Anomaly Identification Algorithms for Indirect Structural Health Monitoring Using a Laboratory-Scale Railroad Track System

Guillermo Montero[a], Jeremy Yin[a], Katherine A. Flanigan*[a], Mario Bergés[a], and James D. Brooks[b]
[a]Department of Civil and Environmental Engineering, Carnegie Mellon University, PA, USA
[b]Advanced Technologies, Wabtec Corporation, PA, USA

ABSTRACT

Presently, railroad monitoring strategies focus on preventative maintenance by detecting wheel anomalies using wayside detection methods (e.g., wheel-impact load detection), and direct detection of track anomalies using onboard systems (e.g., track geometry vehicles). Both approaches are periodic, manual, and do not support real-time track damage detection. Recent research has focused on detecting damage from acceleration signals obtained onboard moving vehicles and identifying anomalies from derived structural dynamic properties. Though promising due to inherent scalability and cost efficiency, its main goal is to detect damage on the supporting infrastructure and has never before been tested for detecting rail crack damage. Among other reasons, a robust anomaly detection algorithm is missing to allow the industry to embrace an automated and more cost-effective monitoring technique. In this work, we leverage a lab-scale track and moving vehicle actuation system that is scaled with the assistance of industry experts, and comprises a vehicle instrumented with two onboard vertical accelerometers. Cracked rails are simulated by introducing discontinuities (longitudinally and transversely). Several types of feature extraction and dimensionality reduction techniques are employed to evaluate their ability to separate damaged and undamaged records. Inspired from previous work, this work tests the ability of existing data-driven damage detection algorithms to detect local damage by using a novel super modular, precise, and realistically scaled down version of a train-track system. The results of the damage sensitivity show that principal component analysis has the highest balanced combination of recall and true negative rate, compared to other techniques.

Keywords: Indirect structural health monitoring, vehicle-rail interaction, damage detection, dimensionality reduction

1. INTRODUCTION

Each year, hundreds of train accidents occur due to damage to railway lines. In 2022, the Federal Railroad Administration reported 339 track-caused accidents, and 1,049 derailments, which combined represent 16% of the overall number of accidents during this year[1]. Railroad monitoring strategies focus on preventative maintenance by detecting wheel anomalies using wayside detection methods (e.g., wheel-impact load detection), and direct detection of track anomalies using onboard systems (e.g., track geometry vehicles)[2]. Both approaches are periodic, manual, and do not support real-time track damage detection. In this paper, we focus on identifying discontinuities in the railway lines by means of the use of accelerometers installed on board a scale model of a railroad car and track system that was developed in collaboration with industry stakeholders.

Recent research has focused on detecting damage from acceleration signals obtained onboard moving vehicles. However, these approaches have not been specifically tested for detecting rail cracks, but instead have a broader focus on detecting and classifying damages in the infrastructure system that supports the rail track. Indirect structural health monitoring techniques have been applied to laboratory-scale bridges by detecting non-destructive damage, such as changes in the supporting conditions and adding masses to the span[3-4]. Instrumentation of operational trains has also been performed by installing accelerometers and a GPS onboard the car to track the measurements. However, not enough positive cases have been collected to characterize the failure, and a baseline obtained from multiple passes of the instrumented trains is used to identify anomalies[5-6]. Vehicle-rail interaction has also been studied using finite element models, but the scope of this research focuses on the stress analysis of the interacting members given the different types of discontinuity scenarios of the rail crack, rather than identifying the presence of damage[7-8].

*kflaniga@andrew.cmu.edu; phone 1 412 268-7274

Health Monitoring of Structural and Biological Systems XVII, edited by
Paul Fromme, Zhongqing Su, Proc. of SPIE Vol. 12488, 124881L

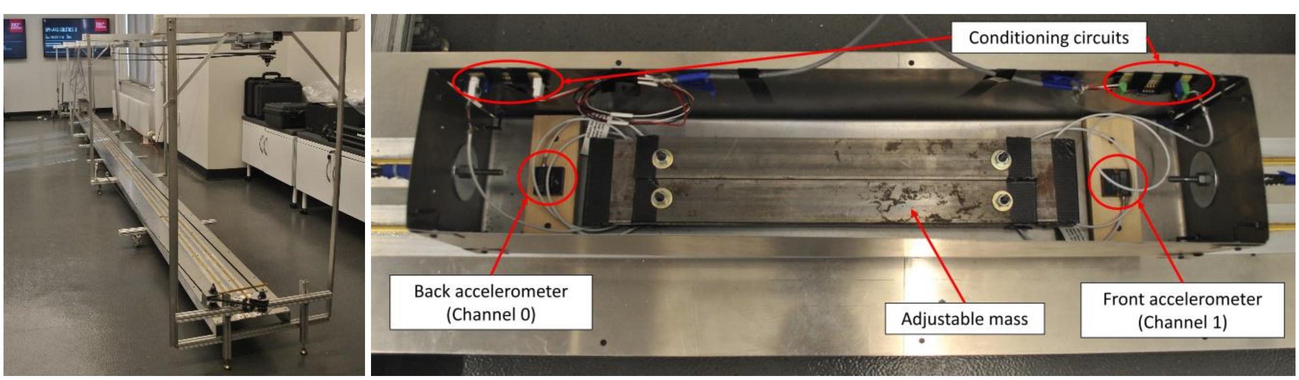
In the present study, we propose the use of a laboratory-scale model[9] of a railway and a railroad car to detect damage. This model allows for a controlled environment that can simulate different kinds of real-world discontinuities and therefore evaluate the effectiveness of the proposed damage detection methods. The use of a scale model offers several advantages, including having a safe environment, where in the case of derailments caused by the inflicted cracks, will not cause damage to a real train or injure a person, and also the ability to identify the best detection strategy before trying to implement the instrumentation of a real railroad car. The testbed model consists of a straight segment of rails that can simulate a localized discontinuity in three dimensions (transverse, vertical, and longitudinal). These discontinuity scenarios are designed to consider that when a crack occurs, the rupture can follow different geometric configurations, and the model setup can simulate a combination of these degrees of freedom. Additionally, the railroad car developed is a down-scaled version of a standard 50-ft boxcar that preserves its geometric proportions and its original dynamic features. The railroad car is propelled using a motorized pulley system, which is driven by a stepper motor. The stepper motor is controlled by software that also synchronously controls the data acquisition system. This synchronous control ensures that the movement of the railroad car and the data acquisition are tightly coupled, enabling accurate and precise measurements of the vehicle's behavior as it moves along the track. The testbed is described in more detail in a separate manuscript[9]. Using this track and moving vehicle actuation system with two onboard accelerometers, several types of feature extraction and dimensionality reduction techniques are employed to evaluate their ability to separate damaged and undamaged records. These techniques have been used widely in related fields, such as highway and bridge monitoring showing positive results, and we derive inspiration from this previous work. The results of damage sensitivity show that principal component analysis and independent component analysis have the highest combination of recall and true negative rate, compared to other techniques. By exploring and utilizing these approaches, this work contributes to the development of more efficient and accurate methods for identifying and assessing damage in railway systems.

