Delft University of Technology

Delft Center for Systems and Control

Technical report 17-003

A Big Data Analysis Approach for Rail Failure Risk Assessment*

A. Jamshidi, S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, R. Dollevoet, Z. Li, and B. De Schutter

To cite this work, please refer to the published version:

A. Jamshidi, S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, R. Dollevoet, Z. Li, and B. De Schutter, "A big data analysis approach for rail failure risk assessment," *Risk Analysis*, vol. 37, no. 8, pp. 1495–1507, Aug. 2017. doi:10.1111/risa.12836

Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
phone: +31-15-278.24.73 (secretary)
URL: https://www.dcsc.tudelft.nl

* This report can also be downloaded via https://dspub.eu/17-003

A Big Data Analysis Approach for Rail Failure Risk Assessment

Ali Jamshidi,[1] Shahrzad Faghih-Roohi,[2] Siamak Hajizadeh,[1] Alfredo Núñez,[1*] Robert Babuska,[2] Rolf Dollevoet,[1] Zili Li,[1] Bart De Schutter[2]

Railway infrastructure monitoring is a vital task to ensure rail transportation safety. A rail failure could result in not only a considerable impact on train delays and maintenance costs, but also on safety of passengers. In this paper, the aim is to assess the risk of a rail failure by analyzing a type of rail surface defects called squats that are detected automatically among the huge amount of records from video cameras. We propose an image processing approach for automatic detection of squats, especially severe types that are prone to rail breaks. We measure the visual length of the squats and use them to model the failure risk. For the assessment of the rail failure risk, we estimate the probability of rail failure based on the growth of squats. Moreover, we perform severity and crack growth analyses to consider the impact of rail traffic loads on defects in three different growth scenarios. The failure risk estimations are provided for several samples of squats with different crack growth lengths on a busy rail track of the Dutch railway network. The results illustrate the practicality and efficiency of the proposed approach.

KEY WORDS: Big data analysis, Rail surface defects, Rail failure risk

1. INTRODUCTION

Among all transportation infrastructure, the railway network is one of the most successful transport systems for reducing transportation cost, traffic congestion, and air pollution emission levels. On the one hand, the increase in usage of the railway network requires a systematic monitoring plan to keep the trains running in a safe way as well as with the least possible disruptions.[(1)] On the other hand, a large amount of data is collected by frequent measurements from the monitoring systems of the infrastructure and the assets involved in the railway operations. This data should be controlled, stored, and processed, such that it can be employed to take all necessary actions to guarantee the rail asset quality level desired by the infrastructure manager.[(2)] The large amount of data should be processed into actionable knowledge within a certain time period.[(3)]

Risk is intuitively connected to decision making under uncertainty.[(4)] Recent developments in big data analytic for uncertainty management and risk assessment of industrial systems have been studied by Wu and Birge,[(5)] and Choi et al.[(6)] Risk assessment of large-scale systems is of current interest across many application domains such as healthcare,[(7)] environmental safety,[(8,9)] transportation,[(10–13)] business,[(14)] and product development.[(15)] In particular for railway applications, risk assessment is critical for the prediction of infrastructure health condition within a given time period. Continuous monitoring of railway systems

[1]Section of Railway Engineering, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands
[2]Delft Center For Systems and Control, Delft University of Technology, The Netherlands
[*] Address correspondence to Alfredo Nunez, Section of Railway Engineering, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands; tel:+31(0)15 27 89355; a.a.nunezvicencio@tudelft.nl

can guarantee the availability of data that can be used to assess the risk of infrastructure failures. Also, the database constructed from continuous monitoring of data will become larger and larger over time. Thus, applying a big data analysis approach is necessary in order to adequately monitor the infrastructure condition.[16]

Among all the railway infrastructure systems involved in the train operation, the rail track plays an important role in the railway networks. In an intensively used network, a considerable amount of the maintenance has to be allocated for the track, e.g., in the Dutch railway network, this amounts for almost half of the annual maintenance budget.[17] As a high percentage of failures occurring in the railway infrastructure is directly related to the rail, it is important to assess the failure risk of rails. The rail risk assessment involves detecting the rail defects that can potentially result in rail break and derailment in extreme cases.[18–20] Rail surface defects are caused by different factors such as fatigue due to large number of trains passing over rail components at especially welds, joints, and switches.[21] Early detection of surface defects is important to mitigate disastrous consequences of rail breaks. There are different methods to diagnose the condition of rail defects, including ultrasonic measurements,[22] eddy current testing,[23] and guided-wave based monitoring.[24] In general, these methods are not able to detect defects in an early stage of growth, i.e. not until the defects are severe. In particular, detection of defects at the late stage of growth imposes extra operation and maintenance costs due to the fact that the only solution is to replace the rail.

