USING EXTREME VALUE THEORY TO IDENTIFY RAILCAR ASYMMETRIC WHEEL WEAR AND ITS BENEFIT ANALYSIS

Yu CUI[1,2], Qing HE[1,3,4*], Zhenhua ZHANG[5], Zhiguo LI[6]

[1,5]*Dept of Civil, Structural and Environmental Engineering; University at Buffalo, Buffalo, United States*
[2]*Turner-Fairbank Highway Research Center, Federal Highway Administration, McLean, United States*
[3]*Dept of Industry Engineering, University at Buffalo, Buffalo, United States*
[4]*MOE Key Laboratory of High-speed Railway Engineering, School of Civil Engineering,*
Southwest Jiaotong University, Chengdu, China
[6]*IBM Thomas J. Watson Research Center, Yorktown Heights, United States*

Abstract. Railcar asymmetric wheel wear leads to severe wear on one wheel but mild wear on the other wheel. The consequences of the asymmetric wheel include accelerated wear, mechanical failure and downtime, and high financial penalties. Therefore, identifying the asymmetric wheel wear is critical not only for cost effective maintenance but also for safe operations. Fortunately, the increasing amount of various wayside detectors is instrumented along the railway that can monitor the health of railcar components and log plenty of detailed information about railroad operations. One can use this information to identify the asymmetric wheel wear in the early stage. However, most elliptically contoured distributions are effective in describing normal events but not in dealing with the outliers, which mainly locate in the tails of the distribution. Asymmetric wheel wear requires effective anomaly detection that mainly focuses on the extreme values in the tail of a right-skewed distribution. In this paper, we employ the Extreme Value Theory (EVT), which handles the unusually high or low data in the distribution, to derive an extreme value score to identify asymmetric wheel wear. Experiment results show that identification of asymmetric wheel wear can generate huge monetary benefit in terms of reducing average maintenance times of railcars.

Keywords: railcar, asymmetric wheel wear, wayside detectors, railway maintenance, extreme value theory, anomaly detection.

Introduction

Rail is one of the essential transportation modes in the United States. According to the Association of American Railroads (AAR), America's railroads are operating over a network of nearly 140000 miles and account for approximately 40% of intercity freight volume that is more than any other mode of transportation (AAR 2016). Safe and efficient railway operation is always the top priority for Federal Railroad Administration (FRA) and corresponding railway accident such as derailment should be prevented. One of the major causes of rail derailments comes from train wheels. The asymmetric wheel wear arises as an important issue during the vehicle/track interaction and it can cause severe wear on one wheel but mild wear on the other wheel (Fröhling 2006). However, asymmetric wheel wear has drawn very little attention in the past literature.

Asymmetric wheel wear has been found in the coal fleet, which is a high production and high mileage rail line segment associated with the railcar (Durham 1997). The primary cause of wheel wear asymmetrically with a consequent degradation in the tracking properties of wheelsets and bogies and resultant accelerated is the action of the brake rigging and brake blocks on the wheel treads (Tournay 2011). This phenomenon on a wheelset will cause rail rolling-contact fatigue, increase lateral wheel/rail forces, accelerate wheel flange wear and damage rail turnouts. Thus, asymmetric wheel wear accelerates the deterioration process of a wheelset and greatly shortens the life of a new wheelset since railroads tend to replace a wheelset instead of a single wheel. Moreover, asymmetric wheel wear will eventually lead to train derailments and result in significant costs to railway operations (Li, He 2015). Therefore, identifying and repairing those asymmetrically worn wheels not only save operation expenses but also improve transportation safety for people and goods.

There are three rolling stock maintenance models that are widely used by railroad operators. Those are mainte-

*Corresponding author. E-mail: qinghe@buffalo.edu

nance scheduled separately by time, by mileage and by Condition-Based Maintenance (CBM). Maintenance by time or by mileage both pulls trucks or vehicles into a maintenance shop after a certain period or mileage. However, the maintenance sometimes is not necessary because it does not detect out asymmetric wheel wear problems most of the time, which waste not only the maintenance downtime but also unnecessary budget. The annual cost of replacement of wheelset in North America is approximately $ 800 mln in recent years, with 27% due to wheel wear, 65% due to high flanges and 22% due to thin flanges (Tournay 2011). One can see that lots of money have been spent on the traditional maintenance of the asymmetric wheel wear and the expenses can possibly be reduced by a more efficient maintenance strategy. Conditioned by the proper wheel wear identification by wayside detectors, we can reschedule the service cycle of a railcar with healthy wheels and thus save a lot of time and money.

In the rail industry, an increasing amount of various wayside detectors is instrumented along the railway that can provide large numbers of detailed real-time information of railroad and railcar components. Such detectors can automatically identify potential railroad car component failures, reduce rolling stock inspection times and maintenance costs, finally improve railway safety (Schlake 2010). Benefit from a large amount of detection data, the causations and forming processes of the wheel wear can be unveiled and the corresponding maintenance strategy can be optimized by devised mathematic algorithms. CBM is a prevailing approach in machinery diagnostics and prognostics (Ellis 2009; Jardine *et al.* 2006). The CBM can be performed after one or more indicators showing that a railcar component is about to fail or its running state is out of order.

In this paper, we aim to identify the asymmetric wheel wear mainly through Machine Vision (MV) detectors (e.g. video cameras), which report the profile of wheels including flange height, flange width, rim thickness, flange angle, etc. Several major challenges of this task lie in (1) MV data are real-time recorded and the large amounts of data need to be effectively processed, (2) MV data could be of low quality and need to be pruned before application due to a variety of reasons (e.g. inclement weather, image resolution, etc.), (3) asymmetric wheel wear is a rare event. It should be kept at a very low false alarm rate, (4) the proper quantification of the cost savings by the new identification of asymmetric wheel wear. To address the above challenges, we develop a statistical identification model based on Extreme Value Theory (EVT) and perform a novel benefit analysis.

