



Adaptive time series representation for out-of-round railway wheels fault diagnosis in wayside monitoring

Afonso Lourenço^{a,*}, Carolina Ferraz^a, Diogo Ribeiro^c, Araliya Mosleh^b, Pedro Montenegro^b, Cecília Vale^b, Andreia Meixedo^b, Goreti Marreiros^a

^a GECAD, Polytechnic of Porto, School of Engineering, Porto, Portugal

^b CONSTRUCT-LESE, University of Porto, Faculty of Engineering, Porto, Portugal

^c CONSTRUCT-LESE, Polytechnic of Porto, School of Engineering, Porto, Portugal



ARTICLE INFO

Keywords:

Machine Learning
Fault diagnosis
Wheel Out-of-roundness
Wayside monitoring
Artificial Intelligence

ABSTRACT

Through the integration of advanced sensor technologies and machine learning algorithms, artificial intelligence has revolutionized wayside monitoring in the railway sector. Although several algorithms have been proposed for detecting out-of-roundness, i.e., flats, wear treads and polygonization, they generally fall short in isolating the root cause of the wheel's issue in a train passage. In this context, the paper presents a novel approach for wheel out-of-roundness diagnosis with (1) detection of aberrant train behavior; (2) isolation of specific defective wheels; (3) identification of the severity. For this, the methodology automatically segments a strain gauge signal, capturing the complex nature and temporal dependence of vibration patterns. This segmentation allows the extraction of localized accelerometer features in both the time and frequency domain, as well as implicit axle count and labelling of each wheel passage. Moreover, a single-value damage indicator based on anomaly detection algorithms was proposed. To validate the effectiveness of the proposed methodology, experiments on a set 3D numerical train-track dynamic interaction simulations are performed for different wheel profiles, track irregularities, train speeds, sensor placement and noise, associated to other environmental and operational variations. This demonstrates the potential of artificial intelligence for real-time assessment of wheels without interfering with normal service conditions, suggesting the possibility of automated fault diagnosis.

© 2023 Elsevier Inc. All rights reserved.

1. Introduction

The railway sector plays a vital role in transporting passengers and goods, and as such, reliability and maintainability are crucial to ensure continuous operations [1]. With trains designed to operate for decades and maintenance costs accounting for a significant portion of the total expenditure, there is an urgent need to keep safety and quality of service at an optimal level while minimizing the operational costs. In particular, the small wheel-rail contact patch has received a lot of attention in the literature, due to its increasing hazard function and associated safety risk [2–4]. The high stick and sliding stresses in rolling contact leads to wheel out-of-roundness, typically manifested in isolated wheel flats, corrugated wheel treads, and polygonal wheels. These can lead to high-impact forces that

* Corresponding author.

E-mail address: fonso@isep.ipp.pt (A. Lourenço).

pose a significant risk of train derailment, generate substantial noise, and aggravate the development of track defects [5,6]. Therefore, neglecting the issue of wheel out-of-roundness can have serious consequences, which is why it is imperative to address it promptly through effective maintenance and monitoring practices. To address this issue, the railway industry often implements a preventive maintenance strategy, which can be based on either tonnage or kilometers traveled [7]. This approach involves routine inspections and maintenance activities to stop wheel abnormalities from developing into more critical problems, putting wheels among the most expensive components to maintain [8]. However, the effectiveness of preventive maintenance relies on various operational and environmental factors such as lighting conditions, meteorological variables, and the physical and mental state of inspectors [9]. Even with extensive experience, identifying defects such as small wheel flats remain a laborious task, and they are easily overlooked during inspections [10].

With the Internet of Things rapidly expanding, many opportunities arise for the implementation of advanced condition monitoring systems in the railway sector. One of the most promising technologies is vibration-based systems [11]. There are two main categories for wheel damage detection: onboard and wayside measurement methods. Onboard techniques involve installing sensors on the vehicle to monitor the condition of the track and the wheels [12,13]. However, this method is rarely used due to the high costs and maintenance problems associated with installing sensors on in-use vehicles. Alternatively, wayside measurement systems are a more feasible solution to identify wheel defects, as the installation of sensors on the track allows the extraction of significant amounts of data referring to all the operating vehicles, thus enabling the methodology to monitor various vehicles with a reduced number of sensors [14–16]. Nonetheless, pinpointing the actual fault in the rail-wheel interface is a complicated process that involves contact mechanics and several modes of material deterioration. The vehicle-track dynamics and forces to support, accelerate, brake, and guide the train leads to unbounded, unlabeled, noisy, and non-stationary time series measurements. In addition, the ever-changing contact surface alteration and friction wear further complicates the process [17].

Thus, to tackle this issue, advanced signal processing methods have been proposed that extract condensed damage-sensitive knowledge from time-series sensor data. Several contributions were proposed with traditional vibration analysis using time-domain and frequency-domain features [18], with the wavelet transform being the most widely used technique [19–21]. However, all these approaches revealed limitations in dealing with the non-linear effects on the suggested time-dependent characteristics. For instance, changes in train speed tend to affect the impact load, but this can vary based on the specific length of wheel flats, not following the pattern of increased impact load with higher train speeds [22]. Some alternatives were then suggested, such as a time-frequency kurtosis to reduce surrounding noise and highlight faulty signal patterns of wheel flats [23] and empirical mode decomposition (EMD) to transform the signal into several intrinsic mode functions that isolate the failure signal mode from interferences. Also, a criterion based on signal lag and envelope spectrum [24], symbolic data [25], and autoregressive models [26] are techniques used to extract the features. Furthermore, all the aforementioned studies have concentrated on the development of techniques for detecting wheel flats and polygonization. However, to the authors' knowledge only a few works have achieved comprehensive and automated fault diagnosis [27,28] with the three necessary objectives: (1) detection of abnormal train behaviour; (2) isolation of specific defective wheels; (3) identification of the severity for the respective defect type [29]. This need for interpretability of detected faults has been confirmed by recent literature reviews on artificial intelligence for railway applications [30,31]. To address this, building a physical-based model provides the opportunity to validate the full diagnosis task and interpret artificial intelligence models, since it would be cost-prohibitive to induce them in a real railway system. In the last two decades, researchers explored the intricate mechanisms of out-of-roundness formation on wheels [32,33], making it possible to induce reliable faults into a simulation [34]. Numerical modelling allows for a deeper understanding on the prediction of anomalies that may not be feasible to reproduce in experimental measurements affected by external factors, such as environmental disturbances, measurement errors, and electromagnetic interferences.

Based on this research gap, the added value of this paper lies in building an automatic and unsupervised fault diagnosis algorithm for out-of-roundness, with two failure modes: wheel flats and wheel tread polygonization. For this, the algorithm must be capable of extracting knowledge of an accelerometer and strain gauge to isolate patterns in different wheel passages. With a Hidden Markov Model (HMM), it was possible to assess the individual condition of wheels for localized feature extraction. For time-frequency analysis, the Short-time Fourier Transform (STFT) was selected for its ability [35] to address the non-linear environmental and operational effects [35]. Then, with Principal Feature Analysis (PFA) a minimal feature subset was used for an Isolation Forest (iForest) to diagnose the defective wheels, while uncovering the causal relations between specific inputs and specific out-of-roundness failure modes. Results were compared with other state-of-the-art anomaly detection algorithms, i.e., mahalanobis distance, support vector machines and local outlier factor. The developed strategy was validated with 3D numerical train-track dynamic interaction simulations, carried out with the in-house software Vehicle-Structure Interaction Analysis [36]. By defining unknown variables one by one, while observing how the methodology deals with the effect of these unknown parameters in the responses, the decision support system can be deemed reliable. Ultimately, the proposed methodology aims to be generic and with successful validation has the potential to be applied to real experimental data considering different types of trains and wheel defects. The main contributions of this work are threefold:

- development of a methodology capable of wheel fault diagnosis without interfering with normal operating conditions, focusing on the challenging task of fault isolation and localization;
- segmentation of the time series measurements, for implicit indexing and guided extraction of localized features, easily correlated with the damage of an individual wheel;
- test the effectiveness of the proposed method with respect to different types and sizes of out-of-roundness defects with high accuracy, using enough data to confirm generalization.

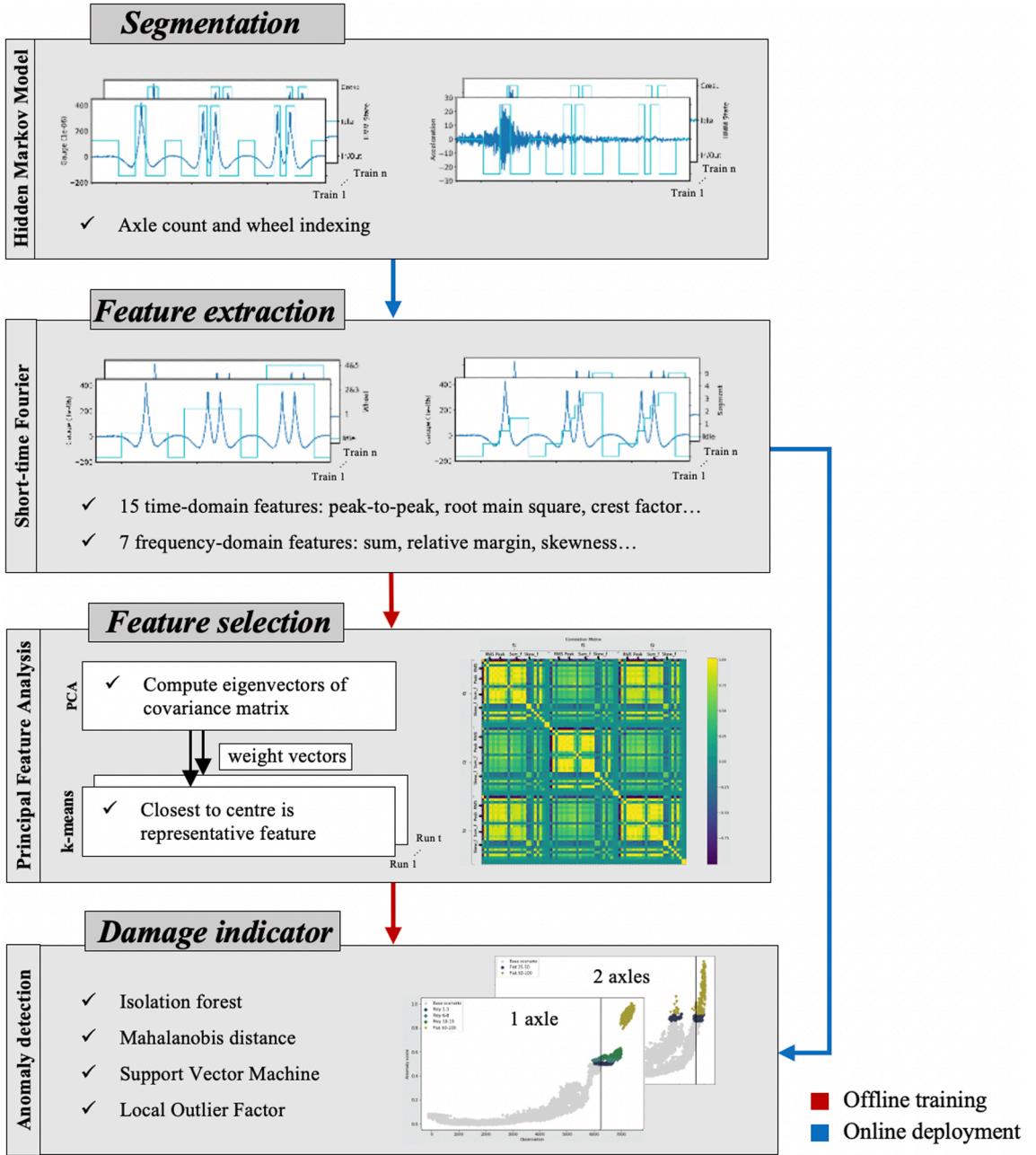


Fig. 1. Flowchart of out-roundness wheel fault diagnosis methodology.

