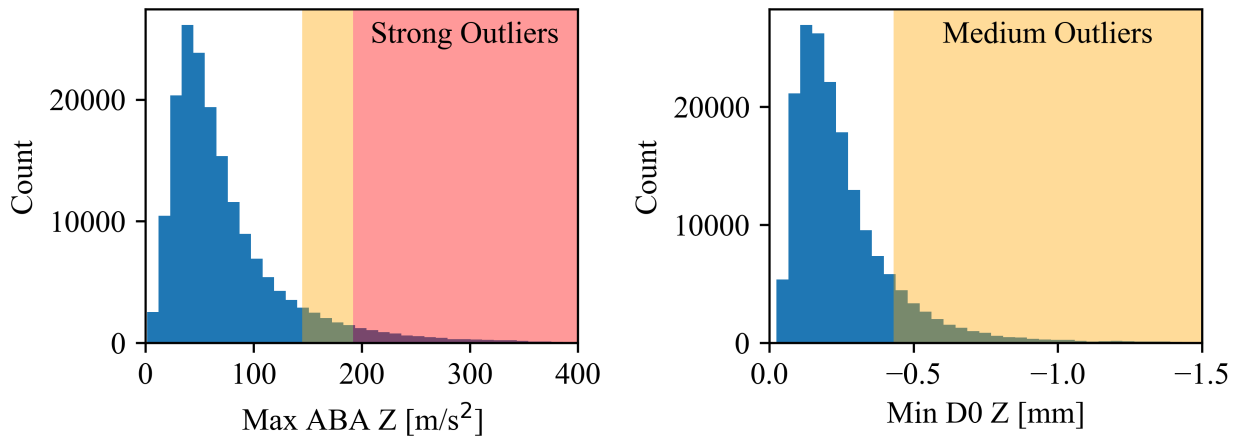
Figure 6.3: Distribution of maximum ABA and vertical D0 dip and corresponding medium and strong outlier regions.

results of the expert assessment are summarized in Tab. 6.1. After two feedback loops with expert assessment, around 11% of the strong outliers were labeled as defect by the in-office experts, while 8% of the medium outliers were labeled as defect by the experts. Fig. 6.4 shows the percentage of defects in the validated samples when the threshold of the medium or strong outliers is fixed to a specific outlier score. The vertical dashed line indicates the threshold between the strong and medium outlier indications. The dotted horizontal line offers a visual indicator for regions where the number of defects in the delivered sample is higher than 10%. One can note that with increasing outlier scores, the ratio of defect welds in the sample grows to 23%. At very high outlier scores, the fluctuation in the ratio of deviations increases due to the fewer remaining samples, resulting in a lower statistic significance. The faulty welds are most commonly identified as low/distorted weldings or as weldings with squats. In the next section, the time history of a faulty weld is assessed.
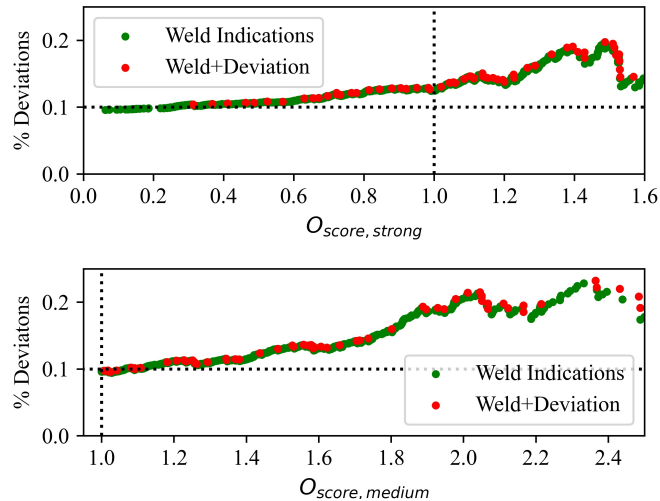
Figure 6.4: Percentage of outliers in validated samples for an outlier score higher than the selected threshold.

6.3.4 Time History analysis

The track sections used in this study are regularly measured by the OBM vehicle of SBB. Fig. 6.5 shows a waterfall plot of a track section with a weld that was labeled as a low / distorted aluminothermic weld during the validation rounds. Two outlier welds are located in

proximity to each other, one weld with a deviation around track offset 2480 m and another one without a deviation around track offset 2500 m. An increase in the peak energy at the location of the welds can be observed over time.
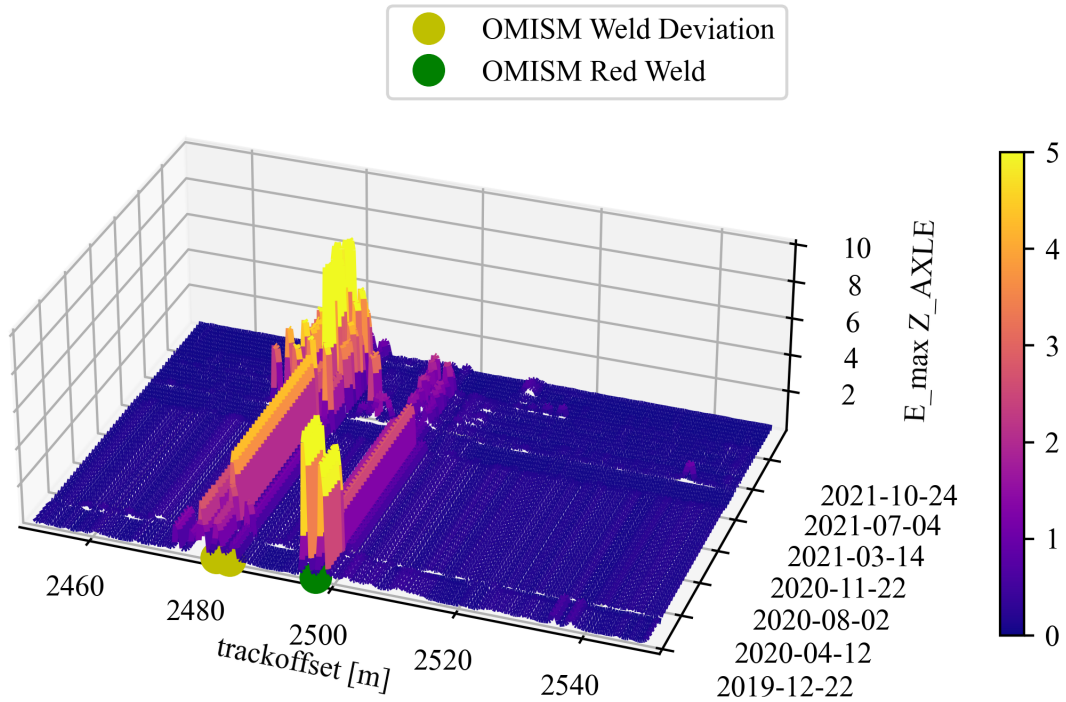
Figure 6.5: Evolution in time of OBM accelerations on a track section with a weld labeled as defect.

The weld that is labeled as defect on the section in Fig. 6.5 is now analyzed more closely. Fig. 6.6 shows a scatter plot of the maximum vertical ABA, the D0 dip and the maximum energy (EN25) in the 200-500 Hz frequency range. The blue line corresponds to the linear regression of the measurements spanning between the identified maintenance windows. The red line corresponds to the 98% quantile threshold used for initiating an expert review. The dashed red line, illustrates the time span between the first (29th January 2020) and the last observation (28th October 2021) of the defect by the track inspectors. Increased maximum ABA and EN25 during the weld crossing have been measured since the end of 2019 (Fig. 6.5). EN25 is significantly lower after the maintenance on this weld at the end of October 2021, however the maximum ABA and D0 dip are largely unchanged. Manually grinding (Fig. 6.1d) the welds with higher longitudinal level D0 does not always improve the observed rail geometry (Fig. 6.1c) and mechanical grinding machines are often more appropriate to achieve the required quality level [72].

6.4 Conclusions

An expert-in-the-loop assessment framework is here suggested, which utilizes acceleration outlier data to gather information regarding weld condition. Extreme Value Analysis models are used to identify outlier welds in an unsupervised fashion on the basis of defined quantile thresholds. The selected outliers are then passed onto the experts for verification on the basis of corresponding rail-head images. During the subsequent validation stage, the experts labeled 10 to 25% of the welds as faulty, with this percentage depending on the outlier score threshold. Internal flaws that would appear during an ultrasonic inspection are not assessed, which limits the faulty detections to visually recognized ones. This approach enables the monitoring of the
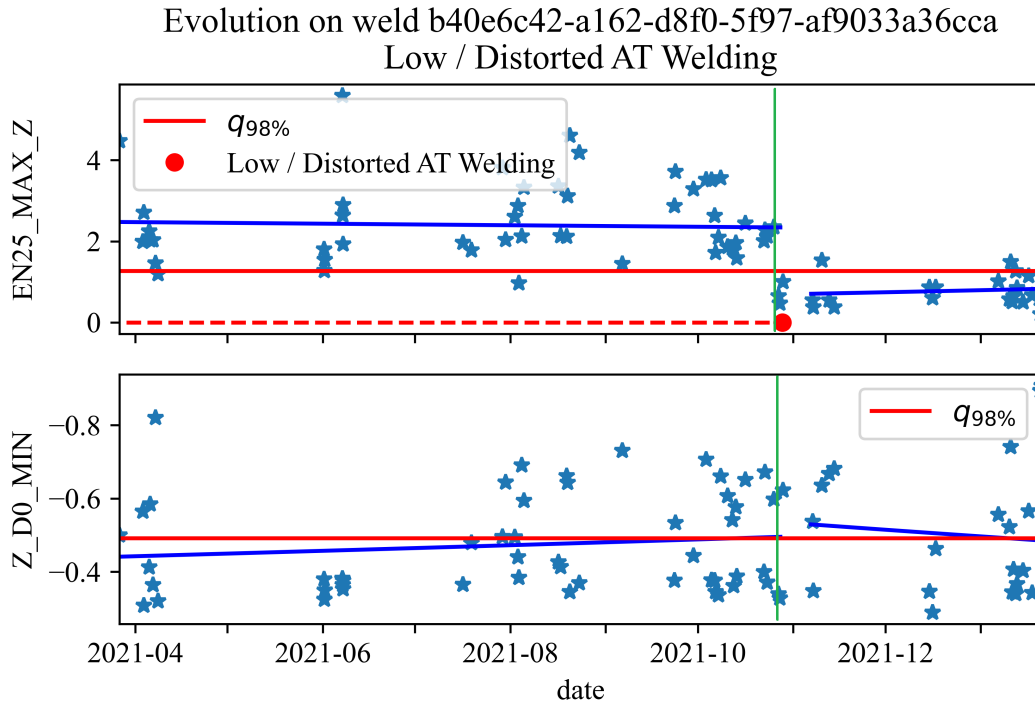
Figure 6.6: Evolution of OBM accelerations on the weld labeled as defect in Fig. 6.5 before and after a manual maintenance action was applied.

weld condition by combining expert knowledge with OBM-based measurements in an offline manner. While this is an initiating approach, the classification of faulty welds can further be improved by exploiting the assigned final labels from the expert feedback in a supervised detection approach [104]. It is important to note that what experts perceive as defects are those faults which have reached a rather advanced (and thus visually discernible) stage. On the other hand, the growth of squats, which can initiate as benign defects, is favored by the resulting increase in the dynamics, even for those defects that are not currently recognized as such by experts. Indeed, we here select a quantile, which can allow for early recognition of such defects at their initiation. The application of regression techniques to the evolution of OBM-based indicators over time is shown to yield parameters that describe the damage growth and the variability in contact conditions. The weld diagnostic model will in future be improved, by combining single measurement indicators with such temporal evolution information, to enable early flaw detection.

Chapter 7

Fusing Expert Knowledge with Monitoring Data for Condition Assessment of Railway Welds

Paper Details

Author and Co-Author Contributions

The author of this thesis conceived and developed the framework and carried out the testing analyses and reporting. G. Arcieri supported the development of the BLR model. L. Ancu and S. Banaszak supported the framework conception, data preparation and model validation. A. Kollros provided supervision and guidance and supported the coordination with all involved partners. Prof. E. N. Chatzi and Dr. V. Dertimanis helped to conceive the framework, provided supervision and guidance.

Key Findings

- The study delivers a holistic framework, which combines data-driven condition indicators with robust outlier analysis and complementary expert feedback for actionable implementation in railway management.

- An automated one-class classification scheme is developed using a combination of Binary Classification (BC), Random Forest (RF) classifiers, and Bayesian Logistic Regression (BLR), which identifies defect welds with significantly higher accuracy compared to the Extreme Value Analysis approach presented in the previous chapter. These models are trained on fused existing information with further evaluation processes, such as on-site inspections and non-destructive evaluations.

- A large challenge in the classification stems from bias or inconsistencies in the labeling of the data, which in practice cannot be considered in a binary approach, since faults grow in a continuously. In light of the high uncertainty caused by faulty ground truth labels, we investigate the BLR model which delivers not only a probability, but further assigns a level of confidence in the prediction.

- Finally, we demonstrate the importance of continuously monitoring the weld condition in order to robustly track the evolution of condition and use this as a guide for preventive maintenance actions.

General comments and Link to the next chapter

This study tackles the last objective of the thesis (see Section 1.4) by presenting a unique and innovative approach for railway infrastructure monitoring, combining both visual and acceleration information in a first of its kind Proof-Of-Concept to provide acceleration-based ratings of the rail condition.

This chapter completes the assessment of shortwave defects on welds and prepares the reader for the Conclusions.

Abstract

Monitoring information can facilitate the condition assessment of railway infrastructure, via delivery of data that is informative on condition. A primary instance of such data is found in Axle Box Accelerations (ABAs), which track the dynamic vehicle/track interaction. Such sensors have been installed on specialized monitoring trains, as well as on in-service On-Board Monitoring (OBM) vehicles across Europe, enabling a continuous assessment of railway track condition. However, ABA measurements come with uncertainties that stem from noise corrupt data and the non-linear rail–wheel contact dynamics, as well as variations in environmental and operational conditions. These uncertainties pose a challenge for the condition assessment of rail welds through existing assessment tools. In this work, we use expert feedback as a complementary information source, which allows the narrowing down of these uncertainties, and, ultimately, refines assessment. Over the past year, with the support of the Swiss Federal Railways (SBB), we have assembled a database of expert evaluations on the condition of rail weld samples that have been diagnosed as critical via ABA monitoring. In this work, we fuse features derived from the ABA data with expert feedback, in order to refine defection of faulty (defect) welds. Three models are employed to this end; Binary Classification and Random Forest (RF) models, as well as a Bayesian Logistic Regression (BLR) scheme. The RF and BLR models proved superior to the Binary Classification model, while the BLR model further delivered a probability of prediction, quantifying the confidence we might attribute to the assigned labels. We explain that the classification task necessarily suffers high uncertainty, which is a result of faulty ground truth labels, and explain the value of continuously tracking the weld condition.

7.1 Introduction

The increasing need for cost reduction and increase in efficiency and safety of railway infrastructure has prompted a surge of data-driven monitoring solutions for optimal management of railway assets [167]. Two essential aspects motivate the need for adoption of automated data-driven track inspection tools [183]: on one hand. the safety of the employees performing visual/on-site field inspections, and, on the other hand, customer comfort and safety [249]. Monitoring-based assessment is achieved by collecting data from specialized diagnostic, as well as from appropriately equipped in-service vehicles, which can provide a network-wide assessment of the railway infrastructure condition [107, 25, 13] and support preventive maintenance schemes [317, 313]. Railway track infrastructure typically consists of continuously welded rails supported by sleepers [97]. Among the critical components of the rail network, welds require particular attention in terms of execution, monitoring, and maintenance [95, 320]. By collecting and analyzing data on the condition of such critical infrastructural components, railway operators can better understand the health of their infrastructure and optimize the course of remedial actions [88].

Material imperfections, often originating near welds, grow into more severe faults over time when subjected to repeated stress [199]. These imperfections can include the following: *surface defects*, which form a broad category of defects caused by factors such as damaged wheels, ballast on the rail surface, or lost goods that induce indentations on the rail; *squats*, which are defined by the International Union of Railways as a "widening and localized depression of the rail/wheel contact band, accompanied by a dark spot containing cracks with a circular arc or V shape" [213, 325]; and *cracks*, which can appear at the head, web, or foot of the weld and which, although less common, can lead to a broken rail [213, 144]. On the railway network operated by the Swiss Federal Railways (SBB), the condition of the track is periodically assessed using data collected from diagnostic measurement vehicles [247]. Diagnostic vehicles are equipped

with sensitive and high-precision measurement systems, such as the Rail-Head Imaging System ($VCUBE$, Mermec Group, Monopoli, Italy) [172] and low-noise piezoelectric accelerometers (Type 4321, Hottinger Brüel & Kjaer, Virum, Denmark) [107]. While such vehicles can only traverse the network at rare pre-planned intervals, On-Board Monitoring (OBM) vehicles, on the other hand, are in-service vehicles equipped with simple and lower-cost sensors, such as microelectromechanical system-based accelerometers, which allow for continuous monitoring [219, 178, 169, 167]. In some implementations, OBM has been scaled down to low-power sensor nodes, further easing installation [32, 58, 228]. Despite, however, their potential to continuously gather data related to track and vehicle condition across the railway network the use of OBM vehicles has not been generalized yet [107] and, thus, the assessment of critical rail components still largely relies on field inspections [245].

ABA measurements can serve for fault identification and classification, since the dynamic properties of a system often closely reflect its condition [109]. Such a task is accomplished on the basis of two main approaches: model-based or data-driven schemes. Model-based methods, also known as hybrid methods, are physics-based models that are combined with data, in order to accomplish identification tasks, such as the recovery of the rail's longitudinal level profile from acceleration measurements [69, 281]. Model-based schemes enable the identification of the crossing of welds with smooth or degraded surface geometry enabling the identification of potentially faulty welds [162]. Simulations of the crossing of welds on high speed lines were used to estimate the relation between ABA and wheel/rail force and to propose a rail-time health detection method for rail welds [9]. This approach was, however, limited to one type of fault and the results may not fully reflect the varying conditions observed in reality. In the Netherlands, the quality of welds is assessed using gradient approaches on the rail geometry [72]. In such a scenario, limit values on the slope of the rail geometry are derived from a simplified vehicle model with an unsprung wheel mass to estimate the relation between the geometry of the rail, the ABA and the rail/wheel contact force during the crossing of welds [269].
Data-driven methods, on the other hand, are freed from a system model and often rely on transformed representations for extracting features. Typical examples of this class are time–frequency domain analysis methods, such as the Short Time Fourier Transform (STFT), and the Discrete or Continuous Wavelet Transform (DWT/CWT). CWT wavelet coefficients were used by Molodova [326] to classify squats and welds using acceleration data, while, in separate studies, the scale-averaged wavelet power, derived from the CWT coefficients, was applied to identify rail corrugation [159].

Time–frequency domain-based approaches were adopted in further studies for the detection of squat defects [56, 203, 327, 302]. The identification of track quality can also be performed directly on the basis of measured acceleration inputs, which are fed into statistical [31], Machine Learning (ML) or Deep Learning (DL) techniques. Such a principle has been exploited in a number of studies to predict geometric anomalies of the track on the basis of ABA measurements, and, thereby, facilitate the early detection of faults, which might otherwise lead to derailment [73, 265].

More recently, Yang et al. [319] demonstrated that both feature-extraction based methods and raw-input based DL methods, such as Convolutional Neural Networks (CNNs), can be applied to detect insulated joints on the basis of acceleration measurements. To further assess the condition of the rail, Tsunashima and Takikawa [283] identified outliers from the CWT spectrograms of ABA. These spectrograms were then analyzed by experts, who noticed that, for small faults, ABA based detection had a higher false positive rate than for larger faults. In their work, Shadfar et al. [250] presented an indicator for condition assessment of rail welds, formulated via coupling of a Fast Fourier Transform (FFT) with Principal Component Analysis (PCA). The authors noted that the performance of this indicator remained to be substantiated in a larger study. A similar approach was proposed by Xiao et al. [312], who combined the

wavelet packet decomposition (WPD) with an adaptive synchro-squeezed short time Fourier transform (ASSTFT) to locate damaged welds on a heavy-haul railway line. Previous studies demonstrated the viability of using the Hilbert–Huang Transform (HHT), which is a tool particularly suited for analysis of non-stationary signals, to characterize abnormal vibrations in damaged welds, as a means of monitoring tramway lines [126]. Availability of large datasets containing a range of Environmental and Operational Parameters [223], such as the DR-train dataset [168], allow for the adoption of more complex classifiers. Lasisi and Attoh-Okine [151] predicted the probability of rail fatigue defects by combining several Machine Learning model predictions via Multilayer Stacking Methodology. Their prediction was based on fault logs for the US Class I freight railroad and a set of parameters, such as the track layout, track type and the Million Gross Tonnes (MGT) [151]. Deep Learning approaches have gained more popularity with the appearance of larger datasets enabling the assessment of railway infrastructure with models, such as Convolutional Neural Networks (CNNs) or recurrent neural Networks (RNNs) [73, 216, 104, 322, 264]. More recently, approaches for fusing imagery/computer vision with inertial measurements have been proposed. Peng et al. used accelerometer, inclinometer and gyroscopic measurements, combined with image sensors, to quantify track alignment and irregularity and combined this information with visual sensors, but the rail condition (squats and surface defects) was evaluated using the computer vision-based assessment only [225]. Purely data-driven approaches are often limited by the quantity and quality of the training data required to learn a reliable representation of the underlying physical dynamics of the vehicle/track system.

In previous works of the authoring team [104], welds, surface defects, squats, and insulated joints were successfully classified using a dataset of over 200,000 instances, via the use of machine learning techniques, namely Random Forests (RFs) and CNN [104]. This was initial work towards the automated classification of essential rail elements on the basis of ABA data. In subsequent work [110], this approach was extended, via the use of an outlier-based detection scheme, to identify potentially faulty welds. The outcome of this investigation was adopted in practice by the SBB in a Proof-of-Concept study, where the suspected faulty instances were delivered to experts for subsequent assessment. The resulting expert-labeled dataset of healthy and defective welds formed the initial dataset, which was exploited in this study. This dataset was, then, further complemented with the welds whose condition labels stemmed from the classical track inspection process. For the SBB, such a process is logged in the form of entries in a so-called condition monitoring database ZMON (ZustandsMONitoring—condition monitoring). As highlighted in Table 7.1, several studies have demonstrated the feasibility of the use of ABAs for identification of squats, welds and insulated joints, usually on a small selection of samples. No study so far has proposed a component-specific and large-scale assessment, which attempts to fuse OBM indicators (acceleration-based ratings) with expert feedback. This study focused on the treatment of welds; a common component of the railway network, whose assessment is critical for alleviating faults that can cause increased costs or compromise safety [247].

This paper addresses this research gap by proposing a method based on ABA data for automatic defect detection on welds. The proposed methodology, which builds on a two-step procedure to assess weld condition, is illustrated in Figure 7.1. First, outlier welds are identified using Extreme Value Analysis (EVA) applied to the ABA indicators, that are extracted from the vehicle track interaction measurement system of the diagnostic vehicle of SBB. The images of these potentially faulty welds, extracted from $VCUBE$, are, then, submitted to experts for visual assessment, on the basis of which a dataset of expert-based condition labels is generated. The feedback from the expert evaluation round is exploited as an extension of the work in [110] to develop a classification system to distinguish between healthy and defective welds based on ABA indicators. The following three alternative analysis methods are examined here for the supervised approach: a pair of more conventional assessment schemes relying on (i) Binary

Table 7.1: Summary of the state-of-the-art of rail condition assessment with a focus on works exploiting ABA.