The rest of the paper is structured as follow: Section 1 summarizes previous studies for asymmetric wheel wear and anomaly detection models. Section 2 introduces the datasets and reveals the underlying relationship between the bad truck and wheel wear. Section 3 presents the methodology developed for identifying asymmetric wheel wear. Section 4 discusses the numerical results and

performs benefits analysis for the identification of asymmetric wheel wear. Finally, this paper concludes and lists several directions for future research in the last section.

1. Literature review

To our best knowledge, there are very few previous studies that specifically address the issues in identifying asymmetric wheel wear. Therefore, we review two kinds of related literature, including modeling and prediction of railcar wheel wear and failure, and anomaly detection and modeling.

1.1. Modeling and prediction of wheel wear and wheel failures

Several studies developed mathematical models, integrating dynamics and wear modeling to predict railway wheel profile evolution due to wear (Braghin *et al.* 2006; Li *et al.* 2011; Lewis *et al.* 2003). Some researchers explored the wheel sets wearing by experiment. They compared cut and flange wearing intensity between locomotives with and without lubrication. The comparison results indicated that the flange wearing intensity is twice as high as the cut. Therefore, the wearing of wheel flanges is the major factor of wheel deformation, which proves the information in the introduction. Moreover, they also found out that the flange wear runs fast at the beginning, then this process will slow down. It will accelerate again after it exceeds a threshold like 150000 km (Mikaliūnas *et al.* 2002). There are also some studies modeling the relationship of contact between wheel and rail. The process of calculation showed that the contact between the wheel and rail should be considered unstable. The higher speed will result in a stronger influence of instability (Dailydka *et al.* 2008). The finite element method also can be used to analyze the "railway vehicle wheel–track" dynamic system. Bogdevičius *et al.* (2015) expressed applying discrete elements to soil and vehicle. Most recently, a variety of wayside detectors were employed to predict railcar wheelset remaining useful life by using random forest (Li, He 2015; Ouyang *et al.* 2009). However, the aforementioned algorithms predict wheel profile or failures due to wear without consideration of asymmetric wheel wear.

Almost all of the models can describe data well as long as they normally behave (Broadwater, Chellappa 2010). However, there still exist some outliers that are taken as anomalies. Asymmetric wheel wear can be considered as anomalies because the data fall into the tail of a distribution. Therefore, we can leverage general anomaly detection methods for identifying asymmetric wheel wear.

1.2. Anomaly detection and modeling

The purpose of anomaly detection is to identify rare or extreme events with minimum delay and fewest false alarm rate as far as possible (Singh *et al.* 2009). Three types of anomaly detection techniques have been developed in

the past few decades: specification-based, string-based and statistical-based (Ye *et al.* 2002). Specification-based and string-based techniques identify abnormal data when they deviate significantly from a norm profile that is built on logical reasoning. Statistical-based anomaly detection techniques detect anomaly when they deviate significantly from the norm profile that builds on statistical properties (e.g. mean and variance). One advantage of the statistical-based techniques over the other two is that statistical-based techniques have the capability of explicitly representing and handling variations and noises inherited in activities of the information system. There are two kinds of algorithms to detect and model anomaly detection with statistical techniques. One is the parametric method, and the other one is t non-parametric method. Parametric methods assume an approximate distribution and estimate the parameters from the historical data. For example, Ye *et al.* (2002) implemented Hoteling's *T*-squared test that is a generalization of student's *t*-statistic used in multivariate hypothesis testing to detect an intrusion action. Non-parametric methods do not assume that the data follows a specific distribution. Akoglu *et al.* (2012) introduced a non-parameter method: COMPREX, that builds a data compression model using multiple dictionaries for encoding, and reports the data points with high encoding cost as anomalous. This algorithm has lower running times for large datasets and higher accuracy. However, this model needs to update coded tables to capture the trending patterns of time-evolving data efficiently.

In this paper, we will use EVT that is a statistical-based method. EVT has been used in many areas to detect abnormal phenomena such as extreme temperature changes (Min *et al.* 2013), extreme storms (Kunkel *et al.* 2013) and extreme waves (Caires *et al.* 2009). However, this method has never been used in railway anomaly detection and modeling.

2. Data preparation and analysis

2.1. Data description

In this paper, wayside detector data were collected from a US Class I Rail Network in 27 months from January 2010 to March 2012. The raw data contain MV data for wheel profile measurements, Optical Geometry Detector (OGD) data for truck hunting and alignment measurements, bad order data, and maintenance (teardown) data. MV systems mainly use a video acquisition system to record digital images of railcar components. Based on these images, the system can detect the irregularity and defects of wheels, brakes, springs and so on (Camargo *et al.* 2011; Li *et al.* 2014). OGD is a laser/vision based system to measure the relationship between the wheel flange and the rail gauge and assess the performance of suspension assembly for axles and wheels of railcars (Tournay 2008; Li *et al.* 2014). Critical measurements of MV and OGD are illustrated in Table 1. We collected bad order data to obtain the failure details, and teardown data to understand actual repair

Table 1. Critical measurements of detectors

Detector	Critical measurements
MV	Wheel flange height
	Wheel flange thickness
	Wheel rim thickness
OGD	Truck hunting peak to peak
	Truck inter-axle misalignment
	Truck rotation measurement
	Truck tracking error
	Truck shift measurement

actions and validate failures identified by bad orders. As soon as a bad order is issued, the equipment is scheduled to set out from the train for further inspection in a workshop. The details of the repairs will be recorded in the teardown data. Such data can be used to acknowledge the true reason for repairs and verify if the bad order is generated as a false alarm. There is total of 6466 records composed of 1832 bad orders and 4634 good records. Therefore, the number of observations is sufficient enough for this study.

