Railcar Bogie Performance Monitoring using Mutual Information and Support Vector Machines

Parham Shahidi[1], Daniel Maraini[1], Brad Hopkins[1], and Andrew Seidel[1]

[1]Amsted Rail Company, Inc. Chicago, IL, 60606, USA

pshahidi@amstedrail.com
dmaraini@amstedrail.com
bhopkins@amstedrail.com
aseidel@amstedrail.com

ABSTRACT

Railcar condition monitoring is an area of high importance and global relevance. The economic and safety concerns of equipment maintenance in North America mandate efforts in prognostics and health management. This paper presents the results from the development of a vibration based condition monitoring algorithm for freight rail, utilizing mutual information feature selection and support vector machine classification of bogie component faults. The algorithm is an implementation of a previously proposed railcar condition monitoring solution by the authors. The proposed monitoring solution is a data-driven method which was developed with measurements taken at a railroad test laboratory under controlled conditions. Vibration data was collected from multiple locations on a railcar over several test runs, each utilizing wheelsets with different levels of wear. The input of controlled wheel wear levels was aimed at varying the system outputs to resemble those of cars with different levels of mileage in revenue service. The generated data sets were processed and a feature set was extracted from the acceleration signals. The data was divided into training and validation partitions using a cross validation scheme to preserve the sequence for both sets. A mutual information (MI) estimation algorithm was used to rank the features based on their similarity to the classified fault state. Both the optimized feature set from the MI feature selection algorithm as well as the full, non-discriminate feature set were used as inputs to the support vector machine to assess classification accuracy. The results of this assessment are presented in the paper along with a presentation of the methods. The paper concludes with a proposal for a monitoring strategy aimed at specifically detecting faulty components and practicing predictive maintenance.

1. INTRODUCTION

The present work is motivated by a need in the freight rail industry to decrease asset maintenance related downtimes and to improve the effectiveness of maintenance schedules. The authors had previously investigated the viability of applying on-board condition monitoring and diagnostics methods to freight rail applications (Shahidi, Maraini, Hopkins, & Seidel, 2014) and had arrived at the conclusion that condition monitoring methods can significantly benefit the current state of railroad maintenance practices. The study of the authors was concerned only with the railcar. In particular, it was focused on the performance of the undercarriage, the bogie system, on which the railcar body traverses the rail network. Figure 1 shows a standard North American three-piece bogie.

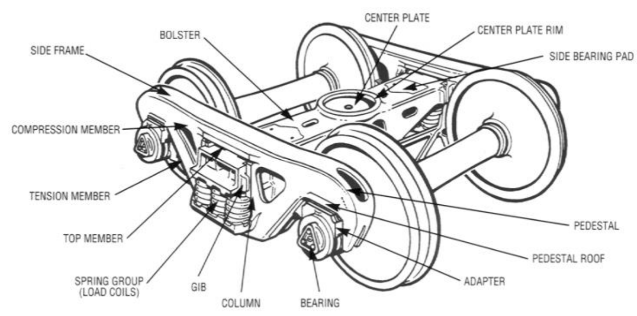


Figure 1. Standard North American three-piece bogie

The focus of the present study remains on the bogie as this is the component of a freight rail car which experiences the most wear and is most susceptible to fault modes.

The trade association tasked with rule-making for railroad transportation, the Association of American Railroads (AAR), has established a set of performance metrics (AAR, 2007) which all bogies have to meet before they can be deployed in service. After they go into service, maintenance is performed either as fixed schedule preventive maintenance or as reactive maintenance following alerts from wayside detectors. In the first case, maintenance downtimes are mostly avoided at the cost of unused capacity and premature component replacements. In the second case, wayside detectors, which are typically installed on the track, monitor passing railcars (Zakharov & Zharov, 2005). The two most common types of wayside detectors for rail car bogie performance are Truck Performance Detectors (TPD) and Truck Hunting Detectors (THD) [1]. Both of these detectors consist of strain gage based instrumentation which is added to the track to measure the lateral and vertical forces that rail car wheels exert on the track. TPDs achieve this through instrumentation of two reverse curves with strain gages to measure the wheel lateral and vertical forces and wheelset angle of attack during curving. THDs use strain gages that are placed on tangent track to measure lateral wheelset oscillations. As of 2013, approximately 15 TPDs and 172 THDs were in service across the 140,000 miles of North American rail network. Other, even less common types of wayside equipment include Acoustics Bearing Detectors (ABD) and laser/vision-based systems. Although these systems are also installed wayside they are aimed at the detection of particular component malfunction vs. the bogie system's performance as a whole. Deployment of these systems is in the low double digit numbers across the North American rail network. The small number of detectors relative to the large size of the US rail network makes it clear that wayside detectors do not provide sufficient coverage to comprehensively monitor freight train bogie performance.

2. BOGIE CONDITION MONITORING

On-board freight rail bogie condition monitoring is an area with large potential for research. As the name implies, the combination of multiple disciplines is the reason that few studies have been completed directly targeting the issue at hand.

