CargoCBM – Feature Generation and Classification for a Condition Monitoring System for Freight Wagons

C Gericke1 and M Hecht2
FG Schienenfahrzeuge, Sekr. SG14, 10587 Berlin, Germany

E-mail: christoph.gericke@tu-berlin.de

Abstract. Despite the fact that rail freight transport is one of the most environmentally friendly matters of transport, its growth has been far behind the growth of freight transport in general. Studies showed that a competitive disadvantage is caused by a low availability of rolling stock, especially freight wagons. Changing from a time based to a condition based maintenance strategy is believed to decrease down times by at least one third. To make condition based maintenance for freight wagons possible the TU Berlin and five industry partners started the research project CargoCBM. One task in this project is to develop algorithms for the automatic on-board diagnosis of wheel flats. The focus of the work is on the process of feature generation and feature selection as well as the application of different classifiers to automatically evaluate the data. Based on the results of measured data, features were selected and tested with different classifiers. Thought advanced classifiers such as neural networks have been analysed in accordance to their classification accuracy. It can be shown that with carefully constructed and selected features comparatively simple classifiers can lead to excellent results.

1. Introduction
In the past decade a major growth in freight traffic took place in the EU27 area. While the percentage of road freight traffic within the modal split increased from 73.7% in 2000 to 77.5% in 2009 the share of rail freight traffic decreased by 3.2% in the same period [1]. This predominant growth of road traffic heavily loads the road network and leads to significant additional greenhouse gas emissions. For a more sustainable development it is necessary to stop this trend and strengthen the competitiveness of rail freight transport. The availability of modern rolling stock is seen to be one of the main constrains to achieve this goal.

Changes in maintenance can be the key for a higher availability and cost reductions. Today freight wagons are maintained in fixed intervals of six to eight years using the so called time-based maintenance. A maintenance strategy helping to reduce the probability of breakdowns is the so called condition-based maintenance for which a periodic or continuous monitoring of critical components has to take place. When indicators show that a certain level of deterioration has been reached equipment is taken into maintenance. [2][3]

To make condition-based maintenance for freight wagons possible, the Chair of Rail Vehicles and six industry partners (Harting, Eckelmann, PCSoft, Vattenfall, Wascoasa and Lenord + Bauer) initialised the research project CargoCBM. Figure 1 shows the structure of the CargoCBM system.

1 Dipl.-Ing. Christoph Gericke, FG Schienenfahrzeuge, Sekr. SG14, 10587 Berlin
2 Prof. Dr.-Ing. Markus Hecht, FG Schienenfahrzeuge, Sekr. SG14, 10587 Berlin

Published under licence by IOP Publishing Ltd
Within the 36 month project not only the necessary hardware but also the condition monitoring algorithms and an interface to integrate this information into an existing maintenance planning software are developed.

Figure 1. CargoCBM system structure.

One of the components to be monitored are the wheels, where wheel flats are one of the main faults to be diagnosed. The process of the development and the resulting algorithms will be described in the following chapters.

2. Theory
As a basis for the algorithm development it is necessary to understand the damages occurring on train wheels and the theoretical approach behind the algorithm development in general. A short overview on both fields will be given below.

2.1. Damages on train wheels
Figure 2 shows a wheelset as it is typically found on freight wagons. Both, the track and the wheel are made out of steel. That means unlike cars where the rubber tires already act as a primary suspension almost no suspension between running surface and the axle exists. Under normal conditions, meaning a perfectly round wheel, this does not cause any major problems. However, when the wheels get out of round major accelerations which are directly transmitted into the axle are the result. A very common type of out of roundness is a wheelflat, it “is caused by unintentional global sliding (without rolling) of the wheel on the rail” [4]. Due to this, a part of the wheel tread is worn off. Wheelflats are seen to be one of the main causes for broken axles which can lead to dangerous accidents [5].