Proposed Approach	Research Findings	Limitations
Time-frequency analysis via Continuous Wavelet Transform (CWT)	Demonstrated identification of rail faults using the Scale-Averaged Wavelet Power of the CWT of ABA [326, 159, 56, 203, 327] or bogie measurements [302].	While the methodology can be extended to other components, these studies are limited to the assessment of squats.
Principal Component Analysis (PCA) on FFT coefficients.	Threshold for evaluating the condition of welds for the prioritization of inspection schedules [250].	Definition of linear weighting on the vehicle speed. The authors note that further studies are necessary to substantiate this indicator.
Wavelet Packet Decomposition (WPD) and Adaptive Synchro-squeezed Short Time Fourier Transform.	The authors identify the 300∼800 Hz frequency range to be indicative for poorly welded joints [312].	Empirical definition of a fixed damage threshold.
Hilbert-Huang Transform (HHT).	Rail joints were detected as impact points and outliers in the ABA signal with the HHT [126]	The methodology does not differentiate between components and only is demonstrated on a couple of samples.
Deep Learning architectures.	Classification of rail faults using Random Forests, Support Vector Machine, Artificial Neural Networks or Convolutional Neural Networks [73, 216, 104, 322, 264].	Requires large training datasets, especially when using acceleration time series instead of features as an input. Complex models require special care since they have a higher risk of overfitting.
Model & simulation based approaches.	Diagnostic thresholds were defined for faulty welds on the basis of ABA or force response simulations [9, 226, 72, 203].	No generalization to generic geometries and track types; simulation may not fully reflect complex site conditions and noisy data.
Current Work: Fusion of ABA-derived indicator with expert feedback.	Statistical methods applied on essential indicators for the identification of weld condition [110].	Uncertainties in the expert assessment cause noise in the input labels for classification algorithms.

Choice classification and (ii) Random Forests, and (iii) a herein proposed approach which adopts a Bayesian Logistic Regression scheme that capitalizes on availability of expert input. The novelty of this approach lies in the integration of human expertise with statistical assessments on ABA indicators, which allows for a more effective and efficient evaluation of welds. By

integrating expert feedback into the ABA-based condition assessment, we achieve a significant advance in the field of condition assessment of railway assets which has the potential to improve the accuracy, consistency, efficiency, and cost-effectiveness of the inspection process. This PoC constitutes a first step towards actionable integration of acceleration-based infrastructure condition assessment into the monitoring process of railway operators.
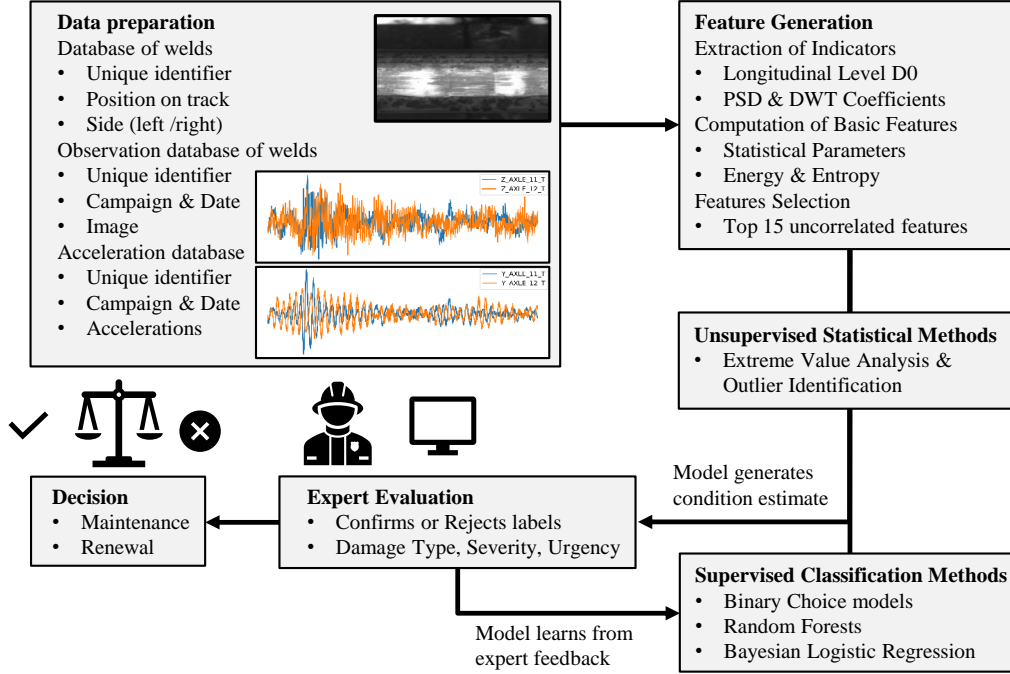
Figure 7.1: Flowchart summarizing the proposed methodology applied for automated weld defect detection and classification from ABA data. The accelerations and rail-head images are continuously collected by the diagnostic vehicle of the SBB (gDFZ) and extracted at the location of welds. Outlier welds, that are statistically identified from the features extracted from the ABA data, are subsequently delivered for expert assessment with complementary input of the available image data. The expert assessment is then used to retrain the models using supervised ML algorithms. Finally, the experts use the improved model to guide their decisions on maintenance and renewal.

7.2   Description of the Measurement Data

To ensure a representative dataset, the welds assessed in this work were selected from rail-track portions throughout the Swiss network, namely the west, south, center and east regions. The selected railway tracks were amongst those that were regularly inspected by the diagnostic vehicle (gDFZ) and were further accessible for on–site inspections by experts, when required. The gDFZ was equipped with vertical and lateral ABA on the left and the right side of the front (axle 1) and rear axle (axle 4) of the vehicle. The sensor range was ± 100 g and the sampling rate was $F_s = 24$ kHz. The naming convention of the associated sensors, complying with railway–specific standards, was D_{AS}, with the related letter-entries explained in Table 7.2.

The location of welds on the railway network was obtained from an automated detection algorithm developed at SBB. It relies on images from $VCUBE$ and determines rail components, such as welds, insulated joints and surface defects [247]. Data obtained between June,

Table 7.2: Meaning of the D_{AS} naming convention of the sensors.

Letter	Explanation	Possible Entries
D	direction	Y for lateral, Z for vertical
A	axle number	1 to 4, starting from the front (leading) axle
S	vehicle side	1 for right, 2 for left (w.r.t vehicle's top view)

2021, and June, 2022, was, herein, used to generate a weld database containing around 25,000 unique welds. While the automated rail-head inspection system does not always detect all existing welds, the repeated measurements increase the probability of individual weld detection. Approximately 10 vehicle runs were conducted within this one year period, implying that the majority of welds were repeatedly measured in this interval, allowing for tracking of their condition over time. The resulting collected ABA samples are illustrated in Figure 7.2, which shows the $VCUBE$ images and the measured ABA for two healthy (Figure 7.2a,b) and two damaged weld (Figure 7.2c,d) cases. The damaged welds showed higher ABA on all channels compared to the healthy welds. Figure 7.2 shows the left and right ABA time series for the vertical ABA on the leading axle Z_1, the vertical ABA on the trailing axle Z_4, the lateral ABA on the leading axle Y_1 and the lateral ABA on the trailing axle Y_4. The ABA data samples corresponding to welds recognized by $VCUBE$ were processed accordingly, as described in Section 7.3.1.

(a) Healthy Weld   (b) Healthy Weld   (c) Damaged Weld   (d) Damaged Weld
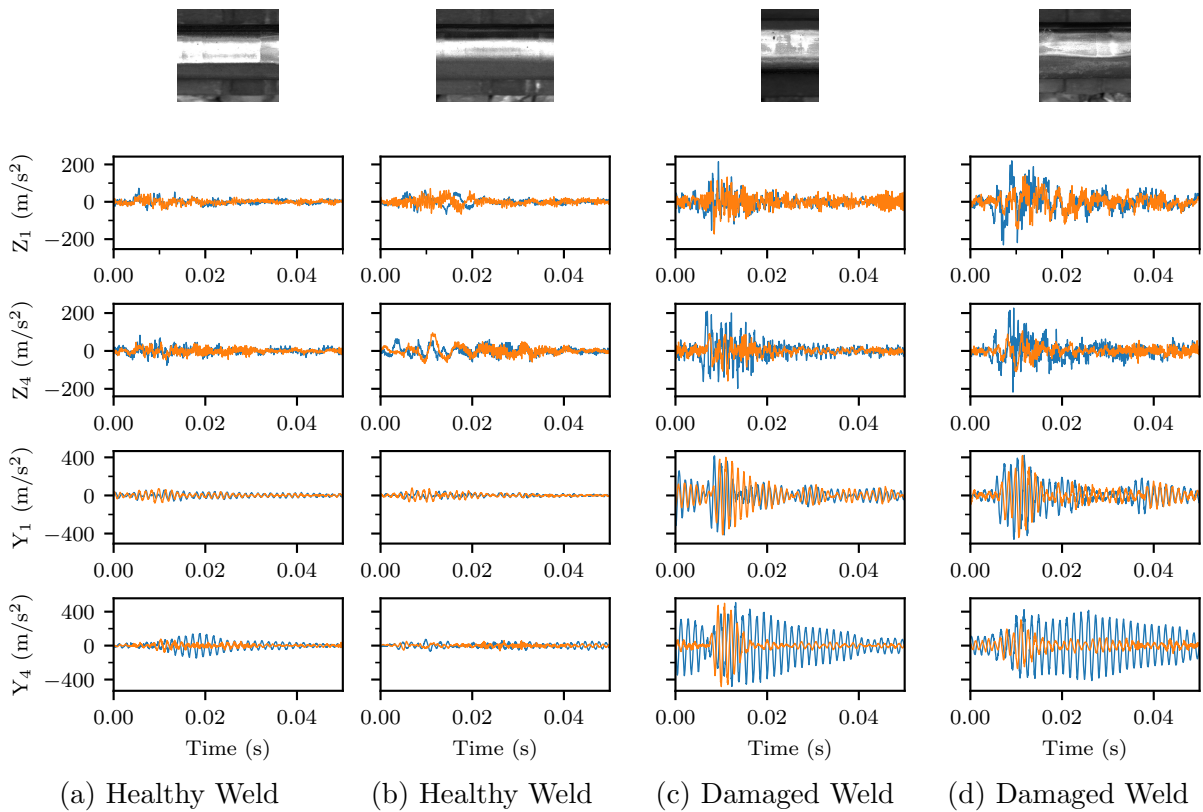
Figure 7.2: Illustration ($VCUBE$) and ABA time series for two healthy (**a,b**) and two damaged (**c,d**) welds. Each figure shows the left (orange) and right (blue) ABA time series, respectively, for the naming convention defined in Table 7.2.

7.3 Methodological Approach

We propose a framework which exploits labeling of rail infrastructure defects on the basis of expert evaluation on outliers in ABA data, acquired from instrumented trains. An automated classification framework for weld condition monitoring is, thus, established, that can be applied to newly acquired measurement data, potentially from in-service trains.

To achieve this goal, we first extracted a range of features from the ABAs of a properly instrumented diagnostic vehicle. These were calculated over the whole data range (one year period). Following a statistical characterization of the features, extreme value analysis was performed and outliers were identified. The latter were submitted to experts, and a large database of fault characteristics was generated. The expert-feedback was finally incorporated into a classification framework, which scored the quality of welds, based on the coupling of ABA features with expert-feedback.

7.3.1 Feature Extraction

The efficient identification and classification of outliers calls for the extraction of representative features from the ABA time series. A critical requirement for an interpretable anomaly detector is the computation of features that provide an intuitive and comprehensive illustration of the state of the assets from the measured time series [181].

The fundamental quantities, on which features were extracted, are listed in Table 7.3 and further explained in Section 7.3.2. The signal length that was used to compute the features was determined on the basis of two criteria. First, the signal length should be sufficiently long to contain the analyzed wavelengths and/or frequencies. For example, a signal length of 5 m around the weld was chosen for the longitudinal level $D0$ (see Section 7.3.2 below), since the filtered wavelengths were up to 3 m. The second criterion stems from uncertainty in vehicle position when crossing the welds. This positional uncertainty was tackled by choosing a rather large signal length around the defect (0.625 s for the DWT features or around 2–3 m for all other features).

The required sparse representation of essential features was achieved by computing statistical indicators on the basis of the extracted features [233]. These included minimum, maximum, mean value, standard deviation, and quantiles, as well as higher statistical moments, such as skewness and kurtosis. Statistics were computed for each quantity of Table 7.3, and for every channel in the vertical Z and lateral Y direction, where applicable. As multiple vehicle axles were equipped with sensors, the statistical indicators were also aggregated between the four sensor locations using the mean, minimum and maximum of the single channel statistics.

Table 7.3: Fundamental quantities used for the feature extraction process.

Feature	Quantity	Signal Length
RAW ABA	Raw accelerations	2 m, 3 m
VS ABA	Vector sum of Y and Z axes	2 m
BP ABA	Band Pass filtered accelerations	3 m
STFT	Short Time Fourier Transform	2 m
DWT	Discrete Wavelet Transform	0.625 s
$D0$ & $D1$	Longitudinal level and lateral displacement	5 m, 20 m

For example, one can consider the maximum vertical acceleration for each of the ABA sensors separately, or the maximum (or mean of the maximum) vertical acceleration over the four sensor locations, in order to form aggregated summary statistics. Such statistics capture the variability in contact conditions mainly arising from defect size, which, for small defects, can lead to greater variance in response between vehicle axles.

Operational parameters, such as the vehicle speed, were additionally included in our assessment framework, since they retain a non-negligible influence on the axle response.

7.3.2 Time Series Analysis

During the crossing of a damaged weld, increased vibration levels were observed in the vertical and lateral ABAs. In particular, impacts on the axle caused high amplitudes in the lateral and vertical vibration of the rail–axle system, whose first resonant frequencies lay, approximately, around 660 Hz and 1 kHz, respectively [176, 36].

Indicatively (in the sequel, all signals were mean value subtracted prior to any processing applied), vertical peak accelerations of up to 200 m/s^2 and lateral peak accelerations of up to 600 m/s^2 were noted during the crossing of the damaged weld shown in Figure 7.3a for both the vertical (Figure 7.3b) and lateral (Figure 7.3c) ABAs. The vertical (s_Z) and the lateral (s_Y) acceleration components were, herein, further combined, by computing the vector sum of both signals, i.e.,

$$s_{\text{vector sum},YZ} = (s_Y^2 + s_Z^2)^{\frac{1}{2}} \tag{7.1}$$

where s is the time series of the respective ABA channel. Figure 7.3d shows the vector sum computed from the lateral and vertical ABA of the damaged weld.

When investigating high frequency vibrations, the signals of Figure 7.3b,c were zero-phase, high-pass filtered using a 6th order digital Butteworth filter at 100 Hz cutoff frequency. Figure 7.4a,b display the spectrograms (Welch's method with $N_{FFT} = 128$, Hanning window and 50% overlap) of the filtered vertical and lateral ABAs, respectively. The previously identified pinned–pinned resonant frequencies of the rail/axle system lay around 700 Hz, laterally, and 1 kHz, vertically, for the standard UIC60 rail [176, 36]. These frequencies are identified from Figure 7.4a,b. The rail–wheel system was less stiff and less damped laterally, which resulted in higher vibration amplitudes compared to the vertical wheel-set response. Moreover, critical vibration response of the damaged weld was observed for frequencies up to 12 kHz. For subsequent analysis of the ABA data, a set of empirical frequency bands was, thus, formulated. These were the following: (i) 200–500 Hz, (ii) 500–800 Hz, (iii) 800 Hz–2 kHz, (iv) 2–4 kHz; and (v) 4–11 kHz. The associated signals, which form the BP ABA entry of Table 7.3, resulted from corresponding filtering of the original ABA, using the same zero phase Butterworth filter as before, applied in band pass mode at the selected ranges.

Similar results for the frequency content of the ABA were extracted by applying the Discrete Wavelet Transform (DWT) [92]. The DWT was, herein, implemented using the Haar mother wavelet and successive filtering operations with two FIR filters: a low-pass filter h_ϕ and a high-pass filter h_ψ. The associated approximation $W_\phi(j,k)$ and detail $W_\psi(j,k)$ coefficients of the j-th scale were computed by convolution [91]

$$W_\phi(j,k) = h_\phi(-n) * W_\phi(j+1,n)|_{n=2k,k\geq0} \tag{7.2}$$

$$W_\psi(j,k) = h_\psi(-n) * W_\phi(j+1,n)|_{n=2k,k\geq0} \tag{7.3}$$

Figure 7.5a,b illustrate the detail coefficients computed for the high pass filtered vertical and lateral ABA time–series of the weld in Figure 7.3a. The same insight, as in the case of the spectrogram, was obtained for the effective frequency bands of the signals.

Geometric features, such as the longitudinal level, or the lateral alignment, reflect the ver-
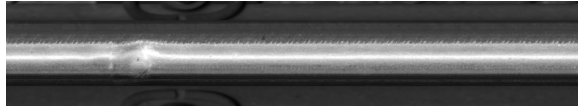
(a) Rail-head image of a weld with a squat identified during the feedback loops (length of image: 0.6 m).
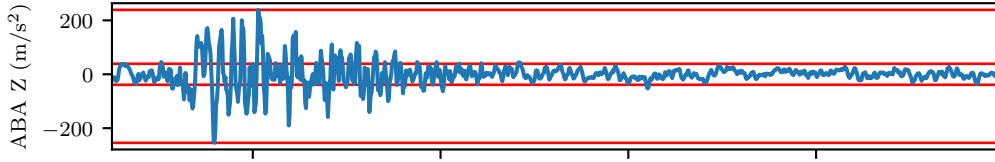
(b) Unfiltered vertical acceleration for channel Z_{12} during the crossing of the weld in Figure 7.3a.
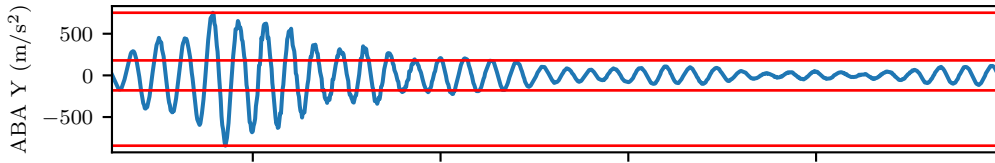
(c) Unfiltered lateral acceleration for channel Y_{12} during the crossing of the weld in Figure 7.3a.
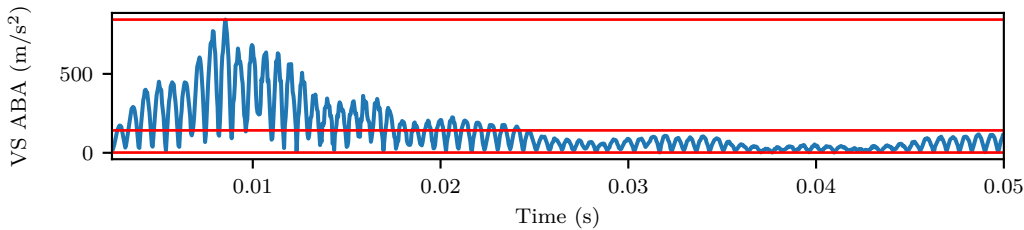
(d) Vector sum of vertical and lateral accelerations for channel YZ_{12} during the crossing of the weld in Figure 7.3a.

Figure 7.3: Feature extraction from the vertical and lateral ABAs. The vertical and lateral ABAs were combined using the vector sum. Where applicable, the continuous red horizontal lines highlight the maximum, minimum and standard deviation of the time series.

tical or lateral smoothness of the rail [295]. These quantities are defined for several wavelength ranges by the railway norm EN13848-1 [52], namely $D0$ (wavelength 1–3 m), $D1$ (wavelength 3–25 m) and $D2$ (wavelength 25–70 m). Using integration and filtering techniques on ABAs, one can obtain robust, speed independent and repeatable indicators corresponding to the longitudinal level and lateral axle displacement [107, 69, 167, 57].

When estimating the longitudinal levels, the ABA signals were initially zero phase, band-pass filtered (0.5 Hz to 75 Hz, 6th order digital Butterworth filter) and resampled at 150 Hz. The cumulative trapezoidal numerical integration method was accordingly applied to yield the double integrated vertical and lateral displacements. Drifts stemming from the integration process were removed by applying a 6th order Butterworth high-pass filter with a cutoff frequency of 0.5 Hz, which corresponded to the minimum frequency response of the sensor. The resulting displacement signals were then transformed from time series to space series, using a wavelength rate of 25 cm. Finally, appropriate band-pass filters were applied, to obtain the longitudinal levels and lateral displacements $D0$, $D1$ and $D2$. This approach for longitudinal level recovery has been successfully applied at the SBB [107] and the German Railways [167].

The longitudinal levels $D0$ and $D1$ computed from the vertical ABA are illustrated in Figure 7.6a,b, respectively. These correspond to the level during the crossing of the damaged
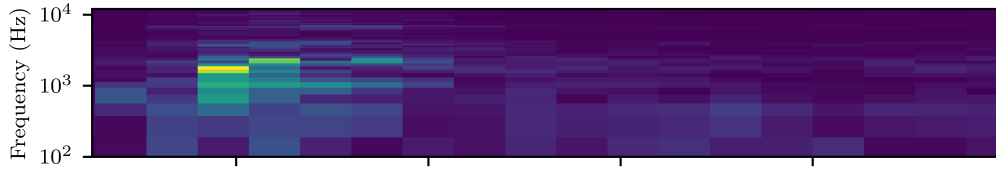
(a) Vertical acceleration (channel Z_{12}).
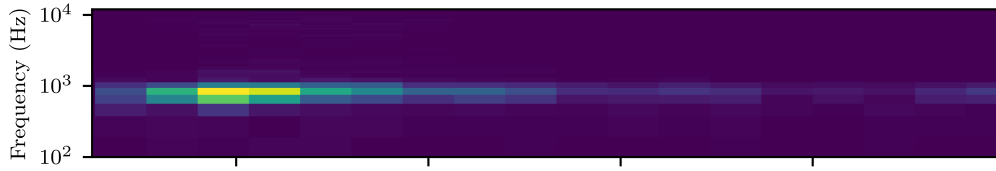
(b) Lateral acceleration (channel Y_{12}).

Figure 7.4: Spectrograms (Welch's method with $N_{FFT} = 128$, Hanning window and 50% overlap) of the high pass filtered (100 Hz cutoff frequency) vertical and lateral ABAs of Figure 7.3b,c.
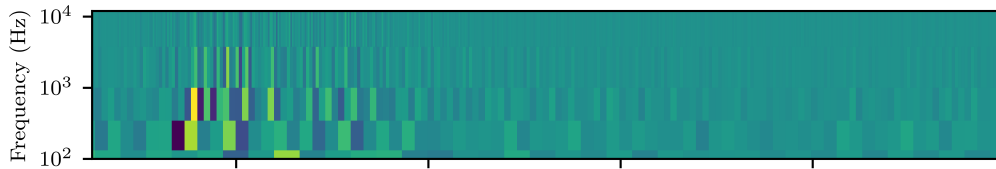
(a) Vertical acceleration (channel Z_{12}).
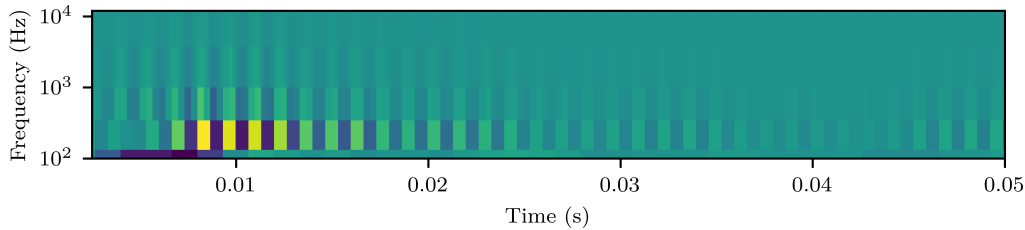
(b) Lateral acceleration (channel Y_{12}).

Figure 7.5: DWT (via Haar wavelet) of the high pass filtered (100 Hz cutoff frequency) vertical and lateral ABAs of Figure 7.3b,c.

weld of Figure 7.3a. The plots revealed settlements, which were commonly observed at the location of the weld in both the short wavelengths under 3 m and the medium wavelengths between 3–25 m. These localized track settlements are the result of the repeated impacts of the vehicle wheels on the damaged weld [243]. The settlements occurring in the wavelength range of 1–25 m could be attributed to further causes, such as changes in the track stiffness and substructure condition [64], thus, forming a less robust indicator of rail and weld condition.