First, on-board condition monitoring has historically not been applied to freight rail applications and is a new technology in the realm of freight rail maintenance. Typically, condition monitoring in the freight rail industry is achieved through wayside equipment and therefore research in this area has traditionally focused on efficiency improvements. Barke and Chiu (Barke, 2005) published a review of existing freight rail bogie condition monitoring technologies but excluded on-board methodologies and solely focused on wayside technologies. Lagnebäck also limited his study of potential cost savings and efficiency improvements through condition monitoring (Lagnebäck, 2007) to wayside techniques.

Second, most on-board condition monitoring studies have been attempted in the area of passenger rail transport (Ward, Goodall, Dixon, & Charles, 2010; Ward et al., 2011). Passenger rail bogies use complete and rigid frames and therefore do not have the issue of non-linearities from the friction based suspension elements of a three-piece bogie. However, passenger bogies still have to deal with other non-linearities such as those from the wheel-rail interface. The difficulty of modeling a friction wedge freight rail suspension was shown in Xia and True's study to model nonlinear dry friction damping with hysteresis and stick-slip action in the friction forces on the contact surfaces of friction wedges (Xia & True, 2003).

Third, condition monitoring of freight rail applications is not limited to bogies and bogie suspension components only. Other areas of interest where significant work has been completed include the wheel-rail interface (Hubbard, Ward, Goodall, & Dixon, 2013), rail car speed inaccuracies due to stick-slip action (Mei & Li, 2008), end-of-car devices (Hopkins, Seidel, Maraini, & Shahidi, 2015) and on-board weighing (Maraini, Shahidi, Hopkins, & Seidel, 2014) applications. It is understandable that the emergence of on-board monitoring technologies and continuous improvements in accuracy lead to a vast scope of interest which includes monitoring strategies for components which have traditionally not been able to be monitored effectively.

With the high cost of both preventive and reactive maintenance, condition-based maintenance can be considered the best solution to the problem at hand. Typically, applications follow one of two paths: either that of model-based condition monitoring or that of data driven condition monitoring.

For model-based condition monitoring, a physics-based model, derived from first principles is used to determine required system parameters. In (Li & Goodall, 2004) a two degrees-of-freedom, half-vehicle model is developed, and simulated to determine parameter deviations. For the data driven case, features are extracted from existing data from field measurements and are then processed with machine learning techniques such as neural networks (Haykin & Network, 2004) and support vector machines (Bishop, 2006; Cortes & Vapnik, 1995) to identify fault modes from measurements.

In both cases, data is required to either compare against the model or else to feed into the machine learning algorithm. Typically, this data is taken from inertial sensors such as accelerometers mounted on the system under test. If prognostics is also part of the monitoring strategy, advanced filtering techniques such as particle filters (Arulampalam, Maskell, Gordon, & Clapp, 2002) or Kalman filters (Kalman, 1960) can be combined with the algorithm to estimate future states from the current state accelerometer measurements.

[1] In the context of railroading and for this paper, the terms bogie and truck can be used interchangeably.

3. FIELD TEST

Data collection was conducted at Transportation Technologies Center, Inc. (TTCI) in Pueblo, CO. TTCI is a transportation research and testing organization which offers a wide range of tests for rail applications. The facility has seven test tracks which are designed to induce a wide variety of fault conditions, including lateral and vertical railcar instability modes.

3.1. Field Test Setup

One of the test tracks at TTCI, the Railroad Test Track (RTT), is a 13.5-mile loop with four 50-minute curves and a single 1-degree, 15-minute reverse curve. The maximum speed on the RTT is 165 mph and all curves have 6-inches of superelevation (difference in rail height on the same section of track). The primary purpose of this track is high speed stability testing which is well suited for exciting lateral vehicle dynamic modes. The selection of lateral instability testing was based on the fact that the main drivers for this instability mode are the suspension parameters and wheel wear levels. Furthermore, increased car loads have resulted in wagon bodies with higher yaw/roll moments of inertia that under faulty suspension conditions can lead to coupled oscillatory resonance modes at speeds as low as 47 mph (Tournay, Wu, & Wilson, 2009). The constant increase in axle loads is certain to affect Mean-Time-To-Failure (MTTF) requirements, and as such poses a particularly well-suited example for an application of condition monitoring strategies.

For this study, one of the 50-minute (0.8 degree) curves with 6-inches of superelevation was used to accelerate the train to a target speed onto a tangent section of track. The target speeds ranged from 40 mph to 80 mph and were broken up into approximately 5 mph increments. Figure 2 shows the profile of the segment of the RTT track that was used.

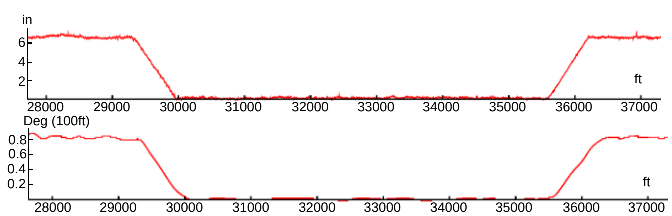


Figure 2. Test segment of RTT track

The upper graph shows the superelevation and the bottom graph shows the curvature. Once the target speed was reached, data acquisition systems began to measure accelerations at multiple locations on the car body and suspension until the test ended. Test runs were aborted once either 80 mph or prescribed maximum acceleration limits per AAR rule MSRP C-II Chapter 11 were reached. The instrumentation setup included accelerometers with various dynamic ranges from ± 5 G to ± 200 G and gyroscopic sensors with rates of 250 °/sec. The sensor specifications were chosen to accommodate signal dynamic ranges that occurred in various measurement locations. A HBM Somat eDAQ rugged data acquisition system was used to acquire the data from the sensors with a sampling rate of 1000 Hz and aliasing protection through analog filtering. The setup was a modified version of the recommended setup from the MSRP C-II Chapter 11 rules, with slightly higher accelerometer bandwidths and dynamic ranges.