The rest of this paper is organized as follows: Section 2 describes the fault diagnosis methodology. Section 3 explains the process of simulating train-track dynamic interaction and numerical modelling, while section 4 presents and discusses the results obtained. Finally, Section 5 summarizes our main conclusions and describes ideas for future work.

2. Fault diagnosis methodology

2.1. Overview

The suggested method for detecting, localizing, and identifying out-of-roundness a monitoring system is a brand-new, all-encompassing strategy that includes a four-stage process for successful fault diagnosis, depicted in Fig. 1.

The method is intended to solve the challenging nature of localizing out-of-round wheels in a train passage with several wagons and

subsequently identify the severity of the defect individually. The initial stage of the approach involves segmentation using Hidden Markov Models (HMM) to capture the temporal structure of the strain gauge signal, which is frequently less impacted by out-of-roundness damage. This step is vital in enabling the extraction of several localized features from the data, conducted in the second stage. In order to handle non-linear environmental and operational influences, guided Short Time Fourier Transform (STFT) is used to retrieve both statistical and frequency domain characteristics in accelerometer data. The third stage of the approach involves determining a minimal subset of features for fault diagnosis using Principal Feature Analysis (PFA). Finally, a single damage indicator is created with Isolation Forest (iForest) for each group of axles to complete the fault diagnosis task, considering the differences in vibration patterns between the train axle groups. It is worth noting that the proposed approach is designed to be practical and efficient in real-world applications, as only the selected features are extracted in online deployment of the monitoring system.

2.2. Segmentation

Segmentation can be effectively used as a front-end signal processing technique to partition vibration signals into distinct segments. These segmented portions of the data can then undergo subsequent analysis to extract dynamic features indicative of the wheel-rail dynamics. In this context, the Hidden Markov Model (HMM) was selected as it enables high computational efficiency and provides a way to comprehend the signal's underlying structure by associating sequences of observations with distinct hidden internal states [37]. To achieve this, the Baum-Welch algorithm is used under an iterative Expectation-Maximization approach to estimate model parameters from observed data [38]. Additionally, the Viterbi algorithm and the Forward-Backward algorithm are used to specify start probability vectors, transition probability matrices, and model likelihoods [39].

Employing this approach results in consistent and meaningful segments that can be indexed automatically to the corresponding wheel. Ultimately this allows for a more efficient and accurate analysis of the wheel-rail dynamics, identifying cutting points in a way that enables direct comparison of similar patterns in individual wheel passages. For hyperparameter tuning of the number of states, between 2 and 5 gaussian emissions were considered. However, between closely connected wagons, the departure of one wheel interferes with the initial approach of the consecutive wheel's signal. To tackle this partial confounding of consecutive wheels with one another, a solution was devised by grouping the axles and developing two separate pipelines. The group of one axle is designed to handle individual axles, specifically the first and last wheels of a train. The group of two axles is dedicated to managing the pairs of axles located between closely connected wagons. With this segmentation technique, three main purposes were achieved:

- guided short-time windowing for the STFT, making the approach less sensitive to environmental and operational variations;
- localized feature extraction in both the statistical and frequency domain. This was particularly important in handling the temporal dependence of the data, which is a crucial factor in ensuring a more comprehensive understanding of the system and accurate fault detection;
- implicit indexing of each wheel passage proved to be a crucial aspect in the subsequent fault isolation and localization process. This allowed for more targeted and efficient maintenance, resulting in reduced downtime and cost savings.
- guided short-time windowing for the STFT, making the approach less sensitive to environmental and operational variations;
- localized feature extraction in both the statistical and frequency domain. This was particularly important in handling the temporal dependence of the data, which is a crucial factor in ensuring a more comprehensive understanding of the system and accurate fault detection;
- implicit indexing of each wheel passage proved to be a crucial aspect in the subsequent fault isolation and localization process. This allowed for more targeted and efficient maintenance, resulting in reduced downtime and cost savings.

2.3. Time-frequency analysis

In damage detection systems, feature extraction is critical in converting the time series data into alternative information sensitive to damage [40]. However, extracting useful features from raw data poses significant challenges in removing environmental and operational effects from the identified dynamic properties, to make it easier to detect potential damage and take preventative measures [41].

2.3.1. Short-time Fourier transform

To address the presence of non-linear factors in the railway environment, the Short-time Fourier Transform (STFT) was adopted [42]. By computing the Fourier transform for overlapping short-time windows of a segmented signal, STFT creates a power spectrum of the signal over time, providing a more accurate analysis of the magnitude spectra of each time segment. STFT has been found to yield promising results in identifying different types of faults and handling noise in various applications related to wheel condition monitoring, exhibiting comparable performance when compared to commonly used alternatives such as wavelet and Wigner-Ville transform [43–45].

In the field of time-frequency representation theory, a well-known trade-off exists between accuracy in one variable and the other. When using wide time windows, the frequency resolution improves, but the time resolution suffers, making it difficult to precisely locate wheel defects. Conversely, narrow time windows offer better time resolution but distort the signal's frequency content within that timeframe. Therefore, determining the window length requires finding a compromise that balances both time and frequency resolutions. To address this resolution trade-off and enhance fault detection, HMM were employed to define the window duration. This approach provided better time and frequency resolution for different parts of the vibration signal, enabling more precise detection of

faults. The signal was split into HMM-based sequences, and overlapping adjacent segments generated windows by 50%. The resulting number of windows is twice the number of segments minus one, providing a more comprehensive and detailed analysis of the signal. The outcome after applying this segmentation was a variable sized time series, where m segments were obtained for the K scenarios of one axle and p segments for the J scenarios of two axles.

2.3.2. Feature extraction

The segments were used to extract 21 features based on the windowing technique. Experimental studies have shown the potential of these features in enhancing vibration-based structural damage detection methods, particularly under environmental effects. [46]. In the time domain, 14 features were extracted from each segment, including mean, max, min, power, and margin. Envelope metrics, such as peak-to-peak amplitude, root mean square, variance, standard deviation, skewness, kurtosis, crest factor, form factor, and pulse indicator, were also selected to capture the overall shape of the data. In the frequency domain, 7 features were extracted for each window, including mean, peak, sum, and envelope metrics, allowing for detecting magnitude variation over frequency. These included variance, skewness, kurtosis, and relative margin.

This led to the automatic construction of two three-dimensional feature matrices, one for a single axle and the other for two axles. The dimensions of the matrix in groups of one axle K -by-21-by- m , where K is the total number of scenarios of one axle, 21 the total number of time-frequency features and m is the number of segments extracted. Simultaneously, J -by-21-by- p is the matrix for groups of two axles, where J is the total number of scenarios of two axles and p is the number of segments extracted. These matrices contained valuable information for further analysis and damage detection.

2.4. Feature selection

Identifying the minimum feature subset that characterizes incipient failures is crucial for effective monitoring and model interpretation. To achieve this, the Principal Feature Analysis (PFA) algorithm was chosen due to its computational efficiency, generalizability, and suitability for unsupervised settings. This filter method uses the k -means method to select a representative feature set based on the structure of a Principal Component Analysis. The goal is to obtain a principal feature subset with maximum variability in a smaller dimensional space [47].

2.4.1. Principal Component analysis

Environmental factors such as temperature and wind, and operational variables like train speed and type, can significantly influence vibration characteristics. Although the variables are not directly measured, their impact can be detected by analyzing changes in the observed features. To surmount this challenge, a latent variable approach of Principal Component Analysis for feature space resolution can be extremely useful [48]. Through analysis of the matrices, containing the features extracted from the dynamic responses of each of the segments, PCA rotates the original coordinate system using a 21-by-21 orthonormal linear transformation matrix. After obtaining the covariance matrix of the extracted features via singular value decomposition, the general rule of cumulative variance reaching 80% is used to select the eigenvalues [49].

2.4.2. k -means Clustering

To leverage the reduced dimension space for feature selection, k -means algorithm is adopted to regroup the data into arbitrary k clusters. The idea is to find k centroids that minimize the Euclidean distance of each vector to its nearest centroid. This method requires k to be defined in advance. However, as we are using it as an auxiliary tool for feature selection, the range to be considered is not critical. Nonetheless, the global silhouette index is used to validate the quality of the resulting feature space, since it revealed a superior performance in previous studies [50]. Then, we choose the feature vector closest to each centroid as a representative feature. To find the n most distinct features, since k -means is a non-deterministic algorithm, 10,000 runs varying k between 2 and 10 were made. It is important to note that PFA returns the minimal subset in terms of redundancy, not accounting for predictive power. The outcome is matrices K -by- n -by- m and J -by- n -by- p , where n is the number of selected features.

2.5. Damage index

Anomaly detection involves ranking instances whose intrinsic properties deviate from normal behavior, in the form of a damage index. These methods typically use distance and density metrics as a ranking measure [51]. However, these have difficulty in detecting both scattered anomalies, i.e., polygonization defects, and clustered anomalies, i.e., wheel flats. To deal with this masking effect, a mass-based approach is more adequate as it has no regard for the characteristics of the regions, e.g., shape and size [52]. In this work separate anomaly detection models were built for groups of one and two axles to deal with the differences in vibration patterns captured with the segment-based features. The former corresponds to matrix of K -by- m for the first and last wheels, while the latter to the matrix J -by- p for consecutive axles of closely connected wagons.