7.3.3 Extreme Value Analysis for Outlier Identification and Expert Labeling

A limitation of supervised machine learning approaches is the limited availability of labels, as well as their quality, since these are often linked to subjective and, therefore, biased assessment. For the dataset we were handling herein, image labels could be extracted. However, the automatic image labeling algorithm of the SBB does not currently output the condition of welds; a task which would be non trivial to effectuate. Capitalizing on the availability of the
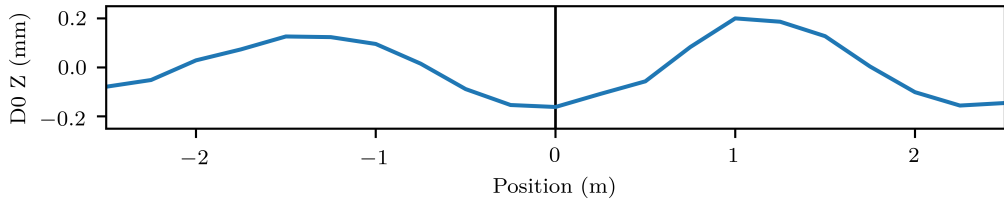
(a) Longitudinal Level $D0$ for the left sensor on axle 1 (sensor 12).
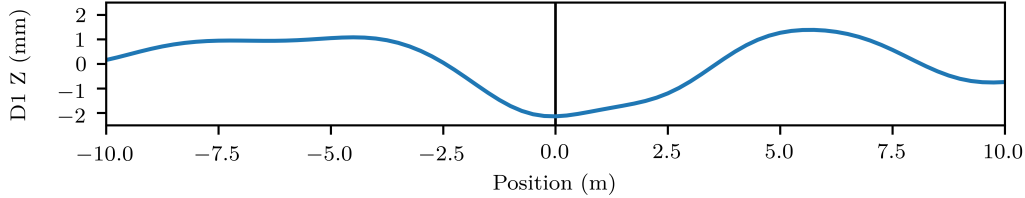
(b) Longitudinal Level $D1$ for the left sensor on axle 1 (sensor 12).

Figure 7.6: Longitudinal Levels $D0$ & $D1$ for the time series of Figure 7.3b. One can observe local settlements occured for both the 1–3 m range of $D0$ and the 3–25 m range of D1.

collected ABA signals, we here adopted an unsupervised scheme, applied directly to raw time history ABA signals, or to the aforementioned features, to detect abnormal, and potentially faulty, welds.

Extreme Value Analysis (EVA) is a statistical technique that is used to analyze the likelihood and impact of the occurrence of extreme events, such as floods, hurricanes, and earthquakes. In practice, EVA often relies on the use of extreme value distributions, such as the Gumbel, Fréchet, and Weibull distributions [98], or even empirical distributions [291], which are formed on the basis of available data. The cumulative density function of an empirical distribution is formulated as [59]

$$\tilde{f}_{\mathrm{ED}}(t) = \frac{i}{n+1} \quad \text{for} \quad x_i \leq x < x_{i+1} \tag{7.4}$$

where $\{x_1, \ldots, x_n\}$ is an ordered sample of n independent observations. $\tilde{f}_{\mathrm{ED}}$ is an estimate of the true probability distribution f, and should be in reasonable agreement with the candidate model (e.g., f_{Gumbel}), provided the candidate model is an adequate estimate of f [59]. The Gumbel distribution is commonly applied for modeling the behavior of extreme events and may have been an alternative to the empirical distribution.

The non-parametric empirical distribution is fitted to the data and is subsequently used to estimate the probability of specific outlier level occurrences. The probability estimates resulting from the empirical distribution enable the estimation of the likelihood of the occurrence of a specific level of an ABA-based feature on a component. The computation of the likelihood of occurrence of an extreme value enables the assessment of the potential damage of a defect (outlier) weld. This further requires the definition of threshold of damage levels for ABA-extracted features. EVA was here adopted as the first step of our proposed assessment framework, in order to identify outlier welds that could be subsequently labeled by experts, who were shown rail-head and track inspection system images extracted from the diagnostic vehicle. The number of samples that could be evaluated by the experts was limited, which must be taken into consideration when setting thresholds for outliers. Expert-based labeling requires significant time, as each sample is checked individually. Therefore, EVA was used in the first step to identify and forward only suspected defect welds for cross-checking and labeling. Section 7.4.1 details the practical implementation of the expert labeling process, the selected thresholds on the ABA features and the results of the expert evaluation. The expert feedback resulted in

a labeled weld condition dataset, which, in turn, enabled the establishment of an automated classification scheme, as described in Section 7.3.4.

7.3.4   Expert-Informed Classification Models

The derived expert labels were exploited for automated weld damage classification, via the use of machine learning classification tools. The binary approach used here, to distinguish between healthy and defective, stemmed primarily from the fact that the magnitude of a defect is not a clearly defined criterion among experts nowadays.

In the most simple scenario, Binary Choice (BC) models assign a choice between two discrete alternatives (in this case defective or healthy) on the basis of a classification rule depending on one variable x. This was used here as an approach to an one-class classification between a defective or healthy weld, in the sense of what expert judgment tried to offer. However, it has to be emphasized that this is not entirely consistent with the goal of continuous monitoring. In reality, the task of characterization of defect welds should also take the damage severity into account, which, however, is a label that is currently missing. A characterization on the basis of severity of the defect is valuable and can be provided via ABA data, which can pick up the initiation and evolution of a defect. Expert labels, on the other hand, tend to only acknowledge quite progressed defects. Section 7.4.3 indicates how such tracking can be accomplished on the basis of ABA measurements.

Returning to one-class classification, when considering a BC model, the threshold which defines the limit for the decision on a healthy or defective weld can be defined as

$$P_{BC}(y|x) = \begin{cases} 1 & x > \gamma \\ 0 & x \leq \gamma \end{cases} \tag{7.5}$$

where y is the label from the expert assessment, x is the statistical indicator and γ is the decision threshold. For indicator values x larger than the decision threshold, the sample is assumed to be defective.

Beyond mere classification, however, alternate models relying on a graph structure, such as Decision Trees (DTs), or their ensembles, Random Forests (RFs), can further point to a root cause analysis path [53, 3]. In other words, they can reveal variable configurations which lead to a specific outcome. DTs are a graph structure, in which each internal node denotes the outcome of a test on an attribute, each branch denotes the result of the test, and each leaf node (end node) denotes a class label. The paths from root to leaf represent the classification rules. Figure 7.7 conceptually illustrates three DTs, which are combined into one RF. RFs are ensemble models that aggregate several DTs to achieve a more robust prediction than an individual DT. The methodology proposed here employed RFs to classify the class label, based on the features extracted from the ABA signals. More formally, given a set of N decision trees $\{T_1, T_2, \ldots, T_N\}$ in the forest, the prediction of the RF was achieved by aggregating the prediction of the individual DTs of the class label y for a set of essential indicators x=$\{x_1, x_2, \ldots, x_k\}$

$$P_{RF}(y) = \sum_{i=1}^{N} \frac{T_i(x)}{N} \tag{7.6}$$

Many models, including RF, benefit from a limited collinearity of the variables [87] by manually discarding variables using the Pearson correlation or combining them via Principal Component Analysis [16]. The essential indicators selected for the evaluation of welds via RF and BLR, summarized in Figure 7.4, were the ones which had the highest $F1$-score in the univariate BC model scenario and which had a Pearson correlation of less than 0.8 to the other indicators.

Figure 7.7 conceptually illustrates the structure of an RF. RFs include a set of hyperpa-

rameters, such as the number of estimators, the minimum number of samples per split, the maximum depth, and the minimum samples per leaf. The optimal parameters of the RF were here estimated using a Cross-Validated Grid Search [224]. The optimal set of hyperparameters identified by the Cross-Validated Grid Search for the Random Forest was a minimum number of samples per split $n_{split} = 20$, a minimum number of samples per leaf $n_{leaf} = 10$, a maximum depth $n_{depth} = 10$, and a number of estimators $N = 100$. The RF model proposed here relied on the implementation of scikit-learn [224], where the Shannon entropy loss $H(X_m)$ was used as the tree node splitting criterion [251]

$$H(X_m) = -p_{\text{healthy},m}log(p_{\text{healthy},m}) - p_{\text{defect},m}log(p_{\text{defect},m}) \tag{7.7}$$

where $p_{n,m}$ is the proportion of observations of each class n at a given node m. The Shannon entropy quantifies the expected uncertainty inherent in the possible outcomes of a discrete random variable; in other words, it quantifies the impurity in a group of observations. Thus, for each node, the tree splitting criterion was set such that the entropy loss $H(X_m)$ was minimized for the data X_m at node m. Each decision tree was obtained by recursively partitioning the feature space using the previously defined entropy loss function until the constraints defined by the hyperparameters (e.g., tree depth) were reached. The RF was obtained by initializing N decision trees with a split composed of a random set of features and random training samples. The aggregation of several DTs in a RF resulted in a more robust prediction compared to single DTs.
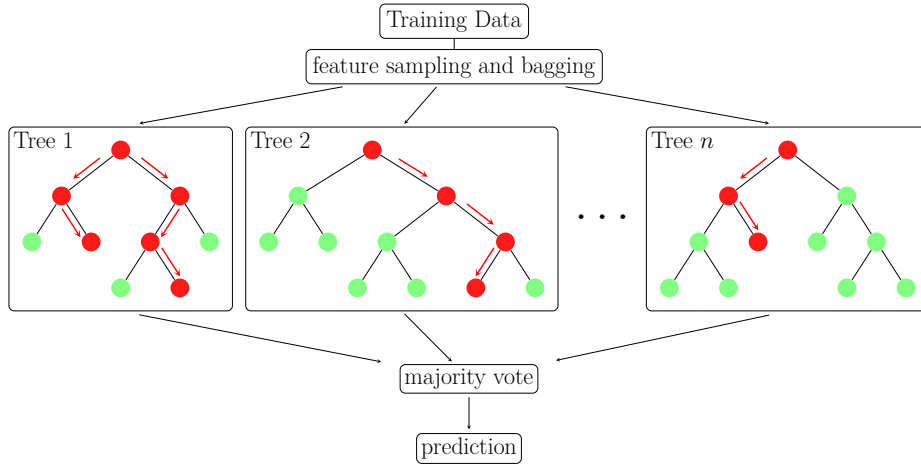
Figure 7.7: Conceptual illustration of a Random Forest composed of n trees for a two class classification problem. Cross-Validated Grid Search was used to determine the optimal hyperparameters of the RF as having a minimum number of samples per split $n_{split} = 20$, a minimum number of samples per leaf $n_{leaf} = 10$, a maximum depth $n_{depth} = 10$, and a number of estimators $n = 100$.

An alternative tool for automated classification lies in the adoption of Bayesian statistical models. In this work, we proposed a Bayesian Logistic Regression (BLR) model. Given our set of features, a logistic regression modeled the probability of the weld being damaged as

$$P_{BLR}(y|\alpha, \beta_{speed}, \beta_1, \ldots, \beta_{15}) = \sigma(\alpha + \beta_{speed}x_{speed} + \sum_{i=1}^{15} \beta_i x_i) \tag{7.8}$$

where $\sigma(t) = \frac{1}{1+e^{-t}}$ is the logistic function, x_{speed} and x_i form our set of predictor variables, and α, β_{speed} and β_i are the parameters of the model to be estimated for the linear transformation of the feature vector (see [35] for an in-depth discussion of the model). The set of essential

indicators x_i and x_{speed} corresponding to the BLR parameters β_i and β_{speed} are summarized in Table 7.4. In order to determine the parameters of the model, for which a closed form solution is not generally available, a maximum likelihood approach can be used. However, this comes at the cost of certain drawbacks. First, a maximum likelihood approach is prone to over-fitting. Second, a set of possible solutions is generally available, but this approach determines a single solution, which highly depends on the adopted optimization algorithm. Furthermore, the assigned labels are often noisy, as is typical in real-world measurements, and as a consequence of the aforementioned subjectivity of the expert assessment. A maximum likelihood approach cannot, however, provide an indication of inherent uncertainty. For these reasons, in this work we adopted a Bayesian estimation of the logistic regression.

Table 7.4: Summary of the 15 indicators with the highest $F1$-score and a cross correlation of under 80% that were input to the BLR and RF models.

Index	Feature
x_1	$\max(\max_{2m}(ABA_Z^{STFT\ 2800Hz}))$
x_2	$\max(\max_{2m}(ABA_Z^{STFT\ 5800Hz}))$
x_3	$\max(\max_{2m}(ABA_Z^{STFT\ 3000Hz}))$
x_4	$\mu(\max_{2m}(ABA_Y^{STFT\ 1400Hz}))$
x_5	$\max(\max_{2m}(ABA_Z^{STFT\ 5400Hz}))$
x_6	$\max(\max_{2m}(ABA_Z^{STFT\ 400Hz}))$
x_7	$\max(\max_{0.625s}(ABA_Z^{DWT\ Haar\ cD5}))$
x_8	$\mu(\max_{3m}(ABA_Y^{BP\ 0.2-0.5kHz}))$
x_9	$\mu(\max_{3m}(ABA_Z^{BP\ 0.1-11kHz}))$
x_{10}	$\max(\max_{0.625s}(ABA_Y^{DWT\ Haar\ cD6}))$
x_{11}	$\max(\max_{2m}(ABA_Z^{STFT\ 1400Hz}))$
x_{12}	$\max(\max_{2m}(ABA_Y^{STFT\ 1400Hz}))$
x_{13}	$\max(\max_{2m}(ABA_Z^{STFT\ 800Hz}))$
x_{14}	$\min(\min_{3m}(ABA_Z^{Long.\ level\ D0}))$
x_{15}	$\max(\max_{3m}(ABA_Z^{Long.\ level\ D0}))$
x_{speed}	vehicle velocity

A BLR model [35] solves the aforementioned issues in the following way: (i) reducing the risk of over-fitting thanks to the regularization of the priors; (ii) producing a distribution of possible model solutions under the model assumptions (i.e., the priors); (iii) providing a more reliable indication of the predictive uncertainty. Again, exact Bayesian inference of the logistic regression is intractable and approximate methods are generally used. We, here, estimated the BLR model by Markov Chain Monte Carlo (MCMC) sampling, exploiting the No-U-Turn Sampler (NUTS) algorithm [111]. The BLR model was implemented with the probabilistic pro-gramming Python package PyMC4, which allows for flexible specification of Bayesian statistical models [241].The model parameters were assigned a Gaussian prior $\mathcal{N}(0, 1)$, while the labels were modeled through a Bernoulli likelihood. Four chains were used in the MCMC inference, with 2000 sample draws and 10,000 tuning samples per chain. A draw refers to a collected sample generated from the posterior distribution of the MCMC inference, while tuning samples are generated before starting to collect posterior samples and are used to tune the sampling algorithm by adjusting the step size of the updated distribution, as well as to ensure the con-vergence of the chains. The graphical model of the implemented BLR is displayed in Figure 7.8.
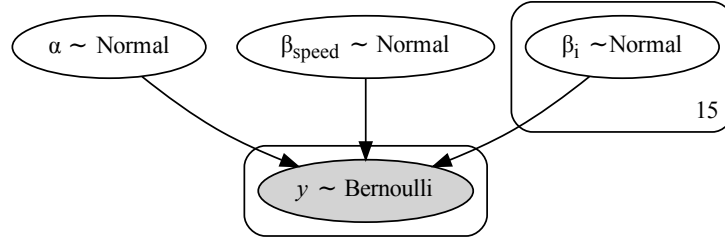
Figure 7.8: Bayesian Logistic Regression model, where the coefficients α, β_{speed}, β_i were assigned a normally distributed prior. The prediction y followed a Bernoulli distribution.

An expert-in-the-loop approach requires the expert feedback to adjust the anomaly detection scheme, such that the outlier detection classifier or decision threshold are more in tune with the expert's understanding of anomalies. Here, the feedback was incorporated using the labels generated by the expert assessment during the training stage. The process used for applying the expert labels in the classification process is illustrated in Figure 7.1, where the expert labels were essentially used to improve the model performance in the supervised classification framework.

7.4 Results and Discussion

In this section, we elaborate on the results from the expert-based evaluation of outlier welds and their influence on classification of their condition. We compared performance on identification of faulty welds for the three schemes outlined above, namely the BC, RF and BLR model.

7.4.1 Expert-Based Evaluation of Outlier Welds

Definition of Capacity-Based Thresholds

Semi-supervised approaches require the definition of suitable outlier metrics. We specified these, here, on the basis of the two main defect types that are encountered for welds, namely geometric defects or surface defects/squats. A third category of defects can be attributed to internal effects, such as cracks, which are, however, not visible, and would not be possible to assess through expert visual inspection. These may, however, be labeled through non-destructive evaluation, which is logged to the ZMON database, as explained in Section 7.4.1. In this subsection, we restricted evaluation tp the visual inspection of experts, for which the first two defect instances were relevant. Geometric defects are linked to decreased longitudinal level values ($D0_{Z,min}$), which point to a degraded weld geometry [72]. On the other hand, surface defects are linked to peaks in acceleration and energy values [189]. Therefore, we used the maximum vertical acceleration ($ABA_{Z,max}$) and the longitudinal level ($D0_{Z,min}$) as the main metrics for selection of outliers.

In order to define thresholds for outlier selection, we employed EVA, as described in Section 7.3.3. To this end, we fitted an empirical distribution to the aggregated values of $D0_{Z,min}$ and $ABA_{Z,max}$, collected on records from all available weld samples. Two thresholds were defined, associated with the 98-th and 95-th percentiles of the fitted EDs, corresponding to strong and weak outliers, respectively. The choice of percentile for the strong outlier case was carried out so as to include instances of weld defects that were discovered in the field through visual inspection, but which had not been picked up by the automated image-based detection system of the diagnostic vehicle ($VCUBE$), which was considered to be a rare incident. Strong outliers were defined only on the basis of maximum vertical ABAs, as we suspected this indicator to be more directly related to the weld defects. The weak outlier definition combined information

from both the maximum vertical ABA and the longitudinal level $D0$, as we suspected that the longitudinal level plays a role, albeit secondary, in the degradation process of welds. The definition of the sets of strong S_s and weak S_w outliers, given an observation k, was formulated as follows:

$$S_s = \left\{ k \middle| O_s = \frac{ABA_{Z,max,k}}{q_{98\%}(ABA_{Z,max})} > 1 \right\} \tag{7.9}$$

$$S_w = \left\{ k \middle| O_w = \left(\frac{D0_{Z,min,k}^2}{q_{95\%}(D0_{Z,min})^2} + \frac{ABA_{Z,max,k}^2}{q_{95\%}(ABA_{Z,max})^2} \right)^{0.5} > 1 \quad \& \right.$$
$$\left. O_s = \frac{ABA_{Z,max,k}}{q_{98\%}(ABA_{Z,max})} < 1 \right\} \tag{7.10}$$

where $ABA_{Z,max,k}$ is the maximum vertical ABA for the k-th observation and $D0_{Z,min,k}$ is the minimum longitudinal level in proximity of the weld for observation k.

An amount of 100 outliers per region and per trimester (evaluation round) was deemed as realistic, to be checked by the assigned experts. The described process resulted in a total of 195 strong outliers, during the first expert evaluation round. However, the evaluation of weak outliers resulted in more samples than could feasibly be evaluated by the experts. Thus, a random selection amongst the weak outlier set was carried out to reach a total of 100 strong and weak outliers per region and per evaluation round (trimester).

Figure 7.9 illustrates the distributions of the selected outlier metrics, namely, the maximum vertical acceleration ($ABA_{Z,max}$) and the minimum longitudinal level ($D0_{Z,min}$) values. Furthermore, the defined outlier regions are highlighted in Figure 7.9. For most welds, it was observed that the peak ABA lay under 100 m/s^2 and that the minimal longitudinal level was lower than 0.3 mm in absolute terms. The defined outlier metrics were used in the next section to deliver samples for expert assessment.
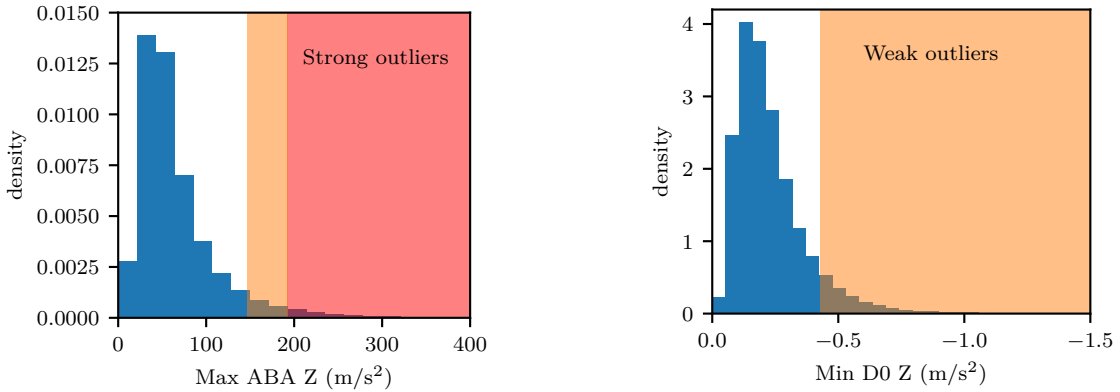
(a) Histogram of the maximum vertical ABA ($ABA_{Z,max}$).

(b) Histogram of the minimum longitudinal level ($D0_{Z,min}$).

Figure 7.9: Distribution of the maximum vertical ABA ($ABA_{Z,max}$) and the minimum longitudinal level ($D0_{Z,min}$). The cutoff thresholds were defined on the basis of the 98-th percentile of the empirical distribution of $ABA_{Z,max}$ for the strong outlier region (highlighted in red), per Equation (7.9), and on the basis of the 95-th percentiles of the minimum longitudinal level $D0_{Z,min}$ and $ABA_{Z,max}$, as formulated in Equation (7.10), for the weak outlier region (highlighted in orange), respectively.

Expert Assessment

In the proposed framework, the outliers defined from the EVA analysis in Section 7.4.1, were forwarded to the experts for cross-check and labeling. The process illustrated in Figure 7.1 was conducted with actual rail monitoring data, as part of a PoC project, enabling a feedback loop between the experts, asset managers and researchers. Outlier welds, as defined in Section 7.4.1, were labeled by experts on the basis of the black & white image feedback offered by rail-head images acquired from the diagnostic vehicle and the $VCUBE$ system. Over the period of one year, four feedback loops were performed, during which a total of 1727 outlier welds were delivered to the experts for evaluation. The thresholds defined prior to the first iteration [110] were kept identical to track the expert evaluations on critical welds over time. The welds submitted for evaluation over all feedback cycles in this study were composed of 911 samples that featured strong outlier scores, and a selection of 816 weak outlier samples. The weak outlier selection corresponded to random samples selected in order to achieve 100 samples per region and per expert evaluation round. The condition of the welds was then assessed based on the criteria defined in the deviation catalog of SBB [244]. Deviations in this catalog are for instance welds featuring a squat, a surface defect or faults in the geometry.

A two stage identification process was carried out. In the first step, the experts were asked to identify if the outlier corresponded to a weld. After four evaluation rounds, 132 samples were not evaluated due to the time-constrains of the inspectors. In 113 cases, the image-based system recognized other faults as welds and for 1491 samples the experts recognized a weld in the image, delivering important information on the performance of the rail-head image-based weld detection (see also Figure 7.11). In the second step, the experts visually assessed the condition of the welds using the $VCUBE$ images. The results of the expert evaluation, upon completion of the four feedback loops, are summarized in Table 7.5. Around 12% of the strong outliers and 6% of the weak outliers were labeled as defective by the in-office experts.

Table 7.5: Number of welds in the database after a monitoring period of one year for the selected portion of tracks on the SBB network.

Condition Label Source	No Defect EL-EVA [1]	Defect EL-EVA [1]	Defect ZMON [2]	No Defect Unlabeled Condition [3]
Low Outliers	0	0	368	19668
Weak Outlier	656	42	141	2242
Strong Outlier	710	83	54	285

[1] EL-EVA stands for Expert Labels from the Extreme Value Analysis. [2] Defect welds from the condition monitoring database ZMON originate from the standard manual and automated track inspection process, or from the ultrasonic assessment of welds. [3] The samples with unlabeled condition correspond to welds that were not submitted for expert evaluation or assigned an existing defect within ZMON and are, here, assumed to be healthy.