To address the limitations of the current measurement methods, the use of video cameras installed on trains has become popular.[25–27] The use of video cameras avoids the error-prone, costly, and time-consuming process of manual rail monitoring. Moreover, the videos taken from side cameras enable the infrastructure manager to capture the real condition of other track components such as fasteners, switches, and sleepers. Using video cameras, one can simply monitor whether the visible defects are at the early or late stage of growth. This means that the infrastructure manager has the opportunity to observe how the defect evolves over time in order to take actions at the right moment and to focus on the most urgent places for maintenance operations. This can lead to a significant reduction in the operation costs induced by the defects and it can prevent potential risks of rail breaks, reducing the risk of derailment. Due to the large amount and the high resolution of the videos taken over the rail, an automatic detection algorithm is required to process the huge amount of images from those videos.

The main contribution of the paper is to assess rail failure risk based on an integrated framework that merges the information of two defect-related variables: visual length and crack growth. There is no similar approach in the literature for risk assessment of rail failure that considers both variables. This is due to the fact that in this case a big data analysis problem has to be faced, as a result of which usually railway maintenance managers look at only one type of data and ignore the other influencing factors. We propose a risk function (equation 1) as a composition of three functions: the probability function, the crack growth function, and the partially inversed severity function. To evaluate these functions, we apply several techniques, including a deep convolutional neural network for image processing and defect detection, an N-step ahead prediction model for defect severity and crack growth analysis, and a Bayesian inference model for failure probability estimation.

To implement our proposed framework, a particular type of surface defect in railway networks called squat is considered in the case study. Furthermore, we give a proposed classification of the squats in terms of the visual length. Thus, squats are classified according to different severities. These classes can be used later for condition-based maintenance where we have different maintenance operations for different stages of the growth (rail grinding for light squats and replacement for severe squats). However, our approach can be generalized and applied for similar cases when there is a need to analyze a huge amount of image data for assessment of failure probability and risk function. For example, in a recent work by Skakun,[28] satellite images have been employed to assess flood hazard risk. Moreover, in the field of health science, abnormality detection using image processing has become very popular.[29] There are many cases in the literature where image data is used to deal with risk assessment problem.[28–32] In all these cases, as long as the focus is to detect abnormalities and failures among a big database of images, the risk assessment approach proposed in the current paper is applicable for merging attained information from images.

This paper is organized as follows. In Section 2, the proposed failure risk assessment model is

presented including the model framework. Section 3 addresses a real-life case study of the Dutch railway network. Section 4 presents the results and discussions. Finally, in Section 5, conclusions are presented.

2. FAILURE RISK ASSESSMENT MODEL

2.1 The Proposed Framework

In this section, we propose a failure risk framework for analyzing the rail surface defects. The proposed framework is depicted in Fig. 1. Video images, ultrasonic detection,[22] and eddy current testing[23] can all be used to detect the defects that can lead to rail break. In this paper, we rely on both ultrasonic detection method and video images. On the one hand, with ultrasonic measurement, we derive a general characteristic of crack growth. One the other hand, with video image, we analyze the growth of the visual length of defects which are detected among huge number of rail images. Then, a sample of the visual length of the detected defects is chosen for the assessment of failure risk model. The approach can be employed for any type of rail defects.

In this framework, a large amount of image data is automatically processed by a deep convolutional neural network to detect squats in Step 1 (see details in §2.4). The visual lengths of defects are measured from the defect detected from the video images, and then used for defect severity analysis in Step 2 (see details in §2.2).

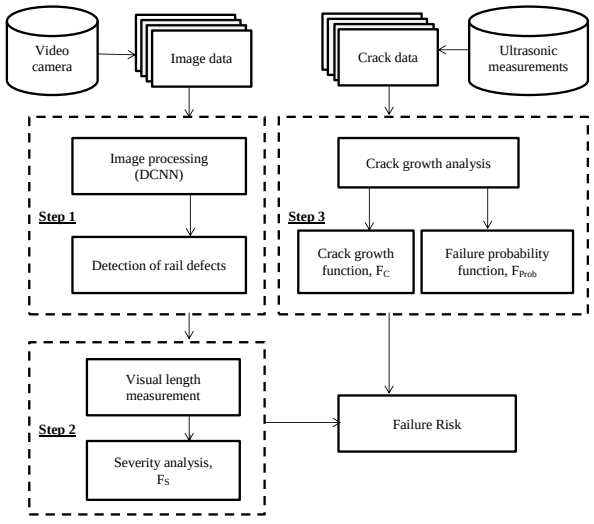

Fig. 1. Flowchart of the proposed methodology

In Step 3, a crack growth analysis is performed to estimate the crack growth as a function of Million Gross Tons (MGT) by using the data from ultrasonic measurements (see details in §2.3). In addition, the probability of rail failure as a function of crack growth is estimated using the crack growth data.

Finally, we propose to assess the risk of rail failure with the composition of the probability function, the crack growth function and the partially inversed of severity function:

$$\text{Risk} \propto F_{\text{Prob}}\left(F_C\left(F_{S,inv.}\left(V_1, V_2\right)\right)\right) \tag{1}$$

where V_1 and V_2 are two consecutive measurements of visual length for a defect, detected by analysis of image data, and $F_{S,inv.}$ relates V_1 and V_2 to MGT. Function F_C relates the estimate of MGT to crack growth, and function F_{Prob} estimates the probability of failure based on the estimate of crack growth. Thus, the risk is approximated relying on the failure probability achieved in equation (1). It means that the failure probability represents the risk of failure within a given MGT.