2. EXPERIMENTAL SETUP

The experimental setup simulates passing a railroad car over a straight section of rail with a localized discontinuity. The vehicle-rail interaction is measured with two uniaxial accelerometers placed onboard the vehicle directly on top of the suspension system located on each truck (Figure 1). The overall system primarily consists of the vehicle, the rail track setup and localized damage simulation, and the motion control and data acquisition system. A more detailed description of the system can be found in Yin et al. 2023[9].

2.1 Vehicle

In this study, a downscaled version of a standard 50-ft boxcar is considered for the design of the laboratory-scale model of the railroad car. The dimensional downscaling factor of 23.7 is selected because it results in a gauge of 2.5 inches, which makes the fabrication of the testbed suitable for a laboratory setting. The overall size of the downscaled railroad car also allows for it to achieve the desired range of speeds in a 24-ft track. The wheels used in the downscaled version are obtained from G-scale train models, which closely match the downscaling factor. The suspension system is 3D printed and



(a) (b)

Figure 1. Downscaled (a) rail track testbed and (b) railroad car with installed accelerometers and added mass.

customized to fit the wheels and springs. The springs are calculated to match the fundamental frequencies of the trains, according to the design guidelines[10], considering the range of masses used in the laboratory railroad car. This design allows for a more controlled and convenient testing environment, while still retaining the dynamic features and the dimensional proportions of a standard railroad car.

In order to measure the vehicle-rail interaction in the downscaled railroad car, two SDI 2012-002 uniaxial accelerometers[11] are mounted on top of each truck. The accelerometers are connected to two conditioning circuit boards, which allow for tuning the baseline, setting the gain, and low-pass filtering the acceleration signals. The conditioned signals are then transmitted via individual cables mounted on top of the testbed, which are designed to move along with the vehicle, avoiding any interference with the dynamic response of the railroad car. The cables connect directly to a National Instruments (NI) data acquisition system, which is connected to a computer capable of synchronously actuating the vehicle and collecting the data.

2.2 Rail track setup

The rail track is designed to make the railroad car travel over a localized discontinuity located at the central point. The discontinuity is modeled in three dimensions: transverse, vertical, and longitudinal, which can be modified with an adjustable bolted connection. This is done to consider the various geometric configurations that can occur when a crack of these characteristics takes place. The rail tracks are obtained from a G-scale train model series, which are chosen to match the wheels of the downscaled railroad car. As a result, the proportions of the dimensions of the rail tracks are similar to the dimensions of the railroad car. The overall length of the rail is 24 ft and it is divided into three sections of the same length. The first and last sections are used for the vehicle to accelerate and decelerate at the start and end of each run, while the middle section is used to modularly simulate the crack and have the vehicle moving at a constant maximum speed. This arrangement allows for a controlled and systematic evaluation of the vehicle-rail interaction under different conditions. The rail track setup also supports a conveyor system attached to the railroad car. The system consists of head and tail pulleys and a stepper motor, which is controlled by the same computer that is connected to the data acquisition system.

3. METHODOLOGY FOR DAMAGE DIAGNOSIS

3.1 Methodology overview

Rail routes typically cover hundreds of kilometers and instrumenting a train and analyzing the signals collected for the entire route can result in a significant amount of data that may require high computational demand. In this paper, we propose to discretize the rail route of interest into sub-segments, which are individually analyzed to detect anomalies. Another benefit of this approach is that analyzing each segment is useful to localize where the detections are located. With this in mind, our experiment considers that the testbed model represents one of the segments of the railway that would be found in reality.

In each of the runs, the acceleration signals are collected onboard the downscaled railroad car. Through the runs, the testbed is adjusted to simulate conditions where the rails in some cases are undamaged, and others are damaged. Acceleration signals are analyzed by simultaneously comparing a single damaged or undamaged run with its corresponding baselines (undamaged runs) in which the mass and the velocity coincide. On each one of these analyses, a dimensionality reduction of the set of runs is performed until a unidimensional scalar is obtained for each run. For practical real implementation, it is worth noting that an online version of the selected optimal algorithm can be employed to enhance the computational performance of the damage detection system. Here, in the case of positive damage detections, it is determined if the damage sensitive feature scalar can be separated from the ones related to the undamaged conditions by setting a threshold, as conceptually shown in Figure 2. If a scalar is beyond the set threshold, then the classification of the processed signal is positive, and on the contrary, if the value is near that of the baseline runs, then it is classified as a negative detection.

In the present study, we employed a Box-Behnken Design[12] (BBD) to design the testing matrix for our testbed. BBD is a widely used experimental design method that is well-suited for the optimization of multiple feature input variables. In our case, the testbed allows the manipulation of several features, including vehicle speed, mass, and rail discontinuity in three different directions. The different combinations of these features reveal which one of them, or which combinations, make the proposed damage detection technique more or less effective.