2.5.1. Isolation Forest

Considering the requirement of real-time monitoring and non-linear relationships represented in the vibration signal, the Isolation Forest (iForest) algorithm's score is leveraged for identifying mass-based anomalies. This approach is preferred for its stability under hyperparameter tuning, computational efficiency, and the absence of assumptions on data distribution [53]. The iForest algorithm operates by constructing multiple binary trees and randomly selecting a feature and split value within a specified range until each

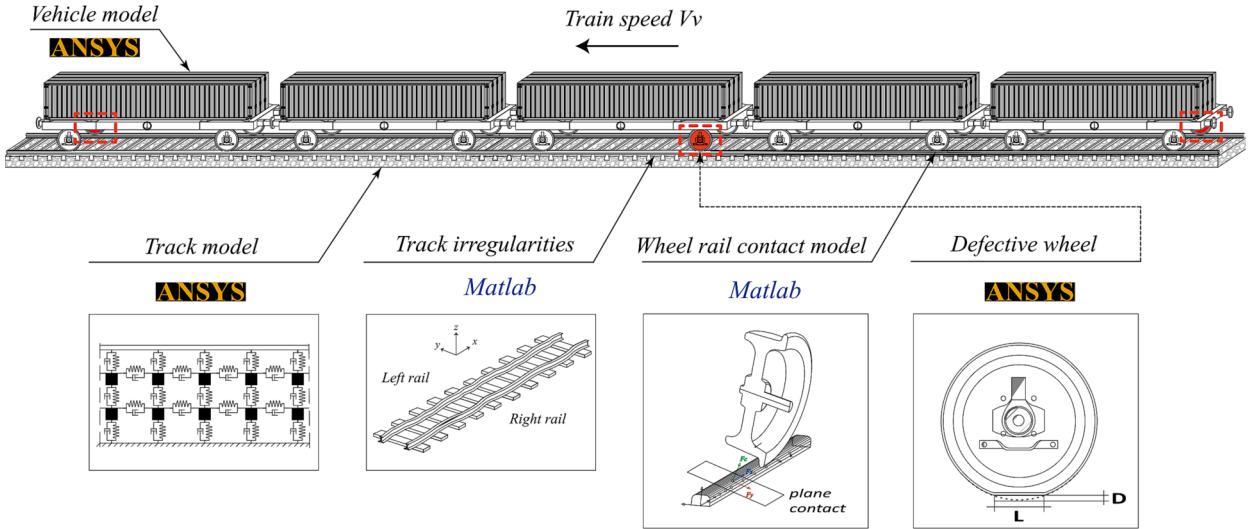


Fig. 2. Train-track dynamic interaction.

terminal leaf contains only one instance or instance with the same value [54]. The number of splits obtained is used to calculate a score that indicates the individual instances' susceptibility to be isolated. Despite its effectiveness, iForest is predisposed to bias towards groups of correlated variables due to the computation of random splits, which was mitigated with the previous step of feature selection [55]. As anomalies represent the events of interest, the iForest ranking score was used for simultaneous fault detection and severity identification while the representative data structure ensured the ability to isolate the location of defects. To inspect the tradeoff between false positives and false negatives, two parameters were considered: the subsampling size ψ from 0.01 to 1 and the number of trees to build t from 50 to 500. Finally, leveraging on the simulated dataset, the isolation paths that led to the decision that a particular data point is an anomaly were analysed. This provides local interpretability by helping uncover the causal relations between specific regions of the feature space and specific fault types, ultimately helping users trust the iForest model.

2.5.2. Benchmark models

To assess the performance of the suggested damage index, its performance was compared to other anomaly detection techniques that have been extensively used in the literature. Three benchmark models were used: the Mahalanobis distance (MD), Support Vector Machine (SVM), and Local Outlier Factor (LOF).

LOF works by examining the local density of each instance in the dataset and comparing it to the density of its neighboring instances [56]. Instances with low density, which are far from other instances, have a higher likelihood of being outliers. However, LOF may be prone to errors in regions with varying densities and skewed distributions, leading to inaccurate anomaly detection. MD is another statistical technique that is commonly used in damage identification studies [26]. It calculates the distance between points in a feature space, considering the correlations between the variables and the scale of each variable [57]. SVM learns a decision boundary that separates normal observations from anomalous observations in high-dimensional feature space [58]. Once the hyperplane is learned, new observations can be classified as normal or anomalous based on their distance to the hyperplane. Despite being robust to imbalanced datasets, this technology has difficulties with complex and nonlinear data distributions.

3. Modeling and simulation

3.1. Train-track dynamic interaction

Numerical simulations of train-track dynamic interaction were conducted using VSI software, which was previously validated and described in detail in the authors' previous publication [36]. The VSI model coupled the train to the track through a 3D wheel-rail contact model that uses the Hertzian theory to compute normal contact forces and the USETAB routine to assess the tangential forces resulting from rolling friction creep [59,60]. The VSI software was implemented in MATLAB® and imports structural matrices from both the vehicle and structure, which were previously modeled separately in a finite element (FE) package.

In this study, the track was modeled using beam elements to represent the rails and sleepers, spring-dashpot elements to simulate the behavior of the flexible layers such as the ballast and fasteners/pad, and mass point elements to consider the mass of the ballast. The train was modeled in ANSYS® through a multibody formulation, using spring-dashpot elements to simulate the flexibility of the primary and secondary suspensions, rigid beams to consider the rigid body movements of the vehicle, and mass point elements located at the center of gravity of each body, including carbody, bogies, and wheelsets, to simulate their mass and inertial effects. A graphical representation of the numerical modeling is presented in Fig. 2.

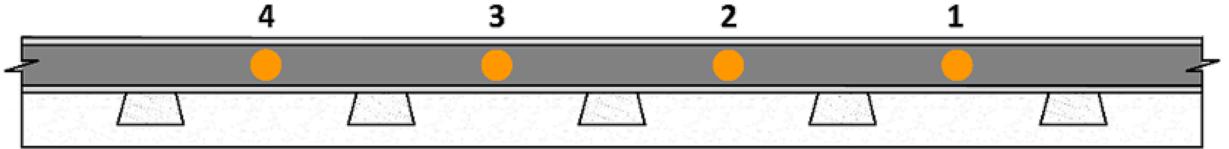


Fig. 3. Virtual wayside monitoring system: sensor placement.

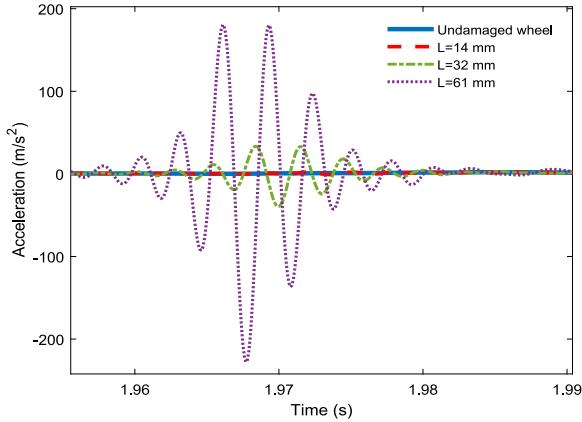


Fig. 4. Acceleration time series for different wheel flat lengths.

3.2. Numerical modelling

The wheel out-of-roundness detection system includes an accelerometer and strain gauge installed along the stretch of the track. To test and validate the methodology presented in this study, a wide range of 3D simulations based on the train-track interaction model were performed. The virtual wayside monitoring system includes measurements evaluated for 24 different locations mounted on the rail between two sleepers. Fig. 3 exemplifies 4 sensors installed along the stretch of the track in midspan location.

A virtual simulation of baseline and damaged wheel scenarios was accomplished to test and validate the automatic out-of-roundness diagnosis presented in this study. After successful validation, this method has the potential to be applied to real experimental data considering different types of trains with several wheel defect conditions. In real conditions, the wheels are not entirely smooth and have imperfections, i.e., flats and polygonization. Despite their small size, these out-of-roundness irregularities can cause extreme variations in the wheel–rail contact forces, producing vibrations on the train and track components. As presented in Fig. 2, three wheels were considered as defective: the first and last on the right side, and the sixth on the left side.

For wheel flats, two intervals of flat lengths (L_w) were considered, nominated as L1 and L2. The uniform distributions U (25, 50) mm, and U (50,100) mm define the lower and upper limits of the wheel flat length for each interval L1 and L2, respectively. The wheel flat depth (D_w) is calculated based on the following equation [61]:

$$D_w = \frac{L_w^2}{16R_w} \quad (1)$$

in which R_w is the radius of the wheel. The vertical profile deviation of the wheel flat is defined as:

$$Z = -\frac{D_w}{2} \left(1 - \cos \frac{2\pi x_w}{L_w} \right) \bullet H(x_w - (2\pi R_w - L_w)), 0 \leq x \leq 2\pi R \quad (2)$$

where H represents the Heaviside function and x_w is the coordinate aligned with the track longitudinal direction. Fig. 4 illustrates the effect of different wheel flat severities.

For wheel polygonization, a periodic radial tread irregularity around the wheel circumference was considered by varying wavelengths (λ) in the function of the harmonic order (θ) and wheel radius.

$$\lambda = \frac{2\pi R_w}{\theta}, \theta = 1, 2, 3, \dots, n \quad (3)$$

The selected wheel profiles were characterized by the wavelengths in the first 20 harmonics, with the 6th to 8th harmonic orders being dominant and different irregularity wheel profiles generated based on the sum of sine functions ($H = 20$) as follows:

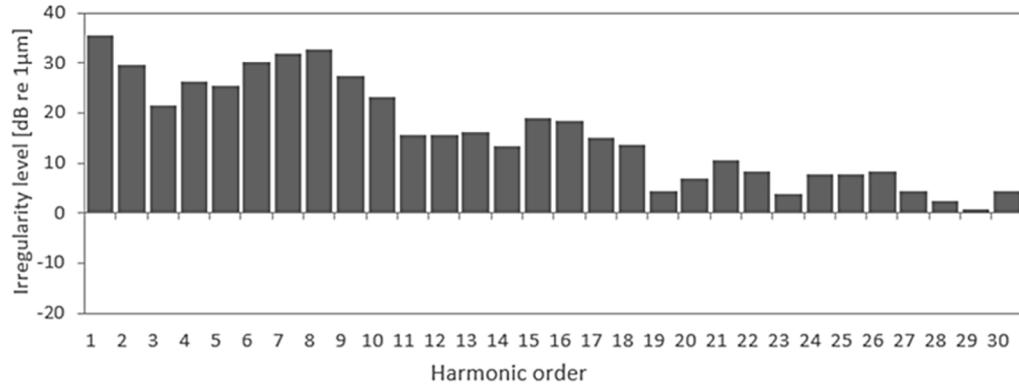


Fig. 5. Wheel irregularity amplitude spectra (L_w) and harmonic order (θ).