Besides this safety relevant effect, wheelflats cause significant acoustic emissions which lead to noise levels strongly affecting the residents along railroad tracks. It is therefore necessary to detect wheelflats as quickly as possible and transfer damaged wagons into maintenance. Unlike wheel wear, wheel flats are not a result of continuous deterioration but single events. This makes it necessary to have algorithms allowing a continuous monitoring.
Figure 2. A railway wheelset.

Wheelflats cause periodic impulses in time domain. These impulses occur with the rotational frequency of the wheel. Typically a Fourier transformation is performed helping to indentify wheelflats in frequency domain where they become visible as periodic spikes. Even though the basic principle can be seen as state of the art since the 1970s, currently no published automatic solution for on-board wheelflat detection is known. [6][7]

2.2. Algorithm development

Figure 3 shows a simplified theoretical approach for the development of a condition monitoring algorithm. In the beginning an analytical investigation of the physical effects caused by the damage type to be diagnosed should be carried out. This will help to get a better understanding of the problem and is necessary for the design of measuring equipment. In case of vibration measurements, the frequency range and the predicted amplitudes are important parameters for the right choice of sensors and a data acquisition system.

Knowing these parameters, measurements covering different wear statuses have to be carried out. This often proved to be one of the difficult tasks within development projects. This is due to the fact that some failures happen very seldom or, in case of railway equipment, expensive test runs have to be done as measurements with damaged equipment are too dangerous to be done under normal operation. Out of the measured data representative samples have to be extracted. The data set should be split up into training and test data for later verification of the algorithm. This can either be done by hand or automatically by using Matlab [8]. The choice and availability of representative samples is a crucial step as every classifier can only be as good as the data used to train it.
After obtaining the data it can be used for feature generations. For condition monitoring purposes a feature can be defined as a mathematical representation of digitalised data which can be used for pattern recognition. The feature generation is necessary as vibration signals are often measured with high sampling frequencies. This leads to a number of single values which is unsuitable to be directly used for classification. The feature generation shall therefore reduce the data by concentrating on the information relevant for classification. The process of feature generation often comprises the use of data analysis functions. [9][10]

The result of the feature generation might be a huge number of features requiring complex classifiers which cause high computational cost. It is therefore necessary to introduce the step of the feature reduction. To do this, two main strategies exist, one is features subset selection, the other one is dimensionality reduction. The feature subset reduction tries to remove irrelevant or redundant features. This can be done by two different methods. The so called filter method tries to select features independently of the later used classifier, for example by their correlation. The wrapper method includes the classifier in the process and selects features in dependence of the classification result. For the second main strategy, the dimensionality reduction, the number of features is reduced by creating new features which are linear combinations of existing ones. Compared to the feature subset selection the dimensionality reduction is advantageous as less information is lost. However, unlike the original features, the combined features are usually not physically interpretable anymore. [9][11]

The last step in the data processing chain is the classification which assigns the features to damage classes. Generally the requirements on the classifier are lower the better the features match the damages classes [12]. For the classification exists a broad variety of classifiers. Some like Bayes classifier or a geometric classifier are rather simple others such as neural networks are a lot more complex to use and to train. In the past few years neural networks became very popular for applications like fault diagnosis on ball bearings [13][14]. For this application they proved to be a very powerful tool. However, especially for railway applications it has hardly been investigated if simple classifiers could lead to similar results.

3. Measurements
The measurements investigated were executed on a freight wagon bogie LEILA [15] and were originally recorded to analyse the running behaviour of this newly developed bogie type. Even though it is not the most commonly used bogie, the results are expected to be similar for the standard freight wagon bogie Y25. This data is especially valuable as the bogie originally was under very good condition and the wheelflat (figure 4) occurred during the measurement. Figure 5 shows an axle-box vibration and a speed signal. The wheelflat developed when braking at around 3000 seconds.

Figure 4. Picture of the wheelflat.
The data is recorded with a sample rate of 500 Hz and a resolution of 16 Bit. For the data processing the measurement is split up into segments with 1024 samples. From the whole dataset the data from 3000 to 4000 seconds is not considered. The rest of the dataset is used for features generation and classification purposes. The condition from 0s-3000s is called “good” the condition from 4000s-6000s is called “bad”. In both cases, only segments with an average speed of at least 20 km/h are considered.

