IOP Conference Series: Materials Science and Engineering

PAPER • **OPEN ACCESS**

Recognition of track defects through measured acceleration - part 2

To cite this article: S Bahamon-Blanco *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **615** 012122

IOP Conf. Series: Materials Science and Engineering **615** (2019) 012122 doi:10.1088/1757-899X/615/1/012122

Recognition of track defects through measured acceleration - part 2

S Bahamon-Blanco, S Rapp*, C Rupp, J Liu and U Martin

Institute of Railway and Transportation Engineering, University of Stuttgart, Pfaffenwaldring 7, 70569 Stuttgart, Germany.

*Corresponding author's email address: sebastian.rapp@ievvwi.uni-stuttgart.de

Abstract. For an optimized maintenance strategy, the early detection of track defects is necessary. Mounted sensors (e.g. acceleration sensors) on in-service trains are very suitable for track monitoring. With the continuous measurement of axle-box acceleration, short wavelength defects can be identified. For example, these defects can be rail breaks or cracks (i.e. rail defects), or local instabilities. Local instabilities can reduce the track quality in a short period of time. For an efficient data analysis of the acceleration signal and classification of different track defects, the development of appropriate methods is necessary. Therefore, a track-vehicle scale model was built to generate acceleration data for typical types of failures. With the generated dataset, developed algorithms for pattern recognition can be easily tested. In the second part of this research, three models created by the supervised learning method are trained and tested for the detection of the local instability in the vertical acceleration signal. The model A is trained with 78 laps and uses a manual classification. The chosen classifier for the model is a bagged tree algorithm implemented in the software MATLAB. The developed models distinguished between no failure, rail defect and local instability. For the training process of the model, the measured acceleration is treated statistically (e.g. Root-mean-square, Standard deviation, Spectral peaks and power). Subsequently test data for different scenarios is generated and used in the prediction model. With this model the track defects in the track-vehicle scale model are detected and classified very reliably. In contrast to existing methods, a machine learning approach is used for the non-destructive detection of the local instability. The results of the model are also improved by the model B and C by using 139 laps as training data, an automatic classification and an optimization of the statistics. The knowledge gained, can be used for acceleration data from in-service trains in regular operation, by adapting the developed model.

1. Introduction

In this day and age, the demand for mobility using railway transport systems is very high. Most trains operate continuously for almost the whole day long. Hence, track maintenance activities are limited to only a few hours per day when the trains are not in operation [1]. Therefore, a good track condition over a long period of time is important to decrease intensive maintenance work. Traditional equipment for inspection such as track recording-vehicles are used by default to determine any track defects. However, these track recording-vehicles are expensive to maintain and may interrupt regular operation. Both the need to occupy the track with regular operations and the small number of track recording cars limit the track geometry measurements. To carry out an efficient maintenance strategy, "continuous track monitoring" is needed. With the use of continuous track monitoring, railway maintenance actions can then be planned in the early stages of track degradation. In order to do this, solutions for the continuous determination of the track condition have to be developed; one solution could be an axle-box accelerometer system mounted on in-service trains. Particularly short track defects can be recognized very well by using continuously-measured axle-box accelerations [2]. The acceleration signals which are the readout of all the acceleration data collected along the model track are represented as a graph of time versus the value of the vertical acceleration and can be analysed based on their time and frequency to determine the characteristics of the different failure types.

In the course of the research project titled "Frühzeitige Erkennung von punktuellen Instabilitäten bei zyklisch dynamischer Einwirkung an bestehenden Bahnkörpern in konventioneller Schotterbauweise bei bindigen Böden im Unterbau/Untergrund" (EPIB), which is funded by the German Research Foundation (DFG), the early detection of local instabilities is investigated. In this research, a track-vehicle scale model is built to generate acceleration data that is used to detect different failure types by running for several laps. In [3], the local instability in the track-vehicle scale model is successfully recognized with the continuously-measured accelerations using five methods: amplitude range investigation, frequency analysis, cross correlation analysis, wavelength analysis and the failure detection method that is based on the classification of peaks using their ranges in amplitude and wavelength. To extend these five methods to track failure detection, machine learning algorithms were introduced.