To test the system with known wear conditions as the input signals into the railcar system, wheels with three different levels of wear (new, intermediate and worn) were used. For each round of testing the wheelsets were swapped out for sets with a higher degree of wear. Figure 3 shows the three different wheel profiles that were used for the three rounds of testing.

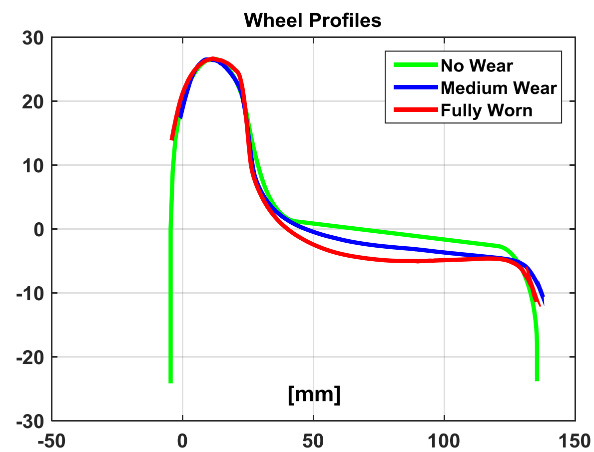


Figure 3. Different wheel wear profiles used as inputs

Every other aspect of the railcar and bogies remained unchanged to ensure that the wheel profiles were the sole factors influencing the stability of the railcar.

3.2. Field Test Results

As mentioned before, each round of testing began with a different level of wheel wear at or below 40 mph and increased gradually until the prescribed maximum acceleration limit per AAR regulations or a test speed of 80 mph was reached. With these limitations, table 1 lists the speeds the rail car was tested with for each wheelset. The green measurements indicate the speeds for the test runs which remained within the AAR limits and the red test speeds indicate where the limits were exceeded.

Table 1. Test speeds [mph] for each wheel wear level

No Wear	Medium Wear	Fully Worn
40	30	40
50	40	50
60	50	55
65	60	60
67	62	62
70	64	64
72		67
75		70

Figure 4 shows the vibration signals for the 64 mph runs for each wheel wear level. Overlaid in red is the averaged signal of each time series signal.

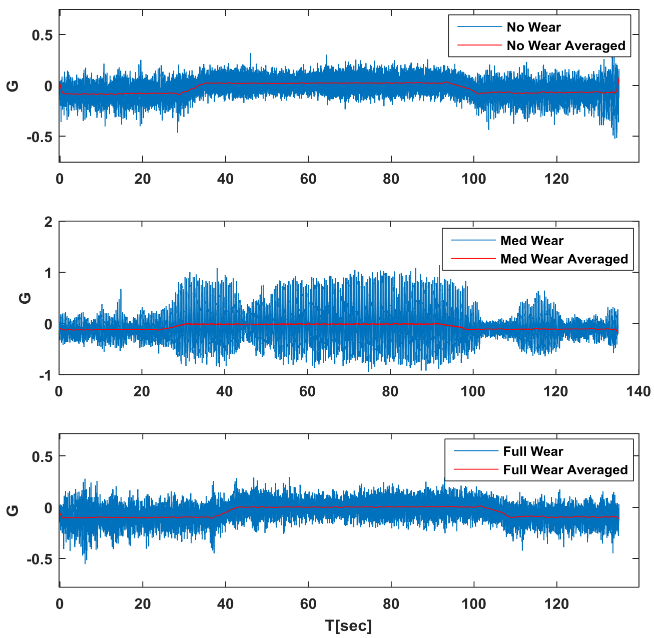


Figure 4. Vibration signals for three wear states at 64 mph.

It can be seen that the signals from figure 4 are reflective of the tabulated data. The vibration signal collected for the medium worn wheel (second subplot) has the highest vibration amplitude amongst the three signals. This corresponds with the test speeds from table 1, where the medium worn wheel set was run up to only 64 mph before the instable situation, pictured in figure 1, began.

4. ANALYSIS

The analysis of the acceleration data was broken down into multiple subtasks which will be explained in this section. The first task was the extraction of the feature set from the data for each wheel wear state and test run. Then, the data sets with the different wear states but same speeds were assembled in a random sequence as the test signal. This was followed by partitioning the assembled data sets into training and validation sets. The training set was used to reduce the dimensionality of the feature vector through a mutual information scheme which ranked the features and thereby allowed to exclude features with information gain below a user defined threshold. Then the reduced dimensionality training set was used to train a multiclass support vector machine. After training was complete, the validation set was used to evaluate the classification performance of the multiclass support vector machine in a one-versus-the-rest classification scheme.