Figure 7.10 illustrates the percentage of defect welds versus the values of the corresponding scores for strong and weak outlier sets defined in Equations (7.9) and (7.10). The vertical dashed line indicates the threshold between strong and weak outlier regions. The dotted horizontal line offers a visual indicator for regions where the number of defects in the delivered sample was higher than 10%. The green markers indicate the welds that were evaluated as healthy on the basis of expert visual assessment, while the red markers show the defective instances. It is evident that, under increasing outlier scores, the ratio of welds that were assigned a defective label versus the complete set of ABA-defined outliers rose from 10% to 22%. At higher outlier scores, a variability was noted in terms of the outlier score due to the fewer remaining samples, resulting in decreased statistical significance. It needs to be here noted, however, that ABA

information can deliver defects in their initiation or formation, which may not be deemed as faults via the expert visual inspection. Heavier weld faults are most commonly attributed to the presence of squats or another distortion in the welding, rather than tied to geometric faults. The outlier scores were, here, formulated so that samples in the weak outlier set included the instances that corresponded to anomalies in the geometry. It should here be noted that, as the evaluation criteria were mainly visually-based, the expert evaluation could not thoroughly capture geometric irregularities, which were smoother (influencing larger wavelengths than a local squat). In addition, the evaluation was affected by inspector bias (see also Figure 7.14). The challenges resulting from the uncertainty in the ground truth are further discussed in Section 7.4.2.
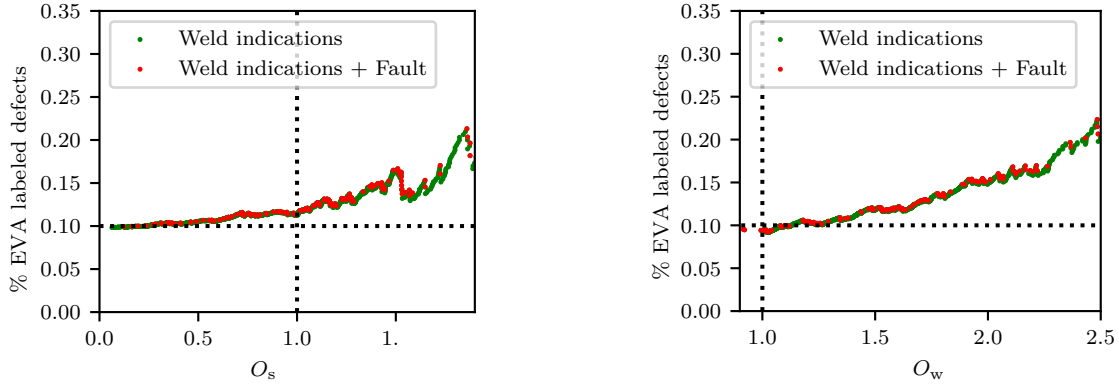
(a) Faulty weld percentage for strong outliers.   (b) Faulty weld percentage for weak outliers.

Figure 7.10: Percentage of damaged welds in the samples evaluated by experts after four evaluation rounds, given outlier scores that were higher than the prescribed thresholds. At higher outlier scores, 22% of the ABA-based outliers were assigned a faulty label by the experts.

Fusion of Data from the Standard Condition Monitoring Database (ZMON)

As part of the standard evaluation procedure, welds that are inspected and deemed as defective by experts are recorded in the ZMON (ZustandsMONitoring—condition monitoring) database. This is a condition logging database containing faults stemming from the following: (i) visual inspections; (ii) more specialized Non-Destructive Evaluations, such as ultrasonic inspections conducted by a system mounted on a dedicated diagnostic vehicle; (iii) automated track inspections of rail faults by means of the $VCUBE$ system, mounted on a specialized diagnostic vehicle. Faulty welds that are picked up by the ultrasonic inspection vehicle are verified by on-site measurements using a handheld device. Furthermore, faults that are picked up by both the automated track inspection system ($VCUBE$) and visual inspection need to be first validated by experts in the office, prior to being added to ZMON. This database serves for efficient planning of maintenance and renewal actions.

During the expert evaluation process that was executed as part of this PoC, 1491 outlier welds were evaluated, and, when classified as faulty, added to ZMON. Due to capacity limitations of the experts, only the majority of strong outliers and a selection of weak outliers were given for evaluation. However, the majority of the 25,000 welds remained unlabeled. To assess the performance of ABA-based classification, exploited here for the first time, faulty welds which were not offered for evaluation were identified from the ZMON condition log using the process illustrated in Figure 7.11. Through this process, the defect dataset was extended to include non-visual inspection sources, such as ultrasonic testing [90], and was not solely composed of ABA-defined outliers. The inclusion of these samples was crucial as the subsequent

classification procedure relied on data which comprised all observed cases of welds, regardless of their ABA status.
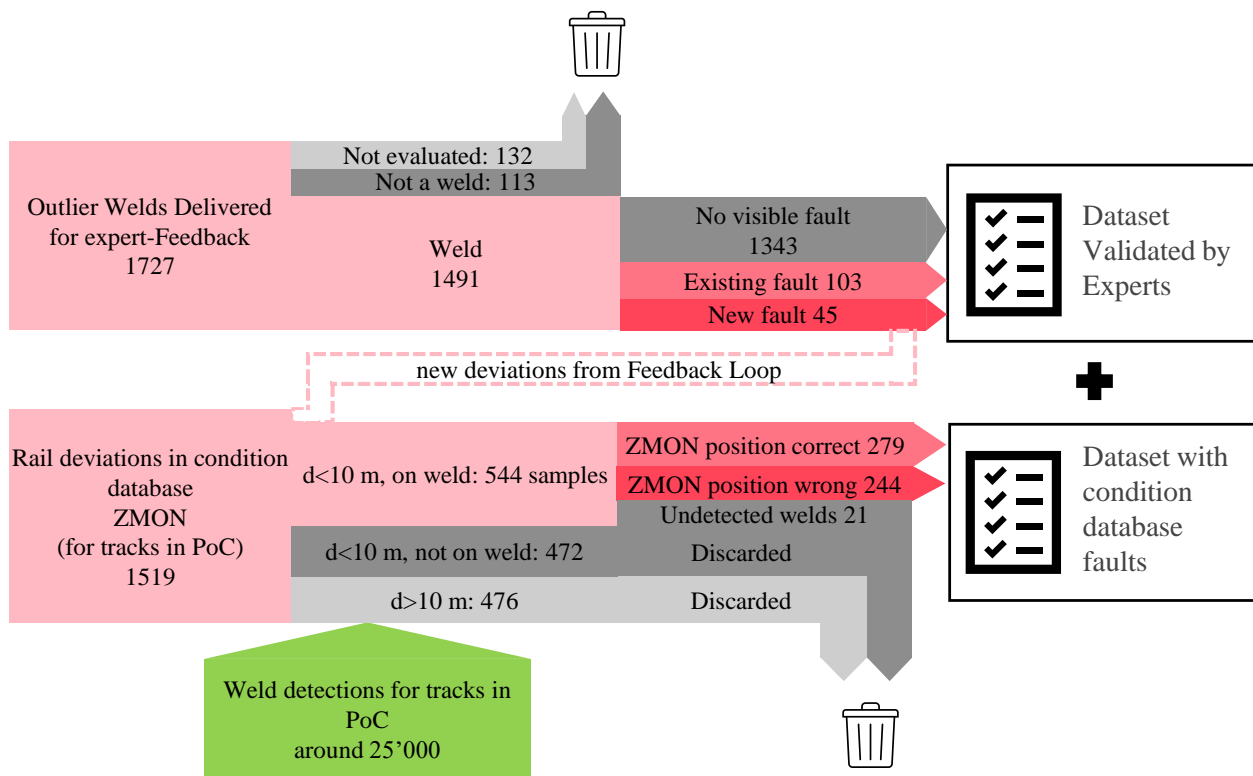
Figure 7.11: Process for generating weld condition labels from the ZMON condition database on the basis of expert feedback conducted on the ABA-derived outliers, as well as through the standard inspection and evaluation processes, which include visual (on-site), $VCUBE$-based, and ultrasonic-based track inspections.

The process of linking rail defects, of a general nature, to weld-specific defects was initiated by extracting all rail faults from the ZMON database for the portions of track evaluated as part of this PoC. These rail defects were then associated to the presence of a weld in their vicinity, allowing for a tolerance of 10 m to account for potential inaccuracy in the reported positions. The resulting matches were then visually verified by the authoring team using the $VCUBE$ image system in order to minimize occurrence of wrong matches due to positioning uncertainties. This process resulted in labels for approximately 544 damaged welds out of the original set of 25,000 labels. In the absence of a label, a weld was assumed to be healthy for the classification process in Section 7.4.2, which was naturally a strong assumption, since certain welds may simply have failed to be picked up as defective.

7.4.2 Classification of Weld Condition

To further improve the aforementioned thresholds for expert feedback, the expert knowledge stemming from EVA and the ZMON logs was taken into account by considering three types of models.

The Bayesian Logistic Regression (BLR) model was compared against a Binary Choice (BC) approach, and the Random Forest (RF) model. As an input, the supervised classification process (see also Figure 7.1) used the condition labels of ZMON, enhanced with the expert evaluation, shown in Table 7.5, to provide superior classification thresholds on the basis of the ABA-derived indicators. Here, it is worth remembering that the longitudinal level indicator is itself an ABA-derived indicator through the process of integration.

The ABA features of Table 7.3 were computed for all samples in the dataset presented in Section 7.4.1. The naming convention of the essential indicators was $T1(T2(ABA_C^M))$, with the related letter entries explained in Table 7.6.

Table 7.6: Meaning of the $T1(T2(ABA_C^M))$ naming convention of the essential indicators.

Letter	Explanation	Possible Entries
$T1$	Summary statistic computed over all (Y/Z) sensor channels	mean μ, standard deviation σ, min, max...
$T2$	Summary statistic computed over time series where the subscript denotes the window size	mean μ, standard deviation σ, min, max...
M	Applied method and parameters	STFT, Longitudinal Level $D0$, DWT...
C	Sensor channel	Z for vertical, Y for lateral

The dataset was divided into a training dataset and a test dataset, with 80% of the samples included in the former and the remaining 20% included in the latter. The models were trained on the processed essential indicators with the assigned condition labels. In some studies, data imbalance and overfitting is prevented by using data augmentation techniques; however, generating such augmented datasets requires special care since the augmented datasets are prone to bias [177]. For this reason, resampling techniques were used instead of generative approaches to account for the large imbalance between the number of healthy welds compared to defective instances. Certain models, such as the RF-based scheme, require the adjustment of the class weights to account for the over-represented category during the training procedure [54]. This was, here, achieved by weighing the minority class proportionally to its count in order to achieve balanced weights

$$w_j = \frac{n}{c \cdot n_j} \tag{7.11}$$

where the weight w_j for class j is weighed inversely proportionally to the ratio between the number of samples n_j of class j and the total number of samples n. For the weld assessment problem, we assigned a binary condition label ("healthy" or "defective") and, thus, the number of classes c is 2. This balancing was only applied during the training phase of the RF.

The model performance assessment metrics used here were accuracy, recall, precision, $F1$-score, and the Area Under the Receiver Operating Characteristic Curve (ROC–AUC). A brief description of these metrics is provided in [118]. The accuracy metric measures the ratio of correct predictions over the total number of evaluated instances. For imbalanced datasets, other metrics are generally preferred, as accuracy is sensitive to imbalance. The precision is defined as the ratio of the correctly classified positives (true defects), also referred to as true postives (TP), versus all classified positive instances (predicted defects), either correctly (TP) or incorrectly (false positives, FP). A low precision score indicates the presence of a high number of false positives, which can be an outcome of imbalanced class or untuned model hyperparameters. The recall, also referred to as the true positive rate (TPR), is calculated as the ratio between the number of correctly classified positive samples (TP) versus the total number of actually positive samples, which includes true positives (TP) and false negatives (FN). Both precision and recall offer metrics on the classification reliability in terms of predicting positives. Low recall rates result in lower safety due to an increased number of false negatives. The $F1$-score is defined as the harmonic mean of precision and recall, and is calculated using the following

equation

$$F1\text{-score} = \frac{2(\text{precision} \cdot \text{recall})}{\text{precision} + \text{recall}} \tag{7.12}$$

The $F1$-score is adopted for assessing models with large class imbalance, as it assigns equal weight to both precision and recall. The Area Under the Receiver Operating Characteristic Curve (ROC–AUC) is another metric which is commonly used to assess the performance of binary classifiers, as it assesses the quality of the distinction between positive and negative classes. The ROC–AUC is calculated using the following equation

$$\text{ROC-AUC} = \int (\text{TPR}(\text{FPR})) d\text{FPR} \tag{7.13}$$

where TPR stands for the true positive rate (or recall) and FPR denotes the false positive rate, defined as $FPR = FP/(TN + FP)$. The ROC–AUC score can be misleading when the dataset is highly imbalanced and, thus, it is best used together with the precision, recall and $F1$-score for the assessment of the model performance.

The model selection proposed in Section 7.3.4 was now trained on the training dataset, which was composed of 80% of the samples of Table 7.5. One should note that the unlabeled condition and the non-defective samples from EVA were here both assumed to be healthy samples. In practice, as previously noted, there is uncertainty both in the expert labeling, as well as regarding the completeness and up-to-dateness of the ZMON database.

Table 7.7 illustrates the classification scores for the formulated models. The BC models offered lower scores than the BLR and RF classifiers, as their univariate nature did not enable them to capture more complex relations. The BLR model offered scores that were nearly as high as the RF classifier, capturing the linear relations of the indicators. The RF classifier offered the highest scores as this model can capture the non-linear relations between indicators. The BC classifier was used to evaluate the single features listed in Table 7.3, such as the maximum lateral acceleration Y measured over a distance of 3m around the weld, averaged across all measurement channels $\mu(\max_{3m}(ABA_Y^{RAW}))$. Table 7.7 furthermore shows that for the univariate BC models, lateral acceleration features performed better, while vertical acceleration features had slightly lower performance. The EVA for outlier detection could have, in hindsight, been performed on such an improved indicator. However, because organizing feedback rounds with many experts and asset managers from different regions was a complex task, expert assessment on the updated indicators will only be performed in future work. In addition, it was observed that increased ABA values in the 0.5 to 2 kHz range, for both the vertical and lateral direction, could indicate the presence of a defect. The single indicator BC models yielded $F1$-scores between 10% and 32% for the best performing single features. From the BC models, we observed that both vertical and lateral acceleration features yielded similar $F1$-scores.

The BLR and RF models were trained using the 15 features yielding the highest $F1$-score in the univariate BC models and a cross-feature correlation of less than 0.8. Figure 7.12 illustrates the inferred posterior distributions for a selection of 6 parameters of the BLR model, inferred by MCMC inference with the samples collected after convergence of the chains. The essential indicators x_i corresponding to the BLR parameters β_i are summarized in Table 7.4. It was observed that certain inferred parameter distributions, such as β_1 and β_{14}, assigned high posterior probability to 0, suggesting low statistical significance. For example, β_{14} corresponded to one of the longitudinal level $D0$ indicators, which further confirmed the limited value in geometric indications for welds, which are, indeed, nowadays mainly assessed visually and not in terms of geometry. Likewise, the other displayed variables considerably differed from the prior, and the 0 value was not contained in the 94% Highest Density Interval (HDI) [182], which supported their statistical significance.
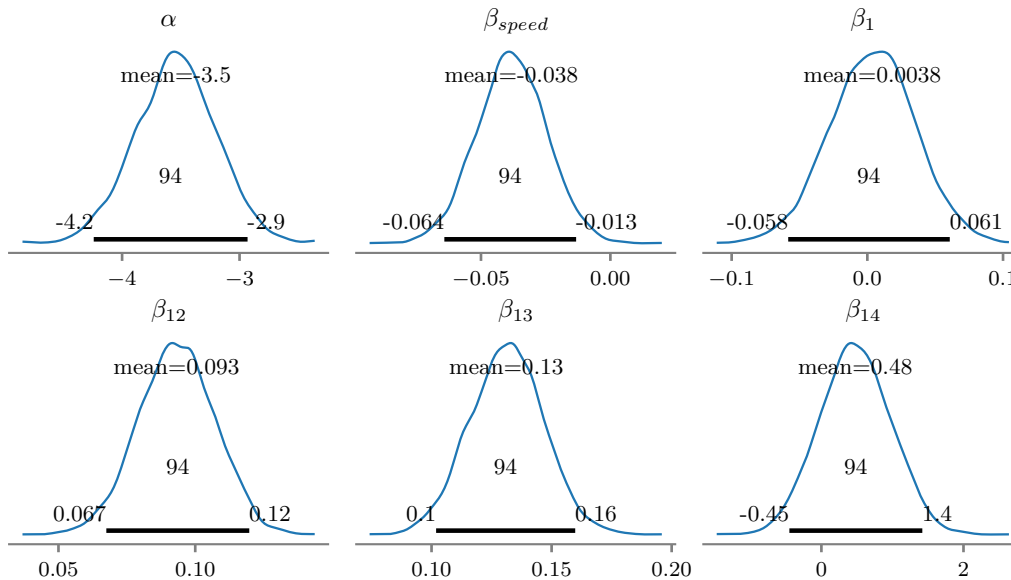
Figure 7.12: Posterior distributions of the BLR parameters α, β_{speed}, β_1, β_{12}, β_{13}, β_{14}. Mean and 94% HDI are reported in the plots. The essential indicators x_i corresponding to the BLR parameters β_i are listed in Table 7.4.

The best BLR model yielded an $F1$-score of around 43 %, while the best-performing RF model delivered an $F1$-score of around 48 %. The models based on multiple indicators performed better than the ones using only a single feature. By further including the speed, which is an important operational parameter, as an input to the RF and the BLR models, the classification metrics were further improved. While the BLR delivered slightly lower metrics than the RF, it presented a reduced risk of over-fitting. Moreover, the BLR model additionally provided estimates of the predictive uncertainty.

For all models, a trade-off always existed between recall and precision. When the recall increased, the precision decreased, resulting in a safer monitoring scheme, which, however, also delivered many false positives. To further illustrate this point, the confusion matrices for the best performing BLR and RF classifiers are shown in Figure 7.13a,b, respectively. While only about 2–3% false positive predictions occurred, the number of false negatives was up to 56%. One can observe in Figure 7.13 that the BLR and RF models had similar recall rates, but different precision, as the RF offered slightly less false positives. The decision threshold was set to maximize the $F1$-score. In practice, this decision threshold can be dynamically adjusted. For example, more ambiguous samples, such as weak outliers, could be considered for expert validation, but this would be associated with a higher false positive rate and a greater validation effort. Concerning the BLR model, the expert-in-the-loop approach could have been further extended to exploit the estimated predictive uncertainty, which offers a quantification of the confidence we may attribute to each judgment [84, 135]. The predictive uncertainty could then be exploited for a second iteration of expert labeling on those samples associated with high uncertainty. The resulting reduced uncertainty in the labels would lead to subsequent improvement in the predictive models. While this extension was not considered in this PoC, for simplicity and to present a fair model comparison, it would certainly be useful to consider this for a practical real-world implementation, in order to fully exploit the BLR model capabilities.

We noticed that the best performing RF model sometimes mistakenly classified healthy samples as defective and vice versa. Indeed, we noticed that the relatively low precision and recall could be attributed to the high uncertainty in the labels. Further examining the instances of correct and misclassified defects, we offer an illustration of some randomly selected instances for true positive (TP), false positive (FP), true negative (TN), and false negative (FN) instances

Table 7.7: Defect weld classification scores on the ZMON test dataset for the best performing features of each feature type of the Binary Choice (BC), as well as the Random Forest (RF) and Bayesian Logistic Regression (BLR) models that yielded the best $F1$-scores (see also the feature naming convention in Table 7.6).

Model	Features	F1	Auc[1]	Acc[2]	Rec[3]	Prec[4]
BC	$\mu(\min_{3m}(ABA_Z^{Long.\ level\ D0}))$	0.109	0.553	0.290	0.846	0.058
BC	$\mu(\max_{3m}(ABA_Z^{Long.\ level\ D0}))$	0.111	0.555	0.371	0.761	0.060
BC	$\min(\min_{3m}(ABA_Z^{Long.\ level\ D0}))$	0.116	0.542	0.799	0.256	0.075
BC	$\max(\max_{3m}(ABA_Z^{Long.\ level\ D0}))$	0.119	0.543	0.822	0.233	0.079
BC	$\mu(\sigma_{3m}(ABA_Y^{RAW}))$	0.250	0.610	0.921	0.265	0.236
BC	$\max(\max_{0.625s}(ABA_Y^{DWT\ Haar\ cD2}))$	0.259	0.647	0.893	0.373	0.198
BC	$\mu(\max_{0.625s}(ABA_Y^{DWT\ Haar\ cD3}))$	0.263	0.634	0.908	0.329	0.219
BC	$\mu(\max_{0.625s}(ABA_Y^{DWT\ Haar\ cD6}))$	0.264	0.645	0.898	0.364	0.207
BC	$\mu(\max_{0.625s}(ABA_Y^{DWT\ Haar\ cD2}))$	0.270	0.635	0.912	0.327	0.230
BC	$max(\max_{3m}(ABA_Y^{RAW}))$	0.293	0.668	0.902	0.408	0.229
BC	$\mu(\max_{3m}(ABA_Y^{RAW}))$	0.307	0.641	0.927	0.322	0.293
BC	$\max(\max_{2m}(ABA_Y^{STFT\ 400Hz}))$	0.307	0.654	0.919	0.360	0.268
BC	$\mu(\max_{3m}(ABA_Y^{BP\ 0.8-2kHz}))$	0.308	0.638	0.930	0.313	0.303
BC	$\mu(\max_{2m}(ABA_{YZ}^{VS\ ABA}))$	0.315	0.650	0.925	0.344	0.291
BC	$\max(\max_{2m}(ABA_Z^{STFT\ 800Hz}))$	0.323	0.655	0.926	0.355	0.295
BLR	15 indicators with highest $F1$-score [5] and cross-feature correlation under 0.8	0.422	0.696	0.937	0.426	0.417
BLR	15 indicators with highest $F1$-score [5] and cross-feature correlation under 0.8 & speed	0.431	0.701	0.938	0.432	0.427
RF	15 indicators with highest $F1$-score [5] and cross-feature correlation under 0.8	0.479	0.711	0.948	0.446	0.517
RF	15 indicators with highest $F1$-score [5] and cross-feature correlation under 0.8 & speed	0.486	0.708	0.950	0.436	0.550

[1] Roc-Auc, [2] Accuracy, [3] Recall, [4] Precision.

[5] The $F1$-scores of the indicators were defined with the univariate BC model.

in Figure 7.14. It is evident that these uncertainties were largely related to challenges in the quality of the dataset labeling. Many samples exhibiting acceleration and visual characteristics that would place them, respectively, in the healthy or defective category might effectively belong to a different condition category than the initially assumed one. Several reasons for these mislabeled samples exist:

1. The assumption that all welds not linked to a fault (from the ZMON database or the PoC expert-based evaluation) are healthy is inaccurate. Figure 7.14b illustrates that some samples assumed as healthy, but classified as defective (categorized as False positives), were indeed defective as they were not recorded in the ZMON database at the time of the inspection.

2. Additionally, errors can occur when matching the potentially outdated time and inaccurate location of defects recorded in the ZMON database with the weld observations. This can result in false negatives, where one can visually observe faults, as shown in Figure 7.14d.
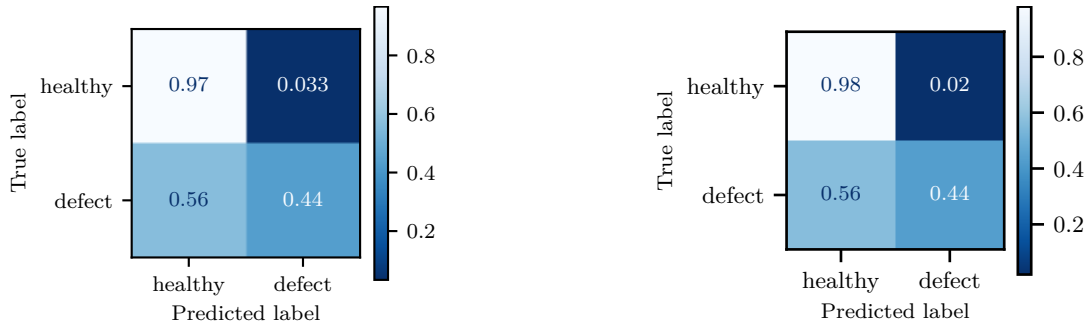
(a) Confusion matrix of the best performing BLR model.