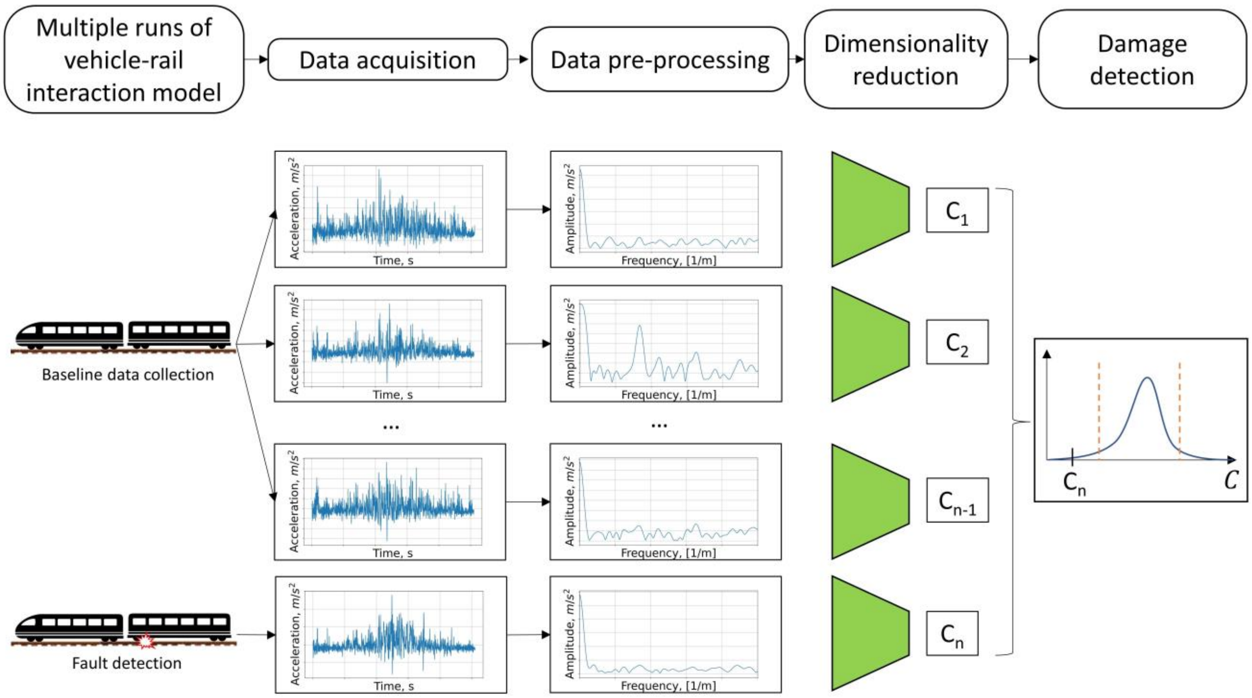


Figure 2. Proposed scheme of baseline and faulty run combinations, signal processing, and dimensionality reduction.

Table 1. Downscaled and upscaled testing features.

	Features	Total mass	Added mass	Speed	Cracks features: Longitudinal	Cracks features: Transverse back rail	Cracks features: Vertical front rail
Downscaled values	Units	Kg	Kg	m/s	mm	mm	mm
	Cases	3.50	1.00	0.9	0	0	0
		5.00	2.50	1.13	1.42	0.20	0.30
		7.50	5.00	1.35	2.55	0.40	0.45
Upscaled values	Units	lbs	lbs	mph	in	in	in
	Cases	103,754	29,644	47.6	0	0	0
		148,220	74,110	59.8	1.32	0.19	0.28
		222,330	148,220	71.4	2.37	0.37	0.42

Ultimately, a total of 9 baseline runs with 25 repetitions and 51 damaged runs with 10 repetitions each are performed to assess the performance of the train's damaged rail detection system as shown in Figure 3. The experimental design considers the features shown in Table 1. Here, each feature has 3 different values according to the BBD test matrix design, where each one of them represents a minimum, mean, and maximum expected value. The first 9 baseline runs evaluate the different combinations of masses and speeds as shown in the top nine rows in green of the table in Figure 3, which are then compared to their corresponding damaged conditions. The 25 baseline runs are compared with one damaged condition run at a time to evaluate the ability of the train to accurately identify a damaged rail in real conditions, as shown in Figure 2. For the case of undamaged detections, 24 baseline runs are compared with the undamaged run left, and this is iterated for each one of them 25 times. Being able to detect a single normal or anomalous run is crucial as the train must be able to detect a damaged rail on the first measurement in order to prevent a potential train accident.

For the data processing, each set of signals (25 baselines and 1 fault run or 24 baselines and 1 undamaged run) is converted from the time domain to the space domain in order to have each acceleration coordinate associated with its

	M	S	Lc	Tc	Vc
1	1.0	0.9	0.0	0.0	0.0
2	1.0	1.13	0.0	0.0	0.0
3	1.0	1.35	0.0	0.0	0.0
4	2.5	0.9	0.0	0.0	0.0
5	2.5	1.13	0.0	0.0	0.0
6	2.5	1.35	0.0	0.0	0.0
7	5.0	0.9	0.0	0.0	0.0
8	5.0	1.13	0.0	0.0	0.0
9	5.0	1.35	0.0	0.0	0.0
10	1.0	1.13	0.0	0.2	0.3
11	2.5	0.9	0.0	0.2	0.3
12	2.5	1.13	0.0	0.0	0.3
13	2.5	1.13	0.0	0.2	0.0
14	2.5	1.13	0.0	0.2	0.45
15	2.5	1.13	0.0	0.4	0.3
16	2.5	1.35	0.0	0.2	0.3
17	5.0	1.13	0.0	0.2	0.3
18	1.0	0.9	1.42	0.2	0.3
19	1.0	1.13	1.42	0.0	0.3
20	1.0	1.13	1.42	0.2	0.0
21	1.0	1.13	1.42	0.2	0.45
22	1.0	1.13	1.42	0.4	0.3
23	1.0	1.35	1.42	0.2	0.3
24	2.5	0.9	1.42	0.0	0.3
25	2.5	0.9	1.42	0.2	0.0
26	2.5	0.9	1.42	0.2	0.45
27	2.5	0.9	1.42	0.4	0.3
28	2.5	1.13	1.42	0.0	0.0
29	2.5	1.13	1.42	0.0	0.45
30	2.5	1.13	1.42	0.2	0.3
31	2.5	1.13	1.42	0.4	0.0
32	2.5	1.13	1.42	0.4	0.45
33	2.5	1.35	1.42	0.0	0.3
34	2.5	1.35	1.42	0.2	0.0
35	2.5	1.35	1.42	0.2	0.45
36	2.5	1.35	1.42	0.4	0.3
37	5.0	0.9	1.42	0.2	0.3
38	5.0	1.13	1.42	0.0	0.3
39	5.0	1.13	1.42	0.2	0.0
40	5.0	1.13	1.42	0.2	0.45
41	5.0	1.13	1.42	0.4	0.3
42	5.0	1.35	1.42	0.2	0.3
43	1.0	1.13	2.55	0.2	0.3
44	2.5	0.9	2.55	0.2	0.3
45	2.5	1.13	2.55	0.0	0.3
46	2.5	1.13	2.55	0.2	0.0
47	2.5	1.13	2.55	0.2	0.45
48	2.5	1.13	2.55	0.4	0.3
49	2.5	1.35	2.55	0.2	0.3
50	5.0	1.13	2.55	0.2	0.3