4.1. Feature Identification and Extraction

In the first analysis step, a set of features had to be identified for extraction and identification of faulty instability modes. The initial feature set was identified as a combination of 14 features including the standard statistical moments, power content in various frequency bands, and spectral measures. The frequency bands were selected based on a qualitative spectrogram analysis in which the bands with the highest frequency content magnitude for faulty conditions were identified. In alignment with previous findings, the most important frequency band was chosen as the band between 2.5 and 3.5 Hz which is the typical range for the track-damaging rigid body rail car oscillation modes. The rigid body modes include yaw, roll, pitch and bounce oscillations which are mainly driven by wheel wear and bogie suspension wear. Since the rigid body oscillation modes exist with new components as well, only at lower frequencies and magnitudes, the first analysis frequency band was selected to be between DC and 5 Hz. Additional frequency bands included 7 – 12 Hz and 25 – 50 Hz. It shall be noted that these bands required frequent changes depending on which one of the measurement locations was chosen for analysis. Since a sampling rate of 1000Hz had been utilized for the testing, the usable bandwidth was from 0 to 500 Hz. The decision to use this bandwidth was based on knowledge of rigid and flexible modes of rail cars experiencing the mentioned oscillation modes. It was also observed that at elevated measurement locations on the carbody, higher frequency content, identifiable on suspension components, became attenuated. This is explicable through the carbody acting as a mechanical filter which attenuated much of the frequency content above 10 Hz.

Since each test run typically lasted longer than 60 seconds and included non-stationary dynamic behavior of the carbody, a windowing approach was selected to compute the feature sets. Multiple window lengths from 2 seconds for statistical features up to 10 seconds for spectral features were selected and incremented in one second intervals to compute the feature set. The complete list of all features is presented in table 2.

Table 2. List of Features

Feature #	Feature Description
1	Band Power (1) 0-5 Hz
2	Band Power (2) 7-12 Hz
3	Band Power (3) 25-50 Hz
4	Magnitude at Fund. Frequency
5	Fundamental Frequency
6	Mean
7	Variance
8	Standard Deviation
9	Peak to Peak
10	Skewness
11	Kurtosis
12	Hyperskewness
13	Hyperflatness
14	Crest Factor

4.2. Cross Validation

After the features were extracted, the wheel wear states were assigned class labels $y \in \{1,2,3\}$, one for each level of wheel wear, and data sets measured at the same speed were assembled as a test sequence with random order. Figure 5 shows the sequenced test signal order after application of the three class labels to the data.

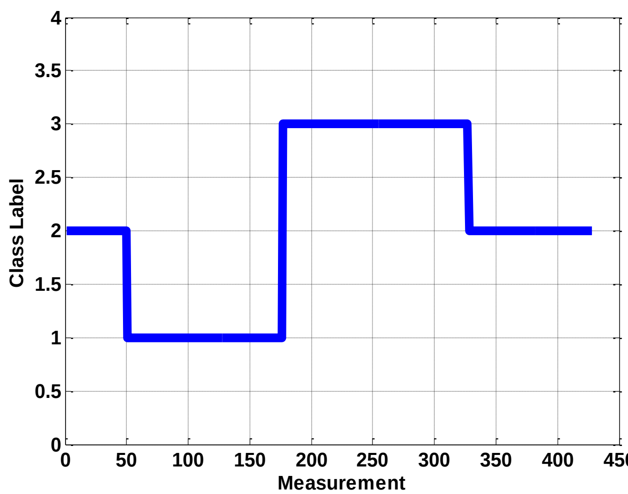


Figure 5. Three-class label classification scheme

A cross validation scheme was applied to the data to divide it into training and validation datasets. In prediction problems it is important to separate training and validation data to avoid overfitting and test generalization for independent datasets. The partitioning scheme was selected as a stratified hold out cross-validation which retained the proportions of the class labels for the training and validation partitions. Additionally, the scheme was reshuffled 10 times to provide additional validation data sets. The length of the validation partitions was selected as approximately one tenth the length of the original set.

4.3. Feature Selection Using Mutual Information

In cases with very large feature sets, a means to find and select only the most relevant features for the classification task is required to improve computational efficiency. Mutual information theory is a frequently used feature selection algorithm to reduce the number of features. The idea is to compute a simple score $S(i)$ which measures how informative each feature x_i is about the predefined class labels y. The information provided by the algorithm can be used then to discard the features with the least amount of relevancy. Mutual information uses the entropy as the amount of information gain provided by each feature. Entropy is defined as

$$H(X) = \sum_{x} p(x) \cdot \log \frac{1}{p(x)} \qquad (1)$$

where p_i is the probability of an event taking place with a certain outcome. An approximation of p_i can be obtained through the probability distribution since the algorithm is dealing with random continous samples. The joint entropy of two events taking place together is defined as

$$H(X,Y) = \sum_{x,y} p(x,y) \cdot \log \frac{1}{p(x,y)}. \qquad (2)$$

Together these quantities can be combined to calculate the mutual information for each feature and the target class as

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \qquad (3)$$

The result is a ranking of the features in the vector together with an information gain score for each feature. Table 3 shows the results for the test sequence of figure 5 and the mutual information based ranking of each feature at 65 mph.