(b) Confusion matrix of the best performing RF model.

Figure 7.13: Confusion matrices for the BLR and RF classification models. The recall rate was around 44% and only 2% of the samples were mislabeled by the classifiers for the healthy scenario.

3. Finally, additional uncertainty arises from expert judgment, which tends to vary considerably as each expert can reach a differing conclusion on the same sample.

Furthermore, experts often only generate new faults in the ZMON database for advanced damage conditions (see also Figure 7.14c), where the resulting maintenance is carried out within a prescribed time horizon. While this is a resource-efficient strategy, it is crucial to realize that accelerations are quite sensitive to moderate-intensity defects that lead to early alarms, as is shown in the next section. A binary classification is less suitable in such a scenario because it does not distinguish between different damage levels; this implies that damage severity and its progression need to be taken into account, which is only feasible under continuous tracking over time.

7.4.3 Continuous Tracking of Health Condition

The ABA-based classification of weld condition in Section 7.4.2 was performed on the basis of individual observations on welds. This work advocates adoption of an OBM paradigm, whose purpose is to deliver regular data collection from either diagnostic vehicles or appropriately instrumented in-service trains. In this way, tracking of the condition and the possible evolution of damage can be accomplished. By combining consecutive ABA measurements over time, the evolution of the rail infrastructure condition can be better estimated [317]. In this section, patterns and trends were identified with respect to the evolution of welds prior to reaching a damaged state, by analyzing the indicators derived from ABA over time. This information could lead to early detection of damage, and, in this way, facilitate the scheduling of timely maintenance activities and the improvement of rail infrastructure reliability.

Figure 7.15 plots the evolution of the ABA-derived indicator observed on a weld with a squat over time from October, 2020, until November, 2022. The indicator $\mu(\max_{2m}(ABA_{YZ}^{VS\ ABA}))$ that was selected for this comparison was the one returning the highest $F1$-score for the BC model (see also Table 7.7), while simultaneously combining both vertical and lateral acceleration information. The ABA feature was normalized to the decision threshold γ defined in Equation (7.5). The normalized ABA feature grew linearly until November, 2021, for the weld of this case study. During the subsequent measurement in April, 2022, the ABA indicator decreased, due to rail grinding maintenance, which took place in March, 2022. The two thresholds highlighted in Figure 7.15 corresponded to the 98-th percentile which was applied during the initial expert evaluation round, and to the limit value for maximizing the $F1$-score during the classification process respectively. The first introduction of the weld of Figure 7.15 in the
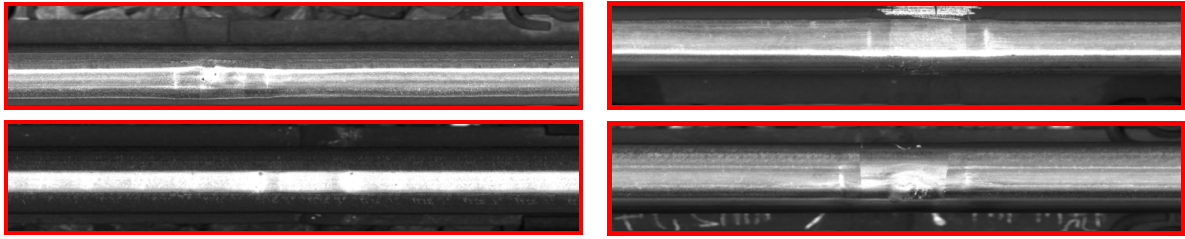
(a) Welds with squat classified by the RF model as faulty
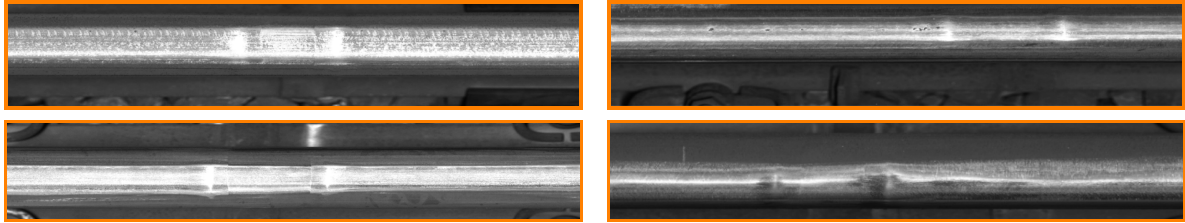
(b) "Healthy" welds classified by RF model as defect
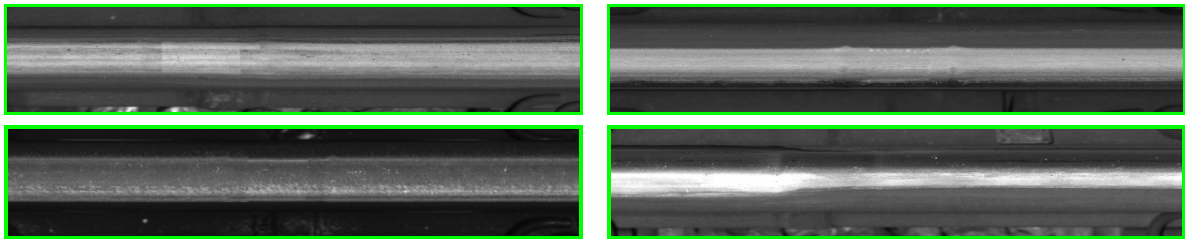
(c) Healthy welds classified by the RF model as healthy
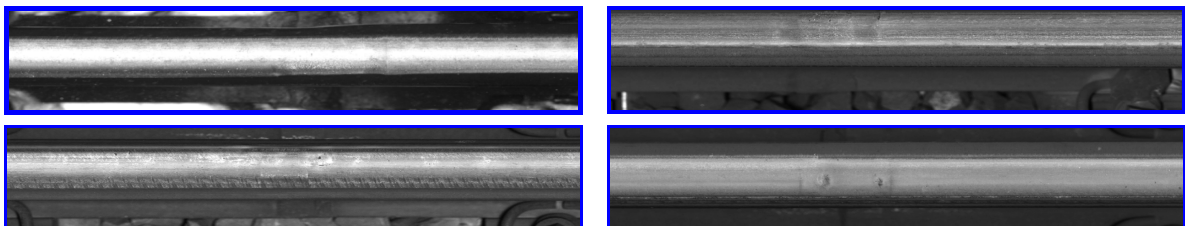
(d) "Faulty" welds classified by RF model as healthy

Figure 7.14: Challenges in the labeling quality of the dataset are showcased with images extracted from $VCUBE$ of characteristic healthy and defective weld component classes for each classification scenario.

ZMON database occurted over a year after the highlighted thresholds were crossed showing the significant potential of using ABA as an early indicator of weld condition. Once the defect was identified, the weld was replaced within months, resulting in a recovery of low indicator values. In conclusion, the ABA-based indicator was able to track the true evolution of condition and deliver an indicator of real time damage progression.

When considering larger scale infrastructure, it may often be more efficient to consider the assets over space and time, as a large number of welds exist on the network and tracking each one individually is resource intensive. The indicator of Figure 7.15 can be computed at any position on the track and visualized as time series plots, or as heatmaps. Asset managers are nowadays trained on interpreting heatmaps to analyze the variation of Track Quality Indicators (TQIs) over large track sections over time [200]. Figure 7.16a,b illustrates the heatmap for two rail sections of 400 m, where the rail condition indicator was normalized to the limit value for decision making defined in Section 7.4.2. The color scale of the heatmap reflects the increasing

129
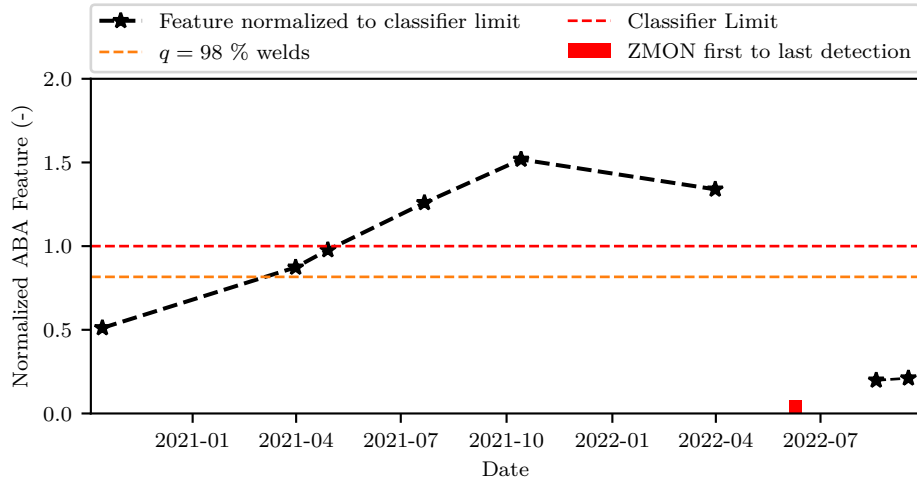
Figure 7.15: Time history of best performing feature of Section 7.4.2 normalized to the decision limit of the classifier. The ABA feature grew linearly until November, 2021. In April, 2022, rail grinding maintenance occurred on the track section, resulting in slightly lower ABA. The weld was labeled as faulty in June, 2022, by the experts and, subsequently, replaced in August, 2022.

level of rail damage. The unvalidated welds, surface defects and insulated joints automatically detected by the $VCUBE$ image-based detection are shown at the bottom of the plot. Increased indicator values are most commonly caused by insulated joints, faulty welds, surface defects or squats [104]. This is supported by Figure 7.16, as the increased indicator values lined up with these infrastructural elements. The start point/time and end point/time of the faults, recorded in the ZMON database, are shown as two stars linked by a dotted line, had lower space and time resolution, as they could be driven by other considerations. Figure 7.16a shows a section with two defect welds, including the one having the time series of the weld at position 542 m from Figure 7.15, while Figure 7.16b shows a section with several defective welds (at positions 525 m, 556 m, 700 m and 706 m). All the damaged welds showed clear growth of the indicator values over time. Beyond the application of ABA for detecting faulty welds only, these indicators characterize the rail surface roughness and, thus, they can be applied for all types of rail surface faults. For instance, a large quantity of surface defects occurred on the rail between position 550 m and 640 m in Figure 7.16b. In such cases, individual consideration of faults is of limited use, as it is most optimal to maintain the entire rail section at once.

In summary, the extraction of regular monitoring observations by means of specialized or in-service measurement vehicles equipped with ABAs bears potential for automating rail fault diagnostics. This can enhance predictive maintenance schemes by presenting asset managers with continual and spatially dense supervision of the rail condition over time.

7.5 Conclusions

In this work, we present a holistic framework for the automated detection of weld defects, by fusing a variety of observations, including on-site and visual inspections, automated diagnostic information extracted from monitoring vehicles and expert assessment. The scheme capitalizes on the availability of ABA information, extracted from accelerometer sensors featured on a diagnostic vehicle of the SBB. Extreme Value Analysis models were initially calibrated on various metrics stemming from the ABA measurements, in order to identify outlier welds in an unsupervised fashion on the basis of defined thresholds. The selected outliers were then passed onto actual field experts, in a first of its kind Proof-of-Concept project in collaboration with
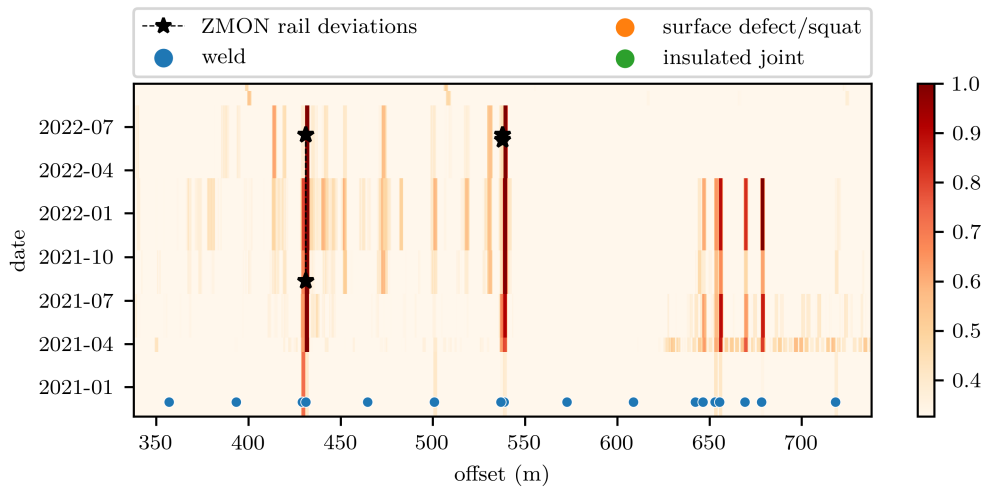
(a) Section with two damaged welds at positions 440 m and 542 m. The time series of the weld at position 542 m shown in Figure 7.15.
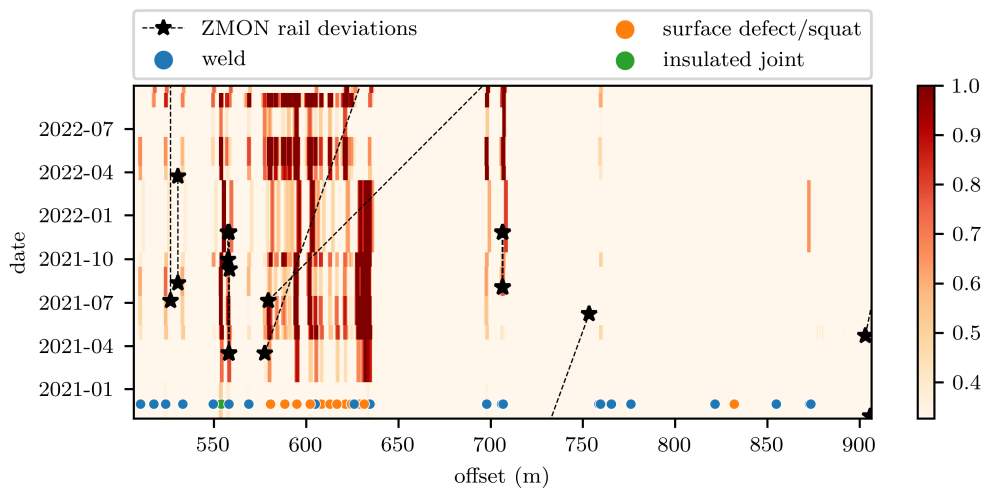
(b) Section with four damaged welds at positions 525 m, 556 m, 700 m and 706 m. Moreover, there were a large number of surface defects and squats between 550 m and 640 m.

Figure 7.16: Heatmaps illustrating the spatial and temporal evolution of the ABA-based rail condition for two 400 m rail sections on different track segments. The detected welds, surface defects and insulated joints corresponded to unvalidated samples from the rail-head image-based detection and are marked with color bullets. The start and end point/time of the faults that are recorded in the ZMON database are shown as two stars linked by a dotted line. Locations with increased ABA indicators matched the welds which were recorded as faulty, or the sections with a high density of rail faults.

the SBB. The experts offered their feedback on rail-head images ($VCUBE$) that corresponded to the EVA-identified outliers. This novel application combined real world data with expert feedback and was executed in four evaluation rounds, carried out over a period of one year. Newly identified defects as part of this PoC were then entered into the condition monitoring database of the SBB, where they wewre fused with existing information from further evaluation processes.

The extended dataset was then used to develop an automated one-class classification scheme, whose purpose was to identify defective welds from the assimilated expert feedback. Three different methods were applied to this end: BC, RF classifiers and BLR. On the basis of the conducted analysis, it is possible to suggest a preferred measurement configuration for ABAs. We recommend the use of both vertical and lateral accelerations in the assessment procedure,

as both are impacted by faults on welds. As such faults tend to induce an impulse-type of vibration response of the axle system, characterized by frequencies in the kilohertz range, this further motivates use of high frequency sensors. In terms of the performance of the suggested classification schemes, both the BLR and the RF models were trained on the same features, comprising the 15 features of the BC analysis that yielded the highest correlation to defect welds, while exhibiting less than 80% cross-feature correlation. The BLR model comes with the further advantage of delivering a prediction probability, which expresses the level of confidence we may attribute to the resulting labels.

These results indicate that component-specific evaluation can be delivered by combining asset type information with acceleration-based indicators and expert evaluation. Such an early detection of defects facilitated by acceleration-based indicators may improve the safety, efficiency, and cost effectiveness of both the inspection and maintenance process of rail welds in the future.

Finally, ABA measurements can detect faults much earlier, at their initiation, and, in this way, yield an estimate on damage severity. The continuous rating of weld condition over time, as opposed to the binary healthy/defective rating, further argues toward the importance of long-term monitoring schemes, which allow for tracking of condition over time. In an effort to demonstrate this, we presented examples which demonstrated that, prior to maintenance actions, significant growth of the ABA indicators was observed over time. This implied that emerging (early) faults were not caught by the experts, but were identified by the ABA and could feasibly be linked to a continuous indicator (rather than a categorical variable - label). Such ABA-derived indicators bear strong potential for effectuating early detection of faults and enabling a more granular and objective assessment of rail infrastructure condition.

While the proposed models are successful at classifying accelerations, further improvements can be obtained by some extensions to the current framework. An unsupervised approach using the Mahalanobis distance will be used in future work in order to allow assignment of labels beyond one-class classification, rather than on the the basis of a continuous scale, i.e., in terms of fault intensity [16]. The probabilistic framework enabled by the BLR can be extended by taking such a distance metric into account, while simultaneously incorporating the uncertainty in the expert labeling [208]. This work shows that such models improve the current paradigm of automated rail-head image-based inspection, but in the long-term pave the path for establishing OBM-based rail condition monitoring. The predictions of the proposed models can be incorporated as observations into a sequential decision-making framework to support optimal maintenance planning of railway assets [12].

Chapter 8

Conclusions and Perspectives

The following concluding chapter provides a summary of the main findings achieved in this thesis and of the research questions and challenges addressed. Moreover, it offers insights into possible future developments of the presented research.

8.1   Conclusions

The research conducted in this dissertation aims to deliver actionable tools for enhancing the robustness in the on board monitoring of railway assets; that is via use simple and low-cost sensing devices that can eventually be mounted on passenger vehicles. We offer an end to end framework for achieving this task in a first-of-its-kind collaborative Proof-of-Concept study with the Swiss Federal Railways (SBB). The overarching goal of this thesis is to address the challenges identified in Section 1.2.7 with the goal of achieving reliable condition monitoring. To this end, ABA-derived indicators are infused in a framework that exploits machine learning tools and statistical methods for uncertainty quantification to enhance robustness, prior to being delivered to experts for further corroboration and assessment; in this way closing the loop.

The thesis begins with an overarching introduction to the field of railway asset monitoring in a data rich environment. Emphasis is given to the systems and methodologies nowadays applied to assess the infrastructure condition, as well as the challenges currently encountered. The introductory part contextualizes the research work and introduces the approaches applied in the different chapters for the railway track assessment.

The first part of the dissertation (Chapter 2) focuses a comprehensive overview of the state-of-the-art of the research in the field of track asset monitoring using data from simple, inexpensive and robust sensors installed in regular rolling stock. We present a typical sensor setup of passenger trains for On-Board-Monitoring setups, with accelerometers mounted to the axle, the bogie and the body that measure the vehicle response at the main degrees of freedom of the vehicle. In this chapter, a suite of parametric and non-parametric methods are presented along with exemplary applications on data from the Swiss Federal Railways (SBB) database. These methods are applied to extract condition indicators using model-based or data-driven approaches, such as Kalman filters or double integration to recover the longitudinal level, or such as time-frequency domain representations for assessing short wavelength faults. In a second step, we perform a short review of classification techniques to characterize asset state and damage. The state of assets can be inferred from both data-driven and physical model-based approaches, where the former is often more straightforward to obtain when sufficient amounts of data are available, while the latter offers a higher degree of transparency and is better able to capture the physics at hand when data amounts are scarce. From this research aspect, we conclude that future monitoring approaches would optimally combine the broad

network coverage at frequent intervals of OBM on in-service vehicles with the less frequent measurements of diagnostic vehicles. Such an approach of combining frequent lower precision measurements with more rare but higher precision measurements from specialized diagnostic vehicles could enable an increase in the predictive capabilities of the demonstrated diagnostic algorithms.

The second part of this research work (Chapter 3) focuses on the identification of the track stiffness by using the Vold-Kalman filter in a novel application for superstructure condition assessment. We show that the VKF-based sleeper passage amplitude can be related to the track stiffness. In turn the dynamic response caused by varying track stiffness can be observed to result in increased maintenance on the corresponding track segments. From this chapter, the following conclusions can be drawn:

- The Vold-Kalman filter is applied to separate the periodic excitations due to the wheel or the sleeper passage from the axle box acceleration signals.

- The components corresponding to the wheel Out-Of-Roundness can be used to reconstruct the theoretical shape of the wheel, thus serving for condition assessment of wheel assets.

- The sleeper passage acceleration amplitude obtained via the VKF correlates to track subsidence measured by the subsidence measurement vehicle, implying that OBM-based ABA measurements could potentially supplement such subsidence recordings, delivering temporally denser measurements.

- Increased sleeper passage acceleration and forces results in higher maintenance requirements on the corresponding track sections. This illustrates that higher acceleration and load amplitudes result in more degradation over time; a phenomenon that can be tracked via ABA recordings.

Chapter 4 proposes the use of physics-based models for the assessment of the vehicle-track dynamics. This chapter demonstrates the fusion of physical models with data-driven approaches, utilizing the MCMC Bayesian updating method to optimize parameters in an ICN RABDe500 wagon model based on actual OBM measurements from an in-service ICN RABDe500 train. The result is a finely calibrated model capable of predicting loads, assessing damage, and optimizing maintenance intervals. In contrast to conventional trial-and-error approaches, MCMC automates parameter determination, albeit with a requirement for prior parameter selection and distribution definition. The posterior probability distribution of the parameters offered by the MCMC-scheme, allows for robust uncertainty quantification in predictions. Ultimately, the calibrated model serves as a valuable asset for future simulations.

While the previous parts investigate the general vehicle track dynamics in wavelengths of over 0.6 m, shorter wavelengths also have a clear influence on the ABA signals, relating to impacts caused by presence of short-wavelength faults on the rail. As the rail bears wheel loads on a contact patch that is smaller than a small coin or a postage stamp. Such a high stress concentration, repeated over several cycles, results in rail wear and in the propagation of cracks at the surface or inside the rail. By regularly maintaining the rail using grinding or milling vehicles, the rail profile geometry and the size of the cracks can be controlled. The efficacy of such interventions and the overall state of the track can be achieved by image recognition drawn from rail head images (e.g. $VCUBE$ system of the SBB) or ultrasonic measurements. However, current systems cannot detect the severity of flaws, and are often unable to distinguish between healthy and defect welds.

In response to the previously identified challenge, the fourth part of this thesis (Chapter 5) builds on the labels that are automatically generated by the image recognition system of the SBB and proposes an ML-based defect classification tool, exploiting ABAs, in order to

automatically distinguish between faults and main system (welds, joints, surface defects and squats). This part compares two approaches: the first one relies on a more traditional feature engineering approach that is combined with a Random Forest classifier algorithm. The second approach exploits a meaningful spectrotemporal mapping, namely the STFT of the ABA signals, whose coefficients are fed as an input to a Convolutional Neural Network. Both algorithms identify insulated joints and non defect rail samples well, but have a lower performance when classifying surface defects and welds. This is likely because these two classes are much broader than reflected by the ground truth labels. Welds can range from pristine and healthy welds to welds with squats or other damages. Similarly, surface defects range from small defects to large squats. From this chapter, the following conclusions can be drawn:

- Random Forests applied to features engineered from ABA achieve classification results similar to Convolutional Neural Networks applied on the time-series themselves.

- Healthy rails and insulated joints are well distinguished from surface defects, squats and welds.

- The condition of welds range from healthy to damaged ones, which results in lower classification scores.

- Surface defects and squats are more broadly defined as visual faults on the rail surface. These faults range from very small indentations to large squats across the full width of the rolling band. While accelerations capture differences in fault size, the visual labels classify these faults with the same defect label.