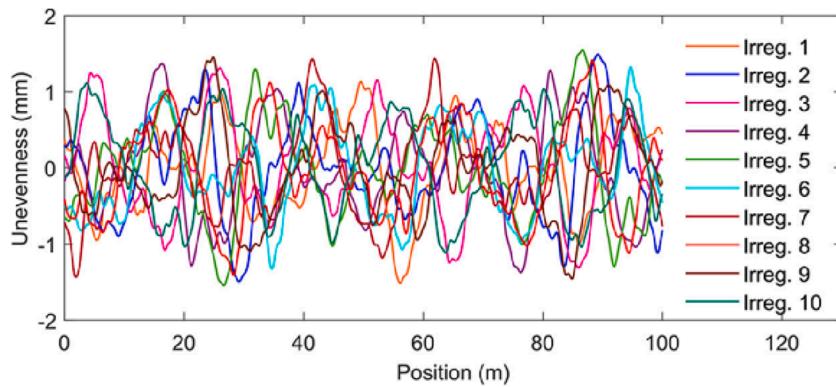


Fig. 6. Irregularities rail profiles.

$$w(x_w) = \sum_{[03B8]=1}^H A_\theta \sin\left(\frac{2\pi}{\lambda}x_w + \varphi_\theta\right) \quad (4)$$

where A_θ is the amplitude of the sine function for each wavelength, which is calculated by the function:

$$A_\theta = \sqrt{2} \cdot 10^{\frac{L_w}{20}} \cdot w_{ref} \quad (5)$$

with $w_{ref} = 1 \mu m$. The wheel irregularity level (L_w) values were selected based on the irregularity spectrum in Fig. 5, produced with measurement values of four wheels with polygonal damage [62]. Considering phase angles to the sine functions that are uniformly and randomly distributed between 0 and 2π , several wheel irregularity profiles were generated with Equation (4) to obtain different damage severities between 0.8 mm and 1.2 mm.

In real conditions, the rails also present small imperfections that can significantly affect the wheel-rail contact force values. In this context, 10 irregularities profiles, shown in Fig. 6, were generated for wavelengths between 1 m and 30 m, which covers the D1 wavelength interval defined by the EN 13848–2 standard [63], with a sampling discretisation of 1 mm.

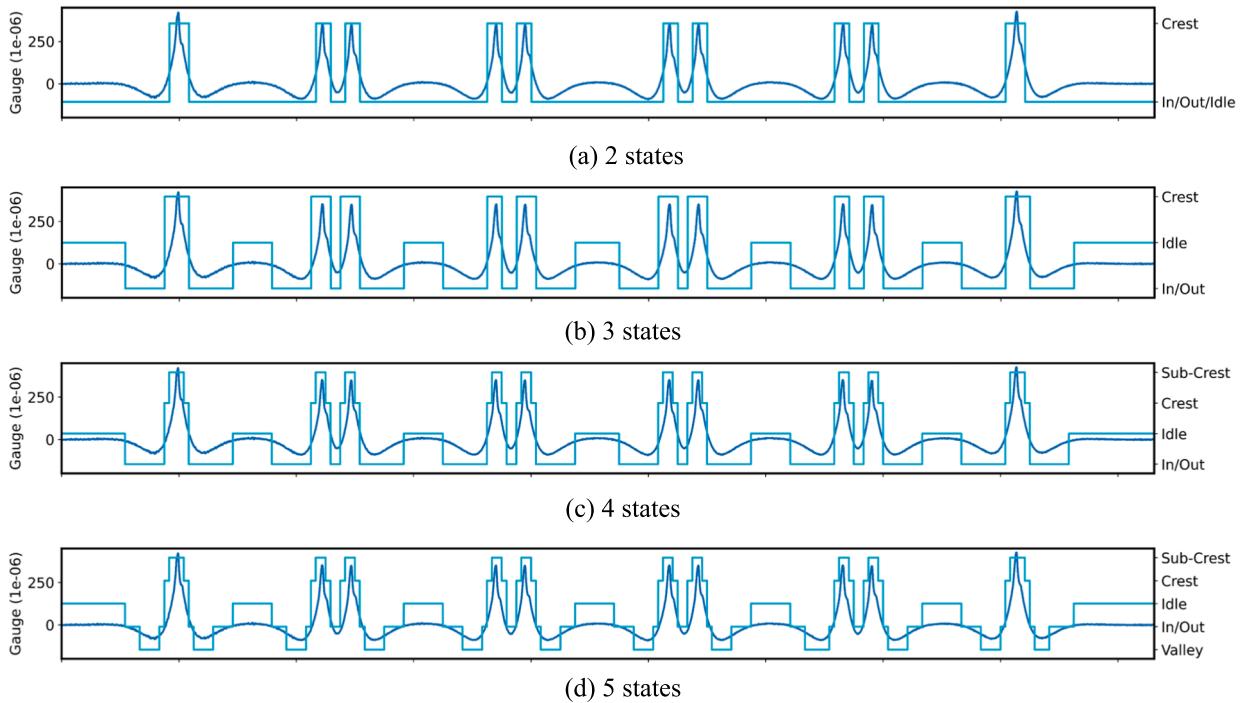
The amplitude of the irregularity profile was varied between –2 mm and 2 mm, for a total simulation length of 100 m. These were based on the real track irregularities of the Northern Line of the Portuguese Railway Network, measured with a track inspection vehicle EM120 every six months. It is important to note that the wavelengths used to generate these track irregularities are significantly longer than the wheel flat and polygonization. Thus, the excited frequencies due to the track are much lower than those of a defective wheel. More details about the generation of unevenness profiles can be found elsewhere [64].

By reproducing the responses of the rail considering several speeds, wheel profiles and rail irregularities, different time-series samples were simulated to capture a range of operating conditions [65]. The dynamic analyses were carried out for one passing vehicle. Table 1 summarizes the 23520-wheel scenario, divided into groups of one axle ($J = 7440$) and two axles ($K = 16080$). Within these groups, further subdivisions were made according to baseline and damaged conditions. The first encompasses the passage of baseline vehicles between 60 and 100 km/h in 10 unevenness profiles, measured by 24 different sets of sensors (accelerometer and strain gauge). The second includes the measurement of vehicles with defective wheels circulating at 80 km/h, for a single unevenness profile and 24 different sets of sensors. To obtain a more reliable reproduction of the rail response, an artificial noise was generated

Table 1

Summary of baseline and damaged scenarios.

	Baseline scenarios	Damaged scenarios	
	Perfect wheel	Polygonized wheel	Flat wheel
Vehicle	5 freight wagons of the Laagrss type		
Irregularity profiles	10	1	1
Speeds	60 – 100 km/h	80 km/h	80 km/h
Noise ratio	5 %	5 %	5%
Location of defect	–	1st wheel right	1st wheel right 6th wheel left 10th wheel right
Amplitude of defect	–	0.8 – 1.2 mm	25 – 50 mm 50 – 100 mm
Total passages of 1 axle	6240	720	480
Total passages of 2 axles	15,360	0	720

**Fig. 7.** Segmented strain gauge for healthy wheels with different gaussian emissions.

based on 5% of the maximum response of the signal and added to all the original signals. Additionally, all the time series are filtered based on a low-pass Chebyshev type II digital filter considering a cut-off frequency of 500 Hz.

4. Fault diagnosis

4.1. Segmentation

To offer a concise discussion of the results, these are illustrated for an exemplary train passage while examining the overall performance of each step. Firstly, the HMM was fitted to the shape of the gauge signal considering between 2 and 5 states of gaussian emissions, as depicted in Fig. 7. All states overcame the partial confounding of consecutive wheels with one another due to the closely connected wagons. However, only with 2 and 3 hidden states there was a consistent segmentation for all wheel passages. The other segmentations, despite adapting well to the shape of the monitoring signal, weren't reliable under a high noise ratio and defective conditions.

Consequently, for automatic segmentation, the HMM with 3 states was selected as it allowed to remove the portion of the wheel signal in between wheel passages. Furthermore, the obtained indexing was granular enough to capture the three stages of a wheel passage: the initial approach towards the sensor, the detection phase where the wheel is on top of the sensor and the departure, as it

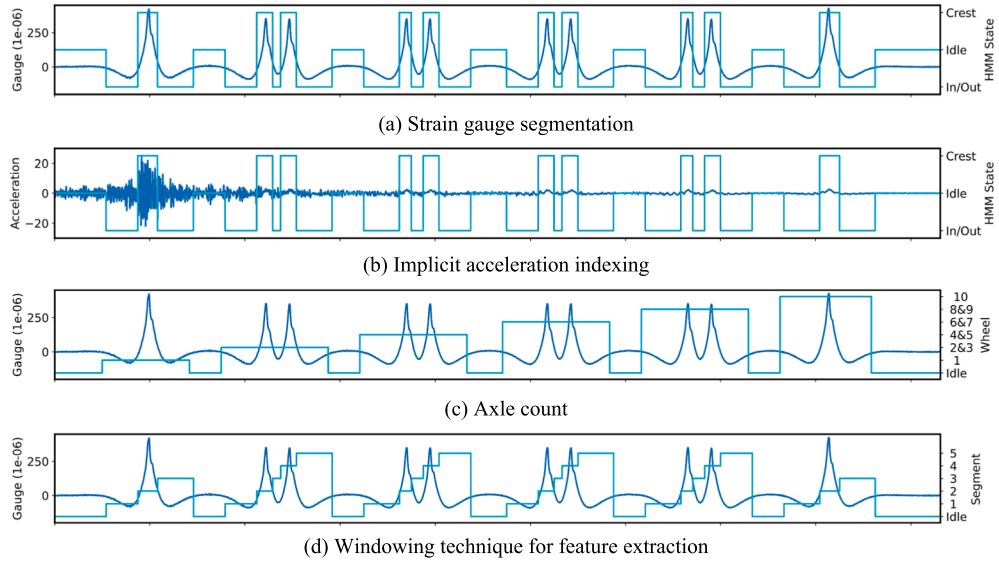


Fig. 8. Segmentation for wheel flat with 3 gaussian emissions.

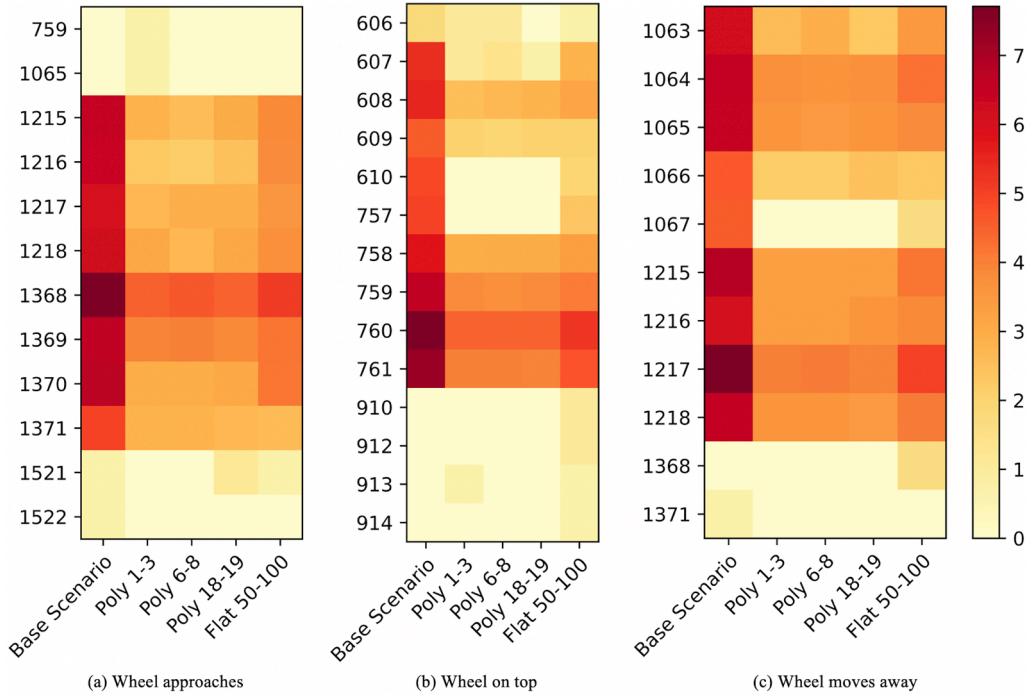


Fig. 9. Distribution of segment lengths in logarithmic scale for groups of one axle.

moves away from the sensor. This granularity proved useful in extracting the impact of an out-of-round wheel on the corresponding accelerometer signal, as shown in Fig. 8.