4. Feature generation
Features for a wheelflat detection are generated either based on the original signal in time domain, its frequency spectrum or its envelope spectrum. For the analysis in time domain, features are statistical values such as standard deviation ($K_{t1}$), absolute maximum ($K_{t2}$), crest factor ($K_{t3}$) or peak to peak value ($K_{t4}$). Using the spectra more complex features are generated. Figure 6 shows the Fourier spectrum and the envelope spectrum (using the Hilbert transform) of a wheel with a wheelflat.

![Figure 5. Time-velocity/time-vertical acceleration diagram of the axle box.](image)

![Figure 6. Fourier spectrum (left), envelope spectrum (right).](image)

Generally wheel flats cause peaks in the spectra which can be found at the excitation frequency ($f_{w1}$) and its harmonics ($f_{wn}$, $n = 2,3,...$). In the Fourier spectrum only higher order harmonics can be
seen whereas in the envelope spectrum already lower orders are clearly visible. Features based on the Fourier spectrum are named \((K_s)\), features based on the envelope spectrum are named \((K_e)\). Usually the principles behind the features can be applied to both spectra.

Feature \((K_{s1}/K_{e1})\) and \((K_{s2}/K_{e2})\) use the spacing between the peaks in the spectrum \((\Delta f_{n-n+1}, n = 1,2,...)\). When a wheelflat occurs, the distance between the \((n = 1,2,...)\) maxima should be equally distributed \((K_{s1}/K_{e1})\) therefore is the variance of these \(n\) distances. This feature only gives general information about harmonics in the spectrum and could lead to similar results for other faults such as a damaged bearing. For a damaged wheel, \(f_{w1}\) should be equal to \(\Delta f_{n-n+1}\). Feature \((K_{s2}/K_{e2})\) therefore compares the distances between the peaks with an excitation frequency calculated out of the wheel diameter and the wheel speed, this way allowing more detailed conclusions about the failure type.

A different approach is used for feature \((K_{s3}/K_{e3})\), where the position of the peaks in the spectrum is compared with the harmonics of the excitation frequency. For a damaged wheel, the frequency of the \(n\) maxima should be integer multiples of the excitation frequency. The cumulated differences form feature \((K_{s3}/K_{e3})\).

Similar to \((K_{s1}/K_{e1})\), the features \((K_{s4}/K_{e4})\) and \((K_{s5}/K_{e5})\) analyse the structure of the signal without any direct reference to the fault type to be diagnosed. For feature \((K_{s4}/K_{e4})\) the spectrum is normalised, afterwards the peaks exceeding a threshold are counted. Feature \((K_{s5}/K_{e5})\) is the kurtosis of the spectra.

5. Feature selection

After the feature generation 14 features are available. For classification purposes a subset of the most relevant features shall be selected. This is done without a feedback loop from the classifier and can therefore be seen as a filter method.

The features are evaluated by two criteria, the information content of a single feature and the correlation with the other features. Generally a feature is more suitable for classification purposes the bigger the distance between its values for different classes (in this case “good” and “bad”) is. To calculate a comparable value describing the distance between two features the approach described in figure 7 is used. It shows the cumulative distribution functions of feature \(K_2\) for the wheel under good and bad condition. As a first step the value of the higher curve under which has a 10% probability that the feature is lower (the 10%-percentile) is calculated. Then the percentile corresponding to this value of the lower curve \((C_{10,H})\) is determined. As a second step the 90%-percentile of the lower curve is calculated and the corresponding percentile of the higher curve \((C_{10,L})\) is determined. The sum \((C_{10})\) of \(C_{10,H}\) and \(C_{10,L}\) symbolises the distance between the two features and can therefore be used for feature selection. Analogue a \(C_{25}\) and a \(C_0\) value, using the 25% percentiles or the min/max are generated.