2.2 Severity Analysis

This section aims to model the visual length of defects based on the Million Gross Tons (MGT). MGT is a measurement unit to show the total weight of freight and passenger trains that pass over a given track in a given time horizon. Thus, the MGT can directly influence the growth of defects in the sense that an increase in the MGT accelerates the defect evolution process and the tracks with a lower train occupation are expected to have a lower degradation rate than the busy tracks.

The defects are automatically detected using the image processing method described in §2.4. We measure the visual lengths of the detected defects to use in severity analysis. We consider visual length as an indicator of a defect severity. Analysis of rail image data shows that the visual length of defects can grow with different rates as the MGT increases.

To capture the dynamics of the growth, we keep track of the growth for each individual squat to determine the increase of the visual length in each MGT step. A generic function is used to model the growth. The function can be applied relying on different methods where two consecutive data measurements are available. We present the benefits of using an N-step ahead prediction model for the prediction of squat's growth in our recent studies. For details, see Jamshidi et al.[33,34]

Thus, considering index as an MGT increment counter, we use a N-step ahead prediction model to describe the growth of visual length at different growth scenarios $h = h_1, h_2, \ldots, h_H$:

$$\begin{cases} \hat{V}_i^h(m+1) = F_S^h\left(\hat{V}_i^h(m), M^h(m)\right) \\ \hat{V}_i^h(0) = V_i^h(0) \\ m = 0, 1, ..., N-1 \end{cases} \quad (2)$$

where $\hat{V}_i^h(m)$ is the estimate of the visual length for each individual squat i at step m assuming scenario h, $M^h(m)$ is the total amount of MGT in step m, $F_S^h(.)$ is the one-step ahead prediction function, and $V_i^h(0)$ is the visual length measurement at the current step.

By partially inversion of $F_S^h(.)$, we get $F_{S,inv.}^h$ as a function of the visual length in two consecutive MGT steps. In case of scarce data for the total amount of MGT in each step, an approximation can be made for the prediction model (2):

$$\hat{V}_i^h(m+1) = F_{S,approx.}^h\left(\hat{V}_i^h(m)\right) \quad (3)$$

A fixed increment of the MGT is selected to keep track the growth of the visual length. Then, we apply function $F_{S,approx.}^h$ in an N-step ahead fashion to reconstruct F_S^h. This yields to formulate the relation between visual length and MGT at step m. Once F_S^h is formulated, we can partially inverse it to get $F_{S,inv.}^h$ as follows:

$$MGT^h(m) = F_{S,inv.}^h\left(\hat{V}_i^h(m+1), \hat{V}_i^h(m)\right) \quad (4)$$

2.3 Crack Growth Analysis

2.3.1 *Crack Growth With MGT*

The crack growth of defects is an important factor in rail breaks. Independent of the defect severity, the growth of the crack length depends on the traffic load (MGT). The idea in this paper is to analyze the data measured by ultrasonic detection technique and to present a function for estimation of the crack growth over the MGT:[33,34]

$$\Delta\hat{L}_i^h(m) = F_C^h\left(\hat{M}^h(m)\right), m = 0, 1, ..., N-1 \quad (5)$$

where $\Delta\hat{L}_i^h(m)$ is the estimate of the crack growth length for defect i at MGT step m assuming scenario h and $F_C^h(.)$ is the crack growth function. We will use a similar approach as described in §2.2 to assess the crack growth function.

2.3.2 *Failure Probability*

Regarding the crack growth data, assume the crack growth length is ΔL, containing total I measurements $(\Delta L_1, \Delta L_2, ..., \Delta L_I)$.Then the failure event can be defined as:

$$\bigcup_{i=1}^{I} \{\Delta L_i > d_i\} \quad (6)$$

where d_i is the critical level for the ith measurement. This formula implies that a failure occurs if the crack growth length exceeds the critical level. Logistic function is appropriate for these data since the variable is binomial meaning that the system fails if the measurement value satisfies (6), otherwise no failure.[35] Therefore, a logistic function is considered for the likelihood of rail failure probability $f(\Delta L|(a,b))$ with parameters a (intercept) and b (slope).

Recently, the Bayesian inference model has been employed extensively to assess model uncertainty and robustness for stochastic data behaviors.[36–38] Using a Bayesian inference model, variations of the model parameters can be considered as a step-wise degradation process. According to Bayes theorem, if prior knowledge about the parameter $\theta = (a,b)$ is represented by its probability density distribution $\pi_0(\theta)$, and if the statistical observations of crack growth length have likelihood $f(\Delta L|\theta)$, then rail failure probability can be expressed as posterior distribution π:

$$\pi(\theta|\Delta L) = \frac{f(\Delta L|\theta)\,\pi_0(\theta)}{f(\theta)} \propto f(\Delta L|\theta)\,\pi_0(\theta) \quad (7)$$

Typically, Monte Carlo methods are used in Bayesian data analysis to derive the posterior distribution.[39,40] The aim of using a Monte Carlo method is to generate random samples from the posterior distribution in order to use them when it is impossible to analytically compute the posterior distribution. Among all the Monte Carlo methods, slice sampling is easier to implement as only the posterior needs to be specified.[41,42] The slice sampling algorithm selects samples uniformly from the region under the density function. Therefore, in this paper, a slice sampling algorithm is selected to capture the failure probability function.