Table 3. Mutual information ranking for 65 mph

Feature	Mutual Information
Peak to Peak	0.6768
Standard Dev.	0.6175
Kurtosis	0.5839
Freq.Magnitude	0.4999
Hyperflatness	0.4905
Variance	0.4881
Band Power 1	0.4607
Fund. Frequency	0.4553
Band Power 3	0.4512
Skewness	0.4056
Band Power 2	0.3677
Hyperskewness	0.3622
Crest Factor	0.3066
Mean	0.1688

It should be noted that since a stratified partitioning scheme was used in the algorithm, the results may slightly differ each time the mutual information algorithm is executed. The reason for this is that for stratification, samples are chosen from the population in no specific order as long as the overall sequence of class labels is maintained. Therefore single values can still vary under the same label and the variation this introduces may influence the probability distribution of the entropy calculation. A threshold can be applied after the ranking to exclude features with an information gain below a desired limit.

4.4. Multiclass Support Vector Machine Classification

A Support Vector Machine (SVM) is a maximum margin classifier that can be used for classifying both separable and non-separable data. This is achieved by finding an optimal hyperplane which defines the maximum margin between two target classes. When the target classes are separable, the equation for the hyperplane is straightforward. However, for non-separable data, kernel based methods must be utilized to transform the data into a space whereby it becomes separable. In the case of only two indicators for each class this is a simple linear line which separates two classes of data. However, when data with more than two features is to be separated the simple line becomes a plane or hyperplane above 3 dimensions. At its core, the classification problem is defined as the decision rule

$$y(\boldsymbol{u}) = \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{u} + b \tag{4}$$

where $\boldsymbol{y(u)}$ is the decision, $\boldsymbol{w}$ a weight vector orthogonal to the decision surface, b a bias and $\boldsymbol{u}$ an unknown input vector. The optimal hyperplane can be found by solving the constrained optimization problem of the form

$$\min \frac{1}{2}\|w\|^2 \tag{5}$$

Limited by the constraint

$$t_i(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x_i} + b) \geq 1 \tag{6}$$

For (6), $\boldsymbol{x_i}$ represents known positive or negative training samples and $t_i \in \{-1,1\}$ is a factor that is either positive negative depending on the sign of $\boldsymbol{x_i}$ so that (6) is always true. To deal with the constraints, we introduce Lagrangian multipliers α_i to find the extremum of equation (5). The Langragian which combines (5) with the constraints from (6) can be expressed as

$$L = \frac{1}{2}\|w\|^2 - \sum_i \alpha_i \left[t_i(\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x_i} + b) - 1\right] \tag{7}$$

Taking the derivative and setting it to zero gives the conditions for the extremum. Those can be plugged back into the original decision rule for a two-class classification problem of the form

$$y(\boldsymbol{u}) = \sum_i \alpha_i\, t_i \boldsymbol{x_i^T}\boldsymbol{u} + b \tag{8}$$

The vectors in the dot product in equation (8) can be transformed for cases when the classes are not linearly separable and in turn make them separable again. This is achieved using a kernel function of the form

$$\phi(\boldsymbol{x}^{\boldsymbol{T}})\phi(\boldsymbol{u}) = k(\boldsymbol{x}, \boldsymbol{u}) \tag{9}$$

For the present study all tests were conducted with a linear kernel, meaning the dot product was used.

The support vector machine is fundamentally a two-class classifier. To deal with the fact that in this case the problem is not only a two-class separation problem but a three-class problem $y \in \{1,2,3\}$ with one class for each wheel wear state, the above introduced support vector machine was modified to be a multiclass support vector machine. A common approach for this is called the *one-versus-the-rest* approach which constructs K separate SVMs in which the k^{th} model $Y_k(x)$ is trained using the data from class y_k as the positive examples and the data from the remaining K-1 classes as the negative examples.

4.5. Analysis Results

The analysis was completed with the above outlined algorithm and data from the field test. The focus of this testing was on identifying the three wheel wear states while testing for robustness of the algorithm against railcar speed and assessing which features contribute most to the accuracy through the mutual information score of each feature.

Figure 6 shows the progression of the features vs the speed for each wheel wear level. The colors in figure 6 were chose in accordance with the colors of the wheel profiles in figure 3: green stands for the no wear wheel profile, blue for the medium wear wheel profile and red for the full wear wheel profile. As presented in table 1, due to the experimental nature of the field data, the data sets for each fault were not always recorded at the exactly same speeds. Hence, the features are also only available at the same speeds (as in table 1). Since for comparison purposes the speed has to be the same for each fault, only three speed levels (65, 60 and 50 Mph), at which data was available for the three faults, were selected for analysis.

For the first case, data from the 65 mph test run for each wheel wear state was used to evaluate classification accuracy of the algorithm. The sequence of the wheel wear levels remained the same as presented in figure 5 in 4.2 and the hold out cross validation scheme was reshuffled 10 times for 10 simulations with the multiclass support vector machine. Features were ranked but none were excluded for the first case. The first simulation of the first case (65 mph) produced a classification accuracy of 93 %.