After examining the dataset carefully, we identify two significant observations. The first one is the *variations of MV measurements*, and the second one is the *correlations between bad trucks and wheel failures*.

1) *The variations of MV measurements.* In the MV dataset, the variations of the measurement are found to be quite large. The MV measurements over time for the same component show significant discrepancies due to both internal (e.g. the wheel and the railway material will have deformations between 0.1 and 0.13 mm when the loading of a wheel is 13 t) and external factors (e.g. weather and temperature) (Li, He 2015; Fröhling 2006). For instance, the wheel rim thickness measurement from MV is usually not accurate during inclement weather. The snow in winter or the mud after rain will adhere to the edge of the rim and affect the accuracy of readings for wheel rim thickness. Figure 1 shows the plot of the rim thickness of 8 wheels in the same equipment. As one can see, the readings fluctuate unpredictably over time. Therefore, identifying asymmetric wheel wear solely by MV measurements will be not reliable. To achieve accurate identification, we need to include more related data, such as truck measurements from OGD. However, a question naturally arises: is there any relationship between truck measurements and wheel measurements?

2) *Correlations between bad trucks and wheel failures.* It is well known that among railcar components, wheels, journal bearings and truck components are not isolated from each other (Li, He 2015). Asymmetric wheel wear results in fatigue of truck components on account of the fact that truck is prone to be loaded more under asymmetric wheel profile conditions and this further deteriorates rapidly.

Vice versa, bad trucks will cause asymmetric load and stress, which cause asymmetric wheel wear. Therefore, we may possibly leverage truck readings to identify asymmetric wheel wear. First, we have to validate the relationship between bad trucks and general wheel failures.

As an initial investigation, we calculate the correlations between wheel profile measurements and truck measurements, shown in Table 2.

One can see that there exist relatively high correlation values, marked by "*", between "Flange thickness" and "Truck tracking error", "Flange height" and "Truck shift", "Truck tracking error" and "Truck inter-axle misalignment", respectively. Therefore, truck defects show a significant relationship with the degradation of the wheel profile.

Furthermore, we employ logistic regression to verify whether or not truck measurements are significant variables in predicting wheel failures:

$$p = \frac{1}{e^{-(\beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \ldots)} + 1}, \quad (1)$$

where: p is the probability of wheel failure in 3 months; x_i is i-th truck measurements; β_i the model coefficient of x_i.

The explanatory variables are monthly aggregated truck measurements, which can be separated into two categories. One is the 90-th percentile of the truck measurements in a month, and the other one is the percentage of truck measurements greater than a threshold that obtained from historical data.

The results of the logistic regression are shown in Table 3. Only "90-th percentile of rotation/speed" and "Percentage
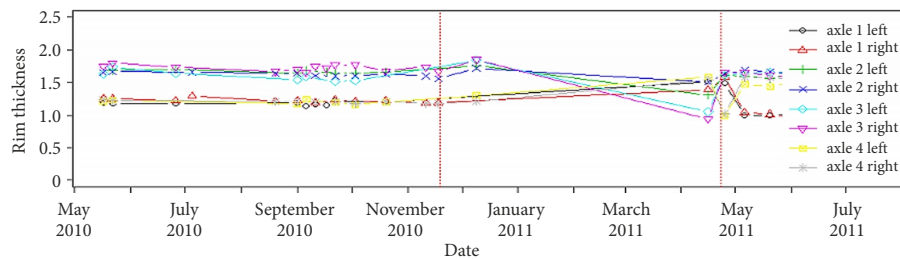

Figure 1. Plot of rim thickness in inches of 8 wheels in the same equipment over time, where:
L – indicates the left wheel; *R* – indicates the right wheel, and red dashed line represents the maintenance time

Table 2. Correlations of wheel profile measurements and truck measurements

	Truck hunting peak-to-peak	Truck shift	Truck tracking error	Truck inter-axle misalignment	Truck rotation error
Flange thickness	−0.102	−0.294	−0.476*	−0.230	−0.151
Flange height	0.094	0.380*	0.430*	0.311*	0.114
Rim thickness	−0.131	−0.212	−0.192	−0.206	−0.040

*Note: * – relatively hight correlation value.*

Table 3. The results of logistic regression between truck measurements (explanatory variables) and wheel failures (dependent variable)

Predictors	Abbreviation	Estimate	Std. error	z	*p*-value	Significance[#]
Description	(intercept)	−4.59108	0.020419	−224.841	2.00×10^{-16}	**
90-th percentile of peak-to-peak/speed	PTP.SPD.per90%	6.476269	0.198645	32.602	2.00×10^{-16}	**
90-th percentile of tracking error/Speed	TE.SPD.per90%	0.633454	0.049483	12.801	2.00×10^{-16}	**
90-th percentile of shift/speed	SPD.per90%	1.279228	0.112054	11.416	2.00×10^{-16}	**
90-th percentile of rotation/speed	SPD.per90%	0.0432	0.584981	0.074	0.94113	*
90-th percentile of inter-axle misalignment/speed	IAM.SPD.per90%	7.438507	1.069041	6.958	3.45×10^{-12}	**
Percentage of peak-to-peak greater than a threshold (>3.9)	HighPTP_per	−0.08309	0.047422	−1.752	0.079743	*
Percentage of inter-axle misalignment greater than a threshold (>0.7)	HighIAM_per	0.210569	0.047615	4.422	9.77×10^{-6}	**
Percentage of rotation greater than a threshold (>0)	HighROT_per	0.459908	0.127678	3.602	0.000316	**
Percentage of tracking error greater than a threshold (>13)	HighTE_per	1.183915	0.035514	33.336	2.00×10^{-16}	**
Percentage of shift greater than a threshold (>6)	HighSHFT_per	0.843808	0.042644	19.787	2.00×10^{-16}	**

Note: [#]predictor significance defined by p-value: 1 for "" and 0 for "**".*

of peak-to-peak greater than a threshold" are not significant with wheel failures. All the other features display very strong significance in predicting the wheel failures. As a consequence, we can draw a conclusion that bad trucks will lead to wheel wear and wheel failures.