2.4 Analysis of Rail Image Data

We consider a railway health monitoring situation where a huge amount of video data is regularly

collected. Subsequently, the video data needs to be analyzed in order to detect defects with a potential risk of rail break. The data is collected by a set of high frame rate cameras that are mounted on a measurement train. The video recordings cover the entire length of the measured distance on the rail track. The mounted cameras capture the rails from several angles to look at different components. The top view camera is aimed at the rail surface defects, with each frame covering a length of 15 cm of the track along the longitudinal direction. The recordings are preprocessed into video compilations where consecutive frames have a few millimeters of overlap and the effects of variations in the train speed are removed. Recordings made from (bi)monthly measurements of roughly 6500 kilometers of rail amount to producing thousands of Gigabytes. Every 4 Gigabytes of data covers 16 kilometers of rail track. As a result, for recording videos of the whole Dutch rail network, almost 10 terabytes of data is required per year.

To be able to automatically extract defect information from the data, we train and apply a deep convolutional neural network (DCNN)[43] to detect and classify the defects. Recently, application of DCNN has become very popular in the domain of big data due to the increases in the size of available training sets and algorithmic advances such as the use of piece-wise linear units and dropout training.[44–46] By passing through a number of convolutional layers, the images are fed to the DCNN to train a set of shared neuron weights, referred to as filters. Convolution filters detect distinguishing features and form what is called a feature map. We use Rectified Linear Unit (ReLU)[47] activation functions after the convolution steps, and max-pooling layers to efficiently down-sample the outcome of each layer. Moreover, to prevent overfitting to the training data, we use dropout layers before each convolutional layer. Overfitting occurs when a classifier is fitted too closely to the sample data set that is unable to accurately describe the entire population, resulting into a high error over the test data. The dropout layer is known to prevent this by randomly disabling some activation from the previous layer.[48] The convolutional and pooling layer are finally attached to a sequence of three fully-connected layers to get class predictions.

The DCNN is trained by iterative feed forward of the training examples through the network and by calculating the error with respect to the desired outcome. The error and its gradient are then evaluated at the last layer of the network and back-propagated through all the layers to adjust all the weights. Repeating this process until decreasing the error to a certain limit is called the gradient descent algorithm.[47] We use a widely applied variation of the algorithm where on each iteration, the error and gradients are calculated using a randomly selected set of training examples usually called a mini-batch.[47]

3. CASE STUDY

In this section, a track from the Dutch railway network is considered to illustrate the capabilities of the proposed methodology. Track availability can be affected by rail surface defects. Among all types of rail surface defects, like rail corrugation, head checks, shatter cracking, vertical splits, head horizontal splits and wheel burns, squats play an important role in having a significant impact on the health condition of the track. Therefore, our main focus is on detecting the squats in this case study.

We select a sample from these data that contains recordings over a track in the North of The Netherlands from Zwolle to Groningen corresponding to approximately 300,000 captured frames. Two successive measurements of the same location along the track are matched together using the available time and geographic data. In total 4220 samples are labeled and used for training and testing of the neural network model. Out of the total set of samples, 3170 are normal rail samples and roughly 1000 are squats.

The proposed DCNN architecture for analyzing this amount of image frames is presented in Fig. 2. Initially the input images are down-scaled to 375×275 pixels and converted into gray-scale. The sequence of three fully-connected layers translates the extracted high-level features from the previous layers, into 3 classes representing the normal rail, trivial defects (seed squats), and squats.

Trivial defects appear in the form of spots or small damages to the rail head, while squats are usually defects that are fully grown indentations and deformations of the rail surface. The normal class includes all other components such as plain rails, switches, welds, possible non-defect contaminations, etc.

To train the network, a set of manually labeled examples is collected from several locations along the measured track and is compiled into a training set for each one of the 3 classes. The network is trained once and then is used for multiple time predictions. The