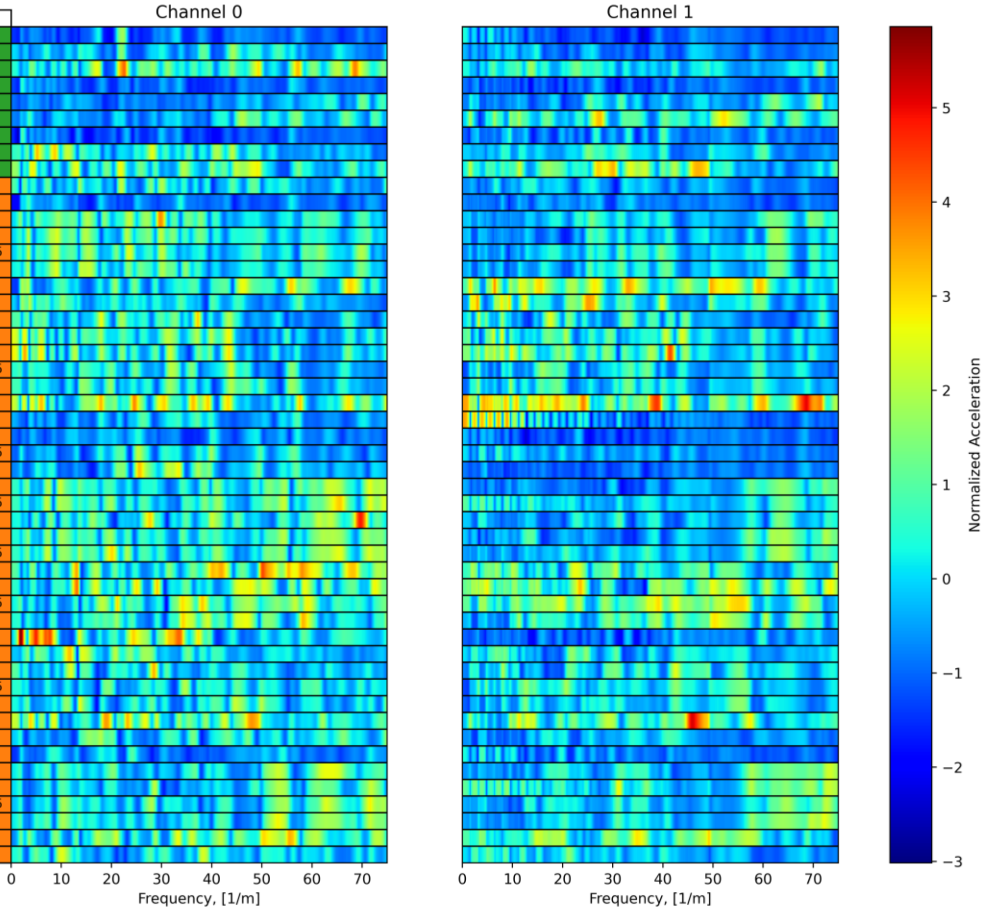


Figure 3. Normalized average spatial frequency response of baseline (green) and damaged (orange) runs for each tested configuration. Table column titles correspond to M: added mass (Kg); S: top speed (m/s); Lc: longitudinal crack (mm); Tc: transverse crack (mm); Vc: vertical crack (mm).

location (Figure 4), which is particularly important if the baselines used for the analysis are made at different speeds. Once this transformation is performed, due to the non-constant speed of the railroad car, there is not a uniform distribution of the measured accelerations, therefore the data is linearly resampled at intervals of 0.01mm. After resampling, the signals are converted to the frequency domain using Fast Fourier Transformation (FFT). Here the phase of each spatial frequency is not considered to avoid the influence of possible misalignments of the data. Finally, the data is normalized as shown in Figure 3, and dimensionality reduction is performed until a single scalar for each run is obtained. In this investigation, several dimensionality reduction techniques are compared to analyze which one of them performs better in separating the damaged from the undamaged signals.

For classifying the different runs, the resulting scalars from the runs are assumed to follow a normal distribution. The mean and standard deviation of these scalars are calculated for each set of signals. Here, the classification threshold is defined by the quantile which maximizes the overall performance of the classification. In each set of signals the last run is classified and the previous ones are assumed to be known, so when the damaged run is being tested, only the faulty run is being classified and the other 25 baseline runs are assumed to be known. On the other hand, for the case of testing an undamaged run, only the last run is classified, and the other 24 runs are assumed to be known. To obtain the optimal quantile classification threshold, all damage sensitive features are respectively standardized with the Z-score according to each set, then the cumulative percentage of each point is obtained. With all data set converted to percentage, the optimal quantile classification threshold is obtained by shuffling the dataset of the damaged and undamaged run testing and 75% of them are selected as a training data set. This process is repeated 100 times, and the mean optimal quantile value is used as the threshold for each channel and dimensionality reduction algorithm. By doing so, we can effectively determine the

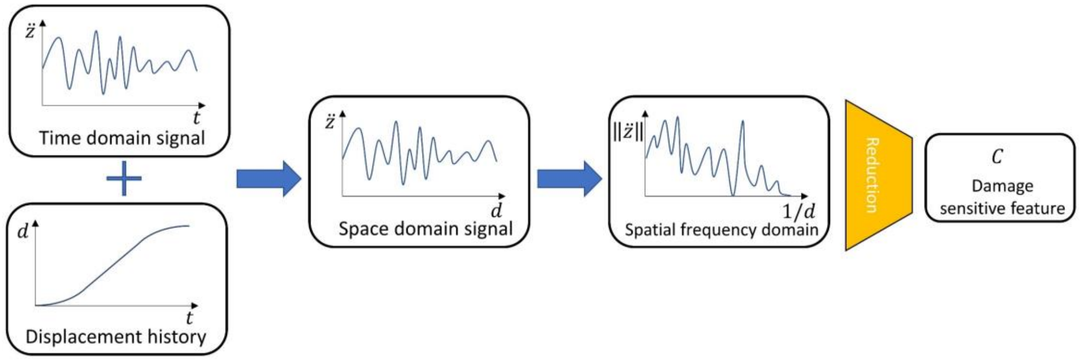


Figure 4. Block diagram of the proposed signal processing and extraction of damage sensitive features.

optimal quantile that leads to a classification threshold and ensure a robust analysis of the scalars while avoiding overfitting. The only dimensionality reduction method that does not follow entirely this classification methodology is the stacked autoencoder; since the activation function "ReLU" is used, the damage scalar is either 0 or a positive number, which does not add variability to the dataset and is justified assuming a normal distribution and using a quantile as a threshold. Therefore the limit is defined in the same units as the damage sensitive feature.