As illustrated in Fig. 8, the shape of accelerometer signal is disrupted upon a damaged scenario, opposed to the strain gauge. Thus, leveraging on the unaltered shape of the strain gauge signal upon a damaged wheel, automatic segmentation could be performed for the attainment of meaningful subsequences in the accelerometer signal of each passage. For groups of one and two axles, 3 and 5 segments were respectively engineered as shown in Fig. 8. These were then used to generate the window duration in STFT and for localized feature extraction in the accelerometer signal, which is more sensitive to damage. To study the robustness of the segmentation, Fig. 9 and Fig. 10 provide the distribution of segment lengths in logarithmic scale for the different scenarios in groups of one and two axles.

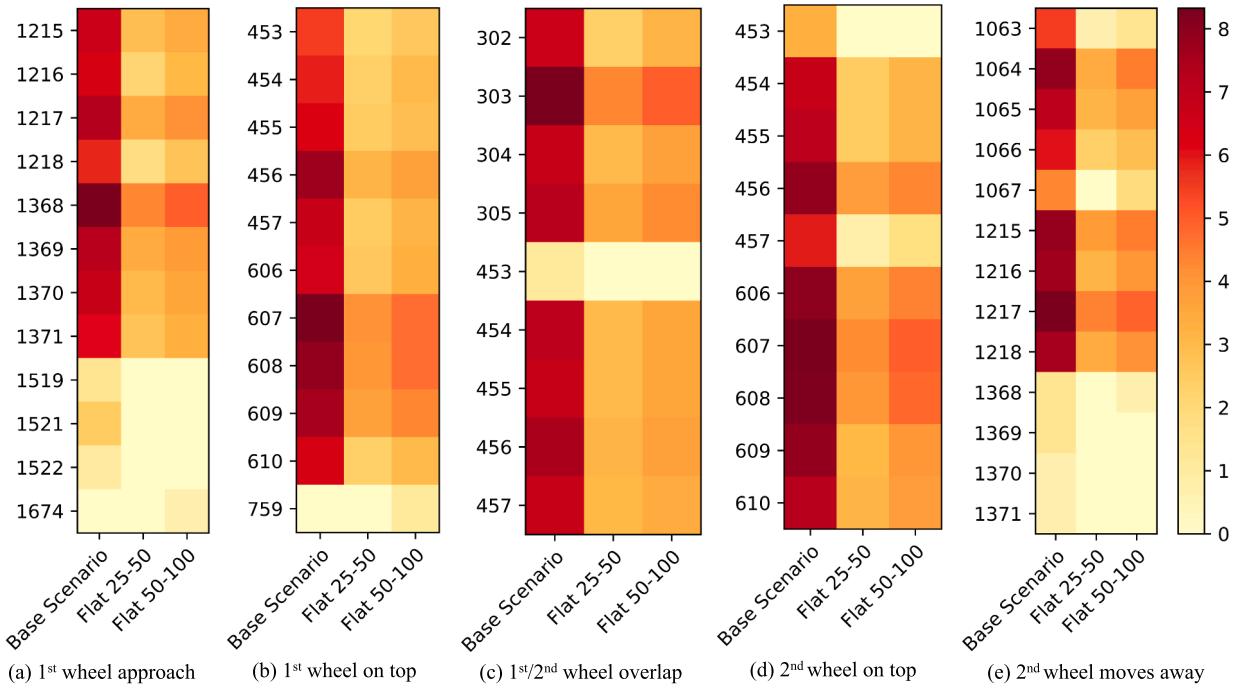


Fig. 10. Distribution of segment lengths in logarithmic scale for groups of two axles.

It is important to note the color bar is in logarithmic scale. In Fig. 9, for the three stages of the wheel passage in groups of one axle, there is a stable segmentation between within ranges 1215–1371, 606–761 and 1063–1218 tenths of a millisecond, corresponding to approximately 10% deviation. Within these ranges, the interaction between the length of segments and the damaged scenarios is merely indicative of the imbalanced dataset. Apart from a few outliers, only in the detection phase with the wheel on top of the accelerometer, considerable larger segments were extracted for a length of 910 to 914 tenths of a millisecond. Regarding the groups of two axles in Fig. 10, both the movement of the first wheel towards and of the second wheel away from the accelerometer exhibit the same pattern of the groups of one axle. However, the partial confounding of the consecutive wheels results in smaller segments from 453 to 610 tenths of a millisecond, when the wheel is on top of the accelerometer. As for segment corresponding to the overlap of the wheel's movement, the range is 302 to 457 tenths of a millisecond.

4.2. Time-frequency analysis

Fig. 10 depicts the sensitivity to damage of 5 out of the 21 features extracted from each subsequence. The features are divided, with the baseline scenarios shown before the black line and the damaged scenarios shown subsequently. In the figure, symbols were assigned to scenarios with the anticipated failure if the feature under consideration was used as a singular damage indicator. In groups of one axle, the symbols, from left to right, denote polygonization, with wavelengths comprising 1–3, 6–8 and 18–19 harmonics, and the wheel flat range of 50–100 mm. For groups of two axles, the symbols, from left to right, denote the wheel flats of ranges 25–50 mm and 50–100 mm.

As shown in the figure, certain features extracted from the data exhibit sensitivity to damage, with higher magnitudes and fluctuations compared to the baseline scenarios. For instance, Fig. 11 demonstrates that the amplitude differences between healthy and damaged scenarios are more pronounced for certain features, e.g., peak, which remains relatively stable results. Conversely, it also presents features, e.g., skewness of total band power, whose amplitude variation due to out-of-round wheels is similar to that resulting from environmental and operational effects in the baseline scenarios. These findings suggest that each feature captures unique aspects of the damage condition, and that a comprehensive approach based on selected features is necessary to fully capture the diagnostic information. Environmental and operational effects may obscure the damage signal and make it challenging to differentiate between baseline and damaged scenarios.

4.3. Principal feature analysis

To identify the most informative and discriminative features that can best differentiate between different conditions or states, PFA was applied to the matrices of both groups of axle 7440-by-21-by-3 and 16080-by-21-by-5. For the first stage of Principal Component Analysis, as explained previously, a cumulative variance of 80% to streamline the analysis. In the second stage, 10,000 runs of k -means clustering were performed, varying k between 2 and 10, to overcome the non-deterministic nature of the algorithm. This step ensured

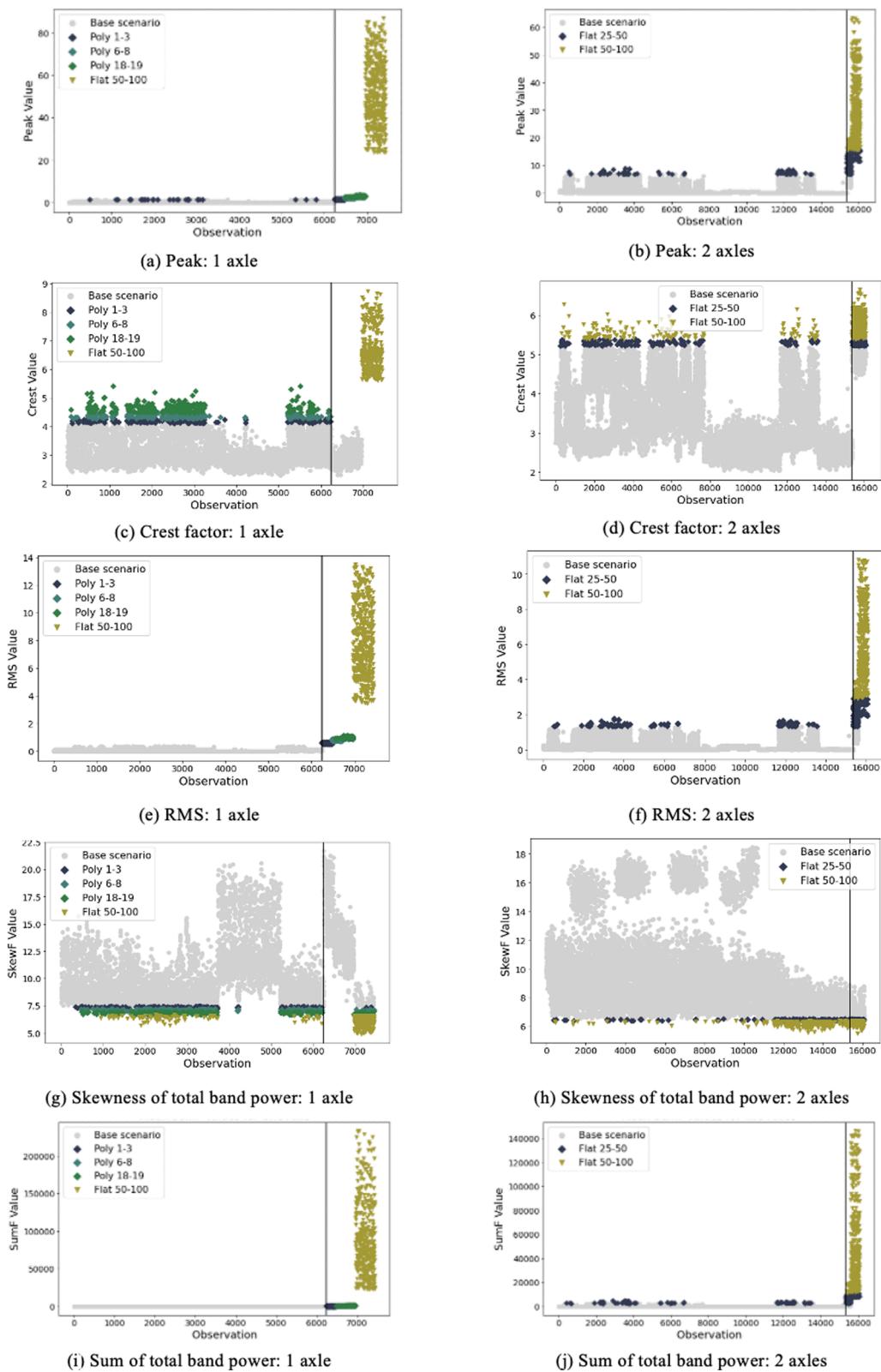
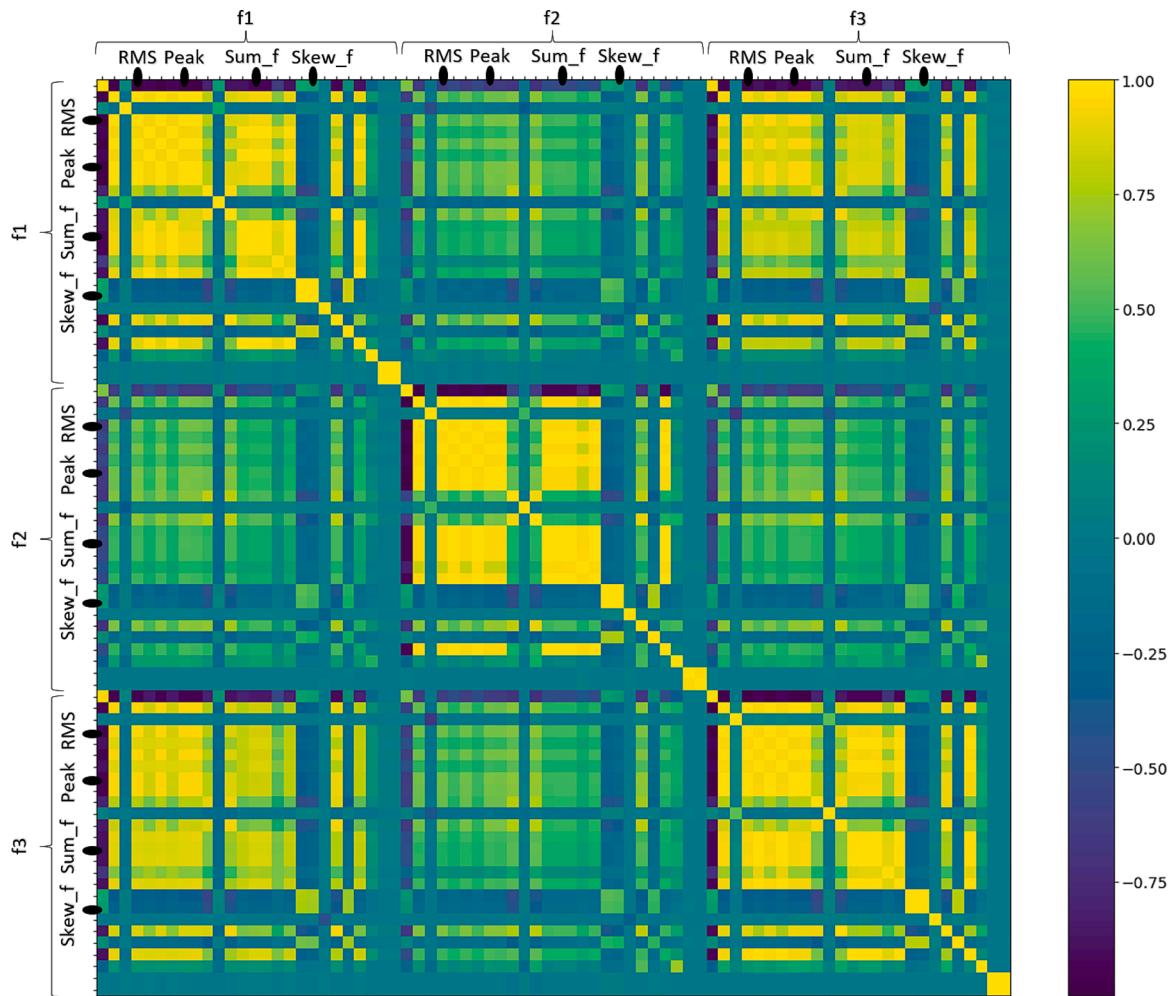


