Damage phenomena characterization in RCF tests using image analysis and vibration-based machine learning

M. Lancini [1], I. Bodini [1], C. Petrogalli [1], L. Provezza [1], M. Faccoli [1], G. Sansoni [1], L. Solazzi [1], A. Mazzù [1]
[1] University of Brescia, Department of Mechanical and Industrial Engineering
via Branze, 38 I-25123, Brescia, Italy
e-mail: **luca.provezza@unibs.it**

Abstract

Rolling Contact Fatigue (RCF) tests are a common effective method to study the behavior of wheel- and rail-steels. The measurements usually performed are discrete and destructive: they can only be performed at each intermediate stop of a test and they result in the alteration or destruction of the examined specimens. This work aims to assess the damage level steel samples during RCF tests, making continuous, non-destructive, and contactless measurements. A machine-learning technique based on vibration and torque measurements, together with 2D image was applied to RCF-dry tests carried out on different railway wheel steels tested according to the same operating parameters.

The proposed algorithm was able to quantitatively estimate the damage level of the samples by calculating the current data distance from specific references, e.g. a defined final damaged state. The used approach ensures a good degree of reliability both in terms of specificity and sensitivity.

1 Introduction

At least 10% of the 2000 train accidents recorded in the EU per year are likely caused by problems related to the rolling contact fatigue (RCF) phenomenon [1]. Victims, investigations, maintenance and replacement are just a few of the high costs involved. These, as well as the need of faster and more reliable railway material, constitute a strong push towards a better characterization of the materials used, and a more detailed study of the evolution of the damage mechanism.

The damage assessment in railway application is, in fact, still an open question. Many parameters affect the assessment in a wheel/rail contact: materials, load history, climatic condition, and sliding ratio. Different damage mechanism can also occur along with wear and rolling contact fatigue, the two principal players [2,3] that can interact and influence each other, even in competition.

The aim of this study is to add new tools and technologies to study wear and fatigue damages due to wheel-rail contact. One of the most common test that can be performed to characterize the wheel and rail steel material involves a twin-disk test bench, where specimens are used to simulate the real operating conditions of wheels and rails, with controlled operating conditions: rotation speed, slip ratio and load [3]. More and less accurate modelling of the damage can be found in the literature: some studies have focused on the mechanical and metallurgical characterization of different steels for wheels and rails, to identify the types of damage phenomena involved; while other researches focused on the development of analytical models.
A typical test procedure [16] using this test bench is to stop the machine at predefined cycles numbers, remove the samples, clean them, and perform some non-destructive tests, such weighing, magnetic

measurement and image analysis; then resume the test. Destructive measurements can be taken only at the end of the test, to identify the damage that let to failure, if any, and measure the cracks, where present. The need to stop the test to perform assessments and the impossibility of repeating a test after a destructive measurement is the main limitation to the study of damage evolution, since staircase approaches are time and resources consuming. Moreover, the interruption of the test, dismounting and mounting of the specimen could lead to different damage evolution, introducing relaxation phases.
The objective of the proposed study is to quantitatively estimate the damage status of the specimens using a continuous non-contact measurement system. This activity is based on the use of k-means clustering, a non-supervised Machine Learning technique, fed with vibration and torque measurement data, suitably pre-processed. Similar techniques have already been used to identify possible variations in the frequency responses of a particular mechanical model of the samples [23], but that approach required extensive post-processing of all tests data and operator-based manual tuning, while the proposed method is fully automated.
In the present work, k-means clustering method is implemented by combining three in different signals in time (two vibrations and one torque) in wheel-rail contact on RCF tests carried out by using a twin-disk bench. With the extraction of statistical and spectral features and a PCA analysis to reduce the number of features, K-means is applied to determine the damage in time on wheel specimens. The damage assessment is obtained with the clustering of the three signals. The novelty of the propose work is that there is a constant evaluation of damage in time with the possibility to have a characterization of materials in railway application with the use of an on-line method. In literature, there is only the determination of a fault condition with machine learning techniques. To have a confirm of the robustness of this method, a correlation with a R_b index, calculated from a binarization of images taken from a speed camera installed on twin-disk bench, it was given.