This current paper is the second of two papers presenting the approach for track defect recognition using a bagged decision tree algorithm. In figure 1, the complete approach for failure detection using the (supervised) machine learning method is shown. This previous paper describes the first two steps: "Data Collection" and "Data Preparation". The current paper describes the last two steps: "Training Process" and "Prediction Process".

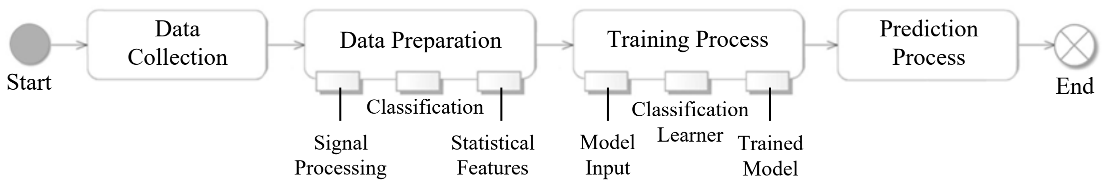


Figure 1. General approach for failure detection using the machine learning method.

2. Training process

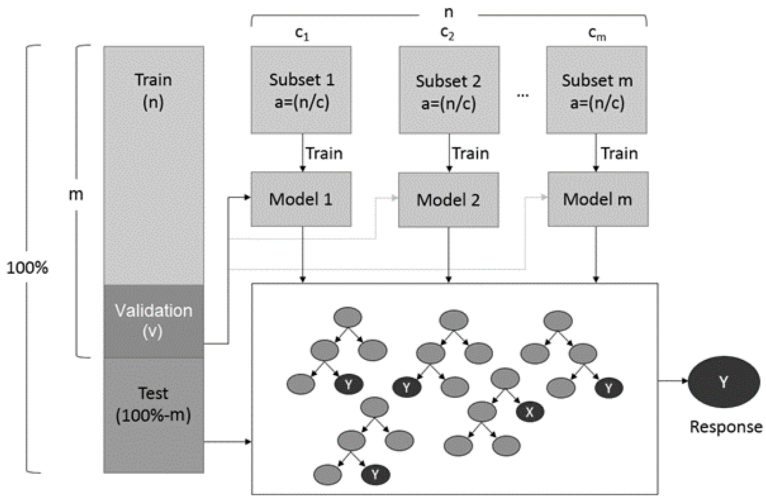
The first paper on the recognition of track defects through measured acceleration presents the data collection and preparation processes. This current paper is based on the first paper and describes the training and prediction processes for the recognition of track defects in the track-vehicle scale model.

2.1. Model input

The training process starts by loading the numerical matrix (Part 1, section 3.3) into the software MATLAB. The data is preprocessed by fixing the information in a format that the machine learning algorithm can work with. The developed machine learning model first learns the programs from the data and thereafter makes iterative regressions till it finds the suitable model [4].

2.2. Classification learner

To create the machine learning model, a bagged decision tree algorithm [5] is used. The main goal of the bagged decision tree is to average noisy and unbiased models in order to create another model with a lower variance in terms of the classifications. The decision trees combine the results of many decision trees, which reduces the effects of overfitting and improves generalization. The result is that the out-of-bag samples have fewer observations from classes with large misclassification costs and more observations from classes with small misclassification costs. In other words, many subsets with random values are created from one part of the sample numbers. Figure 2 shows how the bagged decision tree process is implemented for the recognition of the local instability in the acceleration signal. The initial data is divided in three groups: training data, validation data and testing data. The first one is divided in several bags which are trained with the classification and the validation data, generating multiple models that are compared with the testing data. The average prediction is the final response of the Bagged decision tree model.



n	Percentage of training data	[%]
m	Training and validation data	[%]
v	Percentage of validation data	[%]
c	Number of subsets	[-]
a	Subset percentage	[%]
X	"X" response	[-]
Y	"Y" response	[-]

Figure 2. Bagged decision tree process [6].

2.3. *Trained model*

The results of the trained bagged decision tree model for Model A are shown in figure 4. The number 1 represents Class 1: non-track failure, which is correctly detected in the training process for 91 % of the time. The number 2 represents Class 2: cracks, joints and the entrance/exit to the bridge, which is detected with a 78 % positive predictive value. The number 3 represents Class 3: the local instability in the track-vehicle scale model, which is detected with a 76 % positive predictive value. After being determined, the results of the trained bagged decision tree model are exported and saved to be used for predictions with testing data. In total, three bagged decision tree models (Models A, B and C) were trained to investigate the influence of the size of the data set of the training data, the statistical features and the noise, which is classified as a non-track failure.