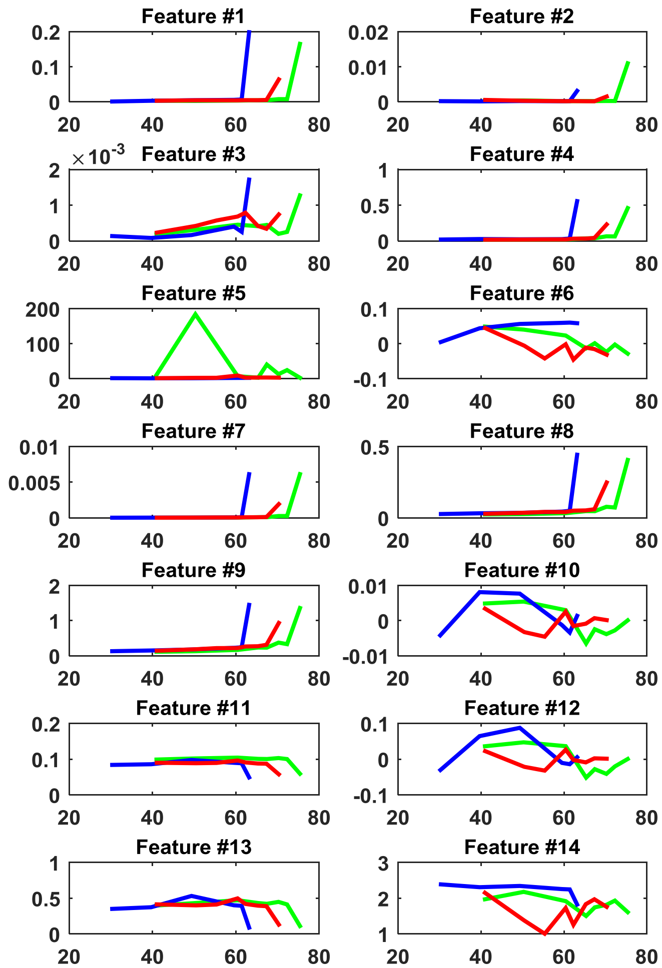


Figure 6. Progression of features versus speed – green stands for the no wear, blue for the medium wear and red for the full wear wheel profile.

Table 4 shows the results in a confusion matrix. As the table shows, only 7 out of 100 samples were incorrectly classified.

Table 4. Confusion Matrix for 65 mph run

		Predicted Class		
Actual Class		No Wear	Med. Wear	Fully worn
	No Wear	29	0	1
	Med.Wear	0	31	4
	Full Wear	1	1	33

The next 10 simulations yielded similar accuracies and the results are presented in table 5. The average classification accuracy for the first case with speeds at 65 mph was 92%.

Table 5. Classification accuracies for 65 mph

Simulation #	Classification Accuracy
1	0.91
2	0.90
3	0.93
4	0.95
5	0.96
6	0.94
7	0.91
8	0.90
9	0.94
10	0.89

For the second case, the same classification runs were repeated for a test speed of 60 mph. The first simulation produced a dramatically decreased classification accuracy of 81%. Table 6 shows results in the confusion matrix.

Table 6. Confusion Matrix for 60 mph run

		Predicted Class		
Actual Class		No Wear	Med. Wear	Fully worn
	No Wear	30	0	0
	Med.Wear	0	31	4
	Full Wear	0	15	20

It can be observed that the majority of the incorrect classifications happened for the full wear class label being incorrectly classified as medium worn. The next 10 simulations yielded accuracies as presented in table 7. The average classification accuracy for the second case at a test speed of 60 mph was 76%

Table 7. Classification accuracies for 60 mph

Simulation #	Classification Accuracy
1	0.73
2	0.73
3	0.74
4	0.77
5	0.79
6	0.77
7	0.75
8	0.77
9	0.78
10	0.76

For the last case, the same simulations were run for a test speed of 50 mph. The first simulation produced an accuracy of 79%. Table 8 shows the results in the confusion matrix. It can be observed again that class label “Full Wear” created the most inaccurate classification events out of the three class labels.

Table 8. Confusion Matrix for 50 mph run

		Predicted Class		
		No Wear	Med. Wear	Fully worn
Actual Class	No Wear	30	0	0
	Med.Wear	0	28	7
	Full Wear	0	14	21

The accuracies of the following reshuffled classifications are presented in table 9 below. The average classification accuracy at 50 mph was 79% which was just slightly above that of the 60 mph simulations.

Table 9. Classification accuracies for 50 mph

Simulation #	Classification Accuracy
1	0.76
2	0.77
3	0.76
4	0.77
5	0.78
6	0.88
7	0.83
8	0.78
9	0.79
10	0.78

Next, the feature rankings were added to the analysis and a minimum information gain threshold for the features to be included in the analysis was enforced. The data from the three test speeds was tested for mutual information thresholds that ranged from 0.5 to 0 where 0 meant that all features will be included for classification. Each speed and mutual information threshold was evaluated for 10 simulations and the approximate number of selected features of the 10 simulations as well the average accuracy with the selected feature set was recorded. The results are presented in table 10. A number of relationships can be observed in the results: first, the speed has a large influence on the number of selected features for each MI cutoff. This can be explained by virtue of the fact that the faster the train moves on the track, the more reflective of the fault condition the features become and hence higher mutual information between the features and classes exist. Second, of course a clear trend towards higher accuracy with more features can be observed for each of the test speeds. In the case of this study, this is not surprising since not a large number of features were used and no contradicting feature trends, which would require exclusion of features, existed.