2 State of the art

Machine learning techniques have a wide use in many research field and they are used in signal vibrations in order to diagnose the health condition of rotary machines. The literature reports many works, in which several machine-learning techniques are used to classify different faults. Bearings of rotary machines are the most commonly studied parts, with research focused on detecting different defects. A lot of works in literature report Convolutional Neural Network (CNN) as a valid technique to extrapolate discriminative features to detect fault condition of beating parts [4-9]. Other works display a comparison with different machine learning methods (ANN, SVM, SOM, fuzzy method, Gaussian Process Regression), or using different machine learning languages with different features extractor algorithm to find higher accuracy in detection of fault conditions [10-13]. For example, Elfojani [14] correlates AE features with corresponding natural wear of slow speed bearing thought a series of laboratory experiments by using neural network model and Gaussian Process Regression. In [15], it was proposed a deep-hybrid model composed of the Convolutional Neural Network (CNN) and stacked denoising-autoencoder for unsupervised features learning and classification on multi-channel vibrations data. This deep hybrid model allows to obtain higher accuracy than single CNN. All these machine-learning techniques are used to handle big amounts of raw data and to detect fault conditions of rotary machines.

K-means algorithm is a simple machine learning technique that allows diagnostics of rotary machine through clustering. K-means is normally used when the number of data is not excessive. With the combination of an algorithm to reduce the number of features, there is the possibility to have a data clustering to find fault status in rotary machine. Several works in literature reports the use of k-means as an algorithm to detect abnormal or normal function of bearing elements in rotary machine [17-21]. In [22], k-means algorithm is used with a combination of vibration signal and thermal images features, to extract shape features using image segmentation.

To our current knowledge only few studies have been focused on quantifying the damage evolution, an improvement of the fault detection, especially without stopping the test.

3 Materials and Methods

3.1 Materials

Six different sets of railway-wheel samples supplied by Lucchini RS were analysed to verify and establish the robustness of the method. All these specimens, obtained directly from the real components, were tested on the twin-disk bench of the University of Brescia. The tests have the same duration in time with a number of cycles that corresponds to 130000 cycles. The tests ranging from 1 to 5 were performed with a periodically stop of the machine where the samples were dismounted, ultrasonically cleaned and weighted to evaluate their weight loss in time. The twin-disk bench was stopped at 10000, 20000, 50000 and 70000 cycles. Test 6 was performed without interruption. The test conditions, reported in table 1, were the same for all the tests carried out and the considered specimens were reported in table 2.

Table 1: Operative conditions of the tests.

Specimens Diameter [mm]	*Specimens thickness [mm]*	*Specimens Speed [r.p.m.]*	*Contact Pressure [MPa]*	*Sliding Ratio [%]*	*Applied Load [N]*	*Lubrication*
60	15	500	1100	1	7557	Air

Table 2: Considered Specimens.

Test Number	*Rail-Specimen material*	*Wheel-specimen material*
1	350HT	ER 8
2	350HT	ER 8
3	350HT	SANDLOS® S
4	350HT	SANDLOS® S
5	350HT	CLASS C
prrrrrrrrrr6	350HT	SANDLOS® H

The steel for the wheel specimens is different for every couple of the considered samples, while the steel for rail specimens is the same for all the tests. The steel considered for the rail specimens was 350HT EN13674-1 and it was a material typically used for rail applications. Thee steel considered for the wheel specimens was three and respectively are:

- ER8 EN13262
- SANDLOS® S, that is a modified AAR CLASS C
- CLASS C (AAR M107/M208 2004)
- SANDLOS® H, that is a modified AAR CLASS D.

The Principal steels properties are summarized in table 3.

Table 3: Principal properties of the steels considered.

STEEL	*350HT*	*ER8*	*SANDLOS® S*	*CLASS C*	*SANDLOS® H*
Rupture Stress [MPa]	≥1175	940	1180	1120	1270
Yield Stress [MPa]		590	750	715	800
Strain [%]	≥9	17	12	12	11
Brinell Hardness [HB]	345	203-255	347	355	354

3.2 Instrumentation

Figure 1 shows a schematic drawing of the twin-disk bench used in this work. The test bench was designed and developed internally and is deeply described in [16]: the specimens are positioned onto independent shafts, one of which can be displaced orthogonally to the shaft axis by a hydraulic cylinder that applies the imposed contact load. The rolling speed of the shafts is measured by encoders, and a load cell located at the piston head measures the contact load between specimens. The bench is equipped with a torque sensor with a full scale of 200 Nm positioned on the displaceable shaft, which can continuously acquire torque values at a synchronous sampling frequency of 5kHz. Piezo-accelerometers with a bandwidth of 26 kHz are fixed near each mandrel support, on the machine frame.