The promising findings from Chapter 5 reveal, however, the need for a more refined investigation of component-based assessment, since the assessment of critical rail components - such as welds or insulated joints - requires deeper information on their condition (and level of deterioration). We here focus on welds as a salient critical component of the Swiss Railway Network, which comprises several hundred thousands of welds at different stages in their life-cycle.

To address this challenge, Chapters 6 and 7 provide an in-depth assessment of welds. These studies are part of a Proof-Of-Concept that was carried out together with asset managers and field experts of the SBB. Machine learning tools and statistical methods are used to assess uncertainty and generate dependable diagnostic indicators in a process that includes experts for additional evaluation and confirmation.

In Chapter 6, we present a process to efficiently deliver acceleration-based outliers identified from Extreme Value Analysis in an unsupervised fashion to field experts for human validation. Experts then review rail-head images corresponding to these outliers, leading to the identification of suspected weld defects, which are entered into a condition monitoring database. The expert labels from this study are subsequently adopted in the framework described in Chapter 7 to deliver actionable condition indicators.

Chapter 7 presents a framework for automated detection of weld defects, combining on-site and visual inspections, diagnostic information from monitoring vehicles and expert assessments.The outliers identified in Chapter 6 are combined with the weld evaluations stemming from other monitoring processes such as ultrasonic testing and field inspections.

The knowledge gained on weld condition through the expert feedback is then used to develop an automated one-class classification scheme to identify defect welds. Three methods are applied: Binary Classification (BC), Random Forest (RF) classifiers and Bayesian Logistic Regression (BLR).
The following conclusions can be drawn regarding the assessment of welds:

- Acceleration-based outliers are validated by experts, based on visual appearance of the welds. Increased accelerations result in a higher probability of defect rating by experts.

- Expert labels from ZMON are used to build improved classification metrics by using BC, RF and BLR schemes. The BLR and the RF models were trained on the top-15 features from the BC analysis that yielded the highest correlation to defect welds, with less than 80% correlation to each other. The speed was additionally taken into account as a parameter in the classification. The RF and BLR models prove superior to the BC model,

- The BLR model offers the advantage of estimating a defect prediction probability and further assigning a level of confidence in the prediction, which is particularly advantageous in light of the high uncertainty caused by faulty ground truth labels.

- Accelerations are tracked over time using heatmaps. We observe continuous growth over time until maintenance measures are applied. While time-histories would enable the early detection of faults on welds, the main requirement is a sufficiently accurate vehicle localization.

- A holistic framework is delivered, combining data-driven condition indicators with robust outlier analysis and complementary expert feedback for actionable implementation in railway management.

In summary, the present thesis has investigated the potential of utilizing cost-efficient and low cost and complexity devices, such as ABA, for comprehensive track asset monitoring applications. In addition to the existing specialized measurement systems (optical, laser-based,...), which are adopted for traditional monitoring of the track components, accelerations measurements further capture the dynamic response resulting from the degradation processes usually observed with specialized monitoring systems. To deliver clear condition labels, the accelerations must be filtered or processed using parametric or non-parametric approaches to generate features, which are then combined with statistical approaches or machine learning techniques to translate acceleration indicators into condition indicators or fault labels.

ABA measurements can be used as a complementary information source on the state of the assets, capturing a large number of common track and vehicle faults. Interpretation challenges were more specifically addressed in this work by extracting sets of readily interpretable indicators. Long wavelength faults (between 0.6 m and 70 m) were characterized in terms of geometric indicators. These indicators were obtained using double integration and filtering techniques. Such indicators are associated with substructure and ballast condition [106]. The wheel geometry and track stiffness affecting wavelengths below 3 m, were characterized in Chapter 3 with indicators obtained with the Vold-Kalman Filter. In the case of short wavelength faults, their characterization can be unsupervised in the case of outlier detection, but requires more complex models when aspects such as the classification of rail fault types and rail component types are taken into account. It was demonstrated that accelerations not only enable the classification of infrastructure components in terms of dynamics, but also allow the dynamic-based assessment of their condition. Such assessments deliver a more refined view of the condition in terms of acceleration-based defect or outlier scores, eventually offering an oversight over possible evolution of deterioration processes and permitting continuous tracking of the growth of faults/defects over time. The continuous condition assessment aspect was demonstrated in a proof-of-concept for the assessment of welds where it is suggested that the continuous tracking of such indicators can serve as a valuable guide for preventive maintenance actions. It should be reminded that due to their nature and the fact that these reflect the interacting vehicle/track system, ABAs can serve for monitoring of both the traversed infrastructure, as well as the instrumented fleet. Even though the latter is not a focus of this thesis, it is touched upon in use of the VKF scheme.

8.2 Research Contributions and Perspectives

8.2.1 Research Contribution

The present research tackles the field of Structural Health Monitoring of railway assets, with some milestones for supporting real-time assessments via ABA. More specifically, component-based (e.g., track geometry, rail condition, welds, insulated joints...) evaluation is proposed by using features extracted from ABA in a data-driven fashion. This research contributes to several aspects of ABA-based assessment by fusing ABA with information from heterogeneous sources: acceleration measurements, measurements from specialized monitoring systems, condition logs, fixed asset information and asset maintenance data. These data sources are combined to assess track geometry, track stiffness, wheel Out-Of-Roundness and the rail condition (rail surface defects and squats, welds and insulated joints). The validation of the proposed ABA processing methods using such auxiliary data sources represents a significant research contribution in the field, since the demonstration on network wide data and information enables realistic assessments in contrast to previous studies that were often limited to very small selected datasets.

In the context of OBM, in addition to the research value of this work, the outcome of the application of this research in the railway sector will serve as an opportunity to achieve broad economic and societal impact. A better understanding of the railway asset condition is essential to improve rail maintenance scheduling, which would reduce the frequency of rail failures and accidents, minimize unscheduled downtime, and prolong the lifespan of rail infrastructure. This in turn would reduce maintenance costs, increase operational efficiency, and potentially enhance the overall competitiveness of the railway industry. We have shown in this work that OBM can be used to reliably assess track condition indicators such as the longitudinal level, stiffness indicators obtained via the VKF or rail condition indicators for the classification and identification of faults on the rail elements.

High quality railway transportation is the result of healthy assets. Combining condition indicators obtained from acceleration-based assessments with expert feedback is essential in improving the knowledge about asset state. From the improved knowledge of asset condition and optimized maintenance, the societal advantages are clear, since a safer and more reliable rail system would lead to increased passenger satisfaction and confidence in rail transportation. In addition, a more efficient rail system can reduce traffic congestion on roads and highways, and decrease air pollution and greenhouse gas emissions.

The outcomes of this thesis have already proven fruitful in both an academic and applied context. First of all, the stiffness assessment of the track that was shown to be an important factor in quantifying the degradation rate of the track. The characterization of the underlying dynamics is key to delivering improved track designs that better take into account the dynamics of the vehicle/track interaction. The detection of faulty welds in an expert-informed framework forms a first step towards actionable decision support algorithms that fuse data from different sources such as $VCUBE$ and ABA. This study was carried out in collaboration with SBB, which enabled an intensive exchange of expertise and knowledge regarding asset and condition databases. Such direct access was a unique opportunity that was instrumental in the research contributions presented herein. Remarkably, this work gave industrial partners hands-on experience and full access to apply the novel approaches developed in this work. Furthermore, demonstration codes, such as the Vold-Kalman filter implementation [101] have been made publicly available, enabling further collaborative advancements in this domain.

8.2.2 Outlook

Building on the research and conclusions of this thesis, this section presents first presents the main challenges and limitations of this work, while proposing several research directions as

extensions or further innovations.

Despite the positive perspectives, some limitation identified in these studies must be highlighted here. First of all, a large number of uncertainties stemming from several sources, such as the vehicle dynamics, the expert assessment ground truth, or the vehicle positioning, are observed in this work. These uncertainties affect both the results but may also pose a challenge for the automated assessment using a fleet of vehicles. The approach presented in Chapter 3 that harnesses the VKF for assessing the track stiffness was shown on a section for which subsidence measurements from the EMW were available. While these results are very promising, hinting that ABA-assessments could complement the measurement of subsidence, the generalization would require further studies containing railway sections with wood sleepers and should further integrate the vehicle speed as a parameter. These drawbacks relate to the lack of data available on the section for which the EMW measurement were provided.

Chapter 4 presents the combination of physics-based models with data-driven approaches through MCMC Bayesian updating optimizes railway vehicle parameters. The calibrated models improve load predictions, damage assessments, and maintenance planning. Future research will focus on enhancing MCMC's performance, including automatic termination criteria and alternative evaluation metrics. Expanding to incorporate acceleration measurements and embracing parameter estimate uncertainty will provide deeper insights. Importantly, this approach is adaptable to various vehicle types and track sections, promising widespread applicability and revolutionizing railway system management.

Uncertainties in the vehicle dynamics such as the hunting oscillation can cause variation in the vehicle response when crossing smaller defects (Chapter 5 and Chapter 7) which can be a source of noise in the predictions. Uncertainties in the labels were observed in the studies presented in Chapter 5, 6 and 7. We recommend that future studies quantify these uncertainties by cross validating the labels using several independent experts. Due to the limited available time from the experts that supported our work, it was chosen here to maximize the number of validated samples, instead of cross-checking a smaller number of samples.

The time-evolution of ABA-based indicators on the defects in Chapter 5 and 6 requires a sufficiently accurate positioning to ensure meaningful results. The uncertainties observed in the vehicle localization have limited the scope of the results concerning the prediction of the evolution of defects in time. An improved positioning was not further investigated in this work, since the focus of this Thesis is not the improvement of the localization system. Uncertainty due to the vehicle speed was an aspect that was accounted for in a data driven way in Chapter 5 and Chapter 7.

Machine learning and statistical approaches are often complex and subject to assumptions and constraints relating to the distribution of the training data, which on different track and in different conditions may result in lower accuracy and reliability, requiring the retraining of the models (via for example adaptive approaches) using an expert feedback process similar to the one presented in Chapter 7. This point is particularly essential for the analysis of the VKF paper, which was shown for a high-speed section on the network of SBB. The algorithms proposed in these studies were obtained on samples with vehicle speeds up to 200 km/h, and for generalizations beyond those conditions requires the retraining of the models to ensure stable predictions. Despite the challenges encountered, I am confident that this work is of high value in terms of proposed approaches on the research side and that if overcome can improve the overall monitoring process of the railway assets.

Generally in asset management there are two aspects which this research can support:

- Understanding the dynamics of the vehicle/track system to optimize the design of the track from a dynamic point of view; especially for high-speed lines, which have much higher requirements.

- Development of diagnostic and prognostic railway track condition indicators to enable more efficient predictive or condition-based maintenance approaches.

- Development of an automated framework, for extraction, processing and analysis of measurement data and the fusion of these indicators with inspection data and complementary expert feedback for practical implementation.

The following research aspects can be explored to build upon or extend the results obtained here:

- VKF was used to assess the track stiffness on a high speed rail-line. This approach is very promising for capturing other faults such as sleeper voids or possibly even substructure condition. More stiffness measurements by the EMW for diverse track types such as superstructures with wood are necessary in order to validate this approach for the most common superstructure configurations on the network of SBB.

- A scheme that utilizes acceleration-based techniques to automate the classification of main rail components and surface faults, including healthy rail, surface defects and squats, insulated joints, and welds is introduced in Chapter 5. Two approaches are compared: Random Forests (RFs) trained on engineered features and Convolutional Neural Networks (CNNs) trained on Short Time Fourier Transform coefficients of ABA signals. Although the validation of the scheme is limited by the lack of validated reference labels, the implementation demonstrates the benefits of automation and encourages the adoption of acceleration-based classifiers for the early detection of faults.

- Probabilistic approaches for assessing the weld condition were applied in Chapter 7. The practical use of such models would entail changes in the currently applied processes of railway maintenance, since the existing approaches for condition assessment are of deterministic nature.

- Bias or errors in the expert labeling leads to increased prediction uncertainties. Models such as BLR capture this uncertainty. The defect probability and uncertainty on the prediction are factors that would potentially improve the sample selection submitted for expert validation. Samples with high uncertainty would be submitted for a second iteration of expert labeling in order to lower the label bias. Although this extension has not been considered in this PoC to present a fair model comparison, it would certainly be useful to consider the prediction uncertainty for a practical real-world implementation to fully exploit the BLR model capabilities.

- One aspect that was only briefly mentioned in the introduction, but not further explored in this work, is the optimal maintenance planning. More regular measurements from in-service vehicles would yield regular observations which could be exploited in the future to develop optimal maintenance planning frameworks. The integration of these assessments into the Life-Cycle Management models of SBB would for example enable the estimation of the benefits of regular in-service measurements with low-cost sensors. Partially observable Markov decision models solved with reinforcement learning algorithms is one approach commonly applied for such investigations.

- Finally, although this work focuses on regular railway tracks, other elements, such as bridges and switches, form critical components of the networks that could benefit from vehicle-based condition assessment. Bridges and switches are complex and critical components that nowadays are still often monitored via on-site inspections. The monitoring

of switches and bridges using on site-sensors has gained popularity in recent years. However, vehicle-based monitoring of these assets would obviously be more scalable compared to local asset monitoring.

List of Figures

List of Tables

Bibliography

[1] Railway applications - testing and simulation for the acceptance of running characteristics of railway vehicles - running behaviour and stationary tests.

[2] S. D. A. Wegner. Digitale abnahme von schienenlängsprofilen. *Infrastruktur Network*, 139, 2015.

[3] I. Abdallah, V. Dertimanis, H. Mylonas, K. Tatsis, E. Chatzi, N. Dervili, K. Worden, and E. Maguire. Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data. In *Safety and Reliability–Safe Societies in a Changing World*, pages 3053–3061. CRC Press, 2018.

[4] C. C. Aggarwal. *Outlier Analysis*. 2017.

[5] C. C. Aggarwal and P. S. Yu. Outlier detection for high dimensional data. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 30, 2001.

[6] C. Ágh. Comparative analysis of axlebox accelerations in correlation with track geometry irregularities. *Acta Technica Jaurinensis*, 12, 2019.

[7] R. Ahmad and S. Kamaruddin. An overview of time-based and condition-based maintenance in industrial application. *Computers &, Industrial Engineering*, 63(1):135–149, Aug. 2012.

[8] A. Alemi, F. Corman, and G. Lodewijks. Condition monitoring approaches for the detection of railway wheel defects. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 231(8):961–981, June 2016.

[9] B. An, P. Wang, J. Xu, R. Chen, and D. Cui. Observation and simulation of axle box acceleration in the presence of rail weld in high-speed railway. *Applied Sciences (Switzerland)*, 7, 2017.

[10] A. R. Andrade and P. F. Teixeira. Exploring different alert limit strategies in the maintenance of railway track geometry. *Journal of Transportation Engineering*, 142(9), Sept. 2016.

[11] A. Aniszewicz and M. Fabijański. Measurements of railway welded rail joints with a laser device. *Welding Technology Review*, 92, 2020.

[12] G. Arcieri, C. Hoelzl, O. Schwery, D. Straub, K. G. Papakonstantinou, and E. Chatzi. Bridging pomdps and bayesian decision making for robust maintenance planning under model uncertainty: An application to railway systems. *arXiv preprint arXiv:2212.07933*, 2022.

[13] S. S. Artagan, L. B. Ciampoli, F. D'Amico, A. Calvi, and F. Tosti. Non-destructive assessment and health monitoring of railway infrastructures. *Surveys in Geophysics*, 41(3):447–483, jul 2019.

[14] F. Auer. Multi-function track recording cars. *Railway Technical Review*, 3-4:32–36, 2013.

[15] L. Avendano-Valencia and S. Fassois. Damage/fault diagnosis in an operating wind turbine under uncertainty via a vibration response Gaussian mixture random coefficient model based framework. *Mechanical Systems and Signal Processing*, 91:326–353, 2016.

[16] L. D. Avendano-Valencia, E. N. Chatzi, and D. Tcherniak. Gaussian process models for mitigation of operational variability in the structural health monitoring of wind turbines. *Mechanical Systems and Signal Processing*, 142, 2020.

[17] L. D. Avendano-Valencia, C. Hoelzl, and E. Chatzi. Operational regime clustering for the construction of pce-based surrogates of operational wind turbines expansions and non-negative matrix factorization. In *7th International Conference on Advances in Experimental Structural Engineering (7AESE)*, pages 961–964, 2017.

[18] S. E. Azam, E. Chatzi, and C. Papadimitriou. A dual kalman filter approach for state estimation via output-only acceleration measurements. *Mechanical Systems and Signal Processing*, 60, 2015.

[19] B. Baasch, J. Heusel, J. Groos, and S. Shankar. Eingebettete zustandsüberwachung der gleisinfrastruktur: Entwicklung und erprobung von eingebetteten multi-sensor-systemen für die kontinuierliche zustandsüberwachung der gleisinfrastruktur im operativen betrieb. *Der Eisenbahningenieur EI*, 2019.

[20] B. Baasch, J. Heusel, M. Roth, and T. Neumann. Train wheel condition monitoring via cepstral analysis of axle box accelerations. *Applied Sciences (Switzerland)*, 11, 2021.

[21] B. Baasch, M. Roth, and J. Groos. In-service condition monitoring of rail tracks. *Internationales Verkehrswesen*, 70(1), 2018.

[22] A. Bakhtiary, J. A. Zakeri, and S. Mohammadzadeh. An opportunistic preventive maintenance policy for tamping scheduling of railway tracks. *International Journal of Rail Transportation*, 9(1):1–22, Mar. 2020.

[23] F. Balouchi, A. Bevan, and R. Formston. Development of railway track condition monitoring from multi-train in-service vehicles. *Vehicle System Dynamics*, 2020.

[24] B. Barahona, C. Hoelzl, and E. Chatzi. Applying design knowledge and machine learning to scada data for classification of wind turbine operating regimes. volume 2018-January, 2018.

[25] D. Barke and W. K. Chiu. Structural health monitoring in the railway industry: A review. *Structural Health Monitoring*, 4(1):81–93, 2005.

[26] D. W. Barke and W. K. Chiu. A review of the effects of out-of-round wheels on track and vehicle components. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 219(3):151–175, 2005.

[27] K. Barrow. Bridging the big data gap. *Railjournal*, 2018.

[28] A. Benedetto, F. Tosti, L. B. Ciampoli, A. Calvi, M. G. Brancadoro, and A. M. Alani. Railway ballast condition assessment using ground-penetrating radar – an experimental, numerical simulation and modelling development. *Construction and Building Materials*, 140, 2017.

[29] E. Berggren. *Railway track stiffness: dynamic measurements and evaluation for efficient maintenance.* PhD thesis, KTH, 2009.

[30] E. G. Berggren, A. M. Kaynia, and B. Dehlbom. Identification of substructure properties of railway tracks by dynamic stiffness measurements and simulations. *Journal of Sound and Vibration*, 329(19):3999 – 4016, 2010.

[31] B. Bergquist and P. Söderholm. Data analysis for condition-based railway infrastructure maintenance. *Quality and Reliability Engineering International*, 31(5):773–781, Mar. 2014.

[32] E. Bernal, M. Spiryagin, and C. Cole. Ultra-low power sensor node for on-board railway wagon monitoring. *IEEE Sensors Journal*, 20(24):15185–15192, 2020.

[33] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14, 1992.

[34] N. Bešinović. Resilience in railway transport systems: a literature review and research agenda. *Transport Reviews*, 40(4):457–478, 2020.

[35] C. M. Bishop. *Pattern Recoginiton and Machine Learning.* 2006.

[36] B. Blanco and A. Alonso. Implementation of timoshenko element local deflection for vertical track modelling. *Vehicle System Dynamics - International Journal of Vehicle Mechanics and Mobility*, 57:10(1421-1444):1421–1444, 2019.

[37] B. Blanco, A. Alonso, L. Kari, N. Gil-Negrete, and J. G. Giménez. Distributed support modelling for vertical track dynamic analysis. *Vehicle System Dynamics*, 56(4):529–552, Nov. 2017.

[38] G. Blatman and B. Sudret. Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. *Reliab. Eng. Sys. Safety*, 95:1216–1229, 2010.

[39] B. Boashash. *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference.* 2015.

[40] M. Bocciolone, A. Caprioli, A. Cigada, and A. Collina. A measurement system for quick rail inspection and effective track maintenance strategy. *Mechanical Systems and Signal Processing*, 21:1242–1254, 4 2007.

[41] C. Boedecker. *Die Wirkungen zwischen Rad und Schiene und ihre Einflüsse auf den Lauf und den Bewegungswiderstand der Fahrzeuge in Eisenbahnzügen.* Hahn'sche Buchhandlung, Hannover, 1887.

[42] Brüel & Kjaer. *Vold-Kalman Order Tracking Filter - Type 7703 for PULSE, the Multi-analyzer System Type 3560*, 4 2008. Rev. 3.

[43] BSL Management Consultants GmbH & Co. KG. "second opinion" zum bericht netzaudit sbb.

[44] Z. A. Bukhsh et al. Network level bridges maintenance planning using multi-attribute utility theory. *Structure and Infrastructure Engineering*, 15(7):872–885, 2019.

[45] Z. A. Bukhsh, I. Stipanovic, G. Klanker, A. O. Connor, and A. G. Doree. Network level bridges maintenance planning using multi-attribute utility theory. *Structure and Infrastructure Engineering*, 15(7):872–885, Jan. 2018.

[46] L. Bull, K. Worden, G. Manson, and N. Dervilis. Active learning for semi-supervised structural health monitoring. *Journal of Sound and Vibration*, 437, 2018.

[47] J. Burghardt, M. Iwaszkiewicz, and J. Madejski. System for the real time railway track geometry parameters evaluation. *International Conference On Signal Processing Applications and Technology, Boston, MA, USA*, pages 1821–1825, 1996.

[48] N. C and S. Farritor. Design of a system to measure track modulus from a moving railcar. *Proceedings from Railway Engineering, London, UK*, 2004.

[49] E. Cabal-Yepez, A. G. Garcia-Ramirez, R. J. Romero-Troncoso, A. Garcia-Perez, and R. A. Osornio-Rios. Reconfigurable monitoring system for time-frequency analysis on industrial equipment through stft and dwt. volume 9, 2013.

[50] A. Caprioli, A. Cigada, and D. Raveglia. Rail inspection in track maintenance: a benchmark between the wavelet approach and the more conventional fourier analysis. *Mechanical Systems & Signal Processing, 21(2)*, 21(2):631–652, 2007.

[51] F. W. Carter. On the stability of running of locomotives. In *Proc. R. Soc. Lond. A. 121*, pages 151–157, 1928.

[52] CEN. En 13848-1, railway applications. track. track geometry quality. characterization of track geometry. *BSI*, 2019.

[53] E. Chatzi, I. Abdallah, K. Tatsis, S. Osmani, and I. Robles. Using interpretable machine learning for data-driven decision support for infrastructure operation & maintenance. In *Bridge Safety, Maintenance, Management, Life-Cycle, Resilience and Sustainability*, pages 837–845. CRC Press, 2022.

[54] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 2002.

[55] X. Chen, X. Chai, and X. Cao. The time-frequency analysis of the train axle box acceleration signals using empirical mode decomposition. *Computer Modelling & New Technologies 2014*, 18, 2014.

[56] H. Cho and J. Park. Study of rail squat characteristics through analysis of train axle box acceleration frequency. *Applied Sciences*, 11(15):7022, July 2021.

[57] A. Chudzikiewicz, R. Bogacz, M. Kostrzewski, and R. Konowrocki. Condition monitoring of railway track systems by using acceleration signals on wheelset axle-boxes. *Transport*, 33, 2018.

[58] S. Cii, G. Tomasini, M. L. Bacci, and D. Tarsitano. Solar wireless sensor nodes for condition monitoring of freight trains. *IEEE Transactions on Intelligent Transportation Systems*, 23(5):3995–4007, 2022.

[59] S. Coles. *Basics of Statistical Modeling*, pages 18–44. Springer London, London, 2001.

[60] F. Corman, E. Quaglietta, and R. M. P. Goverde. Automated real-time railway traffic control: an experimental analysis of reliability, resilience and robustness. *Transportation Planning and Technology*, 41(4):421–447, Mar. 2018.

[61] I. Corni, N. Symonds, R. Wood, A. Wasenczuk, and D. Vincent. *Real-time on-board condition monitoring of train axle bearings*. The Stephenson Conference - Research for Railways, 2015.