Fig. 11. Individual feature values as anticipated damage indicators.

Table 2

Principal feature analysis for 10,000 runs.

Number of axles	Time-domain		Frequency domain	
	Peak	Root main square	Skewness of total band power	Sum of total band power
1	5952	1843	4889	4700
2	4577	2916	3227	4252

**Fig. 12.** Correlation matrix for groups of one axle.

that the resulting clusters were stable and not dependent on the initial random seed value. The features that were most selected as representative of a cluster are presented in **Table 2**.

The most selected features were identified the peak and root main square in the time domain, and the sum and skewness of total band power in the frequency domain. To gain a deeper understanding of this selection process, **Fig. 12** illustrates the correlation analysis of the extracted features in each stage of a wheel passage, with features f1 corresponding to the initial approach towards the sensor, f2 to the detection phase where the wheel is on top of the sensor and f3 to the departure, as the wheel moves away from the sensor.

The results illustrate how these features represent clusters of highly correlated variables within each segment, except for the skewness of the total band power feature. In fact, as described earlier, this statistic was found to be irrelevant to the diagnose of defective wheels and thus removed from the analysis. The presence of the remaining features suggested the existence of underlying physical phenomena. Both the peak and root main square in the time domain were successful discriminators of the higher force exerted on the track by a flat wheel, while for other types of wheel tread defects, i.e., polygonization, there was a clear difference in frequency bands of the measurement, which was captured by the sum of total band power. Thus, these 3 features were selected for the anomaly

Table 3
Iforest hyperparameter tuning.

No. of trees	Samples	F1-score	
		1 axle	2 axles
150	0.01	0.9975	0.8306
300	0.01	0.9925	0.8389
450	0.01	1	0.8389
150	0.21	0.9442	0.8667
300	0.21	0.9383	0.8653
450	0.21	0.9450	0.8708
150	0.41	0.9475	0.8722
300	0.41	0.9425	0.8708
450	0.41	0.9392	0.8764
150	0.61	0.9383	0.8833
300	0.61	0.9342	0.8792
450	0.61	0.9300	0.8819
150	0.81	0.9375	0.8750
300	0.81	0.9342	0.8736
450	0.81	0.9308	0.8778

Table 4
LOF hyperparameter tuning.

No. of neighbours	F1-score	
	1 axle	2 axles
100	0.0883	0.4139
200	0.0850	0.6417
300	0.1925	0.7722
400	0.3308	0.8139
500	0.4017	0.8264

Table 5
SVM hyperparameter tuning.

Gamma	F1-score	
	1 axle	2 axles
0.1	0.4356	0.1793
0.3	0.5830	0.1372
0.5	0.4911	0.1535
0.7	0.5563	0.1338
0.9	0.5619	0.1320

detection stage.

4.4. Damage index

In this study, two separate anomaly detection rankings were built for each matrix of 7440-by-3-by-3 and 16080-by-3-by-5. Building separate models for the two groups helped to deal with the differences in vibration patterns caused by the consecutive wheels in groups of two axles. Regarding hyperparameter tuning of the iForest, the findings of the analysis support the original paper's proposition that a broad range of values for the parameters of number of trees and samples required to split have little impact on detecting instances deemed to be anomalies, as shown in [Table 3](#). In this research study, the parameters were set as 450 and 0.01, respectively, based on the F1-score of overall detection. For this calculation, the contamination percentage, i.e., the threshold to consider instances as anomalies, was defined with the actual value of 1200/7440 for the group of one axle and 720/16080 for two axles.

In addition, to allow for a fair performance benchmark of the iForest, the other algorithms were also submitted to hyperparameter tuning. In LOF, it was studied a range between 100 and 500 of nearest neighbors to consider when calculating the local density of each data point, as depicted in [Table 4](#). A smaller number of neighbors makes the algorithm more sensitive to outliers, as it may classify points with a lower density as outliers if their nearest neighbors also have a low density. While for SVM, the gamma parameter that determines the influence of each training example on the model's decision boundary was varied between 0.1 and 0.9, as shown in [Table 5](#). A smaller gamma value will result in a smoother decision boundary, while a larger gamma value will result in a more complex decision boundary that is more likely to overfit the training data.

The F1-score offers provides an indication of the overall success of the iForest in both fault detection and isolation, due to the

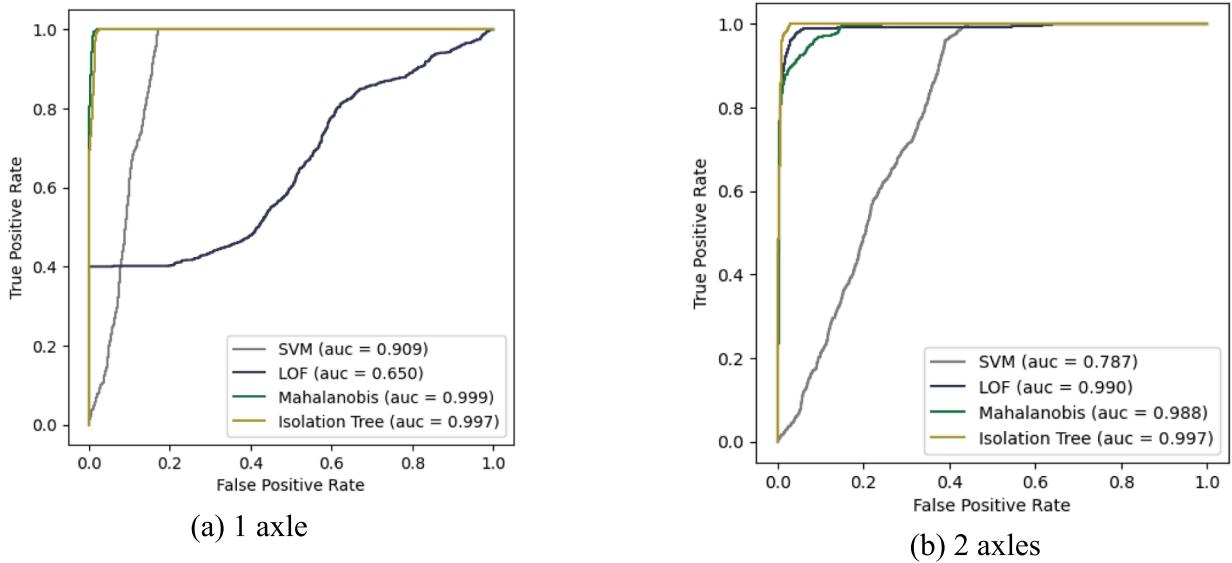


Fig. 13. Area under the receiver operating characteristic curve.

Table 6
Confusion matrix for groups of one axle.

	Baseline	Poly 1–3	Poly 6–8	Poly 18–19	Flat 50–100 mm
Baseline	6218	22	0	0	0
Poly 1–3	22	205	13	0	0
Poly 6–8	0	13	187	40	0
Poly 18–19	0	0	40	200	0
Flat 50–100 mm	0	0	0	0	480

Table 7
Confusion matrix for groups of two axles.

	Healthy	Flat 25–50 mm	Flat 50–100 mm
Healthy	15,255	105	0
Flat 25–50 mm	105	115	20
Flat 50–100 mm	0	20	460

individual treatment of groups of one or two axles. Furthermore, for the selected hyperparameters, it was studied how the contamination percentage used as threshold affected the area under the receiver operating characteristic curve (AUC-ROC) [66]. In practice, this threshold might be limited by various factors, such as the number of defective wheels or the availability of maintenance department resources, i.e., stock of spare parts and workforce. Fig. 13 shows the AUC-ROC obtained for the various anomaly detection models.