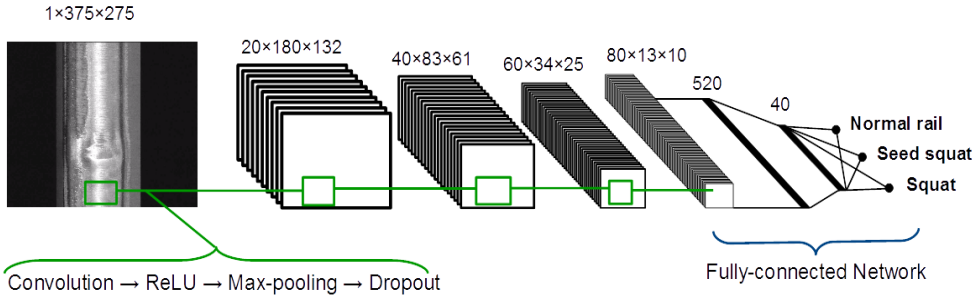

Fig. 2. Architecture of the proposed DCNN model

training time is 40 hours per 1500 examples. Once the network is trained, it is used to find squats in the large pool of previously unseen samples (prediction). These samples are collected from other monitoring sessions. Unlike the training time, the prediction time is insignificant (30 seconds per 15000 examples). The prediction result then has an average binary accuracy of 96.9% (squat vs. normal) when training on 80% of the labeled dataset and testing on the remaining 20%. By putting a high acceptance threshold on the network output response, we opt to detect the correct cases of squats, trivial defects and the normal cases.

Hence, after training and testing, we use the model to predict the severity of squats from the large amounts of available unlabeled data, from which we choose 109 detected squats for manual measurement of visual lengths in the track Zwolle-Groningen. Then, the samples are used in the next step where the growth of visual lengths is considered as described in §2.2. Here, squats with a visual length below 15 mm are considered as light squats, in which cracks have not appeared yet (surface initiation is assumed, and we cannot see beneath the surface from the image). Squats with visual length ranging from 15 to 30 mm are considered to be at the medium stage of growth. The medium squats evolve to severe squats when the network of cracks spreads further. Fig. 3 shows reference photos of squats ranging from light to severe together with crack evolution.

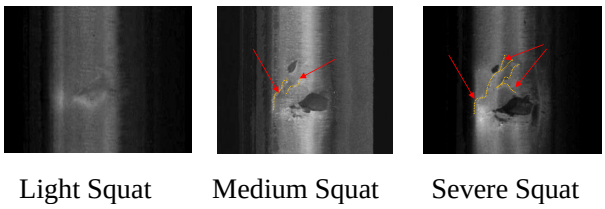

Fig. 3. A sample of squats in different classes of severity, the red arrows show the evolution of the crack when it gets severe.

Light squats will evolve into medium or severe squats after repeated train passes. Once the squat is severe, the squat will evolve into a defect with surface initiated cracks growing along the depth beneath the rail surface.[49]

Following the detection of squats by image processing, we apply the approach as described in §2.2 for this particular case to construct severity function. From real data of visual length, we estimate $F^h_{S,approx.}$ from equation (3).

Fig. 4 shows the relation between two consecutive measurements of visual length for a fixed value of MGT-step ($m = 1$). Relying on the physical understanding of how a squat grows, we fit a polynomial regression model of degree 3, using the least-absolute residual method,[50] to represent the stochasticity of the growth. The residual plot together with the R-square value of 0.9778 determines how well the polynomial model fits the data. We consider the fit model as an average growth scenario, and the 3-sigma control limits as slow and fast scenarios.

We use the estimated function of Fig. 4 for 8-step ahead prediction, and consider a fixed MGT

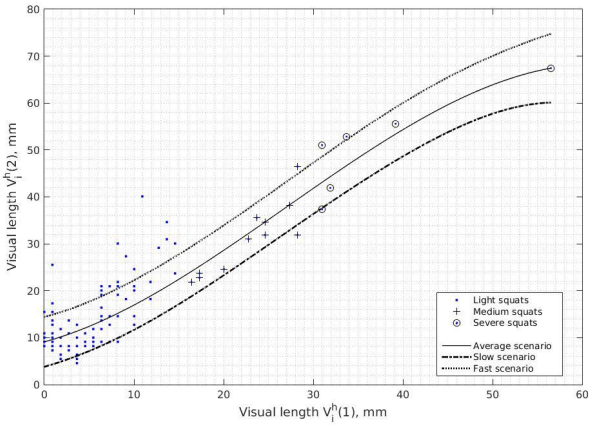

Fig. 4. Estimation of the visual length of the squats for $m = 1$, and based on real data