Figure 7. Creating the \(C_{10}\) value for \(K_2\).
Table 1 shows the $C_0$, $C_{10}$ and $C_{25}$ values, sorted by $C_{10}$, calculated for the features described above. It can be seen that the features generated in time domain lead to worse results than the features using the Fourier or envelope spectrum. In average features based on the envelope spectrum are superior to features using the Fourier spectrum. However, the feature with the biggest distance between the two classes ($K_{s2}$) is based on the Fourier spectrum. Generally, features which compare the peak values with estimated values for the excitation frequency such as ($K_{s2}/K_{e2}$) or ($K_{e3}/K_{e3}$) lead to better results than features analysing the structure of the signal. This becomes especially clear when looking at the $C_0$ value which can be seen as the amount of features in a hundred percent probability range.

However, for classification purposes not only the distance between two classes is important but also the additional information a feature can contribute. Due to this reason the correlation between different features is analysed. Figure 8 shows the correlation matrix for all features used. It can be seen that generally features in one domain are highly correlated with each other. The features with the lowest average correlation are $K_{t1}$, $K_{t3}$, $K_{e1}$ and $K_{s1}$. The two features giving the best results according to their C values ($K_{s2}$ & $K_{e3}$) have a correlation of 0.71 and should therefore not be combined.

Looking at the four features with the best C values ($K_{s2}$, $K_{e3}$, $K_{e4}$, $K_{e5}$), the two features with the lowest correlation of 0.57 are $K_{e3}$ & $K_{e4}$. A combination of these two is therefore likely to produce good classification results. Another feature which might improve the classification result when combined with $K_{e3}$ & $K_{e4}$ is $K_{e1}$ as it shows a very low correlation with the other two features, still producing good results according to its C value. However, the benefit is not clear, therefore primarily $K_{e3}$ & $K_{e4}$ shall be used for a first classification test.

**Table 1. Comparison of the $C_0$, $C_{10}$ and $C_{25}$ values, sorted by $C_{10}$.**

| Classifier | $K_{s2}$ | $K_{e3}$ | $K_{e5}$ | $K_{e4}$ | $K_{e3}$ | $K_{e4}$ | $K_{s3}$ | $K_{s1}$ | $K_{t3}$ | $K_{s2}$ | $K_{e2}$ | $K_{t1}$ |
|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| $C_0$      | 121%     | 115%     | 49%      | 0%       | 90%      | 92%      | 0%       | 48%      | 71%      | 73%      | 40%      | 4%       | 5%       | 3%       |
| $C_{10}$   | 196%     | 195%     | 195%     | 191%     | 183%     | 177%     | 177%     | 176%     | 175%     | 174%     | 132%     | 128%     | 15%      |
| $C_{25}$   | 199%     | 196%     | 198%     | 196%     | 192%     | 192%     | 193%     | 191%     | 191%     | 164%     | 164%     | 80%      |

Figure 8. Correlation Matrix.
6. Classification
With the features selected above, four different classifiers are tested using the routines supplied by Matlab: Bayes classifier, k-nearest neighbour classifier (KNN), decision trees and an artificial neural network (ANN). For training and testing a 10-fold cross-validation is being used. The overall error (good classified as bad and bad classified as good) for different classifiers and features is displayed in table 2.

**Table 2.** Comparison of the classification error of manually selected features.

| Classifier                  | Bayes | KNN | ANN | Tree |
|-----------------------------|-------|-----|-----|------|
| \(K_{e3} \& K_{e4}\)      | 2.6%  | 3.2%| 4.4%| 4.1% |
| \(K_{e3} \& K_{e4} \& K_{e1}\) | 3.6%  | 4.3%| 4.8%| 4.2% |

It can be seen that for the selected features the Bayes classifier leads to particularly good results, the ANN which takes, compared to the other features, a long time to train is inferior to the less complex classifiers. However, changes in the neural network e.g. number of hidden layers might help to improve its classification result.