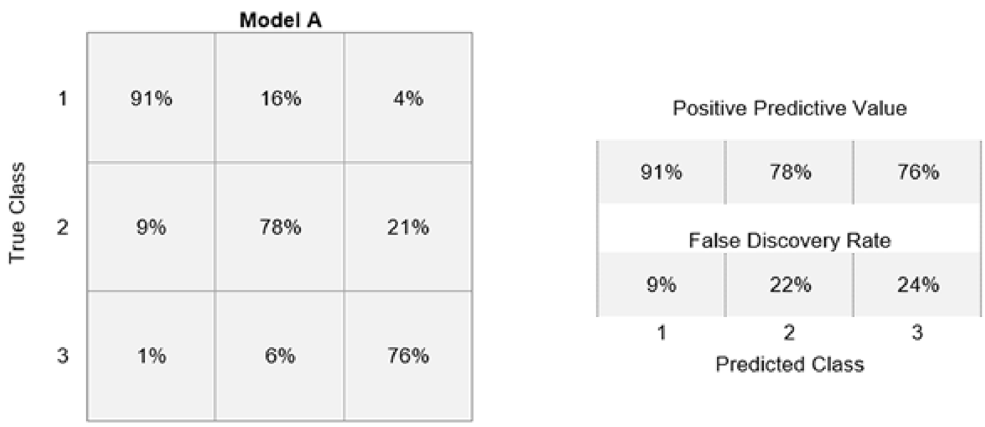
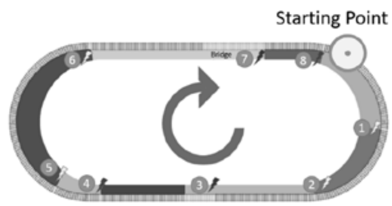
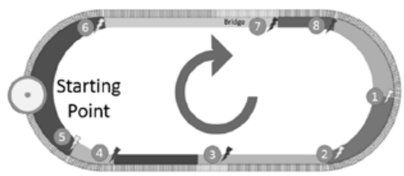
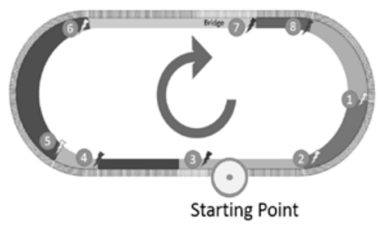
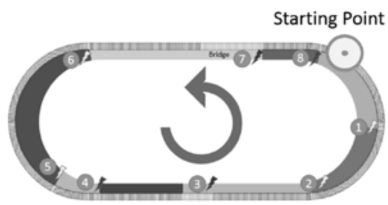


Figure 3. Confusion matrix and predicted classes [6].

3. Prediction process

According to the training process described in Chapter 2, test data must first be generated and preprocessed for the prediction process. 14 tests are defined to create test data that is different from the training data in terms of speed, direction, car sequence, test continuity and starting position. The Models A, B and C are used to classify the failures in each of the 14 test scenarios, these test scenarios are described in detail in table 1.

Table 1. Description of the generated test data for the prediction process [6].

Test description	Sketch
All of the tests numbered 1-6 use the highest speed and the test is stopped momentarily at the end of each lap. Test 7 uses the highest speed (0.37 m/s) without stopping during the 14 laps. Test 10 uses the highest speed and the test is stopped momentarily at laps 1, 3, 6, 9, 13 and 14. Test 11 uses 83 % of the highest speed without stopping during the 12 laps. Test 12 uses 83 % of the highest speed and the test is stopped momentarily at each lap.	 The test starts after the eighth failure in the clockwise direction.
Test 8 uses the highest speed and the test is stopped momentarily at each lap.	 The test starts in the middle of the curve after the fifth failure in the clockwise direction.
Test 9 uses the highest speed and the test is stopped momentarily at each lap.	 The test starts just before the local instability in the clockwise direction.
Test 13 uses the highest speed and the test is stopped momentarily at each lap. In this test, the sensor module is pulling the vehicle. Test 14 uses the highest speed and the test is stopped momentarily at each lap.	 The test starts after the eighth failure in the counterclockwise direction.