3.2 Background of dimensionality reduction and anomaly detection

In this paper, we aim to analyze several dimensionality reduction techniques for the purpose of processing collected signals from the laboratory scaled railroad car. The objective is to obtain a 1-dimensional feature that can distinguish between normal operating conditions and anomalous ones. In the following part of this section, we introduce the techniques that are considered in our analysis. A brief overview of each technique is covered to provide a comprehensive understanding of the methods used in this study.

Principal component analysis (PCA)

Linear dimensionality reduction techniques transform high-dimensional datasets into a new set of orthogonal, linearly uncorrelated components referred to as "principal components". The first principal component is the direction along which the data varies the most, and each subsequent component has the highest possible variance under the constraint of being orthogonal to the previous components. In this research, each time series is considered as an n-dimensional feature point, where n corresponds to the number of data points in the signal.

Independent component analysis (ICA)

Independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals[13]. This technique is used for separating a multivariate signal into independent, non-Gaussian component signals, assuming that the known variables are a linear combination of unknown latent variables. In our experiment, we consider a single-factor reduction, and the obtained weight given the input signal is used as the damage-sensitive feature.

Isomap

Isomap is a nonlinear dimensionality reduction method that uses the graph distance as an approximation of the geodesic distance of a Euclidean manifold[14]. This method intends to preserve the intrinsic geometry of the data by computing a lower-dimensional representation that best approximates the distances between the data points within the manifold. In this application, the collected acceleration measurements are represented in their projection on a one-dimensional manifold.

Laplacian Eigenmaps

Laplacian Eigenmaps construct embeddings based on the properties of the Laplacian matrix. This method provides a low-dimensional representation of the dataset that optimally preserves local neighborhood information[15]. The algorithm generates a discrete approximation to a continuous map that naturally arises from the geometry of the Euclidean manifold.

As in Isomap, the data of each run is represented in the unidimensional projection of the manifold as the damage-sensitive feature.

Self-organizing feature maps (or Kohonen map)

The goal of this method is to represent all points in the source space by Kohonen map points in a target space, such that distance and proximity relationships are preserved as much as possible[16]. This algorithm uses competitive learning where the neurons that are more alike to the input source points (best matching units) are adjusted towards the input vector. For our application, the mapping space is unidimensional and the considered damage-sensitive feature is the resulting score of the single unit given the input signal.

Stacked autoencoder

Autoencoders are neural networks that have the ability to discover low-dimensional representations of high-dimensional data and are able to reconstruct the input from the output[17]. The autoencoder consists of two parts, the encoder and the decoder. The encoder maps the input data into a lower-dimensional representation, also known as the latent representation or encoding. The decoder then maps the latent representation back to the original input space. The goal of the autoencoder is to learn an encoding that preserves the most important information from the input data, while reducing the dimensionality. The encoder and decoder are trained such that the reconstruction error between the input and the reconstructed output is minimized. For our application, the encoding is a single neuron, and its value based on the input features is considered as the damage-sensitive feature.

4. EXPERIMENTS AND RESULTS

After utilizing the experimental setup outlined in Section 2 to create a dataset, the detection method presented in Section 3 is performed. In this section, we present the corresponding results of the presented methods and compare their ability to properly classify the runs. If a reduced signal obtained from a damaged simulation is classified beyond the predefined threshold, it is considered to be a true positive (TP) and if it is below, a false negative (FN). If a damage-sensitive scalar is obtained from an undamaged condition and is beyond the threshold, it is considered as false positive (FP), and if it is below this limit, it is considered as a true negative (TN). Figure 5 shows the receiver operating characteristic (ROC) curve and the area under curve (AUC) of each dimensionality reduction method for Channel 0 and Channel 1. This plot shows the relationship between the true positive rate (TPR) or recall, i.e., $TP/(TP+FN)$, and false positive rate, i.e., $FP/(FP+TN)$ for different threshold levels, which were previously defined as a quantile level for each case scenario, except for the autoencoder as mentioned in Section 3.