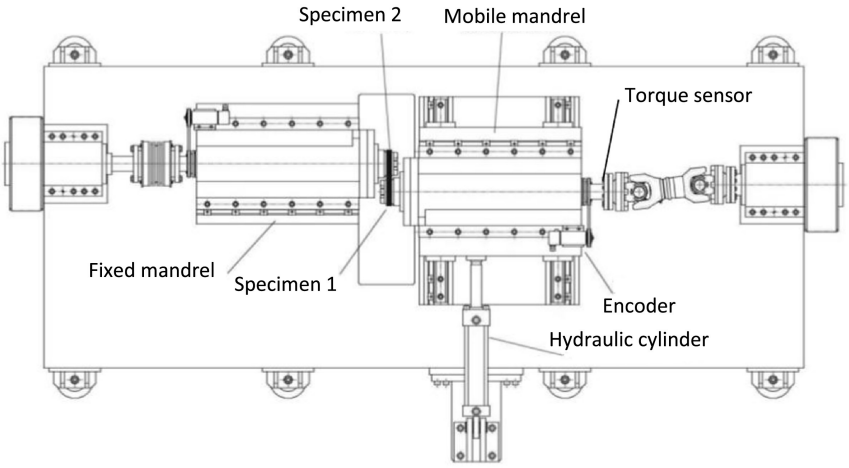

Figure 1: Schematic of the test bench

In addition to the signals of the torque and vibration, the twin-disc bench is equipped with a vision-based measurement system able to acquire frame with a high speed for the duration of all the test. The optical apparatus is based on the use of a speed camera (PROMON 501, AOS Technologies AG), and a system, composed with a defocused laser source and two laser stripes, for the illumination of the interested region (ROI). A detailed description of the optical apparatus for 2D and 3D surface analysis is given in [24]. The images are acquired at a frequency of 377 fps with an exposure time of 40 µs. In this way, there is the possibility to have three different ROI typologies: with the first one, there is the possibility to measure the 3D profile of the light stripe in axial direction of the specimen; with the second one can be measure the profile length of light that correspond at the encoder used for the angular position of the specimen, and the third ROI is used for the 2D elaboration of the surface.

3.3 Data processing

Data acquired from the two piezo-accelerometers and the torque sensor, are recorded in packets of 0.2 s having one thousand samples per channel. The next step in the preparation of the data is the features extraction: for the three sampled signals (two vibration and one torque), statistic-based and spectral-based features were extracted every 0.2 sec. This step is necessary to reduce the amount of data, from 3000 samples to 57 features.

The statistic-based features extracted from the three signals are: mean, standard deviation, variance, RMS, and quartiles. The spectral features are the centroid of the Power Spectrum Density and its quartiles. The signals where further combined to compute the maximum value of the cross-correlation and its associated time delay, the centroid of the Frequency Response Function and its frequency quartiles.

Some extracted features are redundant and were automatically removed by looking at the R^2 value: for each couple of features, the *coefficient of determination* was computed and the set of feature with the

lower overall determination was chosen. After this second cleaning, the number of features was reduced to 44.

As a subsequent step, the features have been normalized with a z-score method, to compare different features with different units of measurements: subtracting the average of its value, and dividing it by its standard deviation.

To have a further reduction of the computation of the algorithm, a Principal Component Analysis, PCA, was employed. This analysis was chosen to allow a reduction of the dimensional data with the transformation of n possible variables in fewer q variables linearly correlated to the initial ones. The criterion adopted for the PCs was to consider the first PCs that are able to describe at least 90% of the variance explained. The result of this analysis is a matrix of the original data projected in the space of the PC's principal components chosen. In conclusion, the dataset described in the PCs space is ready to be processed by the K-means algorithm.

After the reduction, the z-score normalization and the PCA analysis, the dataset is imported in the k-means clustering algorithm. It is important to specify that the mean and standard deviation values used for the normalization derives from the considered test, so the dataset is normalized in relation to the features of the actual test. When the algorithm reaches the convergence, the outputs are different.

To translate the classification information provided by the k-means approach into a quantitative measurement, a probabilistic approach has been chosen. Following this approach the distance between the point (in n-space) identified by the PCs as coordinates X, associated with the current condition, and each cluster centres C_i detected by the algorithm, is used as an indication of similarity between the two states. The membership probability P_i is computed directly as the reciprocal value of the square of this distance, later normalized to guarantee that $\Sigma(Pi)=1$. An example of result can be seen in Figure 2. It is clearly visible how we can clearly detected the phases in which the test bench was offline (in red), and that the system gradually changes from a state similar to the one found at the beginning ("pristine") and the one found at the end of the test ("damaged").