The generated test data must be processed by filtering and calculating the statistical features analog to the process described in Chapter 2, but without the classification process. As expected, the local instability and the rail defects are correctly identified by using a prediction code which contains the trained model and calculates the class of each failure. However in some cases, the response is incorrectly classified by the bagged decision tree due to the variance of the data (see figure 4).

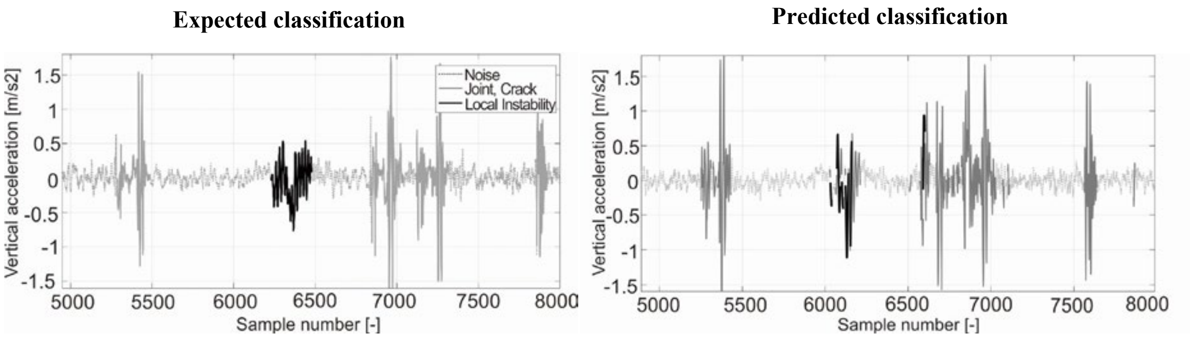


Figure 4. Expected classification vs predicted classification [6].

Initially just one segment of a local instability is expected (left picture), but the prediction has three segments (right picture), to fix this problem, and get one segment, the predictor code is complemented by a special filter. This filter operates on each side of the analyzed acceleration value and forms a buffer value C to optimize the recognition of the track defects. It is found that a window of *C* equal to 30 sample ranges achieves the best results for track failure detection. An example of the effect of the filter is shown in figure 5. The signal in the left graph contains segments of all three classes: noise, joint and local instability. The filter changes them to a unique segment of joint because Class 2 (joint, crack and the entrance/exit to the bridge) is the most frequent value.

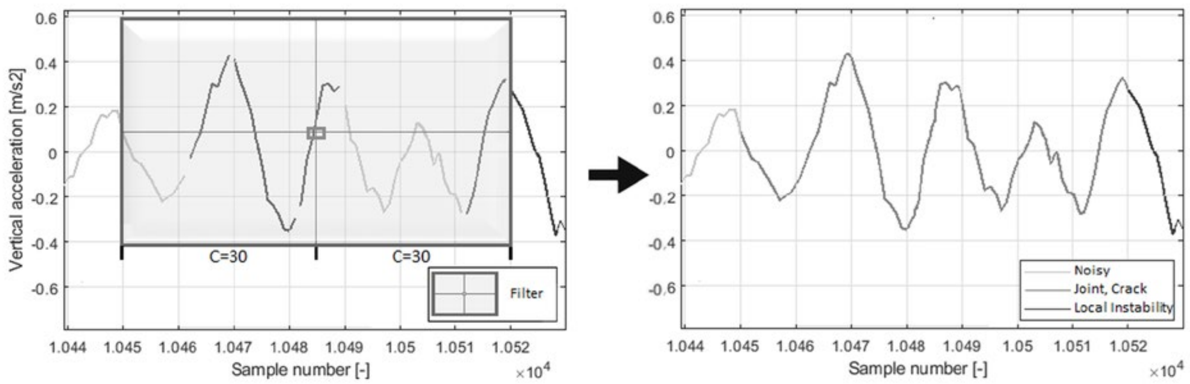


Figure 5. Results of the Filtering Process for the Prediction Process [1, 6].

A general example of the prediction without and with the filter is shown in figure 6 and figure 7, respectively, for 14 laps. As can be seen in figure 7, the filter improves the results because it cleans several of the incorrect predictions.