[62] N. Correa, E. Vadillo, J. Santamaria, and J. Gómez. A rational fraction polynomials model to study vertical dynamic wheelrail interaction. *Journal of Sound and Vibration*, 331(8):1844 – 1858, 2012.

[63] T. Dahlberg. Some railroad settlement models - a critical review. volume 215, 2001.

[64] T. Dahlberg. Railway track stiffness variations - consequences and countermeasures. *International Journal of Civil Engineering*, 8:1–12, 2010.

[65] T. A. Davis. Algorithm 832: Umfpack v4.3—an unsymmetric-pattern multifrontal method. *ACM Trans. Math. Softw.*, 30(2):196–199, jun 2004.

[66] DB Netz AG (DB Systemtechnik unter Mitwirkung von Projektteam Fahrbahnstrategie). Systementscheidung schotteroberbau oder feste fahrbahn. *Bahntechnik*, 139(4), 2003.

[67] A. L. O. de Melo, S. Kaewunruen, M. Papaelias, L. L. B. Bernucci, and R. Motta. Methods to monitor and evaluate the deterioration of track and its components in a railway in-service: A systemic review. *Frontiers in Built Environment*, 6, Sept. 2020.

[68] X. Deng, Z. Li, Z. Qian, W. Zhai, Q. Xiao, and R. Dollevoet. Pre-cracking development of weld-induced squats due to plastic deformation: Five-year field monitoring and numerical analysis. *International Journal of Fatigue*, 127, 2019.

[69] V. K. Dertimanis, M. Zimmermann, F. Corman, and E. N. Chatzi. On-board monitoring of rail roughness via axle box accelerations of revenue trains with uncertain dynamics. 2019.

[70] V. Dubourg and B. Sudret. Meta-model-based importance sampling for reliability sensitivity analysis. *Structural Safety*, 49:27–36, July 2014.

[71] C. Esveld and C. Esveld. *Modern railway track*, volume 385. MRT-productions Zaltbommel, 2001.

[72] C. Esveld and M. Steenbergen. Force-based assessment of weld geometry. *8th International Heavy Haul Conference, Rio de Janeiro, Brazil*, 2005.

[73] A. Falamarzi, S. Moridpour, and M. Nazem. Development of a tram track degradation prediction model based on the acceleration data. *Structure and Infrastructure Engineering*, 15(10):1308–1318, 2019.

[74] C. Farrar and K. Worden. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851):248–253, 2007.

[75] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P. A. Muller. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33, 2019.

[76] C. Feldbauer and R. Höldrich. Realization of a Vold-Kalman Tracking Filter - A Least Squares Problem. In *Proceedings of COST G-6 Conference on Digital Audio Effects (DAFX-00)*, Verona, Italy, December 7–9 2000.

[77] M. Fellinger. Validierung der instandsetzungsmengen der standardelemente gleis der öbb. *FSV aktuell Schiene*, (1):74–75, Mar. 2017. FSV-Preisverleihung 2016 ; Conference date: 10-11-2016 Through 10-11-2016.

[78] Å. Fenander. Frequency dependent stiffness and damping of railpads. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of rail and rapid transit*, 211(1):51–62, 1997.

[79] Z. Feng, S. Qin, and M. Liang. Time-frequency analysis based on vold-kalman filter and higher order energy separation for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions. *Renewable Energy*, 85, 2016.

[80] Z. Feng, W. Zhu, and D. Zhang. Time-frequency demodulation analysis via vold-kalman filter for wind turbine planetary gearbox fault diagnosis under nonstationary speeds. *Mechanical Systems and Signal Processing*, 128, 2019.

[81] M. Fermér and J. Nielsen. Vertical interaction between train and track with soft and stiff railpads—full-scale experiments and theory. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 209(1):39 – 47, 1995.

[82] Fink, Olga. *Failure and degradation prediction by artificial neural networks: Applications to railway systems.* PhD thesis, 2014.

[83] G. Soldati, M. Meier and M. Zurkirchen . Kontinuierliche messung der gleiseinsenkung mit 20 t achslast. *Der Eisenbauingenieur*, 2016.

[84] Y. Gal and Z. Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[85] D. Galar, A. Thaduri, M. Catelani, and L. Ciani. Context awareness for maintenance decision making: A diagnosis and prognosis approach. *Measurement*, 67:137–150, May 2015.

[86] A. Gautam and V. Singh. Parametric versus non-parametric time series forecasting methods: A review, 2020.

[87] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot. Variable selection using random forests. *Pattern Recognition Letters*, 31(14):2225–2236, 2010.

[88] P. C. L. Gerum, A. Altay, and M. Baykal-Gürsoy. Data-driven predictive maintenance scheduling policies for railways. *Transportation Research Part C: Emerging Technologies*, 107:137–154, Oct. 2019.

[89] M. Giunta and F. G. Praticò. Design and maintenance of high-speed rail tracks: A comparison between ballasted and ballast-less solutions based on life cycle cost analysis. 2017.

[90] W. Gong, M. F. Akbar, G. N. Jawad, M. F. P. Mohamed, and M. N. A. Wahab. Nondestructive testing technologies for rail inspection: A review. *Coatings*, 12(11), 2022.

[91] R. C. Gonzalez. *Digital image processing.* Pearson education, 2009.

[92] J. C. Goswami and A. K. Chan. *Fundamentals of Wavelets: Theory, Algorithms, and Applications: Second Edition.* 2010.

[93] S. L. Grassie. Rail corrugation: Characteristics, causes, and treatments. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 223(6):581–596, June 2009.

[94] S. L. Grassie and S. J. Cox. The dynamic response of railway track with unsupported sleepers. *Proceedings of the Institution of Mechanical Engineers, Part D: Transport Engineering*, 199(2):123–136, Apr. 1985.

[95] Group VöV. *Schweissarbeiten an Schienen und Weichenbauteilen*. VöV, 2006.

[96] W. Guo, C. Zeng, H. Gou, Q. Gu, T. Wang, H. Zhou, B. Zhang, and J. Wu. Real-time hybrid simulation of high-speed train-track-bridge interactions using the moving load convolution integral method. *Engineering Structures*, 228:111537, 2021.

[97] Y. Guo, J. Xie, Z. Fan, V. Markine, D. P. Connolly, and G. Jing. Railway ballast material selection and evaluation: A review. *Construction and Building Materials*, 344:128218, aug 2022.

[98] A. Hansen. The three extreme value distributions: An introductory review. *Frontiers in Physics*, 8, 2020.

[99] Y. Hayashi, T. Kojima, H. Tsunashima, and Y. Marumo. Real time fault detection of railway vehicles and tracks. volume 2006, 2006.

[100] H. Herlufsen, S. Gade, H. Konstantin-Hansen, and H. Vold. Characteristics of the vold-kalman order tracking filter. *S V Sound and Vibration*, 33(4):34–44, 1999. cited By 14.

[101] C. Hoelzl. Python Implementation of the second generation Vold Kalman Filter. *GitHub repository*, 2020.

[102] C. Hoelzl. *Data driven condition assessment of railway infrastructure*, pages 693–693. 03 2021.

[103] C. Hoelzl. railFE: Simplified Vehicle Track Interaction Model with discretized supports. *GitHub repository*, 2022.

[104] C. Hoelzl, L. Ancu, H. Grossmann, D. Ferrari, V. Dertimanis, and E. Chatzi. Classification of rail irregularities from axle box accelerations using random forests and convolutional neural networks. In *Data Science in Engineering, Volume 9*, pages 91–97, Cham, 2022. Springer International Publishing.

[105] C. Hoelzl, G. Arcieri, L. Ancu, S. Banaszak, A. Kollros, V. Dertimanis, and E. Chatzi. Fusing expert knowledge with monitoring data for condition assessment of railway welds. *Sensors*, 23(5):2672, Feb. 2023.

[106] C. Hoelzl, V. Dertimanis, E. Chatzi, D. Winklehner, S. Züger, and A. Oprandi. Data driven condition assessment of railway infrastructure. 2021.

[107] C. Hoelzl, V. Dertimanis, M. Landgraf, L. Ancu, M. Zurkirchen, and E. Chatzi. On-board monitoring for smart assessment of railway infrastructure: A systematic review. In *The Rise of Smart Cities*, pages 223–259. Elsevier, 2022.

[108] C. Hoelzl, L. D. A. Valencia, V. Dertimanis, and E. Chatzi. Axle box accelerometer signal identification and modelling. In *Model Validation and Uncertainty Quantification, Volume 3*, pages 87–92, Cham, 2020. Springer International Publishing.

[109] C. A. Hoelzl, V. Dertimanis, L. Ancu, A. Kollros, and E. Chatzi. Vold-kalman filter order tracking of axle box accelerations for railway stiffness assessment. 2022.

[110] C. A. Hoelzl, V. Dertimanis, A. Kollros, L. Ancu, and E. Chatzi. Weld condition monitoring using expert informed extreme value analysis. In *European Workshop on Structural Health Monitoring*, pages 711–720, Cham, 2023. Springer International Publishing.

[111] M. D. Hoffman, A. Gelman, et al. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.

[112] J. Holzfeind. Anlagenstrategie Fahrbahn 2040. SBB Infrastruktur, I-AT-FB-AMM, 2014.

[113] J. Holzfeind, I. Nerlich, and Z. Kull. Anabel – ein detailliertes belastungs- und beanspruchungsmonitoringsystem mit großem potential. *ZEVrail*, 140(11,12), 2016.

[114] J. Hong, W. Guangrui, Z. Xiangfeng, S. Yongfang, and X. Bin. Application of spectral kurtosis and vold-kalman filter based order tracking in wind turbine gearbox fault diagnosis. volume 2018-March, 2018.

[115] W. Hongrui. *Data-based Dynamic Condition Assessment of Railway Catenaries*. PhD thesis, Doctoral Thesis, TU Delft, 2019.

[116] M. Hoshiya and O. Maruyama. Identification of running load and beam system. *Journal of Engineering Mechanics*, 113:813–824, 6 1987.

[117] M. Hosseingholian, M. Froumentin, and A. Robinet. Feasibility of a continuous method to measure track stiffness. *Proceedings from Railway Foundations conference, Birmingham, UK*, 2006.

[118] M. Hossin and S. M.N. A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5:01–11, 03 2015.

[119] J. Huang, X. Yin, and S. Kaewunruen. Quantification of dynamic track stiffness using machine learning. *IEEE Access*, 10:78747–78753, 2022.

[120] Z. Huang, Z. Liang, and J. Zou. Optimization of vehicle suspension parameters based on simpack/isight simulation of wheel eccentric out-of-roundness. In *2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI)*, pages 270–274. IEEE, 2022.

[121] I-FW-PS-FB and P. Güldenapfel. *Einbau, Kontrollen und Unterhalt von Gleisen, R I-22070*. Infrastruktur SBB, 2009.

[122] M. Ishida, M. Akama, K. Kashiwaya, and A. Kapoor. The current status of theory and practice on rail integrity in japanese railways-rolling contact fatigue and corrugations. *Fatigue and Fracture of Engineering Materials and Structures*, 26, 2003.

[123] S. Iwnicki, editor. *Handbook of railway vehicle dynamics*. CRC Press, 1 2006.

[124] M. V. e. a. J. Rajamaeki. Limitations of eddy current inspection in railway rail evaluation. *Proc IMechE Part F: J Rail and Rapid Transit*, 232:121–129, 2018.

[125] M. Jesussek and K. Ellermann. Fault detection and isolation for a full-scale railway vehicle suspension with multiple kalman filters. *Vehicle System Dynamics*, 52:1695–1715, 12 2014.

[126] Y. Ji, J. Zeng, and W. Sun. Research on wheel-rail local impact identification based on axle box acceleration. *Shock and Vibration*, 2022:1–17, feb 2022.

[127] C. Jia, W. Xu, L. Wei, and H. Wang. Study of railway track irregularity standard deviation time series based on data mining and linear model. *Mathematical Problems in Engineering*, 2013, 2013.

[128] G. Jodlbauer and J. Neubert". "digitales aufzeichnungssystem für gleisbaumaschinen". *Der Eisenbahningenieur EI*, 2014.

[129] A. Johansson and C. Andersson. Out-of-round railway wheels—a study of wheel polygonalization through simulation of three-dimensional wheel–rail interaction and wear. *Vehicle System Dynamics*, 43(8):539–559, Aug. 2005.

[130] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte. New approach for filtering nonlinear systems. volume 3, 1995.

[131] S. Kaewunruen and A. M. Remennikov. Field trials for dynamic characteristics of railway track and its components using impact excitation technique. *NDT and E International*, 40(7):510 – 519, 2007.

[132] K. Kaiser and N. Gebraeel. Predictive maintenance management using sensor-based degradation models. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 39(4):840–849, July 2009.

[133] M. Kamargianni, W. Li, M. Matyas, and A. Schäfer. A critical review of new mobility services for urban transport. *Transportation Research Procedia*, 14:3294–3303, 2016. Transport Research Arena TRA2016.

[134] T. Karis, M. Berg, S. Stichel, M. Li, D. Thomas, and B. Dirks. Correlation of track irregularities and vehicle responses based on measured data. *Vehicle System Dynamics*, 56:967–981, 6 2018.

[135] A. Kendall and Y. Gal. What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, 30, 2017.

[136] H. Khajehei, A. Ahmadi, I. Soleimanmeigouni, and A. Nissen. Allocation of effective maintenance limit for railway track geometry. *Structure and Infrastructure Engineering*, 15(12):1597–1612, June 2019.

[137] J. Klingel. Über den lauf von eisenbahnwagen auf gerader bahn. *Organ für die Fortschritte des Eisenbahnwesens*, 20:113–123, 1883.

[138] E. M. Knorr and R. T. Ng. Finding intensional knowledge of distance-based outliers. *Proceedings of the 25th International Conference on Very Large Data Bases*, 1999.

[139] K. Knothe. *Gleisdynamik*. Ernst & Sohn, 2001.

[140] K. L. Knothe and S. L. Grassie. Modelling of railway track and vehicle/track interaction at high frequencies. *Vehicle System Dynamics*, 22(3-4):209–262, 1993.

[141] K. L. Knothe and S. Stichel. Rail vehicle dynamics. *Springer*, (20-200), 2017.

[142] S. Koroma, M. Hussein, and J. Owen. The effects of railpad nonlinearity on the dynamic behaviour of railway track. *Annual Spring Conference, Nottingham, UK*, 2013.

[143] C. Kossmann. SIMPACK vehicle model of ICN (RABDe500) SBB Infrastructure, 2021.

[144] A. Królicka, G. Lesiuk, K. Radwański, R. Kuziak, A. Janik, R. Mech, and T. Zygmunt. Comparison of fatigue crack growth rate: Pearlitic rail versus bainitic rail. *International Journal of Fatigue*, 149:106280, aug 2021.

[145] E. Kubera, A. Wieczorkowska, A. Kuranc, and T. Słowik. Discovering speed changes of vehicles from audio data. *Sensors (Switzerland)*, 19, 7 2019.

[146] Z. Lai, C. Mylonas, S. Nagarajaiah, and E. Chatzi. Structural identification with physics-informed neural ordinary differential equations. *Journal of Sound and Vibration*, 508, 2021.

[147] H.-F. Lam, S. A. Alabi, and J.-H. Yang. Identification of rail-sleeper-ballast system through time-domain markov chain monte carlo-based bayesian approach. *Engineering Structures*, 140:421 – 436, 2017.

[148] M. Landgraf. *Zustandsbeschreibung des Fahrwegs der Eisenbahn - Von der Messdatenanalyse zum Anlagenmanagement*. PhD thesis, 2016.

[149] M. Landgraf and M. Enzi. Smart Data for a Pro-active Railway Asset Management. *Transport Research Arena*, 2018.

[150] M. Landgraf and F. Hansmann. Fractal analysis as an innovative approach for evaluating the condition of railway tracks. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 233, 2019.

[151] A. Lasisi and N. Attoh-Okine. Machine learning ensembles and rail defects prediction: Multilayer stacking methodology. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 5(4):04019016, 2019.

[152] A. S. Lathe and A. Gautam. Estimating vertical profile irregularities from vehicle dynamics measurements. *IEEE Sensors Journal*, 20:377–385, 1 2020.

[153] G. Lederman, S. Chen, J. Garrett, J. Kovačević, H. Y. Noh, and J. Bielak. Track-monitoring from the dynamic response of an operational train. *Mechanical Systems and Signal Processing,*, 87(A):1–16, 2015.

[154] G. Lederman, S. Chen, J. H. Garrett, J. Kovačević, H. Y. Noh, and J. Bielak. Track monitoring from the dynamic response of a passing train: A sparse approach. *Mechanical Systems and Signal Processing*, 90:141–153, 6 2017.

[155] C. K. M. Lee, Y. Cao, and K. H. Ng. Big data analytics for predictive maintenance strategies. In *Supply Chain Management in the Big Data Era*, pages 50–74. IGI Global, 2017.

[156] N. Lestoille, C. Soize, and C. Funfschilling. Stochastic prediction of high-speed train dynamics to long-term evolution of track irregularities. *Mechanics Research Communications*, 75, 2016.

[157] C. Li, S. Luo, C. Cole, and M. Spiryagin. An overview: modern techniques for railway vehicle on-board health monitoring systems. *Vehicle System Dynamics*, 55:1045–1070, 7 2017.

[158] Q. Li, D. J. Thompson, and M. G. Toward. Estimation of track parameters and wheel–rail combined roughness from rail vibration. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 232(4):1149 – 1167, 2018.

[159] S. Li, A. N. nez, Z. Li, and R. Dollevoet. Automatic detection of corrugation: Preliminary results in the dutch network using axle box acceleration measurements. 2015.

[160] Y. Li, K. Feng, X. Liang, and M. J. Zuo. A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved vold-kalman filter and multi-scale sample entropy. *Journal of Sound and Vibration*, 439, 2019.

[161] Z. Li, R. Dollevoet, M. Molodova, and X. Zhao. Squat growth-some observations and the validation of numerical predictions. *Wear*, 271, 2011.

[162] Z. Li, M. Molodova, and R. Dollevoet. An investigation of the possibility to use axle box acceleration for condition monitoring of welds. In P. Sas and B. Bergen, editors, *Proceedings of ISMA2008*, pages 2879–2886. Katholieke Universiteit Leuven, Sept. 2008. 2008 International Conference on Noise and Vibration Engineering, ISMA 2008, ISMA ; Conference date: 15-09-2008 Through 17-09-2008.

[163] Z. Li, X. Zhao, R. Dollevoet, and M. Molodova. Differential wear and plastic deformation as causes of squat at track local stiffness change combined with other track short defects. *Vehicle System Dynamics*, 46(sup1):237–246, Sept. 2008.

[164] Z. J. Li, L. Wei, H. Y. Dai, J. Zeng, and Y. J. Wang. Identification method of wheel flat based on hilbert-huang transform. *Jiaotong Yunshu Gongcheng Xuebao/Journal of Traffic and Transportation Engineering*, 12, 2012.

[165] B. Lichtberger. Track compendium. *Hamburg, Eurailpress Tetzlaff-Hestra GmbH & Co. Publ*, 2005.

[166] N. Lillin and S. Rapp. Frühzeitige detektion von punktuellen instabilitäten am bahnkörper. *EI, Der Eisenbahningenieur*, 2018.

[167] P. Linke, D.-I. V. Hoksch, K. Wolter, B. Werning, H. Mösken, and L. Uebel. Monitoring und zustandsorientierte instandhaltung von schienenfahrzeugen und -fahrweg mittels mustererkennung in ereignisdaten. *Ausgabe Sonderheft Graz 2016*, 140, 2016.

[168] J. Liu, S. Chen, G. Lederman, D. B. Kramer, H. Y. Noh, J. Bielak, J. H. Garrett, J. Kovacevic, and M. Berges. The DR-Train dataset: dynamic responses, GPS positions and environmental conditions of two light rail vehicles in Pittsburgh, 10 2018. This material is also based on work supported by a University Transportation Center grant (DTRT12-G-UTC11) from the US Department of Transportation.

[169] J. Liu, S. Chen, G. Lederman, D. B. Kramer, H. Y. Noh, J. Bielak, J. H. Garrett, J. Kovačević, and M. Bergés. Dynamic responses, GPS positions and environmental conditions of two light rail vehicles in pittsburgh. *Scientific Data*, 6(1), aug 2019.

[170] J. Liu and Y. Wei. Detecting anomalies in longitudinal elevation of track geometry using train dynamic responses via a variational autoencoder. *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, 10970, 2019.

[171] X. Y. Liu, S. Alfi, and S. Bruni. An efficient recursive least square-based condition monitoring approach for a rail vehicle suspension system. *Vehicle System Dynamics*, 54:814–830, 6 2016.

[172] Y. Liu and X. Wei. Track surface defect detection based on image processing. In *Lecture Notes in Electrical Engineering*, pages 225–232. Springer Singapore, 2018.

[173] L. Ljung. Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control*, 24, 1979.

[174] A. Lye, A. Cicirello, and E. Patelli. Sampling methods for solving bayesian model updating problems: A tutorial. *Mechanical Systems and Signal Processing*, 159:107760, 2021.

[175] M. Molodova, L. Zili and et al. Automatic detection of squats in railway infrastructure. *IEEE T INTELL TRANS*, 15(5)(1980–1990), 2014.

[176] J. Maes and H. Sol. A double tuned rail damper–increased damping at the two first pinned–pinned frequencies. *Journal of Sound and Vibration*, 267(3):721–737, 2003. Proceedings of the Seventh International Workshop on Railway Noise.

[177] K. Maharana, S. Mondal, and B. Nemade. A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3(1):91–99, June 2022.

[178] A. Malekjafarian, E. OBrien, P. Quirke, and C. Bowe. Railway track monitoring using train measurements: An experimental case study. *Applied Sciences*, 9(22):4859, nov 2019.

[179] S. Marschnig, M. Fellinger, M. Quirchmair, and H. Loy. Optimising the track bedding stiffness and settlement behaviour at insulated rail joints. *Rail Engineering International*, 2020(04):9–14, 2020.

[180] S. Marschnig and P. Veit. *SBB Standardelemente Gleise und Weichen.* ., 2015.

[181] P. Martakis, A. Movsessian, Y. Reuland, S. G. S. Pai, S. Quqa, D. Garcia Cava, D. Tcherniak, and E. Chatzi. A semi-supervised interpretable machine learning framework for sensor fault detection. *Smart Structures and Systems, An International Journal*, Sept. 2021.

[182] R. McElreath. *Statistical rethinking: A Bayesian course with examples in R and Stan.* Chapman and Hall/CRC, 2020.

[183] P. Mcmahon, T. Zhang, and R. Dwight. Requirements for big data adoption for railway asset management. *IEEE Access*, 8:15543–15564, 2020.

[184] M. Mehrali, M. Esmaeili, and S. Mohammadzadeh. Application of data mining techniques for the investigation of track geometry and stiffness variation. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 234(5):439–453, Apr. 2019.

[185] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[186] W. Michels. Theoretical investigations of the running stability of a wheelset. *Status Seminar III. Track guided long-distance traffic. Wheel/rail-technique. Reports*, (13):1–31, 1976.

[187] M. Miskolczi, D. Földes, A. Munkácsy, and M. Jászberényi. Urban mobility scenarios until the 2030s. *Sustainable Cities and Society*, 72:103029, Sept. 2021.

[188] M. Molodova, Z. Li, and R. Dollevoet. Axle box acceleration: Measurement and simulation for detection of short track defects. *Wear*, 271:349–356, 5 2011.

[189] M. Molodova, Z. Li, A. N. nez, and R. Dollevoet. Parametric study of axle box acceleration at squats. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 229, 2015.

[190] M. Molodova, M. Oregui, A. N. nez, Z. Li, and R. Dollevoet. Health condition monitoring of insulated joints based on axle box acceleration measurements. *Engineering Structures*, 123, 2016.

[191] C. Monitoring. of railway track systems by using acceleration signals on wheelset axle-boxes. *Transport*, 32(2):555–566, 2018.

[192] G. Morel and A. Quibel. The portancemetre: a new continuous control apparatus for capping layers and subgrades. *Revue Générale des Routes*, (768), 1998.

[193] C. Mylonas, I. Abdallah, and E. Chatzi. Surrogate modelling for fatigue damage of wind-turbine blades using polynomial chaos expansions and non-negative matrix factorization. *39th IABSE Symposium*, pages 19–23, 2017.