The AUC-ROC and its corresponding score provide insights into the success of the fault identification task in terms of severity. A higher AUC indicates that the system is better at distinguishing between faulty and non-faulty instances. It implies that the ranking obtained from the anomaly scores is effective in capturing the severity of faults, as the system can differentiate between them with a higher degree of accuracy. As shown, MD and LOF were successful in fault identification for groups of one and two axles, respectively. However, only iForest presented a stable performance, with scores of 0.997 in both groups. To provide a more detailed analysis of the success of the full fault diagnosis task with the iForest algorithm, Table 6 depicts the confusion matrix with contamination equal to the proportion of defective wheels. Out of 7440 wheels, only 22 (0.2%) polygonization treads of low harmonics were confounded with undamaged scenarios. As for wheels with flats of 50–100 mm, these were identified as most anomalous with perfect precision. Furthermore, the progression of false positives across only consecutive classes suggests that the model is successfully ranking the

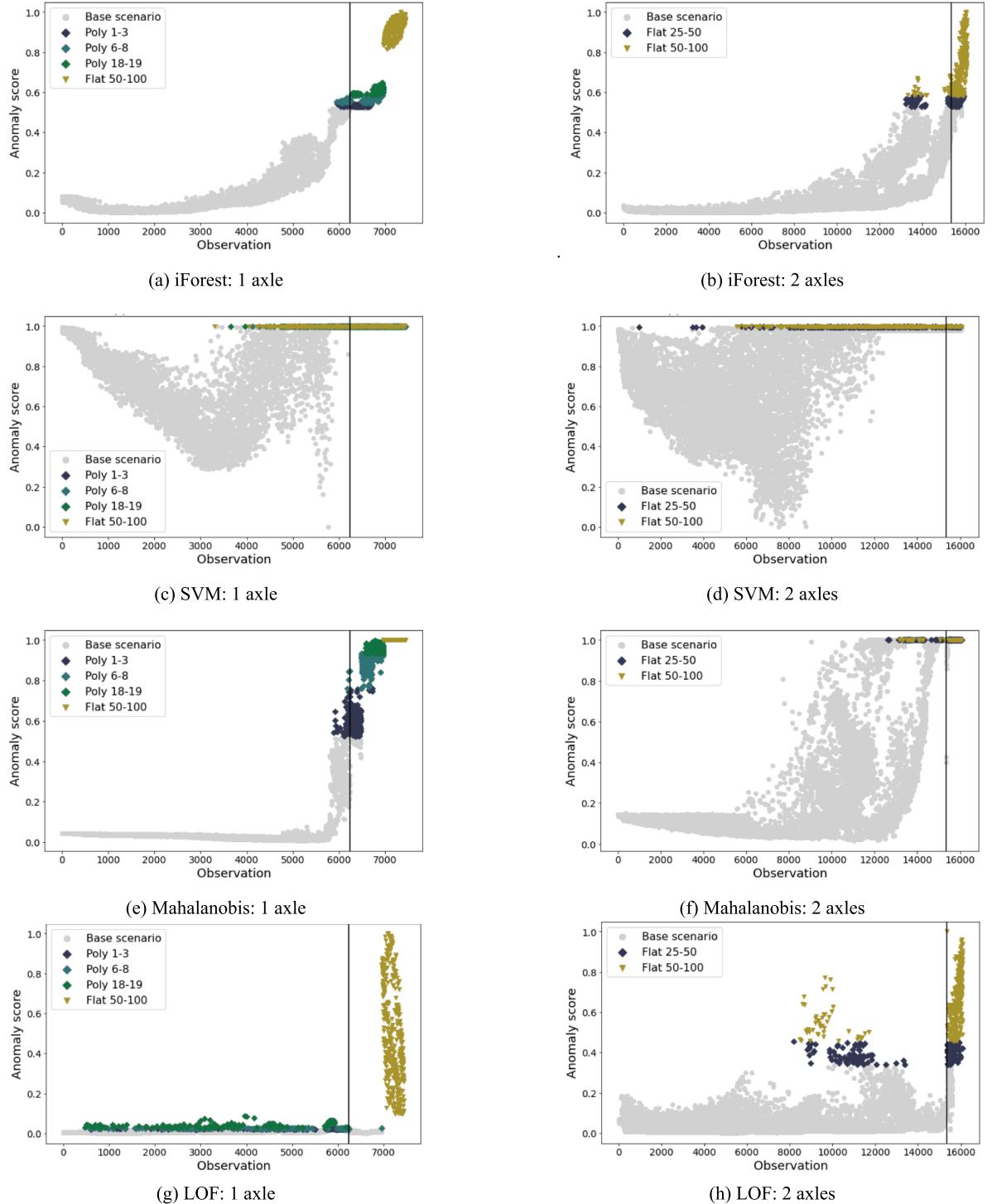


Fig. 14. Damage indicator based on various anomaly detection algorithms.

severities.

Regarding groups of two axles, results were also promising in terms of the most severe flats, as illustrated in Table 7. Only 105 out of 240 wheels with small flat depths were confounded as non-defective wheels. Notably, these wheels exhibited a flat length below the replacement threshold recommended by the General Contract for the Use of Wagons [67]. For instance, a wheelset with a diameter larger than 840 mm should only be replaced if the wheel flat length exceeds 60 mm. Therefore, the proposed strategy not only

demonstrated complete accuracy in detecting faults, but also had the potential to predict the emergence of flats before they reach a critical state. In practice, these findings suggest that when tackling the unsupervised context of the real-world application, if the maintenance technician follows the predicted rank and starts to be consistently advised to replace wheels then he can be confident that all wheels are in good conditions.

Finally, Fig. 14 summarizes the damage indicator for all 23520-wheel passages, with benchmark models being outperformed by the iForest. The scenarios were separated by a black line into two groups, as in Fig. 10.

5. Conclusions

In this paper, an automatic fault diagnosis algorithm for wheel out-of-roundness was developed, focusing on two failure modes: wheel flats and wheel tread polygonization. The proposed methodology consisted in applying a HMM to guide signal preprocessing with variable STFT to ultimately extract localized features in the statistical and frequency domain, overcoming the limitations of addressing non-linear environmental and operational effects. Using PFA for a subset of the attained time series representation, it was possible to select damage-sensitive features with minimal redundancy. With the selected features a single severity indicator was computed based on the iForest anomaly score. This completed the fault diagnosis with (1) detection of aberrant train behavior; (2) isolation of specific defective wheels; (3) identification of the severity. The main achievements, including contributions from this study, can be summarized as follow:

- A novel method of signal segmentation, leveraging the sensitivity of each type of sensor. The stable temporal structure of the strain gauge signal was used to segment both signals, while feature extraction was focused on the most damage-sensitive accelerometer, less impacted by non-linear operational and environmental factors;
- Fault localization method proposed in this study offers a wheel-based approach with implicit indexing and axle count, thereby providing infrastructure managers and maintenance personnel with more comprehensive and valuable information;
- The methodology doesn't require the installment of a series of sensors on the rail, reducing the cost of installation and maintenance. By only using one accelerometer and strain gauge data quality wasn't compromised;
- The final anomaly detection stage not only provided a comprehensive context to the success of the fault localization method in distinguishing a healthy wheel from defective ones, but also provided a damage indicator for severity identification, regardless of different wheel profiles, track irregularities, train speeds, position of the sensors and artificial noise, associated to other environmental and operational variations.

Such results demonstrate the high potential of this innovative application of artificial intelligence in the railway wheels maintenance. As this method was validated with physical-based modelling, the next efforts should lie in ensuring there are not too many unfulfilled assumptions and rough approximations. Thus, future work includes a field trial to further evaluate the proposed methodology based on on-site measurements. Furthermore, the robustness and efficiency of the methodology will be assessed under distinct track environments, particularly in the presence of bridges, tunnels and other under passing structures, will be evaluated.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The present work has been developed under project FERROVIA 4.0 (POCI/01/0247/FEDER/046111) and WAY4SafeRail (NORTE/01/0247/FEDER/069595). It has received Portuguese National Funds through FCT (Portuguese Foundation for Science and Technology) under project UIDB/00760/2020 of the GECAD and UIDB/04708/2020 of the CONSTRUCT – Instituto de I&D em Estruturas e Construções.

References

- [1] S. Hidirov, H. Guler, Reliability, availability and maintainability analyses for railway infrastructure management, *Struct. Infrastruct. Eng.* 15 (9) (2019) 1221–1233.
- [2] V. Gonçalves, A. Mosleh, C. Vale, P.A. Montenegro, Wheel out-of-roundness detection using an envelope spectrum analysis, *Sensors* 23 (4) (2023) 2138.
- [3] A. Mosleh, P. Montenegro, P. Alves Costa, R. Calçada, An approach for wheel flat detection of railway train wheels using envelope spectrum analysis, *Struct. Infrastruct. Eng.* 17 (12) (Dec. 2021) 1710–1729, <https://doi.org/10.1080/15732479.2020.1832536>.
- [4] A. Guedes, et al., Detection of wheel polygonization based on wayside monitoring and artificial intelligence, *Sensors* 23 (4) (2023) 2188.
- [5] S.Y. Chong, J.-R. Lee, H.-J. Shin, A review of health and operation monitoring technologies for trains, *Smart Struct. Syst.* 6 (9) (2010) 1079–1105.
- [6] M. Mohammadi, A. Mosleh, C. Vale, D. Ribeiro, P. Montenegro, A. Meixedo, An unsupervised learning approach for wayside train wheel flat detection, *Sensors* 23 (4) (2023) 1910.