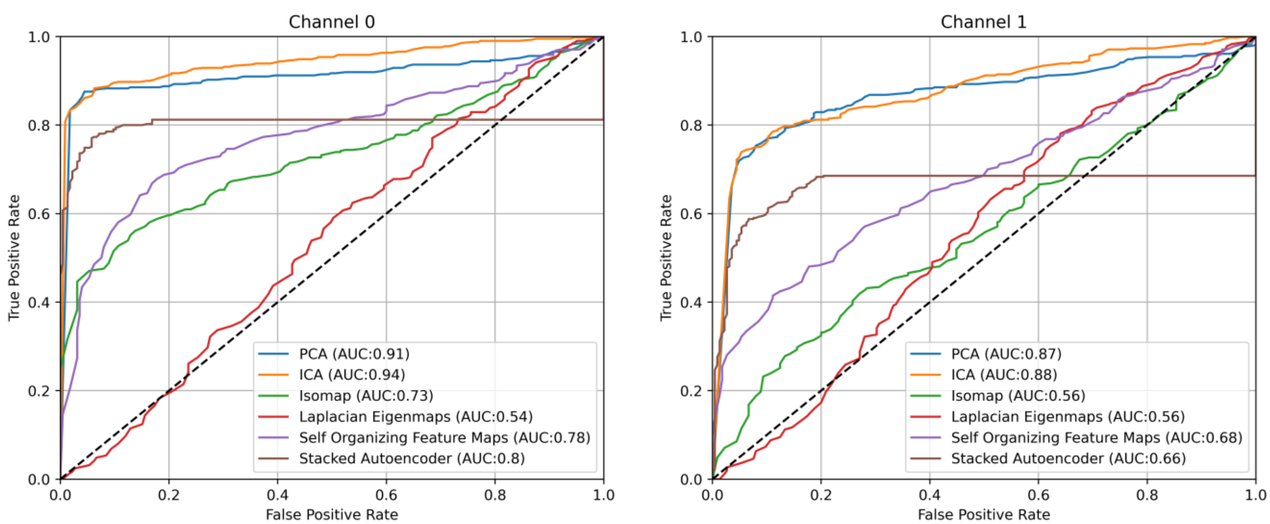
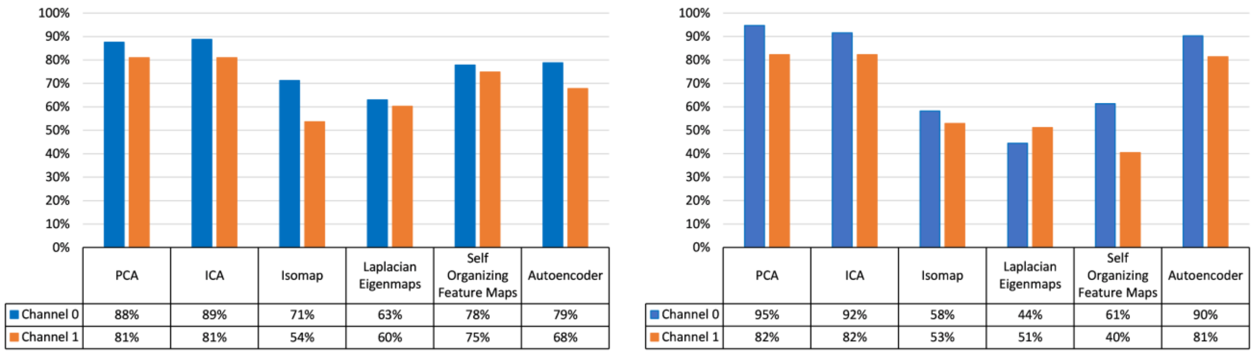
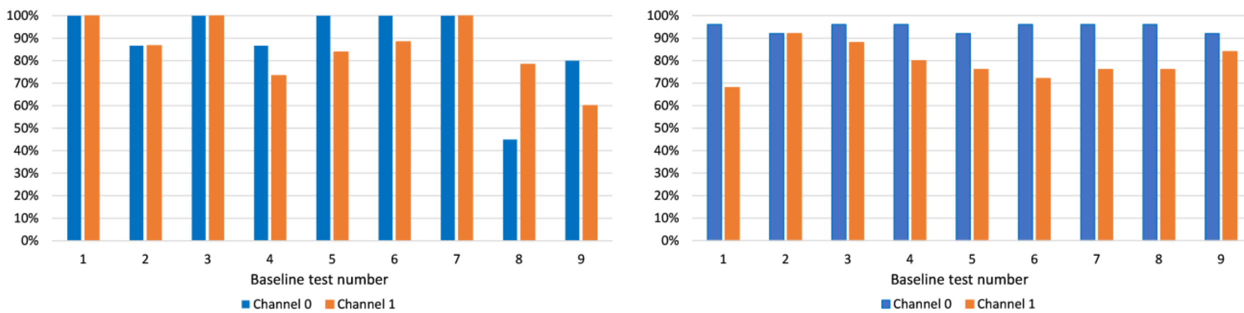


Figure 5. ROC curve and AUC of tested dimensionality reduction techniques for the front (Channel 1) and back (Channel 0) accelerometers.



(a) Recall (b) True Negative Rate

Figure 6. (a) Recall and (b) true negative rate of tested dimensionality reduction techniques for the front (Channel 1) and back (Channel 0) accelerometers.



(a) Recall (b) True Negative Rate

Figure 7. PCA mean (a) recall and (b) true negative rates for different baseline cases.

For both Channel 0 and Channel 1, the performance of the tested methods is classified in the same order according to AUC. However, it is worth noting that in Channel 0, the AUC values are higher than for Channel 1 for every tested method, except for Laplacian Eigenmaps where the difference between accelerometers is neglectable. Based on this analysis, ICA and PCA are the methods that are able to better separate damaged and undamaged conditions. These methods show consistent performance across both channels and had higher AUC values compared to other methods tested.

For each analyzed method the optimal threshold to classify damaged and undamaged signals is obtained according to the procedure mentioned in Section 3. Figure 6 shows the recall and true negative rate, i.e., $TN/(TN + FP)$, results obtained from the experiments using the proposed anomaly detection approaches. Recall measures the proportion of actual positives that are correctly identified, and true negative rate quantifies the proportion of actual negatives that are correctly identified. By adopting this approach, it becomes possible to assess the effectiveness of each method in accurately detecting both positive and negative signals.

Regarding the recall metric, PCA and ICA perform the best for Channel 0 with a difference of 1% between each other. For Channel 1, the PCA and ICA methods also perform best, showing the same percentage as Channel 0. The similarity for the results achieved with both PCA and ICA rely on the fact that both are linear methods in which only the principal component was used; therefore, the fact that PCA has the restriction that the projections must be orthogonal does not have almost any influence on the results.

In evaluating the true negative rate for the two channels, it is observed that again both PCA and ICA outperform other methods in Channel 0 and Channel 1. In this case, the autoencoder shows a similar percentage to these methods, but for the case of the recall its performance is considerably lower than ICA and PCA. PCA and ICA achieve the best overall performance considering recall and true negative rate, however, since PCA achieves a higher performance considering both recall and true negative rate, this method is further analyzed. Figure 7 presents more specific results, showing the average of the previously mentioned metrics for each one of the baselines combined with the different tests. Baseline test numbers and their respective mass and speed are shown in the top nine rows in green of the table in Figure 3.

The results indicate that the combinations of faulty runs with baseline 8 have the most significant impact on the average recall results. These combinations show a lower recall rate, indicating that these combinations result in signals that are more difficult to detect with this method. However, it is worth noting that there are also combinations for Channel 0 that achieve perfect recall. This suggests that the test set ups that align with baseline 8 have a combination of parameters resulting in the vehicle exhibiting a dynamic response that is not being adequately captured by the sensors or in vibrations that are challenging to classify with this methodology. On the other hand, the evaluation of the true negative rate for the different tests achieves percentages that are consistent across baselines and accelerometers for Channel 0, resulting in relatively high percentages ranging from 92% to 96%, and for Channel 1 the percentages range from 68% to 92%. These findings suggest that the selected fault detection methods are effective in identifying normal behavior in the system for Channel 0, which is crucial for the implementation of a reliable system.