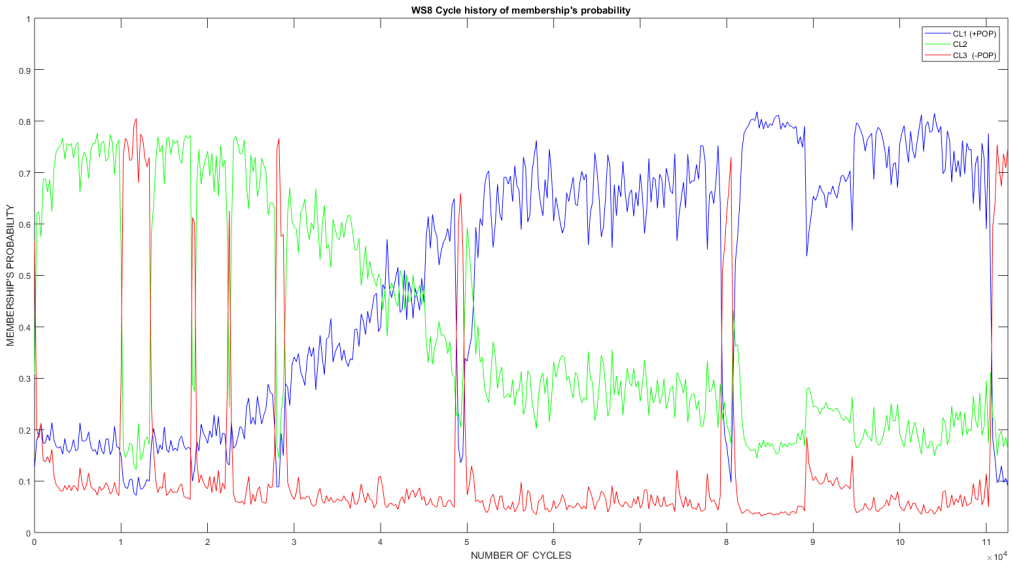

Figure 2: Example of Membership probability chart, for test number 1. In red the probability of belonging to the "test offline" cluster, in green the probability of belonging to the same cluster of the first cycles, and in blue the membership probability of belonging to the condition found at the end of the test.

The images acquired with the high-speed camera are analysed using different ROI. One of the ROI is used to perform a *blob analysis* to count small craters that form on the surface of the specimen during the test, likely due to fatigue and wear phenomena. The acquired images are elaborated with a binarization of the

images to have images clearer and, after this elaboration, it can be calculated synthetic indices to evaluate the trend of these blob.

The most relevant index that is possible to correlate to the k-means probability graph is the $R_b\%$ calculated with this formulation:

$$R_b\% = \frac{\sum_{i=1}^{n} A_{bi}}{A_{ROI}} \times 100$$

where n is the number of blob found in the considered ROI, A_{ROI} is the total area of the interested region and the numerator is the sum of the n Area blob found in the ROI. This index is calculated for every frames acquired during the test and then, it is averaged at defined number of cycles in order to find an averaged index. This averaged $R_b\%$ represents the evolution of the damage ratio expressed as fraction.

Figure 3 shows an example of how the surface image was analysed and the Rb coefficient computed.