[194] C. Mylonas and E. Chatzi. Remaining useful life estimation under uncertainty with causal graphnets, 2020.

[195] A. Myronenko and X. Song. Point set registration: Coherent point drifts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32, 2010.

[196] L. Müller. Effiziente instandhaltung: Überwachung der infrastruktur mit regelzügen. *DB Systemtechnik*, pages 1–6, 2015.

[197] J. B. Nagel and B. Sudret. A unified framework for multilevel uncertainty quantification in bayesian inverse problems. *Probabilistic Engineering Mechanics*, 43:68–84, Jan. 2016.

[198] M. C. Nakhaee, D. Hiemstra, M. Stoelinga, and M. van Noort. The recent applications of machine learning in rail track maintenance: A survey. In *Reliability, Safety, and Security of Railway Systems. Modelling, Analysis, Verification, and Certification*, pages 91–105. Springer International Publishing, 2019.

[199] R. M. Nejad and F. Berto. Fatigue fracture and fatigue life assessment of railway wheel using non-linear model for fatigue crack growth. *International Journal of Fatigue*, 153:106516, dec 2021.

[200] I. Nerlich, J. Holzfeind, and K. Wilczek. Swisstamp – big data in proactive track asset management. *European Railway Review*, 2(6), 2016.

[201] NeTIRail-WP4. Development of technology for track and ride quality monitoring. *H2020-MG-2015*, GA-636237, 2016.

[202] NextSense. *Calipri C4X*, 2022.

[203] A. K. Ng, L. Martua, and G. Sun. Dynamic modelling and acceleration signal analysis of rail surface defects for enhanced rail condition monitoring and diagnosis. In *2019 4th International Conference on Intelligent Transportation Engineering (ICITE)*, pages 69–73, 2019.

[204] J. C. Nielsen, E. G. Berggren, A. Hammar, F. Jansson, and R. Bolmsvik. Degradation of railway track geometry – correlation between track stiffness gradient and differential settlement. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 234:108–119, 1 2020.

[205] J. C. Nielsen, A. Ekberg, and R. Lundén. Influence of short-pitch wheel/rail corrugation on rolling contact fatigue of railway wheels. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 219, 2005.

[206] J. C. Nielsen and A. Johansson. Out-of-round railway wheels-a literature survey. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 214, 2000.

[207] L. D. A. no Valencia and E. N. Chatzi. Sensitivity driven robust vibration-based damage diagnosis under uncertainty through hierarchical bayes time-series representations. volume 199, 2017.

[208] C. Northcutt, L. Jiang, and I. Chuang. Confident learning: Estimating uncertainty in dataset labels. *J. Artif. Int. Res.*, 70:1373–1411, may 2021.

[209] T. Oba, K. Yamada, N. Okada, H. Soma, and K. Tanifuji. Condition monitoring for shinkansen bogies based on vibration analysis. *Nihon Kikai Gakkai Ronbunshu, C Hen/Transactions of the Japan Society of Mechanical Engineers, Part C*, 75, 2009.

[210] E. J. O'Brien, C. Bowe, and P. Quirke. Determination of vertical alignment of track using accelerometer readings. *International Mechanical Engineering Stephenson Conference for Railways: Research for Railways*, pages 21–23, 2015.

[211] E. J. OBrien, C. Bowe, P. Quirke, and D. Cantero. Determination of longitudinal profile of railway track using vehicle-based inertial readings. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 231, 5 2017.

[212] E. J. OBrien, P. Quirke, C. Bowe, and D. Cantero. Determination of railway track longitudinal profile using measured inertial response of an in-service railway vehicle. *Structural Health Monitoring*, 17:1425–1440, 11 2018.

[213] P. I. U. of Railways. Rail defects. *UIC Code 712*, 4th edition:106–107, 2002.

[214] S. Offenbacher, J. Neuhold, P. Veit, and M. Landgraf. Analyzing major track quality indices and introducing a universally applicable tqi. *Applied Sciences (Switzerland)*, 10, 2020.

[215] M. Ögren. Noise emission from railway traffic. *Swedish National Road and Transport Research Institute (VTI)*, 559A, 2006.

[216] K. Oh, M. Yoo, N. Jin, J. Ko, J. Seo, H. Joo, and M. Ko. A review of deep learning applications for railway safety. *Applied Sciences*, 12(20):10572, Oct. 2022.

[217] A. Onat, P. Voltr, and M. Lata. A new friction condition identification approach for wheel–rail interface. *International Journal of Rail Transportation*, 5, 2017.

[218] M. Oregui, Z. Li, and R. Dollevoet. Identification of characteristic frequencies of damaged railway tracks using field hammer test measurements. *Mechanical Systems and Signal Processing*, 54:224 – 242, 2015.

[219] A. Paixão, E. Fortunato, and R. Calçada. Smartphone's sensing capabilities for onboard railway track monitoring: Structural performance and geometrical degradation assessment. *Advances in Civil Engineering*, 2019:1–13, feb 2019.

[220] R. D. Palmer. *Maintenance planning and scheduling handbook*. McGraw-Hill Education, 2013.

[221] C. Papadimitriou and D.-C. Papadioti. Component mode synthesis techniques for finite element model updating. *Computers and Structures*, 126:15–28, 2013. Uncertainty Quantification in structural analysis and design: To commemorate Professor Gerhart I. Schueller for his life-time contribution in the area of computational stochastic mechanics.

[222] M. P. Papaelias, C. Roberts, and C. L. Davis. A review on non-destructive evaluation of rails: State-of-the-art and future development. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 222, 2008.

[223] M. J. Pappaterra, F. Flammini, V. Vittorini, and N. Bešinović. A systematic review of artificial intelligence public datasets for railway applications. *Infrastructures*, 6(10), 2021.

[224] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[225] L. Peng, S. Zheng, P. Li, Y. Wang, and Q. Zhong. A comprehensive detection system for track geometry using fused vision and inertia, 2021.

[226] A. Pieringer and W. Kropp. Model-based estimation of rail roughness from axle box acceleration. *Applied Acoustics*, 193:108760, May 2022.

[227] O. Polach. Influence of locomotive tractive effort on the forces between wheel and rail. *Vehicle System Dynamics*, 35(1):7–22, 2001.

[228] Y. Qin, M. Liu, and H. Fu. Small excitation self-powered sensing energy harvester for rail traffic condition monitoring. *Journal of Physics: Conference Series*, 2369(1):012087, nov 2022.

[229] A. Quibel. Deliverable d2. 1.2 adapted "portancemetre" for track structure stiffness measurement on existing tracks. *Innotrack*, TIP5-CT-2006-031415, 2007.

[230] P. Quirke, D. Cantero, E. J. Obrien, and C. Bowe. Drive-by detection of railway track stiffness variation using in-service vehicles. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 231:498–514, 4 2017.

[231] J. Real, P. Salvador, L. Montalban, and M. Bueno. Determination of rail vertical profile through inertial methods. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit,*, 225(1):14–23, 2011.

[232] J. I. Real, L. Gómez, L. Montalbán, and T. Real. Study of the influence of geometrical and mechanical parameters on ballasted railway tracks design. *Journal of Mechanical Science and Technology*, 26:2837–2844, 9 2012.

[233] D. Rees. Summarizing data by numerical measures, 2020.

[234] S. Rezvanizaniani, M. Valibeigloo, M. Asghari, J. Barabady, and U. Kumar. Reliability centered maintenance for rolling stock: A case study in coaches; wheel sets of passenger trains of iranian railway. In *2008 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, Dec. 2008.

[235] B. Ripke and K. Knothe. Simulation of high frequency vehicle-track interactions. *Vehicle System Dynamics*, 24(sup1):72–85, 1995.

[236] A. Rohrbeck. Der Messradsatz – ein Sensor wie jeder andere? *ZEVrail*, 136(9), 2012.

[237] M. Roth, B. Baasch, P. Havrila, and J. Groos. Map-supported positioning enables in-service condition monitoring of railway tracks. 2018.

[238] A. Rytter. Vibrational based inspection of civil engineering structures. 1993.

[239] M. Sadri, T. Lu, and M. Steenbergen. Railway track degradation: The contribution of a spatially variant support stiffness - local variation. *Journal of Sound and Vibration*, 455:203–220, Sept. 2019.

[240] J. Sainz-Aja, J. Pombo, D. Tholken, I. Carrascal, J. Polanco, D. Ferreño, J. Casado, S. Diego, A. Perez, J. E. A. Filho, A. Esen, T. M. Cebasek, O. Laghrouche, and P. Woodward. Dynamic calibration of slab track models for railway applications using full-scale testing. *Computers & Structures*, 228:106180, Feb. 2020.

[241] J. Salvatier, T. Wiecki, and C. Fonnesbeck. Probabilistic programming in python using pymc, 2015.

[242] M. Z. Sarwar and D. Cantero. Deep autoencoder architecture for bridge damage assessment using responses from several vehicles. *Engineering Structures*, 246, 2021.

[243] G. Saussine, C. Cholet, P. E. Gautier, F. Dubois, C. Bohatier, and J. J. Moreau. Modelling ballast behaviour under dynamic loading. part 1: A 2d polygonal discrete element method approach. *Computer Methods in Applied Mechanics and Engineering*, 195:2841–2859, 4 2006.

[244] SBB. Catalog of deviations for track inspectors. *Surveillance of installations, ZMON*, 2019.

[245] SBB/CFF/FFS. Einbau, kontrollen und unterhalt von gleisen. *Regelwerk Technik Eisenbahn RTE*, 2022.

[246] R. Schenkendorf and B. Dutschk. Improved railway track irregularities classification by a model inversion approach. *Third European Conference of the Prognostics and Health Management Society 2016*, 2016.

[247] P. Schmid, J. Casutt, and M. Zurkirchen. Künstliche intelligenz auf schienen. *Bulletin*, 9, 2019.

[248] O. Schwabe and H. Berg. Messradsatz für schienenfahrzeuge, May 2007.

[249] Schweizerische Bundesbahnen AG. Sbb geschäftsbericht 2021. *PressSBB.ch*, pages 29–31, 2021.

[250] M. Shadfar, H. Molatefi, and A. Nasr. An index for rail weld health assessment in urban metro using in-service train. *Mathematical Problems in Engineering*, 2022:1–10, Dec. 2022.

[251] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[252] C. Shen, R. Dollevoet, and Z. Li. Fast and robust identification of railway track stiffness from simple field measurement. *Mechanical Systems and Signal Processing*, 152:107431, May 2021.

[253] Y. Shokrani, V. K. Dertimanis, E. N. Chatzi, and M. N. Savoia. On the use of mode shape curvatures for damage localization under varying environmental conditions. *Structural Control and Health Monitoring*, 25, 2018.

[254] K. Shridhar, F. Laumann, and M. Liwicki. Uncertainty estimations by softplus normalization in Bayesian convolutional neural networks with variational inference. *arXiv*, 2018.

[255] M. L. Sichitiu and C. Veerarittiphan. Simple, accurate time synchronization for wireless sensor networks. volume 2, 2003.

[256] SIMPACK. SIMPACK documentation, 2021. DASSAULT Système Simulia Corp.

[257] T. Simpson, V. K. Dertimanis, and E. N. Chatzi. Towards data-driven real-time hybrid simulation: Adaptive modeling of control plants. *Frontiers in Built Environment*, 6, Sept. 2020.

[258] R. Skrypnyk, U. Ossberger, B. A. Pålsson, M. Ekh, and J. C. Nielsen. Long-term rail profile damage in a railway crossing: Field measurements and numerical simulations. *Wear*, 472:203331, 2021.

[259] S. Smith. Smart infrastructure for future urban mobility. *AI Magazine*, 41(1):5–18, Apr. 2020.

[260] S. Sob. *Südostbahn mit Onboard-Monitoring für die Infrastruktur*. DB News, 2018.

[261] H. Sohn, D. W. Allen, K. Worden, and C. R. Farrar. Structural damage classification using extreme value statistics. *Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME*, 127, 2005.

[262] G. Soldati. EMW Kiesen 2006 – Gleiseinsenkungen mit dem Einsenkungsmesswagen Achslast 20t. *SBB-report*, 2007.

[263] Y. Song, L. Liang, Y. Du, and B. Sun. Railway polygonized wheel detection based on numerical time-frequency analysis of axle-box acceleration. *Applied Sciences (Switzerland)*, 10, 2020.

[264] J. Sresakoolchai and S. Kaewunruen. Detection and severity evaluation of combined rail defects using deep learning. *Vibration*, 4(2):341–356, Apr. 2021.

[265] J. Sresakoolchai and S. Kaewunruen. Railway defect detection based on track geometry using supervised and unsupervised machine learning. *Structural Health Monitoring*, 21(4):1757–1767, 2022.

[266] M. Steenbergen. Physics of railroad degradation: The role of a varying dynamic stiffness and transition radiation processes. *Computers & Structures*, 124(102-111), 2013.

[267] M. Steenbergen and E. de Jong. Railway track degradation: The contribution of rolling stock. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 230(1164-1171), 2016.

[268] M. Steenbergen and R. Dollevoet. On the mechanism of squat formation on train rails – part i: Origination. *International Journal of Fatigue*, 47:361–372, 2013.

[269] M. J. M. M. Steenbergen and C. Esveld. Rail weld geometry and assessment concepts. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 220(3):257–271, 2006.

[270] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Sys. Safety*, 93:964–979, 2008.

[271] T. R. Sussman, W. Ebersöhn, and E. T. Selig. Fundamental nonlinear track load-deflection behavior for condition evaluation. *Transportation Research Record: Journal of the Transportation Research Board*, 1742(1):61–67, Jan. 2001.

[272] H. Tanaka and M. Miwa. Modeling the development of rail corrugation to schedule a more economical rail grinding. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 234, 2020.

[273] K. Tatsis, V. Dertimanis, C. Papadimitriou, E. Lourens, and E. Chatzi. A general substructure-based framework for input-state estimation using limited output measurements. *Mechanical Systems and Signal Processing*, 150:107223, 2021.

[274] H. Thomas and T. Heckel. Advantage of a combined ultrasonic and eddy current examination for railway inspection trains. *OR Insight*, 49(6):341–344, 2006.

[275] W. Thomson and A. E. Roth. The shapley value: Essays in honor of lloyd s. shapley. *Economica*, 58, 1991.

[276] Y. Tong, G. Liu, K. Yousefian, and G. Jing. Track vertical stiffness –value, measurement methods, effective parameters and challenges: A review. *Transportation Geotechnics*, 37:100833, Nov. 2022.

[277] P. T. Torstensson and J. C. Nielsen. Simulation of dynamic vehicle-track interaction on small radius curves. volume 49, 2011.

[278] P. T. Torstensson, G. Squicciarini, M. Krüger, B. A. Pålsson, J. C. Nielsen, and D. J. Thompson. Wheel–rail impact loads and noise generated at railway crossings – influence of vehicle speed and crossing dip angle. *Journal of Sound and Vibration*, 456, 2019.

[279] G. Tsialiamanis, C. Mylonas, E. N. Chatzi, D. J. Wagg, N. Dervilis, and K. Worden. On an application of graph neural networks in population-based SHM. In *Data Science in Engineering, Volume 9*, pages 47–63. Springer International Publishing, May 2021.

[280] H. Tsunashima. Condition monitoring of railway tracks from car-body vibration using a machine learning technique. *Applied Sciences 9*, 13(2734), 2019.

[281] H. Tsunashima and R. Hirose. Condition monitoring of railway track from car-body vibration using time–frequency analysis. *Vehicle System Dynamics*, 2020.

[282] H. Tsunashima, Y. Naganuma, and T. Kobayashi. Track geometry estimation from car-body vibration. *Vehicle System Dynamics*, 52:207–219, 5 2014.

[283] H. Tsunashima and M. Takikawa. Monitoring the condition of railway tracks using a convolutional neural network. In F. Bulnes, editor, *Recent Advances in Wavelet Transforms and Their Applications*, chapter 6. IntechOpen, Rijeka, 2022.

[284] J. Tuma. Setting the passband width in the vold-kalman order tracking filter. *Proceedings of ICSV, Lisbon, Portugal*, 12, 2005.

[285] K. Tzanakakis. The effect of track stiffness on track performance. In *Springer Tracts on Transportation and Traffic*, pages 79–87. Springer Berlin Heidelberg, 2013.

[286] K. Tzanakakis. *The Railway Track and Its Long Term Behaviour, The Effect of Track Stiffness on Track Performance*. Number 79-87. Springer, 2013.

[287] M. D. Ulriksen, D. Tcherniak, and L. Damkilde. Damage detection in an operating vestas v27 wind turbine blade by use of outlier analysis. pages 50–55. Institute of Electrical and Electronics Engineers Inc., 8 2015.

[288] UVEK. Ausführungsbestimmungen zur eisenbahnverordnung (ab-ebv). *Bundesamt für Bauten und Logistik (BBL)*, SR 742.141.11, 2020.

[289] C. Vale, I. M. Ribeiro, and R. Calçada. Integer programming to optimize tamping in railway tracks as preventive maintenance. *Journal of Transportation Engineering*, 138(1):123–131, Jan. 2012.

[290] C. Vale and M. L. Simões. Prediction of railway track condition for preventive maintenance by using a data-driven approach. *Infrastructures*, 7(3):34, Mar. 2022.

[291] K. van der Wiel, N. Wanders, F. M. Selten, and M. F. Bierkens. Added value of large ensemble simulations for assessing extreme river discharge in a 2 °c warmer world. *Geophysical Research Letters*, 46:2093–2102, 2 2019.

[292] H. Vold, H. Herlufsen, M. Mains, and D. Corwin-Renner. Multi axle order tracking with the vold-kalman tracking filter. *Sound and Vibration 31*, 5(30–35), 1997.

[293] VöV. *Einbau, Kontrollen und Unterhalt der Weichen, R RTE-22066*. Regelwerk Technik Eisenbahn, Bundesamt für Verkehr BAV, 2015.

[294] P.-R. Wagner, J. Nagel, S. Marelli, and B. Sudret. UQLab user manual – Bayesian inversion for model calibration and validation. Technical report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2022. Report UQLab-V2.0-113.

[295] H. Wang, J. Berkers, N. Hurk, and N. Farsad Layegh. Study of loaded versus unloaded measurements in railway track inspection. *Measurement*, 169, 10 2020.

[296] K. S. Wang, D. Guo, and P. S. Heyns. The application of order tracking for vibration analysis of a varying speed rotor with a propagating transverse crack. *Engineering Failure Analysis*, 21, 2012.

[297] K. S. Wang and P. S. Heyns. Vold-kalman filter order tracking in vibration monitoring of electrical machines. *JVC/Journal of Vibration and Control*, 15, 2009.

[298] P. Wang, L. Wang, R. Chen, J. Xu, J. Xu, and M. Gao. Overview and outlook on railway track stiffness measurement. *Journal of Modern Transportation*, 24, 2016.

[299] W. Wangqing, Z. Geming, Z. Kaiming, and L. Lin. Development of inspection car for measuring railway track elasticity. In *Proceedings from 6th international heavy haul conference, Cape Town*, 1997.

[300] X. Wei, L. Jia, and H. Liu. A comparative study on fault detection methods of rail vehicle suspension systems based on acceleration measurements. *Vehicle System Dynamics*, 51:700–720, 2013.

[301] X. Wei, Y. Liu, and X. Yin. Detection of rail squats based on hilbert-huang transform by using bogie acceleration measurement. 2016.

[302] X. Wei, X. Yin, Y. Hu, Y. He, and L. Jia. Squats and corrugation detection of railway track based on time-frequency analysis by using bogie acceleration measurements. *Vehicle System Dynamics*, 58(8):1167–1188, Apr. 2019.

[303] T. Weinold and A. Grimm-Pitzinger. Die lagerung der gleisvermessungen der öbb. *Vermessung & Geoinformation*, 3(10), 2012.

[304] P. Weston, C. Roberts, G. Yeo, and E. Stewart. Perspectives on railway track geometry condition monitoring from in-service railway vehicles. *Vehicle System Dynamics*, 53:1063–1091, 7 2015.

[305] K. Wilczek, S. Züger, and J. Holzfeind. SwissTAMP – Big Data für ein proaktives Anlagenmanagement Fahrweg. *ZEVrail*, 141(5), 2017.

[306] D. Winklehner. Monitoring und zustandsorientierte instandhaltung von schienenfahrzeugen und -fahrweg mittels mustererkennung in ereignisdaten. *ZEVrail*, 139(4), 2015.

[307] H. Wold. A study in the analysis of stationary time series. *Ph.D. thesis, Stockholm*, 1938.

[308] Wold S., Esbensen K., and Geladi P. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(37–52), 1987.

[309] K. Worden, G. Manson, and D. Allman. Experimental validation of a structural health monitoring methodology: Part i. novelty detection on a laboratory structure. *Journal of Sound and Vibration*, 259, 2003.

[310] Q. Wu, C. Cole, and T. McSweeney. Applications of particle swarm optimization in the railway domain. *International Journal of Rail Transportation*, 4(3):167–190, 2016.

[311] X. Wu, W. Cai, M. Chi, L. Wei, H. Shi, and M. Zhu. Investigation of the effects of sleeper-passing impacts on the high-speed train. *Vehicle System Dynamics*, 53(12):1902–1917, 2015.

[312] B. Xiao, X. Mao, J. Liu, L. Niu, X. Xu, and M. Zhang. An improved marginal index method to diagnose poor welded joints of heavy-haul railway. In *2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing)*, pages 1–7, 2021.

[313] J. Xie, J. Huang, C. Zeng, S. H. Jiang, and N. Podlich. Systematic literature review on data-driven models for predictive maintenance of railway track: Implications in geotechnical engineering. *Geosciences (Switzerland)*, 10, 2020.

[314] K. Xue, T. Nagayama, and B. Zhao. Road profile estimation and half-car model identification through the automated processing of smartphone data. *Mechanical Systems and Signal Processing*, 142.

[315] A. M. Y. Sato and K. Knothe. Review on rail corrugation studies. *Wear*, 253:130–139, 2002.

[316] T. H. Yan and F. Corman. Assessing and extending track quality index for novel measurement techniques in railway systems. *Transportation Research Record*, 2674, 2020.

[317] T.-H. Yan, M. De Almeida Costa, and F. Corman. Developing and extending status prediction models for railway tracks based on on-board monitoring data. 2023-01. 102nd Annual Meeting of the Transportation Research Board (TRB 2023); Conference Location: Washington, DC, USA; Conference Date: January 8-12, 2023; OMISM (MI Mobility Initiative project).

[318] C. Yang, Y. Sun, C. Ladubec, and Y. Liu. Article developing machine learning-based models for railway inspection. *Applied Sciences (Switzerland)*, 11, 2021.

[319] C. Yang, Y. Sun, C. Ladubec, and Y. Liu. Developing machine learning-based models for railway inspection. *Applied Sciences*, 11(1), 2021.

[320] N. Yao, Y. Jia, and K. Tao. Rail weld defect prediction and related condition-based maintenance. *IEEE Access*, 8:103746–103758, 2020.

[321] T. C. Yeh and M. C. Pan. Online real-time monitoring system through using adaptive angular-velocity vkf order tracking. 2017.

[322] Z. Yuan, S. Zhu, X. Yuan, and W. Zhai. Vibration-based damage detection of rail fastener clip using convolutional neural network: Experiment and simulation. *Engineering Failure Analysis*, 119:104906, Jan. 2021.

[323] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13, 1994.

[324] Y. Zhao, B. Liang, and S. Iwnicki. Friction coefficient estimation using an unscented kalman filter. volume 52, 2014.

[325] H. Zhu, H. Li, A. Al-Juboori, D. Wexler, C. Lu, A. McCusker, J. McLeod, S. Pannila, and J. Barnes. Understanding and treatment of squat defects in a railway network. *Wear*, 442-443:203139, feb 2020.

[326] L. Zili, M. Molodova, A. Nunez, and R. Dollevoet. Improvements in axle box acceleration measurements for the detection of light squats in railway infrastructure. *IEEE Transactions on Industrial Electronics*, 62(7):4385–4396, 2011.

[327] Y. Zuo, F. Thiery, P. Chandran, J. Odelius, and M. Rantatalo. Squat detection of railway switches and crossings using wavelets and isolation forest. *Sensors*, 22(17):6357, aug 2022.