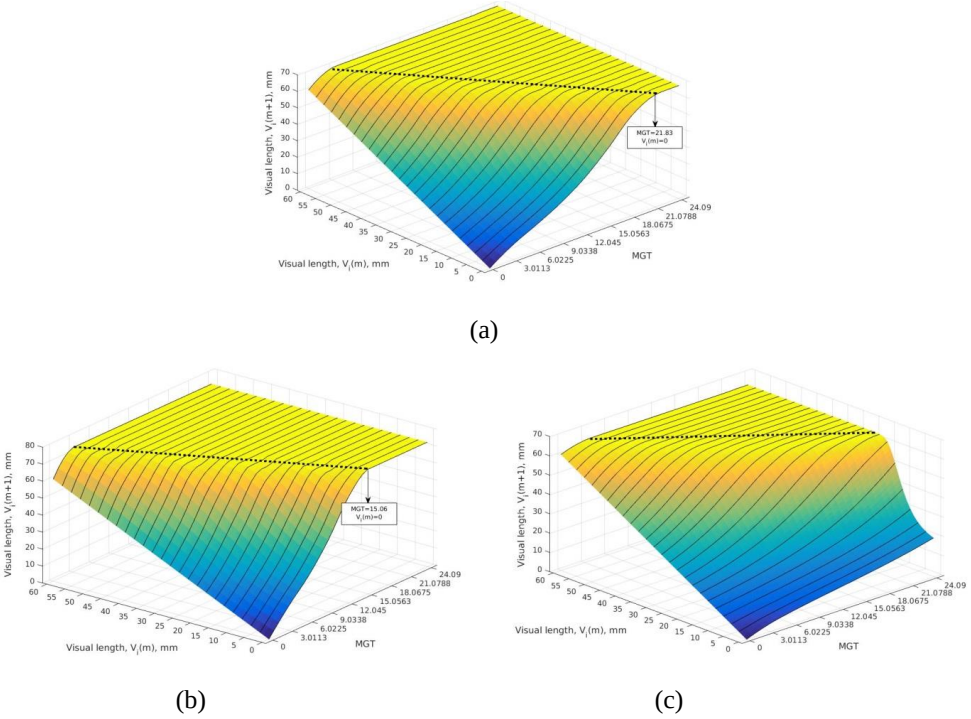

Fig. 5. Growth of squat visual length over MGT for the following model-based growth scenarios: (a) average, (b) fast, (c) slow, the dotted line depicts an upper bound of squat visual length.

increment of 3.01 in each step. As a result, a model-based prediction function for the visual lengths versus MGT is depicted in Fig. 5, considering the three scenarios of average (a), fast (b), and slow (c).

The dotted line shows the upper bound of the estimation for visual length, i.e. it is very rare to observe a squat with a length over the upper bound in reality.

Assuming $V_i^h(m) = 0$, the visual length at MGT step $m + 1$ at the fast scenario, reaches the upper bound with a MGT ($\text{MGT}_{h_1} = 15.06$) lower than at the average scenario ($\text{MGT}_{h_3} = 21.83$) and at the slow scenario ($\text{MGT}_{h_2} = 51.32$). It means that the degradation process in the fast scenario is more accelerated than in the average and slow scenarios as the traffic load on rail increases.

As described in §2.3.1, we estimate the crack growth function, $F_C^h(\cdot)$ by relying on ultrasonic measurement data. The model-based relation between the crack growth length and MGT is shown in Fig. 6. In addition, three different scenarios are considered to capture the crack growth dynamics, including the average scenario, the slow scenario, and the fast scenario. As seen in the figure, at the fast scenario, crack propagation of the squat at a given MGT is significantly faster than squats that are at the average and slow scenario. For example, at MGT = 10.36, it is estimated that the crack length of a squat grows 1 mm at the slow scenario, 2 mm at the average scenario, and 8 mm at the fast scenario. We can assess the risk of rail failure considering any of the different scenarios of crack growth length.

In the failure probability model, we consider that a rail is prone to fail when a squat reaches a crack length of 9 mm. The crack length of each squat is

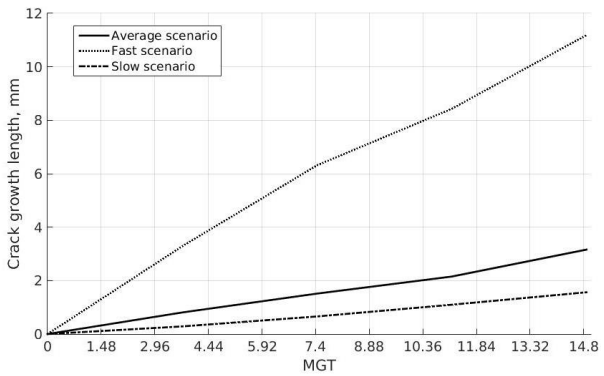

Fig. 6. A model-based relation between crack growth length and MGT

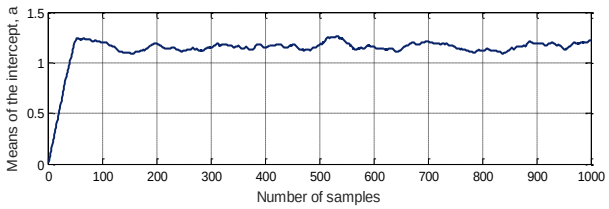

Fig. 7. Posterior distributions of regression parameter a

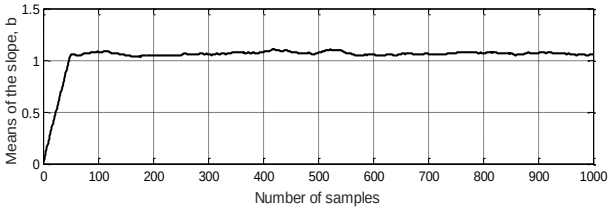

Fig. 8. Posterior distributions of regression parameter b

measured to see how it has grown over MGT, and how many cracks have reached a length of 9 mm or even more.

We use normal priors for the regression parameters (a, b). Relying on the data for the crack growth length, the parameters are estimated by a slice sampling algorithm considering 1000 samples. Respectively, Fig. 7 and Fig. 8 shows how the mean of the parameter a and b varies over the samples and converges to a constant value. As seen in the figures, the posterior means of parameters converge to a stationary status after the first 50 samples.