To be able to evaluate the quality of the manual feature selection, the classification errors shall be compared to the average error using an automatic feature selection (wrapper method). Table 3 shows the average error using a forward or backward propagation to select the features.

**Table 3.** Comparison of the classification error of automatically selected features.

| Classifier                  | Bayes | KNN | ANN | Tree |
|-----------------------------|-------|-----|-----|------|
| Forward propagation         | 3.2%  | 2.4%| 4.7%| 3.0% |
| Backward propagation        | 3.2%  | 2.7%| 1.5%| 3.1% |

Generally the automatic feature selection does not lead to significantly better results than those reached with a manual selection. In some cases, such as for the Bayes classifier it is in average even worse than the manual selection. The biggest improvement happened for the ANN using backward propagation which in average improved by 3.3% giving the best result of all options tested. However, the time which is necessary to train the ANN is at least 10 times longer than the time needed for the training of the other classifiers.

Table 4 and table 5 show the percentage of which features were selected in multiple iterations of the forward and backward propagation process. Even though major differences can be seen using the forward and backward propagation for almost every classifier and training pattern \(K_{e3}\) and \(K_{e4}\) lead to very good results. The automatic feature selection therefore confirms the results of the manual selection. Yet, the good results concerning \(K_{e1}\) where not expected analysing the criteria used for manual feature selection. Additionally, the differences between using forward and backward propagation for \(K_{e1}\) and \(K_{e4}\) will be subject of further investigations.

**Table 4.** Selected features using forward propagation.

| Classifier | \(K_{e1}\) | \(K_{e2}\) | \(K_{e3}\) | \(K_{e4}\) | \(K_{e5}\) | \(K_{e6}\) |
|------------|------------|------------|------------|------------|------------|------------|
| Bayes      | 0%         | 0%         | 96%        | 4%         | 0%         | 4%         | 96%        | 100%       | 0%         | 100%       | 0%         | 88%        | 0%         |
| KNN        | 52%        | 0%         | 80%        | 0%         | 0%         | 0%         | 100%       | 52%        | 32%        | 0%         | 24%        | 4%         | 12%        | 72%        |
| ANN        | 50%        | 35%        | 60%        | 30%        | 10%        | 15%        | 90%        | 60%        | 45%        | 10%        | 40%        | 0%         | 50%        | 5%         |
| Tree       | 56%        | 24%        | 24%        | 8%         | 12%        | 32%        | 100%       | 36%        | 84%        | 44%        | 20%        | 0%         | 8%         | 28%        |
Table 5. Selected features using back propagation.

| Classifier | $K_{t1}$ | $K_{t2}$ | $K_{t3}$ | $K_{t4}$ | $K_{e1}$ | $K_{e2}$ | $K_{e3}$ | $K_{e4}$ | $K_{e5}$ | $K_{s1}$ | $K_{s2}$ | $K_{s3}$ | $K_{s4}$ |
|------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Bayes      | 24%      | 76%      | 100%     | 72%      | 0%       | 96%      | 100%     | 100%     | 0%       | 100%     | 52%      | 100%     | 0%       |
| KNN        | 88%      | 0%       | 48%      | 12%      | 0%       | 4%       | 100%     | 100%     | 48%      | 0%       | 4%       | 68%      | 96%      | 56%      |
| ANN        | 85%      | 90%      | 75%      | 65%      | 85%      | 95%      | 95%      | 100%     | 90%      | 75%      | 90%      | 85%      | 100%     | 70%      |
| Tree       | 96%      | 16%      | 68%      | 0%       | 8%       | 76%      | 24%      | 48%      | 72%      | 8%       | 96%      | 56%      | 76%      | 20%      |

7. Concluding remarks
A brief introduction into the problem of wheelflats on railway vehicle wheels was given. Afterwards a general approach for the development of condition monitoring algorithms was presented. Following these steps measurements were evaluated, features generated and selected according to their suitability for classification. The feature selection was done manually without the feedback from a classifier. Out of fourteen features only two features were selected. One feature compares the position of the peaks in the frequency spectrum with predicted positions. The other feature counts the amount of frequencies above a certain threshold.