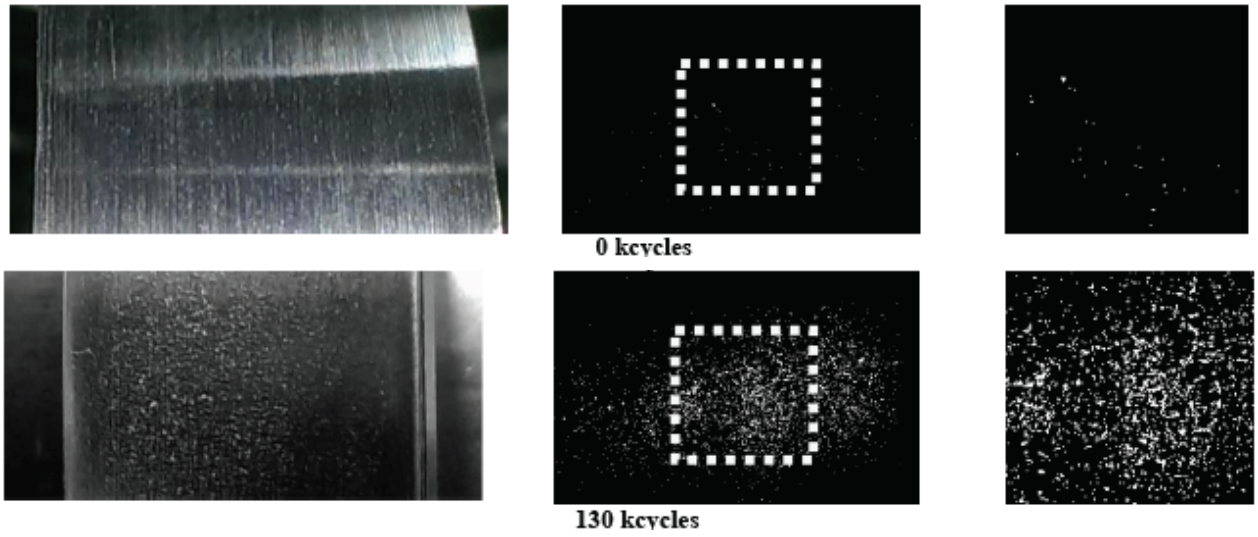

Figure 3: Example of images analysis of the specimen sample, at the beginning of the test (top) and at 130k cycles (bottom). On the left column a photo of the specimen, in the central column the particle analysis and in the right column the ROI zoomed in.

4 Results

4.1 Vibration analysis validation

As can be seen in figure 2, the change between the pristine and damaged state of the specimen happens at a specific point in time, which could be associated with a change in the damage phenomena. The change is gradual, allowing to propose a quantitative evaluation of the phenomena progression. To demonstrate that the developed method is reliable during the clustering elaboration, a K-fold cross-validation was used. The datasets of two tests were divided in five parts and each part was used as ground-truth to validate the other parts. The fold used as groundtruth is the part used for the training and the other parts are used for the validation. The results of this cross-validation are positive, with values of sensitivity, specificity and accuracy in a range of 98% - 100%.

To demonstrate that the developed method is robust and not subjected to overfitting, an exhaustive cross-validation was used, taking into account all the tests under examination. In this cross-validation each test was used validated against each other, using as ground-truth the dataset used to train the system, obtaining the results shown in Table 2. The reliability of the algorithm in classifying states based on extracted features is good, in terms of sensitivity (min.19% - max.100%), specificity (min.25% – max.99%) and accuracy (min.59% – max.98%).

Table 2. Exhaustive cross-validation results

Status	Pristine / undamaged						Bench halted						Damaged					
Specimen	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**	**1**	**2**	**3**	**4**	**5**	**6**
Sensitivity [%]	19	55	25	53	61	79	36	72	97	69	85	54	100	99	90	98	85	95
Specificity [%]	99	98	98	96	87	93	99	99	94	99	98	98	25	62	49	62	65	83
Accuracy [%]	64	83	85	80	76	86	92	97	94	94	98	97	59	83	79	80	76	88

4.2 Comparison with image data

Out of the six trials, three where chosen and image data was also analysed, to get confirmation that the vibration-based machine-learning assessment was able to detect a phenomena whose effects could also be visible using an high speed camera focusing on the specimen.

The Ra coefficient was computed for every image available, and plotted against the number of cycles. The membership probability was assigned a 63% level (using 1 sigma approximation) to be used as a threshold, and the data representing the machine stops was remove.

Figure 3, 4 and 5 respectively present the test number 5, 2 and 4. Except for test 2, where the variation in slope of the image-based index is less clear, the other charts clearly show a good superimposition between the change in state in the vibration state, and the change in the Ra image-based index