For the implementation of PCA for damage detection, we evaluate optimizing the algorithm by analyzing the variability of the principal component vectors across tests. This analysis aims to determine if the principal components axis can be preserved without the need for recalculating them each time a new set of data is analyzed. However, prior to this analysis, it is necessary to confirm that within each test configuration the principal component vector has low variability. Figure 8 presents the standard deviation of the different principal components within each test configuration for the different runs, with the dot product distance function used to evaluate the difference between vectors in terms of their relative angle.

Although there are variations in the standard deviation of the principal components between test configurations, the value tends to exceed 0.5, indicating a deviation of 60° or less between different runs within each test and its corresponding mean. Consequently, we assess the dot product of the mean vector for each test number to evaluate the dissimilarities between them across different test configurations, which is shown in Figure 9.

In terms of consistency with the direction of the principal component, the respective mean values for Channel 0 and Channel 1 are 0.21 and 0.22, with corresponding standard deviations of 0.21 and 0.19. While these metrics suggest that Channel 1 performs slightly better due to its higher mean and lower standard deviation, the difference is not significant. To optimize computational performance, we recommend using the online version of PCA[18] if this method is to be implemented. However, given the significant variability in the principal component vectors, it is crucial to update them constantly to maintain the separability of the classes.

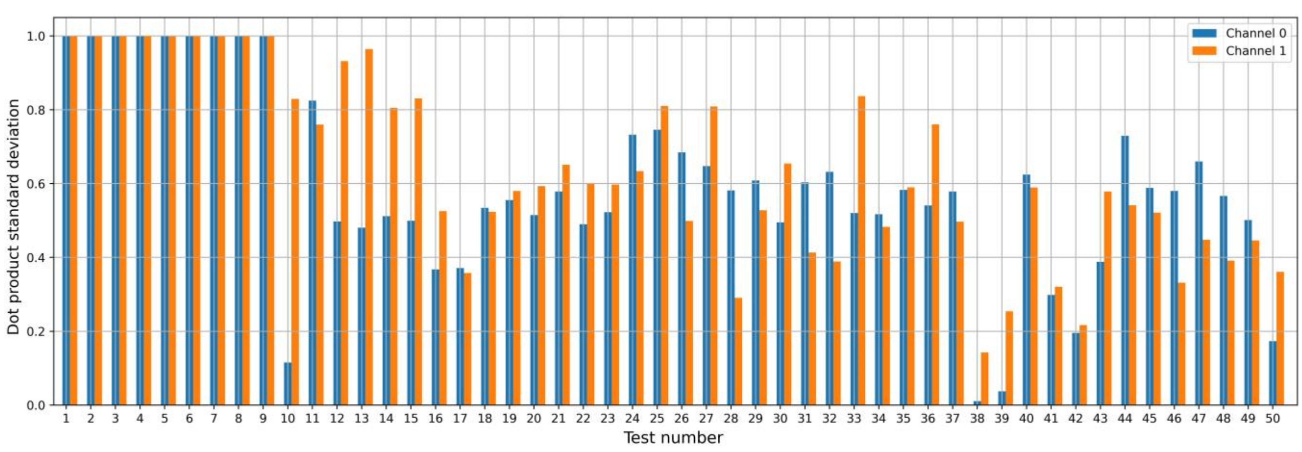


Figure 8. Standard deviation of principal components for test configurations, considering the dot product as the distance function.

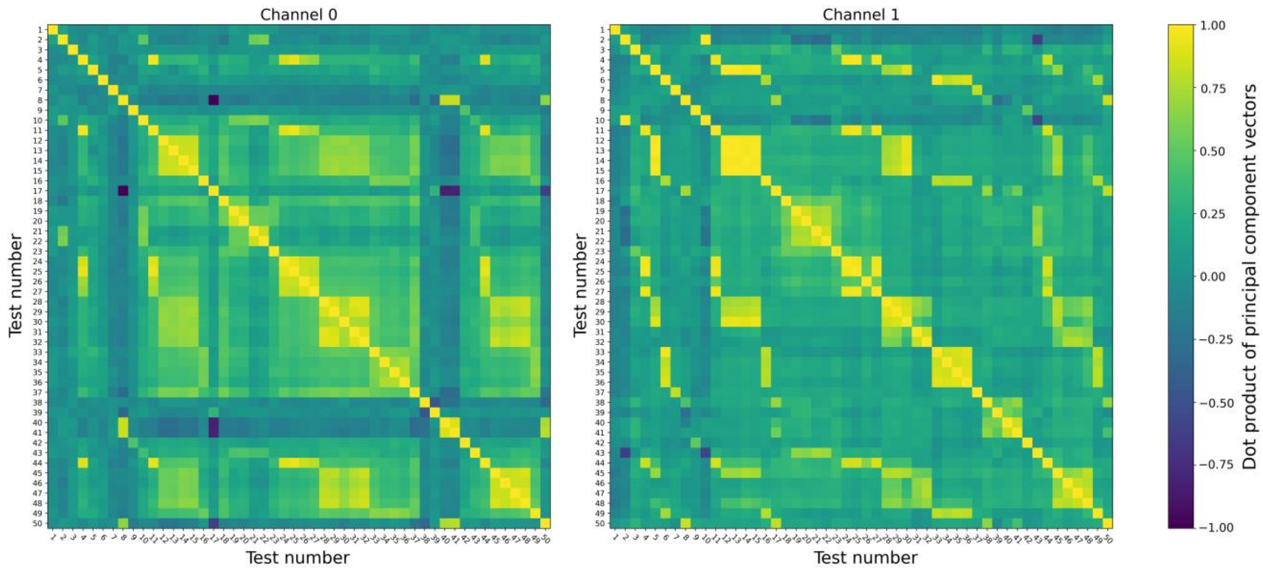


Figure 9. Dot product correlation matrix of principal component between test configurations.