2.2. Data pre-processing

The overview of the proposed algorithm for the identification of asymmetric wheel wear is presented in Figure 2. In this section, we process the unorganized data from three datasets (MV, OGD and maintenance) to obtain the potential features in the identification model.

In order to reduce the errors from locating incorrect wheel positions, we only focus on 2-truck equipment that has 4 axles and 8 wheels. Raw MV data distinguishes the left and right wheels. Since we shall examine the asymmetry of the wheel profile, the first step of data pre-processing is to check data variability for the differences between two wheels on the same axle and combine them to the axle level. As discussed above, bad trucks will lead to wheel failures. Therefore, it is necessary to combine MV data with truck measurements from OGD together. We aggregate data into monthly level (using min, max, percentile functions, etc.) to ensure the accuracy of identification. On the one hand, in order to describe MV data with small bias, we aggregate individual MV measurements into monthly data by taking the median. One the other hand, we aggregate OGD readings into monthly data by taking the 95-th percentile of absolute values of readings to study the worst case of truck components. Then we extract the unique truck ID between MV and OGD data. After this, we combine MV and OGD of the same truck together by truck ID. Finally, we merge maintenance data with combined MV and OGD data to obtain the validation results. In this paper, we assume that the asymmetric wheel wear is validated if a wheel failure is observed within a specific time window of the identification event by the proposed model.
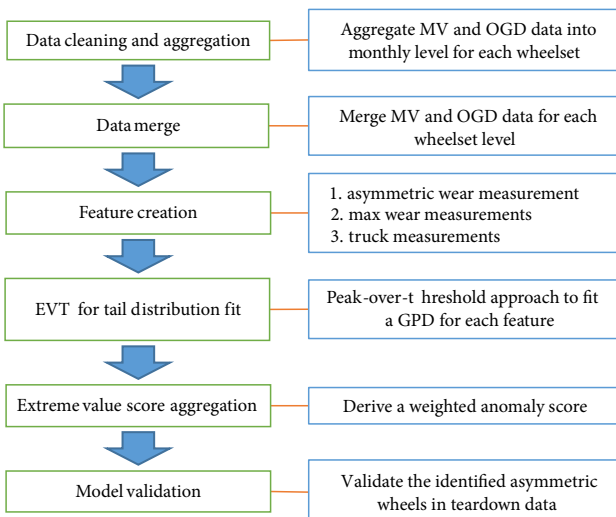

Figure 2. Flow chart of the proposed algorithm
for identification asymmetric wheel wear

3. Proposed methodology

3.1. Model of methodology

EVT is a branch of statistics that cares about the unusually high or low data of distribution (e.g. data locates in the tails of a distribution) (Markou, Singh 2003). Since the extreme data points can represent the outlying regions of normal events against anomaly detections, these points are important for further investigation (Roberts 1999).

Assume X_i are independence identically distributions with a distribution:

$$F(x) = P\{X_i \le x\}. \tag{2}$$

Let:

$$M_n = \max\{X_1, X_2, ..., X_n\} \tag{3}$$

represents the maximum value of the n observations. It follows that the nth order statistic is given as follow:

$$\Pr\{M_n \le x\} = F^n(x). \tag{4}$$

Under certain conditions, there exist normalizing constants $a_n > 0$ and b_n, which can normalize M_n as follow:

$$M_n^* = \frac{M_n - b_n}{a_n}. \tag{5}$$

Then the distribution function becomes:

$$\Pr\left\{\frac{M_n - b_n}{a_n} \le x\right\} = \Pr\{M_n \le a_n + b_n\} =$$
$$F^n(a_n \cdot x + b_n) \to H(x). \tag{6}$$

The three types theorem (Fisher–Tippett–Gnedenko) asserts that it must be one of three types if nondegenerate H exists:

$$H(x) = \exp(-e^{-x}), \quad x \in R; \tag{7}$$

$$H(x) = \begin{cases} 0, & x < 0; \\ \exp(-x^{-\alpha}), & x > 0; \end{cases} \tag{8}$$

$$H(x) = \begin{cases} \exp(-|x|^\alpha), & x < 0; \\ 1, & x > 0. \end{cases} \tag{9}$$

Equations (7)–(9) are called Gumbel distribution, Fréchet distribution, and Weibull distribution, respectively. Moreover, in Fréchet distribution and Weibull distribution, $\alpha > 0$. All three distributions are called General Extreme Value (GEV) distributions, and their general form is:

$$G(x; \mu, \sigma, \xi) = \exp\left(-\left(1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right), \tag{10}$$

where: μ is the location parameter $(\mu \in R)$; σ is the scale parameter $(\sigma > 0)$; ξ is the shape parameter $(\xi \in R)$.

Asymmetric wheel wear lies in the right tail part of a distribution, which means it is the extremely bad case of data and only has a small portion. In order to describe asymmetric wheel wear, we employ perk-over-threshold of EVT to fit a Generalized Pareto Distribution (GPD) for each feature.