Table 10. Results with applied feature selection

Speed [mph]	Simulations	MI cutoff	Selected features (approx)	Average Accuracy
65	10	0.5	5	0.85
	10	0.25	12	0.90
	10	0.1	13	0.92
	10	0	14	0.92
60	10	0.5	0	0
	10	0.25	6	0.68
	10	0.1	12	0.76
	10	0	14	0.76
50	10	0.5	1	0.34
	10	0.25	5	0.52
	10	0.1	13	0.80
	10	0	14	0.80

As an example, the features excluded by the MI feature ranking algorithm for low cut offs typically include the *mean*, *crest factor* and *hyper skewness*. However, it must be noted that this should be only cautiously considered as representative, since the stratified partitioning scheme here too causes a level of variation as explained in section 4.3.

5. DISCUSSION

Analysis for the detection of wheel wear states from the vibration signature of acceleration data taken on the rail car was completed. A success rate of 92% was achieved for the ideal case of high test speeds. Particularly, the “1” class label achieved ideal classification accuracy which can be interpreted that the identification of normal operation would be most reliable in an implementation. Performance of the algorithm at lower speeds was worse but still acceptable and of high value with a success rate of approximately 80% classification accuracy.

A few interesting points emerged from the analysis which require a deeper discussion. The first and most important observation was that speed influences the classification accuracy. In the case of this testing the test runs were completed with incrementally increasing test speeds until failure occurred. Failure was considered a lateral instability mode which was more likely to occur with worn wheels than with new wheels at a certain speed. This instability mode, which is linked to wheel wear and entered by the train above a certain speed, creates high amplitude lateral oscillations which severely change the vibration characteristics of the acceleration signal. Therefore it is not surprising that the directly linked wheel wear state was clearly discernible at higher speeds in the analysis. The simulations supported this conclusion with an average classification accuracy of 92% at 65 mph.

In the second set of simulations with reduced test speeds of approximately 60 mph, the average accuracy dropped to 76%. The plot in figure 6 shows that majority of the misclassification occurred for the samples labeled “3” in the test set. These data samples were typically incorrectly classified as having the label “2”. Conversely, the opposite misclassification of label “2” values as label “3” values did not occur, which leads to the question of why these unidirectional classification inaccuracies occur. One explanation may be that at speeds below 65 mph which do not excite the lateral instability mode, distinction between wear states of medium to fully worn states can be a challenge. Interestingly, this trend does not prevail when the speed is further reduced to 50 mph. Although speeds lower than 65 mph had significantly lower classification accuracies, the 50 mph had higher classification accuracy than 60 mph. This non-linear dependency on speed remains subject to further research. However, in preliminary experiments it has been discovered that by implementing support vector machines with kernels that are more sophisticated such as polynomial or Gaussian radial basis function kernels, improved accuracies can be achieved. Another aspect which will require future work is the addition of a probability score for the *one-versus-the-rest* multiclass support vector machine. Without a probability, score samples may be assigned to multiple states and training sets will be imbalanced. In the above example, the test set had a length of 100 samples but less than one third belong to each class therefore giving the *rest* class label an undue overweight and loss of symmetry may occur. Additional techniques for multiclass support vector machines shall be explored as alternatives.

Other improvements for future iterations of the algorithm include the addition and grouping of features, such as ratios and derivatives. Furthermore, the investigation of additional measurement locations and the addition of more components for wear estimation can provide insight into additional failure modes. Finally, an extension of the algorithm to also include non-parametric data into the feature set will further be able to enhance classification accuracy.

6. CONCLUSION

On-board condition-based maintenance for North American freight rail applications is an underdeveloped yet promising field for the application of condition monitoring and machine learning techniques. Past efforts were mainly focused on passenger rail and wayside detection technologies. In this study an algorithm to estimate wear levels of freight rail bogie components based on mutual information and multiclass support vector machines was developed and tested with field data. Promising results were achieved with classification accuracy above 90% for test speeds which excite relevant failure modes. Lower speeds still yield an accuracy of approximately 80% with the full feature set. Very high potential for improved results in the future exists based on the proposed improvements for the algorithm and the expansion of test locations. For the algorithm, feature set extension and improved kernel function will most likely yield the highest improvements.

REFERENCES

AAR. (2007). Design, Fabrication, and Construction of Freight Cars *Manual of Standards and Recommended Practices C-II* (Vol. [M-1001]).

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *Signal Processing, IEEE Transactions on, 50*(2), 174-188.

Barke, D. a. C. K. W. (2005). Structural health monitoring in the railway industry: A review. *Structural Health Monitoring, 4*(1), 81 - 94.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 4): springer New York.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297.

Haykin, S., & Network, N. (2004). A comprehensive foundation. *Neural Networks, 2*(2004).