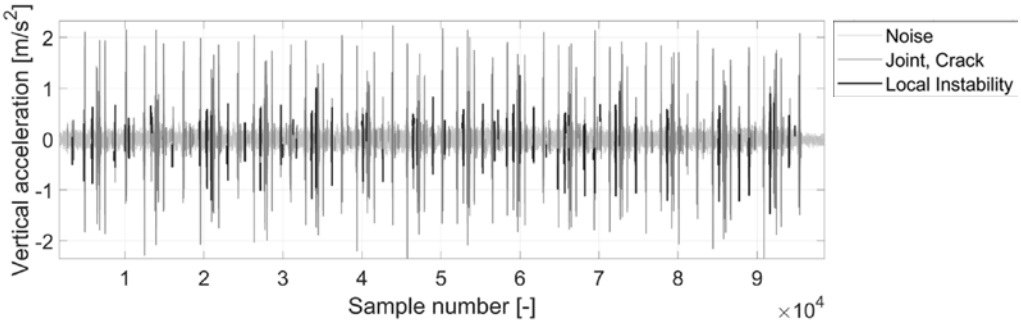


Figure 6. Predicted classification without filter for 14 laps [6].

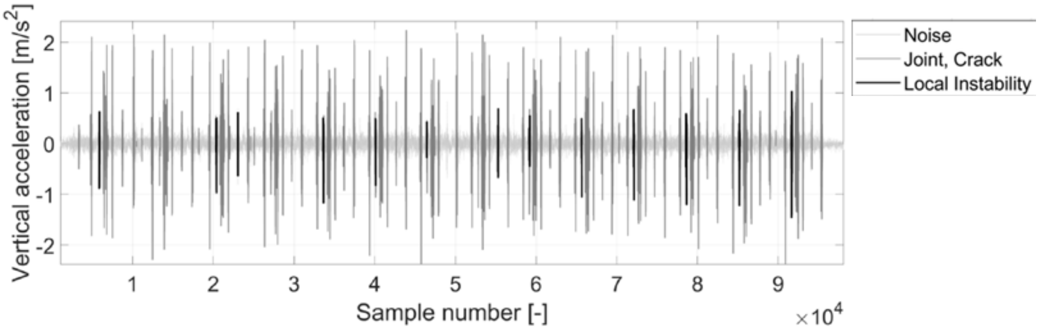


Figure 7. Predicted classification with filter for 14 laps [6].

The results of the prediction process for the three models A, B and C are shown in table 2. Model A is trained with 78 laps of training data and determined by using manual classification. Both Model A and Model B use statistical features. Model B and Model C are each trained with additional data in total 139 laps and are determined by using automatic classification. Model C uses only four features (mean value, root mean square, 12 spectral peaks and 5 spectral power) and the original acceleration values for Class 1 are replaced by random values between ±0.4 m/s².

The table 2 summarizes the results of the prediction process for the three different trained models A, B and C by comparing the positive predictive value for each test and model. In summary, Model C provides the best results with a 88 % positive predictive value.

The results of table 2 are convincing: with 14 tests, 195 laps and approximately 1.8 million sample numbers, the trained Model A can correctly predict 183 of the 195 local instabilities. However, 24 sections are incorrectly classified. Despite the variations of the tests (in terms of speed, direction, car sequence, test continuity and starting position), an average of 84 % positive predictive value is achieved in both bagged decision tree Models A and B and an average of 88 % of the times the positive predictive value is achieved for Model C.

The low performance of some of the tests is due to the following reasons. In the case of Test 7, the continual testing generates slightly higher accelerations after the second lap. In Tests 11 and 12, the speed is reduced and therefore the acceleration decreased. The low speed effected the tests, when the vehicle tried to pass over some of the failures. On the other hand, several tests had adequate results. For example the first six tests, which logically achieved high positive predictive value between 85 % and 100 % because the model is trained with six tests conducted with nearly the same conditions.

As for Test 8, the result has a 100 % positive predictive value in Models A and C. Test 13 reached a 94 % two times and 100 % one time as a positive predictive value, even though it is one of the tests carried out counterclockwise and the sensor module pulls the vehicle. Those conditions are quite different to those of the training data. Nevertheless the results show that those conditions don't affect the good results and allow to see how effective the trained model is for scenarios that deviate from the training data.