For the distribution $G(x; \mu, \sigma, \xi)$ has a high threshold u. X_i is peak over value, when $X_i > u$. When $X_i < u$, X_i follows GEV distribution. When $X_i > u$, let $y = x - u \,|\, x > u$,

$$F(y) = \frac{F(u+y) - F(u)}{1 - F(u)}. \tag{11}$$

As $u \to w_F = \sup\{x : F(x) < 1\}$ often finds a limit:

$$F(y) \approx G(y; \sigma, \xi), \tag{12}$$

where G is called GPD with parameters σ – scale and ξ – shape,

$$G(y; \sigma, \xi) = \begin{cases} 1 - \left(1 + \xi \cdot \dfrac{y}{\sigma}\right)^{-\frac{1}{\xi}}, & \xi \neq 0; \\ 1 - \exp\left(-\dfrac{y}{\sigma}\right), & \xi = 0. \end{cases} \tag{13}$$

We consider a X_i serves as an input for parameter estimation of GPD once it excesses the threshold (Ortiz *et al.* 2009). Figure 3a depicts the concept of GPD and Figure 3b displays the Cumulative Distribution Function (CDF) of GPD.

Then we identify the asymmetric wheel wear by using the EV scores. We compute EV scores CDF for the fitted GPD for each feature:

$$EV(y) = CDF\big(G(y; \sigma, \xi)\big). \tag{14}$$

After this, we aggregate the scores of all features. Weights w_i for feature i could be determined by preferences between false alarm rate and the number of identified asymmetric wheelsets (Dykes 2012):

$$\lambda = \sum_i w_i \cdot EV_i(y). \tag{15}$$

Then we define an anomaly detection threshold u for asymmetric wheel wear identification:

$$\text{Asymmetric wheel wear} = \begin{cases} 1, & \lambda \geq u; \\ 0, & \lambda < u. \end{cases} \tag{16}$$

Once λ of a wheelset measurement exceeds u, it is identified as asymmetric wheel wear.
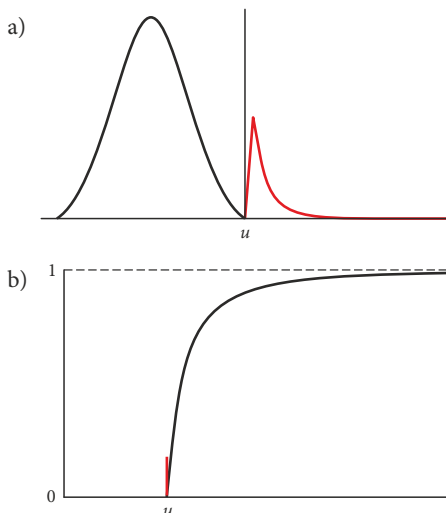

Figure 3. Distributions: a – probability distribution F (left) and extreme value distribution (GPD) (right); b – the CDF of GPD

3.2. Identification analysis and modeling

Data pre-processing generates many potential features that fall into two categories: respectively the wheel level data from MV and truck level data from OGD. In this section, we create several new features by using these two feature types. Then we use appropriate features to obtain the identification model, shown in Table 4.

1) *Identification features selection.* MV measures the information in wheel level, and we cannot use raw MV data directly. In order to describe the asymmetry of wheel wear and obtain identification model, two types of new features are created. They are asymmetric wheel wear and max wheel wear in a wheelset. Where x_L and x_R are the data of left and right side wheels, respectively. We use the absolute value of the difference between the left wheel and the right wheel to describe the degree of asymmetric wheel wear. The max wheel wear, representing the worst case between two wheels, is calculated based on the dimensions of a new wheel (rim thickness = 1.75 inch, flange thickness = 1.375 inch, and flange height = 1.1 inch).

From Section 3.1, it is found that wheel wear correlates with truck measurements and bad trucks will lead to potential wheel failures. Therefore, we should also consider truck level features in the model, shown in Table 4.

2) *Identification modeling.* Choosing an appropriate threshold u is a challenge when we model the GPD for identifying asymmetric wheel wear. If u is too large, there will be few data points exceeding this threshold. Thus, the number of identified wheels is very small.

Table 4. Features for identification modeling

| Feature category	Feature name	Equation
Asymmetric wheel wear	wear differences of rim thickness	$\|x_L - x_R\|$
	wear differences of flange thickness	$\|x_L - x_R\|$
	wear differences of flange height	$\|x_L - x_R\|$
Max wheel wear in a wheelset	max wear of rim thickness	$\max\big(1.75 - \min(x_L, x_R), 0\big)$
	max wear of flange thickness	$\max\big(1.375 - \min(x_L, x_R), 0\big)$
	max wear of flange height	$\max\big(\max(x_L, x_R) - 1.1, 0\big)$
Truck measurements	truck hunting peak-to-peak	normalized by truck speed
	truck shift	
	truck tracking error	
	truck rotation error	
	truck inter-axle misalignment	
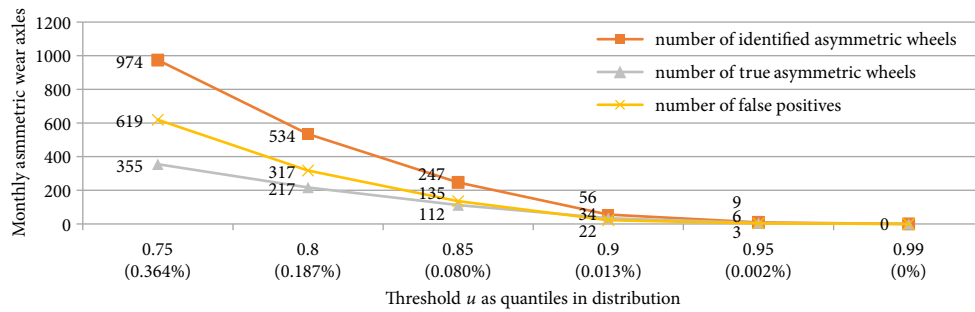
Figure 4. Identification results with different threshold *u* (the false alarm rates are indicated in the labels of *x*-axis)

If *u* is too small, the estimated model for asymmetric wheel contains lots of false alarms. Therefore, we shall balance between hit rates and false alarms.