Hopkins, B., Seidel, A., Maraini, D., & Shahidi, P. (2015). *End-of-car Device Condition Monitoring with Onboard Sensors*. Paper presented at the *ASME Joint Rail Conference*, San Jose, CA.

Hubbard, P., Ward, C., Goodall, R., & Dixon, R. (2013). Real time detection of low adhesion in the wheel/rail contact. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 0954409713503634.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering, 82*(1), 35-45.

Lagnebäck, R. (2007). *Evaluation of wayside condition monitoring technologies for condition-based maintenance of railway vehicles*: Luleå University of Technology Luleå.

Li, P., & Goodall, R. (2004). *Model-based condition monitoring for railway vehicle systems.* Paper presented at the Proceedings of the UKACC international conference on control, Bath, UK.

Maraini, D., Shahidi, P., Hopkins, B. M., & Seidel, A. (2014). *Development of a Bogie-Mounted Vehicle On-Board Weighing System.* Paper presented at the 2014 Joint Rail Conference.

Mei, T., & Li, H. (2008). Measurement of vehicle ground speed using bogie-based inertial sensors. *Proceedings of the Institution of Mechanical*

Engineers, Part F: Journal of Rail and Rapid Transit, 222(2), 107-116.

Shahidi, P., Maraini, D., Hopkins, B., & Seidel, A. (2014). *Estimation of Bogie Performance Criteria Through On-Board Condition Monitoring*. Paper presented at the Annual Conference of the Prognostics and Health Management Society 2014, Fort Worth, TX.

Tournay, H., Wu, H., & Wilson, N. (2009). A Review of the Root Causes for Loaded Car Hunting, Technology Digest TD-09-014: AAR, TTCI, Pueblo, CO.

Ward, C. P., Goodall, R. M., Dixon, R., & Charles, G. (2010). Condition monitoring of rail vehicle bogies.

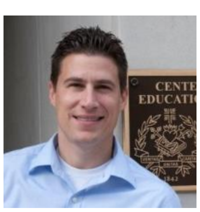
Ward, C. P., Weston, P., Stewart, E., Li, H., Goodall, R. M., Roberts, C., . . . Dixon, R. (2011). Condition monitoring opportunities using vehicle-based sensors. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 225*(2), 202-218.

Xia, F., & True, H. (2003, 22-24 April 2003). *On the dynamics of the three-piece-freight truck.* Paper presented at the Rail Conference, 2003. Proceedings of the 2003 IEEE/ASME Joint.

Zakharov, S. M., & Zharov, I. A. (2005). Criteria of bogie performance and wheel/rail wear prediction based on wayside measurements. *Wear, 258*(7–8), 1135-1141. doi: http://dx.doi.org/10.1016/j.wear.2004.03.025

BIOGRAPHIES

Parham Shahidi is a Project Engineer in the Bogie Systems Engineering group at Amsted Rail Company, Inc. He holds a BS degree in Mechanical & Process Engineering from TU Darmstadt in Germany and a PhD in Mechanical Engineering from Virginia Tech. Parham has been working in the area of condition monitoring since 2007, starting with the development of a speech based fatigue estimation system for train conductors in graduate school. In 2011 he joined Amsted Rail where his major projects in the area of health monitoring include the development of systems for vehicle instability detection, and bearing condition monitoring. He has published more than 10 technical articles related to his research. He is also an adjunct professor at Virginia Commonwealth University's Mechanical Engineering department and a board member of the German Engineers Association (VDI) in North America.



Daniel Maraini is the Manager of Bogie Systems Engineering for Amsted Rail Company, Inc. He leads a team of Project Engineers focusing on condition monitoring solutions for freight railcars. Daniel has been a part of Amsted Rail since 2008. In various positions, he has worked on wireless sensing, remote asset monitoring, and condition monitoring for freight rail applications. Daniel is currently a PhD student in the department of Mechanical Engineering at Villanova University. His research is in the field of machinery diagnostics, with a focus on nonlinear model based techniques for rolling element bearings. He holds a MS in Mechanical Engineering from Villanova University and a BS in Physics from West Chester University of PA.

Brad Hopkins is a Project Engineer in the Bogie Systems Condition Monitoring group at Amsted Rail and a Lecturer in the Mechanical Engineering Department at Southern Illinois University at Edwardsville. He holds a BS, MS, and PhD in Mechanical Engineering from Virginia Tech. He has been working on track and railcar condition monitoring since 2010, with a focus on accelerometer-based monitoring and algorithm development. His current work includes broken rail detection, wheel defect monitoring, and end-of-car system monitoring. His additional research interests are vehicle dynamics, modeling, and simulation, controls, and vibration analysis.

Andrew Seidel is a Project Engineer in the Bogie Systems Engineering group at Amsted Rail Company, Inc. He holds a BS degree in Mechanical Engineering from Pennsylvania State University in State College, Pennsylvania. Andrew has been working in the area of railroad engineering since 2012 when he began working with end of car impact detection systems and railcar condition monitoring projects. He is a member of the American Society of Mechanical Engineers (ASME) and sits on the Mechanical Engineering Technical Advisory Board for Penn State University (Harrisburg campus).