5. CONCLUSIONS AND FUTURE WORK

The study presented examines the ability of an indirect structural health monitoring approach to diagnose damage, using acceleration signals obtained from a downscaled railroad car passing over simulated cracks of varying type and severity. To detect damage with high accuracy, signal processing and dimensionality reduction methods are employed and compared. The efficacy of various methods for detecting damage in rails using acceleration signals obtained from a testbed is analyzed. The results indicate that PCA is the most balanced method for accurately detecting positive and negative cases and performs closely to the best method in detecting damaged rails, achieving an 88% recall and 95% true negative rate for Channel 0 (back accelerometer). While the method performs well overall, for recall, combinations of damaged conditions with baseline 8 prove to be more challenging to detect. We hypothesize that the combination of mass and speed of the vehicle may be contributing generating signals that cannot be differentiated from the baseline conditions. To further explore the effectiveness of PCA, testing smaller damage with less severity can be explored, though we acknowledge that smaller downscaled cracks are challenging to simulate accurately, and a higher or full-scale railroad car should be employed.

We explore the possibility of optimizing detection processes by analyzing the consistency of the direction of the principal component across different configurations. This analysis aims to eliminate the need for recalculation of the main component vector every time a detection is evaluated if the direction remains consistent. Our results reveal that the principal component vectors within each experiment configuration remain closely aligned across all of the runs. However, when comparing the mean principal component vectors across different test configurations, we observe considerable variability for both Channel 0 and Channel 1. This suggests that if PCA is to be applied for real-time damage detection, it should be recalculated or used online[18] to maintain the accuracy and reliability of the detection process.

The findings of this study suggest that future research should explore damage severity estimation, by correlating damage-sensitive features with the severity level of the damage. We also recommend exploring the analysis of time-domain signals with the same framework. However, this brings additional preprocessing challenges such as synchronizing and scaling the signals. Lastly, multidimensional feature reduction can be explored to evaluate if different detection conditions can be successfully clustered with equal or better results than the ones obtained in this experiment. These avenues of research will help to advance the field of damage detection in dynamic systems and enhance the overall accuracy of machine learning methods, which have the potential of being fully deployed in full-scale railroad cars.

ACKNOWLEDGEMENTS

The authors would like to gratefully acknowledge Brian Belowich for his support in the development of the testbed used for this research. This research is sponsored in part by the Pennsylvania Infrastructure Technology Alliance (PITA). Information developed under this Award was or is sponsored by the U.S. ARMY Contracting Command under Contract Number W911NF20D0002 and W911NF22F0014 delivery order No. 4.

REFERENCES

[1] Federal Railroad Administration, "1.12 - Ten Year Accident/Incident Overview", <https://safetydata.fra.dot.gov/officeofsafety/publicsite/query/TenYearAccidentIncidentOverview.aspx> (21 February 2023).

[2] Flanigan, K. A., Aguero, M., Nasimi, R., Moreu, F., Lynch, J. P. and Ettouney, M., "Objective Resilience Monitoring for Railroad Systems," ASCE Objective Resilience Manual of Practice, Objective Resilience: Technology, (ed. M. Ettouney), ASCE Press (2022).

[3] Cerda, F., Garrett, J., Bielak, J., Rizzo, P., Barrera, J., Zhuang, Z., Chen, S., McCann, M. and Kova, J., "Indirect structural health monitoring in bridges: scale experiments," Proceedings of the International Conference on Bridge Maintenance, Safety, and Management (July 2012).

[4] Liu, J., "HierMUD: Hierarchical multi-task unsupervised domain adaptation between bridges for drive-by damage diagnosis," Structural Health Monitoring, (2022).

[5] Lederman, G., Chen, S., Garrett, J. H., Kovačević, J., Noh, H. Y. and Bielak, J., "Track monitoring from the dynamic response of a passing train: A sparse approach," Mechanical Systems and Signal Processing, **90**, 141–153 (2017).

[6] Liu, J., Wei, Y., Bergés, M., Bielak, J., Jr, J. H. G. and Noh, H., "Detecting anomalies in longitudinal elevation of track geometry using train dynamic responses via a variational autoencoder," SPIE Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems, **10970**, 389–401 (2019).

[7] Gao, Y., Wang, P., Wang, K., Xu, J. and Dong, Z., "Damage tolerance of fractured rails on continuous welded rail track for high-speed railways," Railroad Engineering Science, **29**(1), 59–73 (2021).

[8] Suárez, B., Rodriguez, P., Vázquez, M. and Fernández, I., "Safety assessment of underground vehicles passing over highly resilient curved tracks in the presence of a broken rail," Vehicle System Dynamics, **50**(1), 59–78 (2012).

[9] Yin, J., Montero, G., Flanigan, K. A., Bergés, M. and Brooks, J. D., "Open-source hardware for a laboratory-scale track and moving vehicle actuation system used for indirect broken rail detection," SPIE Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems (Mar. 2023).

[10] Harris, C. M., [Shock and vibration handbook Vol. 3], eds. C. M. Harris and C. E. Crede, McGraw-Hill Companies, Inc. (1961).

[11] Silicon Designs, Inc., "SDI 2012 / 2422 Accelerometers," <https://www.silicondesigns.com/2012-2422> (21 February 2023).

[12] Ferreira, S. L. C., Bruns, R. E., Ferreira, H. S., Matos, G. D., David, J. M., Brandão, G. C., da Silva, E. G. P., Portugal, L. A., dos Reis, P. S., Souza, A. S. and dos Santos, W. N. L., "Box-Behnken design: an alternative for the optimization of analytical methods," Analytica Chimica Acta, **597**(2), 179–186 (2007).

[13] Hyvärinen, A., Karhunen, J. and Oja, E., [Independent Component Analysis] (2001).

[14] Lee, J. A. and Verleysen, M., [Nonlinear dimensionality reduction], Springer, New York (2007).

[15] Belkin, M. and Niyogi, P., "Laplacian Eigenmaps for dimensionality reduction and data representation," Neural Computation, **15**(6), 1373–1396 (2003).

[16] Duda, R. O., Hart, P. E. and Stork, D. G., [Pattern Classification, 2nd edition], Wiley-Interscience, New York (2000).

[17] Alla, S. and Adari, S. K., [Beginning Anomaly Detection Using Python-Based Deep Learning: With Keras and PyTorch], Apress, Berkeley, CA (2019).

[18] Boutsidis, C., Garber, D., Liberty, E. and Karnin, Z., "Online Principal Component Analysis," ACM-SIAM Symposium on Discrete Algorithms (Jan. 2015).