In this paper, we set threshold equals to several possible values first, and then determine the scale and shape parameters for GPD distribution for each feature with each threshold. After this, we calculate CDF for every feature and each threshold and obtained the weighted sum λ with Equation (15). Weight w_i for feature *i* is determined by preferences between *false alarm rate* and the number of identified asymmetric wheelsets. Once λ is greater than the threshold *u*, we regard it as asymmetric wheel wear. Further, we calculate false alarm rate for each threshold:

false alarm rate =

$$\frac{number\ of\ false\ postivie}{number\ of\ false\ positive + number\ of\ true\ negative},$$

(17)

where: the *number of false positive* is the difference between the number of identified asymmetric wheels and true asymmetric wheelsets, and the *number of true negative* is the difference between the number of measured wheelsets and the number of validated wheelsets in teardown data.

For the false alarm rate, the smaller the better.

The entire algorithm can be summarized as follows:
– step 1: set threshold equals to predefined values;
– step 2: determine scale and shape parameters for GPD distribution for each feature with each threshold;
– step 3: calculate CDF for every feature and each threshold;
– step 4: calculate summation of extreme value score of each feature by weight;
– step 5: once λ is greater than the threshold *u*, asymmetric wheel wear is identified;
– step 6: choose appropriate threshold *u* that has both low false alarm rate and reasonable identified asymmetric wheelsets.

4. Numerical examples

4.1. Identification results

Figure 4 illustrates the number of identified asymmetric wheels and the number of true asymmetric wheels with different threshold *u*. As one can anticipate, the false

alarm rate keeps decreasing when *u* increases. It reaches a desirable value when threshold equals to 0.9 or 0.95. However, with threshold 0.95, the number of identified is very small to keep a low variance of GPD model. In addition, when the threshold is equal to 0.90, the results show that the inspection team needs to examine around 50 wheels in a month, which sounds very reasonable for a Class I railroad. Therefore, 0.9 has been selected as the threshold.

4.2. Pareto frontier

Pareto frontier is a line consisting of different Pareto efficient values that are all the potentially optimal solutions. The Pareto frontier for asymmetric wheel wear represents the lowest false alarm rate that can be obtained for a certain number of identified asymmetric wheelsets, shown in Figure 5.

In Figure 5, the *x*-axle represents the reciprocal of the number of identified asymmetric wheelsets and *y*-axle represents the false alarm rate. When the number of identified asymmetric wheelsets is small, the false alarm rate is also small. When the number of identified asymmetric wheelsets approaches positive infinity, the false alarm rate goes up rapidly. Railroads can choose their preferred number of monthly alerts based on their available inspection resources.

4.3. Benefit analysis for asymmetric wheel wear identification

We make the assumption for the benefit analysis as follows – the railcars with asymmetric wheel wear require more annual maintenance times than normal railcars.
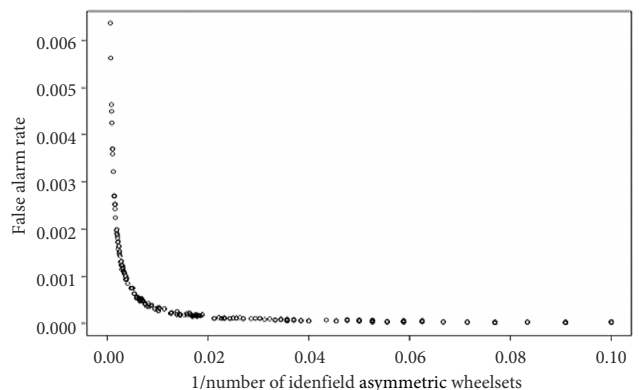

Figure 5. Pareto frontier for identified asymmetric wheelsets

Table 5. The results of benefit analysis

Maintenance railcar type	No of unique railcars N_A	No of maintenance times M	Maintenance rate $MR = \dfrac{M}{N_A}$	Benefit
Railcars identified with asymmetric wheel wear annually	602 (50 per month)	1285	2.134	$ 635 k
	961 (80 per month)	1997	2.078	$ 959 k
Randomly selected railcars	8219	8876	1.080 (MR_0)	n/a

According to the above assumption, if we can identify asymmetrically worn wheelsets and repair them, we can reduce the total annual maintenance times for a railcar with asymmetric wheel wear to be the same level of normal railcars, so that the yearly maintenance costs will be reduced eventually. In this paper, the benefit is defined as the annual amount of dollars saved by the reduction of maintenance times:

$$Benefit = N_A \cdot \left(MR - MR_0 \right) \cdot c, \qquad (18)$$

where: N_A is the number of railcars identified with asymmetric wheel wear; MR represents maintenance rate (the number of maintenance times divided by N_A) for railcars with asymmetric wheelsets; MR_0 is basic maintenance rate for normal railcars, which is equal to the number of maintenance times divided by the number of randomly selected railcars from database; c stands for the average cost of pulling a railcar through a maintenance shop.

Table 5 validates the above assumption from the maintenance data. It is found that the number of maintenance rate (maintenance times divided by unique railcar) for railcars with asymmetrically worn wheelsets is almost twice as high as the randomly selected ones. If we identify 50 (80) unique railcars with asymmetric wheel wear every month, their maintenance rate is 2.134 (2.078). In contrast, the maintenance rate for randomly selected railcars is 1.08. One can see that more frequent wheel failures are observed for railcars with asymmetric wheel wear.