The selected features were tested with four different classifiers. The best classification result with an overall error of 2.6% was reached with a Bayes classifier. This result was 1.8% better than the result with an artificial neural network. For comparative purposes an automatic feature selection was tested. This showed that improvements could be achieved with an automatic feature selection. However, for most classifiers the classification error was in the same range for both selection methods. In addition to that it, became clear that the manually selected features were predominantly used within the automatic selection as well.

For the given problem it can therefore be seen that the approach used for the manual feature selection can give a good idea about the suitability of a feature and that carefully selected features together with simple classifiers such as a Bayes classifier can lead to almost as good results as complex classifiers like artificial neural networks.

References
[1] Eurostat. Güterverkehr nach Verkehrszweig [Internet]. Available from: http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=tran_hv_frmod&lang=de

[2] Swanson L. Linking maintenance strategies to performance [Internet]. International Journal of Production Economics. 2001 Apr ;70(3): p. 237-244.Available from: http://linkinghub.elsevier.com/retrieve/pii/S0925527300000670

[3] Endrenyi J, Aboresheid S, Allan RN, Anders GJ, Asgarpoor S, Billinton R, et al. The present status of maintenance strategies and the impact of maintenance on reliability [Internet]. IEEE Transactions on Power Systems. 2001 ;16(4): p. 638-646.Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=962408

[4] Johansson A. Out-of-Round Railway Wheels - Causes and Consequences. Göteborg: Chalmers University of Technology, Dissertation; 2005.

[5] Schöbel A, Walter M. Sicherheitsrelevante Parameter im Eisenbahnwesen. ETR-Eisenbahntechnische Rundschau. 2009 ;(12): p. 743-747.

[6] Hershey JE, Kenneth BW. Patent US5433111A - Apparatus and Method for Detecting Defective Conditions in Railway Vehicle Wheels and Railtracks. 1995;
[7] Bajorat K. Offenlegungsschrift DE2105869A - Verfahren zur Laufschadenerkennung an Eisenbahnradern. 1972;

[8] Willey R. Computational Statistics: Getting Started with Classification Using MATLAB [Internet]. [cited 2011 Dec 5] Available from: http://www.mathworks.de/company/events/webinars/wbnr51468.html

[9] Rühl T. Studien zur verbesserten Ausnutzung des Informationsgehaltes von Multisensorsystemen. Gießen: Justus-Liebig-Universität, Dissertation; 2001.

[10] Niemann H. Klassifikation von Mustern [Internet]. Berlin: Springer; 1983. Available from: http://www5.informatik.uni-erlangen.de/de/mitarbeiter/niemann-heinrich/buecher/

[11] Janecek AGK, Gansterer WN, Deme MA, Ecker GF. On the Relationship Between Feature Selection and Classification Accuracy. In: JMLR Workshop and Conference Proceedings: Volume 4. 2008. p. 90-105.

[12] Hoppe D. Nutzung der statistischen Informationstheorie und eines tensororientierten Klassifizierungsverfahrens zur Signalerkennung an einem chemischen Prozeß. Dresden: Forschungszentrum Rossendorf, Arbeitsbericht; 1998.

[13] Yang D-M, Stronach AF, Macconnell P, Penman J. Third-order spectral techniques for the diagnosis of motor bearing condition using artificial neural networks [Internet]. Mechanical systems and signal processing. 2002;16(2-3): p. 391-411. Available from: http://cat.inist.fr/?aModele=afficheN&cpsidt=13764314

[14] Subrahmanyam M. Using neural networks for the diagnosis of localized defects in ball bearings [Internet]. Tribology International. 1997 Oct;30(10): p. 739-752. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0301679X9700056X

[15] Grote K-H, Antonsson EK. Springer Handbook of Mechanical Engineering. Heidelberg: Springer; 2009.