- [7] R. Lagnebäck, "Evaluation of wayside condition monitoring technologies for condition-based maintenance of railway vehicles".
- [8] R.D. Fröhling, G. Hettasch, *Wheel-rail interface management: a rolling stock perspective*, Proc. Inst. Mech. Eng. F J Rail Rapid Transit. 224 (5) (2010) 491–497.
- [9] A. Mosleh, P.A. Montenegro, P.A. Costa, R. Calçada, 10 - Approaches for weigh-in-motion and wheel defect detection of railway vehicles, in: R. Calçada, S. Kaewunrunworn (Eds.), *Rail Infrastructure Resilience*, Woodhead Publishing, 2022, pp. 183–207, <https://doi.org/10.1016/B978-0-12-821042-0.00009-5>.
- [10] D. Barke, W.K. Chiu, Structural health monitoring in the railway industry: a review, Struct. Health Monit. 4 (1) (2005) 81–93.
- [11] A. Alemi, F. Corman, G. Lodewijks, Condition monitoring approaches for the detection of railway wheel defects, Proc. Inst. Mech. Eng. F J Rail Rapid Transit. 231 (8) (2017) 961–981.
- [12] A. Meixedo, J. Santos, D. Ribeiro, R. Calçada, M.D. Todd, Online unsupervised detection of structural changes using train-induced dynamic responses, Mech. Syst. Sig. Process. 165 (2022), 108268.
- [13] A. Lourenço, J. Meira, and G. Marreiros, "Online Adaptive Learning for Out-of-Round Railway Wheels Detection," in *Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing*, in SAC '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 418–421.
- [14] B. Pintao, A. Mosleh, C. Vale, P. Montenegro, P. Costa, Development and validation of a weigh-in-motion methodology for railway tracks, Sensors 22 (5) (2022) 1976.
- [15] R. Silva, et al., Early identification of unbalanced freight traffic loads based on wayside monitoring and artificial intelligence, Sensors 23 (3) (2023) 1544.
- [16] A. Mosleh, A. Meixedo, D. Ribeiro, P. Montenegro, R. Calçada, "Automatic clustering-based approach for train wheels condition monitoring", Int. J. Rail Transp. (2022) 1–26.
- [17] G. Charles, R. Dixon, and R. Goodall, "Condition monitoring approaches to estimating wheel-rail profile," in *Proceedings of UKACC Control Conference, Manchester*, 2008.
- [18] J.C.O. Nielsen, A. Johansson, Out-of-round railway wheels-a literature survey, Proc. Inst. Mech. Eng. F J Rail Rapid Transit. 214 (2) (2000) 79–91.
- [19] P. Jayawal, S.N. Verma, A.K. Wadhwan, Development of EBP-Artificial neural network expert system for rolling element bearing fault diagnosis, J. Vib. Control 17 (8) (2011) 1131–1148.
- [20] S. Jia, M. Dhanasekar, Detection of rail wheel flats using wavelet approaches, Struct. Health Monit. 6 (2) (2007) 121–131.
- [21] G. Krummenacher, C.S. Ong, S. Koller, S. Kobayashi, J.M. Buhmann, Wheel defect detection with machine learning, IEEE Trans. Intell. Transp. Syst. 19 (4) (2017) 1176–1187.
- [22] R.V. Dukkipati, R. Dong, Impact loads due to wheel flats and shells, Veh. Syst. Dyn. 31 (1) (1999) 1–22.
- [23] A. Amini, M. Entezami, Z. Huang, H. Rowshandel, and M. Papaelias, "Wayside detection of faults in railway axle bearings using time spectral kurtosis analysis on high-frequency acoustic emission signals," *Advances in Mechanical Engineering*, vol. 8, no. 11, p. 1687814016676000, 2016.
- [24] A. Mosleh, P.A. Montenegro, P.A. Costa, R. Calçada, Railway vehicle wheel flat detection with multiple records using spectral kurtosis analysis, Appl. Sci. 11 (9) (2021) 4002.
- [25] V. Alves, A. Cury, N. Roitman, C. Magluta, C. Cremona, Novelty detection for SHM using raw acceleration measurements, Struct. Control Health Monit. 22 (9) (2015) 1193–1207.
- [26] A. Meixedo, J. Santos, D. Ribeiro, R. Calçada, M. Todd, Damage detection in railway bridges using traffic-induced dynamic responses, Eng. Struct. 238 (2021), 112189.
- [27] S. Chen, K. Wang, Z. Zhou, Y. Yang, Z. Chen, W. Zhai, Quantitative detection of locomotive wheel polygonization under non-stationary conditions by adaptive chirp mode decomposition, Railway Eng. Sci. 30 (2) (2022) 129–147.
- [28] H. Li, B. Qian, D. Parikh, A. Hampapur, "Alarm prediction in large-scale sensor networks—A case study in railroad", in, IEEE international conference on big data, IEEE (2013) 7–14.
- [29] A. Lourenço, M. Fernandes, A. Canito, A. Almeida, G. Marreiros, Using an explainable machine learning approach to minimize opportunistic maintenance interventions, in: A. González-Briones, A. Almeida, A. Fernandez, A. El Bolock, D. Durães, J. Jordán, F. Lopes (Eds.), *Highlights in Practical Applications of Agents, Multi-Agent Systems, and Complex Systems Simulation*. The PAAMS Collection, Springer International Publishing, Cham, 2022, pp. 41–54.
- [30] M. Chenaryan Nakhaee, D. Hiemstra, M. Stoelinga, 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: third international conference, RSSRail Lille, france, June 4–6, 2019, proceedings 3, Springer (2019) 91–105.
- [31] J. Xie, J. Huang, C. Zeng, S.-H. Jiang, N. Podlich, Systematic literature review on data-driven models for predictive maintenance of railway track: Implications in geotechnical engineering, Geosciences (Basel) 10 (11) (2020) 425.
- [32] A. Johansson, C. Andersson, Out-of-round railway wheels—a study of wheel polygonalization through simulation of three-dimensional wheel-rail interaction and wear, Veh. Syst. Dyn. 43 (8) (2005) 539–559.
- [33] W. Cai, M. Chi, G. Tao, X. Wu, Z. Wen, Experimental and numerical investigation into formation of metro wheel polygonalization, Shock Vib. (2019).
- [34] A. Lourenço, M. Fernandes, G. Marreiros, J.M. Corchado, "Using simulation to evaluate a concept drift detector for condition based maintenance", in *IECON 2022–48th Annual Conference of the IEEE Industrial Electronics Society*, IEEE (2022) 1–7.
- [35] L. Cohen, *Time-frequency analysis*, vol. 778. Prentice hall New Jersey, 1995.
- [36] P.A. Montenegro, S.G.M. Neves, R. Calçada, M. Tanabe, M. Sogabe, Wheel-rail contact formulation for analyzing the lateral train-structure dynamic interaction, Comput. Struct. 152 (May 2015) 200–214, <https://doi.org/10.1016/J.COMPSTRUC.2015.01.004>.
- [37] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [38] J.A. Bilmes, A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, Int. computer Sci. institute 4 (510) (1998) 126.
- [39] M. Stamp, A revealing introduction to hidden Markov models, Department of Computer Science San Jose State University, 2004, pp. 26–56.
- [40] K. Javed, R. Gouriveau, N. Zerhouni, P. Nectoux, Enabling health monitoring approach based on vibration data for accurate prognostics, IEEE Trans. Ind. Electron. 62 (1) (2014) 647–656.
- [41] G.G. Yen, K.-C. Lin, Wavelet packet feature extraction for vibration monitoring, IEEE Trans. Ind. Electron. 47 (3) (2000) 650–667.
- [42] D. Griffin, J. Lim, Signal estimation from modified short-time Fourier transform, IEEE Trans Acoust 32 (2) (1984) 236–243.
- [43] B. Liang, S.D. Iwnicki, Y. Zhao, D. Crosbee, Railway wheel-flat and rail surface defect modelling and analysis by time-frequency techniques, Veh. Syst. Dyn. 51 (9) (2013) 1403–1421.
- [44] G. Xin, et al., Fault diagnosis of wheelset bearings in high-speed trains using logarithmic short-time Fourier transform and modified self-calibrated residual network, IEEE Trans Industr Inform 18 (10) (2021) 7285–7295.
- [45] B. Liang, S. Iwnicki, A. Ball, A.E. Young, Adaptive noise cancelling and time-frequency techniques for rail surface defect detection, Mech. Syst. Sig. Process. 54 (2015) 41–51.
- [46] C. Zhang, A.A. Mousavi, S.F. Masri, G. Gholipour, K. Yan, X. Li, Vibration feature extraction using signal processing techniques for structural health monitoring: a review, Mech. Syst. Sig. Process. 177 (2022), 109175.
- [47] Y. Lu, I. Cohen, X. S. Zhou, and Q. Tian, "Feature selection using principal feature analysis," in *Proceedings of the 15th ACM international conference on Multimedia*, 2007, pp. 301–304.
- [48] I. Jolliffe, "Principal component analysis," *Encyclopedia of statistics in behavioral science*, 2005.
- [49] W.K. Härdle, L. Simar, *Applied multivariate statistical analysis*, Springer Nature, 2019.
- [50] J. P. de Oliveira Dias Prudente dos Santos, C. Crémona, A. P. C. da Silveira, and L. C. de Oliveira Martins, "Real-time damage detection based on pattern recognition," *Structural Concrete*, vol. 17, no. 3, pp. 338–354, 2016.
- [51] M. Goldstein, S. Uchida, Behavior Analysis Using Unsupervised Anomaly Detection. (2014), <https://doi.org/10.13140/2.1.3349.7605>.
- [52] K.M. Ting, S.C. Tam, F.T. Liu Mass: a new ranking measure for anomaly detection, Monash University, Gippsland School of Information Technology, 2009.
- [53] J. Lesouïgle, C. Baudoin, M. Spigai, J.-Y. Tourneret, Generalized isolation forest for anomaly detection, Pattern Recogn. Lett. 149 (2021) 109–119.
- [54] F.T. Liu, K.M. Ting, Z.-H. Zhou, Isolation-based anomaly detection, ACM Trans. Knowledge Discovery from Data (TKDD) 6 (1) (2012) 1–39.

- [55] L. Puggini, S. McLoone, An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data, *Eng. Appl. Artif. Intel.* 67 (2018) 126–135.
- [56] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “LOF: identifying density-based local outliers,” in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 93–104.
- [57] R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart, The mahalanobis distance, *Chemom. Intel. Lab. Syst.* 50 (1) (2000) 1–18.
- [58] V.A. Sotiris, W.T. Peter, M.G. Pecht, Anomaly detection through a bayesian support vector machine, *IEEE Trans. Reliab.* 59 (2) (2010) 277–286.
- [59] H. Hertz, “Ueber die Berührung fester elastischer Körper,” 1882.
- [60] J.J. Kalker, *Book of Tables for the Herzian Creep-force Law*, Delft University of Technology, Faculty of Technical Mathematics and Informatics, 1996.
- [61] W.M. Zhai, Q.C. Wang, Z.W. Lu, X.S. Wu, Dynamic effects of vehicles on tracks in the case of raising train speeds, *Proc. Inst. Mech. Eng. F J Rail. Rapid. Transit.* 215 (2) (2001) 125–135.
- [62] W. Cai, M. Chi, G. Tao, X. Wu, Z. Wen, Experimental and numerical investigation into formation of metro wheel polygonalization, *Shock Vib.* (2019) 1538273, <https://doi.org/10.1155/2019/1538273>.
- [63] B. Standard, Railway applications/Track-Track geometry quality, *BS EN* (2003) 13841–13848.
- [64] A. Mosleh, P.A. Costa, R. Calçada, A new strategy to estimate static loads for the dynamic weighing in motion of railway vehicles, *Proc. Inst. Mech. Eng. F J Rail. Rapid. Transit.* 234 (2) (2020) 183–200.
- [65] eixedo, D. Ribeiro, P. Montenegro, R. Calçada, Early wheel flat detection: an automatic data-driven wavelet-based approach for railways, *Veh. Syst. Dyn.* (Jul. 2022) 1–30, <https://doi.org/10.1080/00423114.2022.2103436>.
- [66] R. Baeza-Yates and B. Ribeiro-Neto, *Modern information retrieval*, vol. 463. ACM press New York, 1999.
- [67] GCU, *General contract of use for wagons – GCU*, Edition dated. 2018.