Based on a rough estimation from a Class I railroad, we assume that the average cost of pulling a railcar through a maintenance shop is approximate $ 1000 each time. Without loss of generality, if one has another accurate estimation for different railroads, the approach proposed for benefit-cost analysis still stands. Table 5 shows that the railroad can save almost $ 1 mln annually by fixing 80 identified railcars with asymmetric wheel wear monthly.

Conclusions

Railroad wayside detectors can automatically capture rich information of locomotive and railcar components, including wheels, axles, and trucks. In this paper, based on data from wayside detectors, we present a mathematical model for identifying asymmetric wheel wear. A benefit analysis is performed to verify that our model can greatly reduce the expenses accordingly. Several important findings are worth discussing:

First, the readings of the MV detector are not always accurate. The variations of the MV measurement are found to be quite large. The readings of the rim thickness of 8 wheels in the same equipment fluctuate unpredictably over time. The correlation study shows that there exist a relatively high correlation between "Flange thickness" and "Truck tracking error", "Flange height" and "Truck shift", "Truck tracking error" and "Truck inter-axle misalignment", respectively.

Second, a logistic regression analysis shows that almost all the features of truck measurements show strong significant levels in predicting the wheel failures. The results clearly show that bad trucks will lead to wheel wear even wheel failures.

Third, we leverage inspection data from MV and OGD and aggregate them together on the wheelset (axle) level to identify asymmetric wheel wear with EVT. We introduce peak over threshold theorem to model the tail of the distribution, which is corresponding to asymmetric wheel wear. The GPD is proved to be the proper distribution to model the tail for each feature.

Fourth, we develop an extreme value score to identify asymmetric wheel wear. Wheelsets with high scores are identified as asymmetric wheel wear. The false positive rate is generally very low. We also obtain a Pareto frontier, which represents the lowest false alarm rate under a certain number of identified asymmetric wheelsets. Based on the Pareto frontier, the railroads can choose their preferred number of monthly alerts based on their available inspection resources. After maintenance, the number of maintenance times of railcar with asymmetric wheels will be lowered to the same level as normal railcars.

Finally, we perform a novel benefit analysis to calculate the savings by comparing the maintenance times of normal railcars and railcars with asymmetric wheel wear. Assuming that the average cost of pulling a railcar through a maintenance shop is approximate $ 1000 each time, early identification and repair of asymmetric wheel wear will approximately save $ 1 mln each year for fixing 80 railcars with asymmetric wheelsets monthly. The advantages of doing this not only save maintenance expenses but also improve railway safety.

For future work, with the rapid development of wayside detectors and other monitoring technologies, we can obtain more data to improve accuracy and decrease the false alarm rate. Moreover, with additional detailed maintenance and repair data, the truck maintenance study can expand from the asymmetric wheel wear to other rail safety related topics such as the rail track maintenance etc. The proper selection of repair type between replacement and turning for wheel maintenance is also a worthy topic in the future study.

Contribution

In this paper, based on data from wayside detectors, we present a mathematical model for identifying asymmetric wheel wear with EVT. We associated truck readings with wheel readings to predict wheel failure, and found that bad trucks will lead to bad wheels. A novel maintenance-times based benefit analysis is performed to verify that our model can greatly reduce the expenses accordingly.

Disclosure statement

We claim that we do not have any competing financial, professional, or personal interests from other parties.

References

AAR. 2016. *Economic and Public Benefits*. Association of American Railroads (AAR). Available from Internet:
https://www.aar.org

Akoglu, L.; Tong, H.; Vreeken, J.; Faloutsos, C. 2012. Fast and reliable anomaly detection in categorical data, in *CIKM'12: Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, 29 October – 2 November 2012, Maui, Hawaii, US, 415–424.
https://doi.org/10.1145/2396761.2396816

Bogdevičius, M.; Žygienė, R.; Dailydka, S.; Bartulis, V.; Skrickij, V.; Pukalskas, S. 2015. The dynamic behaviour of a wheel flat of a railway vehicle and rail irregularities, *Transport* 30(2): 217–232.
https://doi.org/10.3846/16484142.2015.1051108

Braghin, F.; Lewis, R.; Dwyer-Joyce, R. S.; Bruni, S. 2006. A mathematical model to predict railway wheel profile evolution due to wear, *Wear* 261(11–12): 1253–1264.
https://doi.org/10.1016/j.wear.2006.03.025

Broadwater, J. B.; Chellappa, R. 2010. Adaptive threshold estimation via extreme value theory, *IEEE Transactions on Signal Processing* 58(2): 490–500.
https://doi.org/10.1109/TSP.2009.2031285

Caires, S.; Groeneweg, J.; Sterl, A. 2009. Past and future changes in the North Sea extreme waves, in *Proceedings of the 31st International Conference on Coastal Engineering*, 31 August – 5 September 2008, Hamburg, Germany, 547–559.
https://doi.org/10.1142/9789814277426_0046

Camargo, L. F. M.; Resendiz, E.; Hart, J.; Edwards, J. R.; Ahuja, N.; Barkan, C. P. L. 2011. *Machine Vision Inspection of Railroad Track*. USDOT Region V Regional University Transportation Center Final Report. NEXTRANS Project No 0281Y02. 46 p.

Dailydka, S.; Lingaitis, L. P.; Myamlin, S.; Prichodko, V. 2008. Modelling the interaction between railway wheel and rail, *Transport* 23(3): 236–239.
https://doi.org/10.3846/1648-4142.2008.23.236-239

Durham, A. 1997. Case study: the coal line wheel and rail interaction strategy, in *Proceedings of the Sixth International Heavy Haul Railway Conference*, 6–10 April 1997, Cape Town, South Africa, 405–415.