4. RESULTS AND DISCUSSION

For a detected squat with measured visual lengths in one MGT step, we estimate the risk of rail failure as follows: From the model in Fig. 5, we estimate the MGT for the visual lengths in two consecutive measurements. Then, from the model in Fig. 6, we find the crack growth length for the estimated MGT. Finally, we estimate the failure probability from the crack growth length in Fig. 9.

The failure probability plot represents how probable a squat fails in the next MGT step when the crack growth length is given. As an example, if the crack length of a squat increases 6 mm for MGT =7.04, the probability that the squat could lead to a rail break is roughly 0.82.

In Fig. 10, a sample of 5 squats is visualized, and the estimates of failure probability from the given visual lengths are presented.

For instance, the squat with $V_1 = 42$ mm and $V_2 = 57$ mm will cause a rail break with a probability

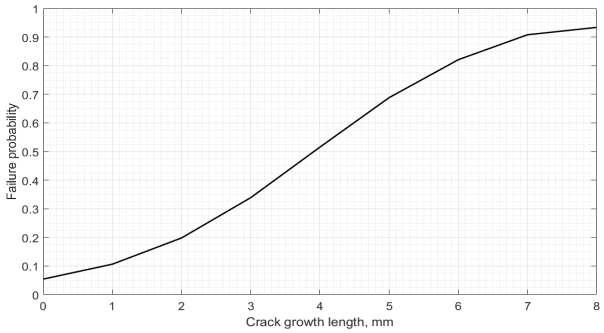

Fig. 9. Probability of rail failure based on the growth of crack length

of 28.9% in next MGT step, if no maintenance action is operated. However, no serious failure threatens the squat at the early stage and the failure probability is then almost 10% (see the squat with 16 mm in visual length). In Table I , more samples of squats are presented.

The table includes 64 samples of squats with

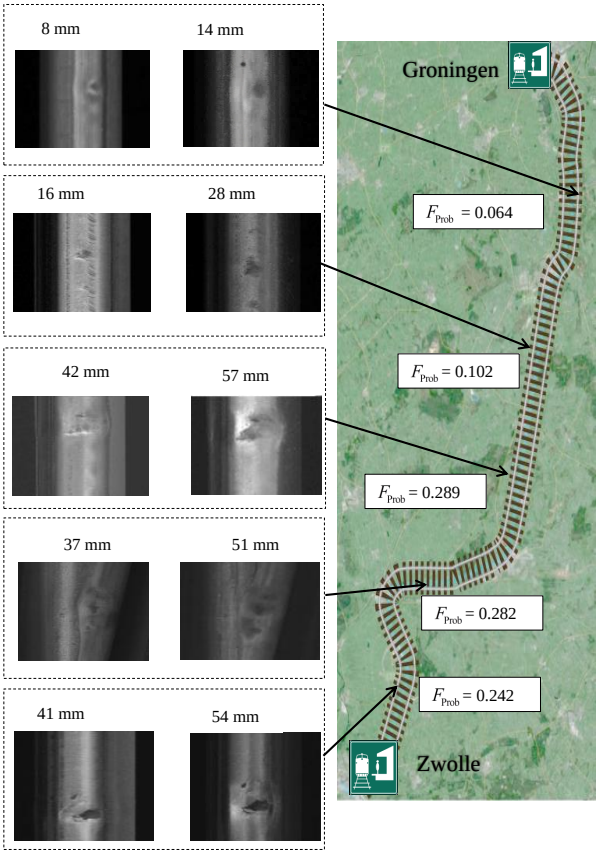

Fig. 10. A sample of failure risk estimates for 5 squats over the track

Table I . Failure risk estimation for a sample of squats, detected on the track Zwelle-Groningen