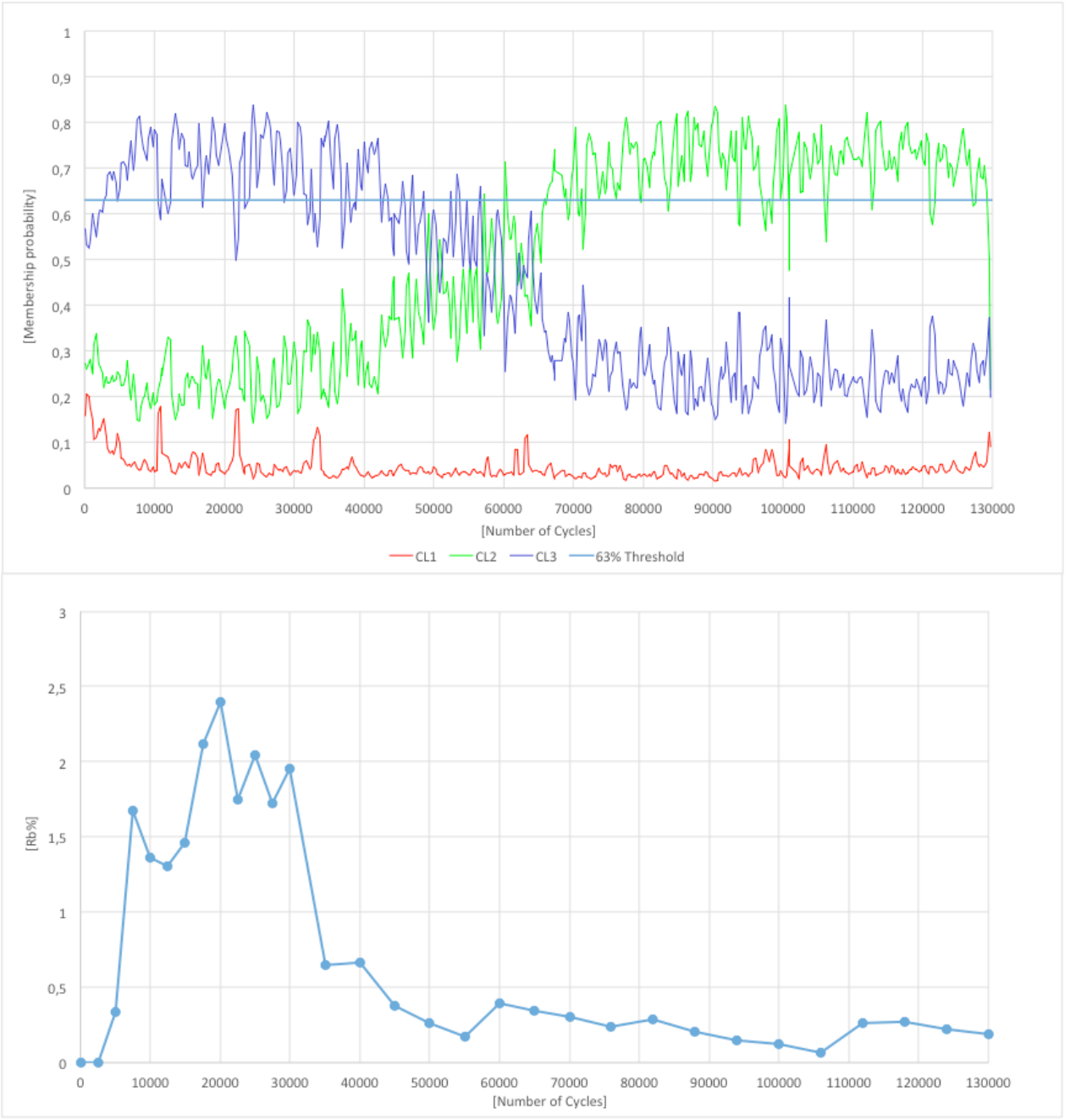

Figure 3: K-means clustering results for test number 5 (top) with blue indicating membership of the "pristine state" cluster, green the "damaged" cluster and red (removed here) the bench stops due to maintenance. Ra index for the same test (bottom).

Figure 4: K-means clustering results for test number 2 (top) with blue indicating membership of the "pristine state" cluster, green the "damaged" cluster and red (removed here) the bench stops due to maintenance. Ra index for the same test (bottom).