Dykes, S. G. 2012. An extreme value theory approach to anomaly detection (EVT-AD). Poster, in *IEEE Symposium on Security & Privacy*, 20–23 May 2012, San Francisco, CA, US, 1–2.

Ellis, B. A. 2009. *The Challenges of Condition Based Maintenance*. The Jethro Project Consulting Group. 4 p.

Fröhling, R. D. 2006. Analysis of asymmetric wheel profile wear and its consequences, *Vehicle System Dynamics: International Journal of Vehicle Mechanics and Mobility* 44: 590–600.
https://doi.org/10.1080/00423110600879296

Jardine, A. K. S.; Lin, D.; Banjevic, D. 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing* 20(7): 1483–1510.
https://doi.org/10.1016/j.ymssp.2005.09.012

Kunkel, K. E.; Karl, T. R.; Brooks, H.; Kossin, J.; Lawrimore, J. H.; Arndt, D.; Bosart, L.; Changnon, D.; Cutter, S. L.; Doesken, N.; Emanuel, K.; Groisman, P. Y.; Katz, R. W.; Knutson, T.; O'Brien, J.; Paciorek, C. J.; Peterson, T. C.; Redmond, K.; Robinson, D.; Trapp, J.; Vose, R.; Weaver, S.; Wehner, M.; Wolter, K.; Wuebbles, D. 2013. Monitoring and understanding trends in extreme storms: state of knowledge, *Bulletin of the American Meteorological Society* 94(4): 499–514.
https://doi.org/10.1175/BAMS-D-11-00262.1

Lewis, R.; Braghin, F.; Ward, A.; Bruni, S.; Dwyer-Joyce, R. S.; Bel Knani, K.; Bologna, P. 2003. Integrating dynamics and wear modelling to predict railway wheel profile evolution, in *6th International Conference on Contact Mechanics and Wear of Rail/Wheel Systems* (*CM2003*), 10–13 June 2003, Gothenburg, Sweden, 1–11.

Li, H.; Parikh, D.; He, Q.; Qian, B.; Li, Z.; Fang, D.; Hampapur, A. 2014. Improving rail network velocity: a machine learning approach to predictive maintenance, *Transportation Research Part C: Emerging Technologies* 45: 17–26.
https://doi.org/10.1016/j.trc.2014.04.013

Li, X.; Jin, X.; Wen, Z.; Cui, D.; Zhang, W. 2011. A new integrated model to predict wheel profile evolution due to wear, *Wear* 271(1–2): 227–237. https://doi.org/10.1016/j.wear.2010.10.043

Li, Z.; He, Q. 2015. Prediction of railcar remaining useful life by multiple data source fusion, *IEEE Transactions on Intelligent Transportation Systems* 16(4): 2226–2235.
https://doi.org/10.1109/TITS.2015.2400424

Markou, M.; Singh, S. 2003. Novelty detection: a review – part 1: statistical approaches, *Signal Processing* 83(12): 2481–2497.
https://doi.org/10.1016/j.sigpro.2003.07.018

Mikaliūnas, Š.; Lingaitis, L. P.; Subačius, R. 2002. Analysis of locomotive wheel sets wearing, *Transport* 17(1): 3–7.
https://doi.org/10.3846/16483480.2002.10414003

Min, S.-K.; Zhang, X.; Zwiers, F.; Shiogama, H.; Tung, Y.-S.; Wehner, M. 2013. Multimodel detection and attribution of extreme temperature changes, *Journal of Climate* 26(19): 7430–7451. https://doi.org/10.1175/JCLI-D-12-00551.1

Ortiz, E.; Babbar, A.; Syrmos, V. L. 2009. Extreme Value Theory for engine health monitoring and diagnosis, in *2009 IEEE Control Applications*, (*CCA*) & *Intelligent Control*, (*ISIC*), 8–10 July 2009, St. Petersburg, Russia, 1069–1074.
https://doi.org/10.1109/CCA.2009.5280957

Ouyang, Y.; Li, X.; Barkan, C. P. L.; Kawprasert, A.; Lai, Y.-C. 2009. Optimal locations of railroad wayside defect detection installations, *Computer-Aided Civil and Infrastructure Engineering* 24(5): 309–319.
https://doi.org/10.1111/j.1467-8667.2008.00584.x

Roberts, S. J. 1999. Novelty detection using extreme value statistics, *IEE Proceedings – Vision, Image and Signal Processing* 146(3): 124–129. https://doi.org/10.1049/ip-vis:19990428

Schlake, B. 2010. *Impact of Automated Condition Monitoring Technologies on Railroad Safety and Efficiency*. MSc Thesis. University of Illinois at Urbana-Champaign, Champaign, IL, US 143 p.

Singh, S.; Tu, H.; Donat, W.; Pattipati, K.; Willett, P. 2009. Anomaly detection via feature-aided tracking and hidden Markov models, *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans* 39(1): 144–159. https://doi.org/10.1109/TSMCA.2008.2007944

Tournay, H. 2011. Integrated freight car truck design concept, in *WCRR 2011: 9th World Congress on Railway Research*, 22–26 May 2011, Lille, France, 1–11.

Tournay, H. 2008. The development of algorithms to detect poorly performing vehicles at wayside detectors, in *WCRR 2008: 8th World Congress on Railway Research*, 18–22 May 2008, Seoul, Korea, 1–11.

Ye, N.; Emran, S. M.; Chen, Q.; Vilbert, S. 2002. Multivariate statistical analysis of audit trails for host-based intrusion detection, *IEEE Transactions on Computers* 51(7): 810–820. https://doi.org/10.1109/TC.2002.1017701