As for Test 8, the result has a 100 % positive predictive value in Models A and C. Test 13 reached a 94 % two times and 100 % one time as a positive predictive value, even though it is one of the tests carried out counterclockwise and the sensor module pulls the vehicle. However the conditions of this test are quite different to those of the training data. The other 13 tests, which have different conditions than the training data, allow to see how effective the trained model is for scenarios that deviate from the training data.

Table 2. Summary of the Prediction Results for Models A, B and C for the Local Instability [6, 7].

Test	Actual number of local instabilities	Detection of Local Instabilities						Positive Predictive Value [%]		
		Correct			Wrong					
		A	B	C	A	B	C	A	B	C
1	17	17	17	17	3	0	1	85	100	94
2	16	16	16	16	2	2	0	89	89	100
3	15	15	15	15	1	1	2	94	94	88
4	16	16	16	16	1	3	2	94	84	89
5	21	21	21	21	3	1	4	88	95	84
6	17	17	17	17	1	2	1	94	89	94
7	14	11	11	13	3	1	3	65	73	76
8	8	8	8	8	0	1	0	100	88	100
9	8	7	8	6	2	2	0	70	80	75
10	14	14	12	14	1	1	0	93	80	100
11	12	7	8	10	2	1	1	50	62	77
12	8	6	2	7	1	0	0	67	25	88
13	15	15	15	15	1	0	1	94	100	94
14	14	13	13	13	3	2	4	76	81	72
Total	195	183	179	188	24	17	19	84	84	88

$$\text{Positive Predictive Value}=\frac{\text{Correct}}{(\text{Expected+Wrong})}*100\ \ [\%]$$

4. Conclusions

The three models, Model A, Model B and Model C allow for the reliable detection of the local instability in the track-vehicle scale model.

- The trained models are tested with 14 tests (see table 1) that have different configurations: speed, direction, car sequence, test continuity and starting position. Despite these differences, the recognition of the failures is successful, as shown in the successful detection results of the local instability where in two of the models a positive predictive value of 84 % is achieved.
- Model C, created by using a random value between ±0.4 m/s² instead of the original class 1 value and optimizing the features (mean value, root mean square level, 12 spectral peaks features and 5 spectral power features), improves the results of the initial model A achieving 88 % of positive predictive value.
- The three models do not have difficulties with different configurations of the test data. For example, the change of car sequence, direction and start position. However, it can be recognized that the positive response percentage decreases when the speed is decreased and the test is continuously-run.
- It is expected that the positive predictive value of the model can be improved when the sensor is able to simultaneously-measure the vertical, horizontal and lateral accelerations and when a

gyroscope that can measure the number of spins in the three axes is added to the measuring vehicle.

- For future research, it is recommended that the model not only be trained with more classified data but with tests with different configurations in order for the model to adjust to more diverse conditions.

Acknowledgment
Thank you to Mrs. Maureen Bridget Kösters from the Institute of Transportation Research Stuttgart for the linguistic revision of this paper.

Funding
The research described in this paper was financed by the DFG (German Research Foundation).

References

[1] Real J Montalbán L Real T and Puig V 2012 Development of a system to obtain vertical track geometry measuring axle-box accelerations from in-service trains *J. of Vibroengineering* **14**

[2] Nielsen 2013 Overview of methods for measurement of track irregularities important for ground-borne vibration *Railway Induced Vibration Abatement Solutions Collaborative project*

[3] Rapp S, Martin U, Strähle M and Scheffbuch M 2019 Track-vehicle scale model for evaluating local track defects detection methods *Transportation Geotechnics* **19**

[4] Domingos P 2012 A Few Useful Things to Know About Machine Learning *Communications of the ACM* **10**

[5] Hagenauer J 2017 A comparative study of machine learning classifiers for modeling travel mode choice *J. of expert systems with applications* **78**

[6] Bahamon S 2018 *Detection of local instabilities on a scale vehicle-track model through measured accelerations of the vehicle* Master Thesis (Germany: Institute of Railway and Transportation Engineering of the University of Stuttgart)

[7] Liu J 2019 *Work Summary Machine Learning* Institute of Railway and Transportation Engineering of the University of Stuttgart