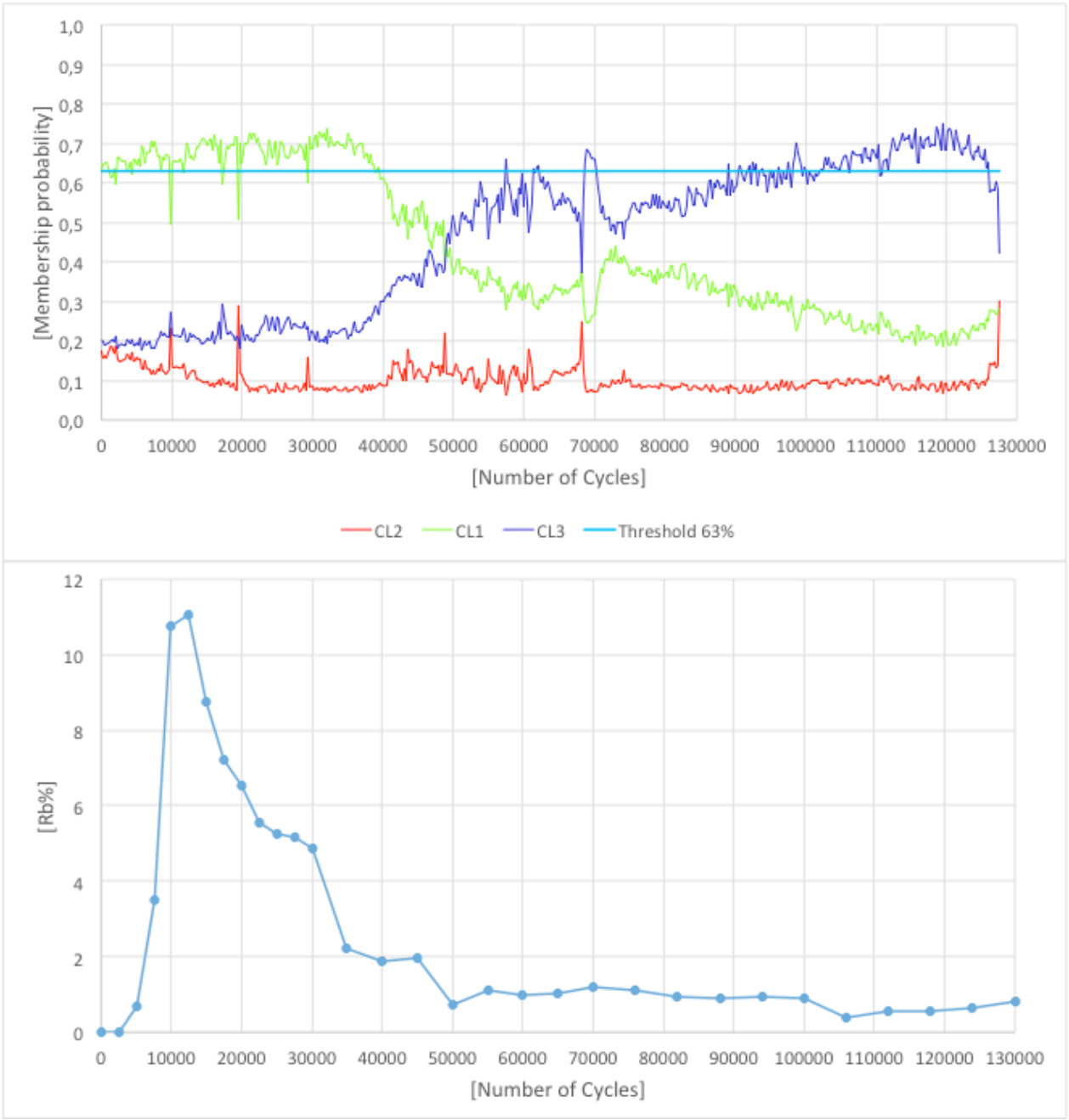

Figure 5: K-means clustering results for test number 4 (top) with blue indicating membership of the "pristine state" cluster, green the "damaged" cluster and red (removed here) the bench stops due to maintenance. Ra index for the same test (bottom).

5 Discussion

The vibration-based clustering method presented is consistent in terms of results, and portable between different material conditions, as the correlation validation pointed out. More tests are needed to understand if the image-based index will corroborate the results obtained. A proper validation could be achieved only by performing destructive tests at given intervals, but such solution will take time. For this reason a comparison between different continuous and non-destructive solutions could be sought.

This method, could lead to the ability to carry out a quick analysis of how the single test is performing in real time, which in turn will provide the ability to automatically detect the test bench working conditions, and to detect abnormalities, asymmetries or data structures that a traditional test could miss.

6 Conclusion

The method proposed is simple to implement and could greatly decrease the costs of fatigue characterization in wheel and rail steel. The membership probability obtained, if validated by further study, could prove to be a metric of interest to understand what damage phenomena is occurring and when. The ability to detect the change between different states will also open up the possibility of designing tests to be stopped at a particular state change, driven by the machine-learning information, to allow for destructive tests at a specific progression stage of the damage phenomena.
This work lays the ground for a more detailed study on the application of machine learning techniques in material characterization tests. The correlation of the presented results with traditional damage progression assessment (wear, cracks etc..) is not yet fully proven, although validation tests are currently under evaluation and preliminary results are positive.

Acknowledgements

Authors would like to thank Mr. Coffetti and Mr. Bonometti for their help in the experimental setup. Thanks also should be extended to Ing. Bini Chiesa for his support during the development of the algorithm.

References

[1] Eurostat, 2018, Annual number of accidents by type of accident

[2] A. Ekberg, E. Kabo, Fatigue of railway wheels and rails under rolling contact, Wear 258 (2005) 1288–1300.

[3] Bharat Bhushan (Ed.), Modern Tribology Handbook Vol. II, CRC Press, Boca Raton, Florida, USA, 2001, pp. 1271–1326.

[4] Kankar, P. K., Sharma, S. C., & Harsha, S. P. (2011). Rolling element bearing fault diagnosis using wavelet transform. *Neurocomputing*, *74*(10), 1638–1645.

[5] Elforjani, M. (2018). Diagnosis and prognosis of slow speed bearing behavior under grease starvation condition. Structural Health Monitoring, 17(3), 532–548.

[6] Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., … Van Hoecke, S. (2016). Convolutional Neural Network Based Fault Detection for Rotating Machinery. Journal of Sound and Vibration, 377, 331–345.

[7] Hayashi, S., Asakura, T., & Zhang, S. Z. S. (2002). Study of machine fault diagnosis system using neural networks. *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, *1*, 956–961.

[8] Kankar, P. K., Sharma, S. C., & Harsha, S. P. (2011). Fault diagnosis of ball bearings using machine learning methods. *Expert Systems with Applications*, *38*(3), 1876–1886.

[9] Prieto, M. D., Cirrincione, G., Espinosa, A. G., Ortega, J. A., & Henao, H. (2013). Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks. *IEEE Transactions on Industrial Electronics*, *60*(8), 3398–3407.

[10] Kankar, P. K., Sharma, S. C., & Harsha, S. P. (2011). Fault diagnosis of ball bearings using continuous wavelet transform. *Applied Soft Computing Journal*, *11*(2), 2300–2312.

[11] Wong, P. K., Yang, Z., Vong, C. M., & Zhong, J. (2014). Real-time fault diagnosis for gas turbine generator systems using extreme learning machine. *Neurocomputing*, *128*, 249–257.

[12] Yu, X., Dong, F., Ding, E., Wu, S., & Fan, C. (2017). Rolling Bearing Fault Diagnosis Using Modified LFDA and EMD with Sensitive Feature Selection. *IEEE Access*, *PP*(99), 1.

[13] Xie, J. P., Yang, Y., Li, T. R., & Jin, W. D. (2014). Learning Features from High Speed Train Vibration Signals with Deep Belief Networks. *Proceedings of the 2014 International Joint Conference on Neural Networks (Ijcnn)*, (2012), 2205–2210.

[14] Elforjani, M., & Shanbr, S. (2017). Prognosis of Bearing Acoustic Emission Signals Using Supervised Machine Learning. *IEEE Transactions on Industrial Electronics*, *PP*(99), 1.

[15] Shaheryar, A., Yin, X.-C., & Ramay, W. Y. (2017). Robust Feature Extraction on Vibration Data under Deep-Learning Framework: An Application for Fault Identification in Rotary Machines. *International Journal of Computer Applications*, *167*(4), 975–8887.

[16] G. Donzella, M. Faccoli, A. Mazzù, C. Petrogalli, R. Roberti (2011) Progressive damage assessment in the near-surface layer of railway wheel–rail couple under cyclic contact Wear, 271 (1–2), pp. 408-416

[17] Danner, C., Gov, K., & Xu, S. (n.d.). Condition Monitoring Using Accelerometer Readings. *Cs229.Stanford.Edu*.

[18] Amruthnath, N., & Gupta, T. (2018). Fault Class Prediction in Unsupervised Learning using Model-Based Clustering Approach. *2018 International Conference on Information and Computer Technologies (ICICT)*, 5–12.

[19] Jia, F., Lei, Y. G., Xing, S. B., & Lin, J. (2016). A Method of Automatic Feature Extraction from Massive Vibration Signals of Machines. *2016 Ieee International Instrumentation and Measurement Technology Conference Proceedings*, 714–719.

[20] Mai, C., & Engineering, I. (2016). A Comparative Investigation of the Robustness of Unsupervised Clustering Techniques for Rotating Machine Fault Diagnosis with Poorly-Separated Data, 165–172.

[21] Bouguerne, A. (2015). K- Means Clustering Based on Vibration Signal for Classification Induction Machine Faults, (2), 1–5.

[22] Hidayat, A. Y., Widodo, A., & Dwi Haryadi, G. (2018). Fault Diagnostic System Bearing Centrifugal Pump Using K-Means Method for Thermography Image and Signal Analysis Vibrations. *MATEC Web of Conferences*, *159*.

[23] A. Mazzù et al. (2015) An experimental procedure for surface damage assessment in railway wheel and rail steels. Wear 342-343 p 22–32

[24] I. Bodini et al. (2018) Evaluation of wear in rolling contact tests by means of 2D image analysis Wear 400-401 p 156-168