Squat	V_1	V_2	ΔL	F_{Prob}	Squat	V_1	V_2	ΔL	F_{Prob}
1	3.65	4.56	0.02	0.055	33	16.41	21.87	0.48	0.079
2	8.20	9.11	0.03	0.056	34	10.03	14.58	0.42	0.076
3	3.65	5.47	0.05	0.057	35	6.38	19.14	0.46	0.078
4	7.29	9.11	0.05	0.057	36	8.20	21.87	0.48	0.079
5	3.65	6.38	0.08	0.058	37	17.32	23.70	0.55	0.083
6	5.47	8.20	0.09	0.059	38	7.29	20.96	0.49	0.080
7	6.38	9.11	0.09	0.059	39	6.38	20.96	0.53	0.082
8	4.56	8.20	0.10	0.060	40	9.11	27.34	0.63	0.087
9	5.47	9.11	0.11	0.060	41	11.85	18.23	0.60	0.085
10	2.73	7.29	0.13	0.061	42	8.20	30.08	0.78	0.095
11	3.65	8.20	0.13	0.061	43	14.58	23.70	0.77	0.094
12	4.56	9.11	0.14	0.061	44	28.25	31.90	0.95	0.104
13	2.73	8.20	0.15	0.062	45	11.85	21.87	0.90	0.101
14	5.47	10.03	0.15	0.062	46	10.03	20.96	0.94	0.103
15	6.38	11.85	0.17	0.063	47	14.58	30.08	1.17	0.122
16	7.29	12.76	0.19	0.064	48	30.99	37.37	1.55	0.156
17	3.65	10.03	0.19	0.064	49	13.67	30.99	1.31	0.134
18	4.56	10.94	0.19	0.064	50	12.76	29.16	1.29	0.133
19	5.47	11.85	0.21	0.065	51	10.03	24.61	1.20	0.125
20	8.20	14.58	0.21	0.065	52	20.05	24.61	1.48	0.151
21	10.03	12.76	0.25	0.067	53	13.67	34.63	1.48	0.151
22	2.73	10.94	0.24	0.067	54	24.61	31.90	1.91	0.190
23	6.38	13.67	0.24	0.067	55	31.90	41.92	2.23	0.231
24	7.29	14.58	0.24	0.067	56	10.94	40.10	1.95	0.194
25	6.38	14.58	0.27	0.068	57	22.78	30.99	2.35	0.248
26	3.65	12.76	0.27	0.068	58	24.61	34.63	2.56	0.277
27	9.11	18.23	0.29	0.069	59	27.34	38.28	2.62	0.286
28	2.73	13.67	0.34	0.072	60	39.19	55.59	3.05	0.348
29	6.38	16.41	0.34	0.072	61	23.70	35.54	3.09	0.355
30	8.20	19.14	0.38	0.074	62	33.72	52.86	3.69	0.461
31	17.32	22.78	0.46	0.078	63	28.25	46.48	3.87	0.493
32	8.20	20.96	0.44	0.077	64	30.99	51.04	4.10	0.532

their measurements of visual length for two MGT steps. As expected, the squat at the severe stage will be prone to a rail break if no operation is carried out on the rail within a given MGT step. For example, there is a 53% chance of failure for the 64th squat in which the crack growth length is 4.10 mm within the given MGT step. The estimated risk values for the squats at the late stage indicate the need for immediate rail replacements. For the squats at early stage, a grinding operation is suggested to postpone rail failure by treating the squats.

5. CONCLUSIONS

In this paper, we present a methodology for the risk assessment of rail failure for a type of rail surface defects called squats. A big data analysis approach is used to automatically detect squats from rail images. The visual lengths of squats are measured in order to use them in the severity analysis model, which captures the growth of visual length over MGT increments. In addition, due to the influence of crack growth on estimation of the failure risk, a crack growth analysis based on MGT has been performed. At the end, a Bayesian model is employed to estimate the failure probability. By relying on the estimated failure risk, the infrastructure manager is able to take actions at the right time and the right place in order to prevent unexpected consequences induced by rail breaks. While this paper is focused on the analysis of squats, the results can also be applicable for the analysis of other type of rail defects.

ACKNOWLEDGMENT

This research is part of the NWO/ProRail project (Multi-party risk management and key performance indicator design at the whole system level, PYRAMIDS), project code 438-12-300, and the STW/ProRail project (Advanced monitoring of intelligent rail infrastructure, ADMIRE), project 12235, which are partly funded by the Ministry of

Economic Affairs. The authors also would like to thank INSPECTATION for providing us with image data and technical support.

REFERENCES

1. Greenberg M, Lioy P, Ozbas B, Mantell N, Isukapalli S, Lahr M, Altiok T, Bober J, Lacy C, Lowrie K, Mayer H. Passenger rail security, planning, and resilience: application of network, plume, and economic simulation models as decision support tools. Risk Analysis, 2013; 33(11):1969-1986.
2. Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. International Journal of Information Management, 2015; 35(2):137-144.
3. Sheppard JW, Kaufman M, Wilmer TJ. IEEE standards for prognostics and health management. Aerospace and Electronic Systems Magazine, 2009; 24(9):34-41.
4. She S, Ma C, Wu DD. General probability-time tradeoff and intertemporal risk-value model. Risk Analysis, 2010; 30(3):421-431.
5. Wu D, Birge JR. Risk intelligence in big data era: a review and introduction to special issue. IEEE Transactions on Cybernetics, 2016; 46(8): 1718-1720.
6. Choi TM, Chan HK, Yue X. Recent development in big data analytics for business operations and risk management. IEEE Transactions on Cybernetics, 2016; 99:1-12.
7. Hofstetter P, Hammitt JK. Selecting human health metrics for environmental decision-support tools. Risk Analysis, 2002; 22(5):965-983.
8. Thekdi SA, Lambert JH. Decision analysis and risk models for land development affecting infrastructure systems. Risk analysis, 2012; 32(7):1253-1269.
9. Zhou Q, Lambert JH, Karvetski CW, Keisler JM, Linkov I. Flood protection diversification to reduce probabilities of extreme losses. Risk Analysis, 2012; 32(11):1873-1287.
10. Xu J, Lambert JH. Risk-cost-benefit analysis for transportation corridors with interval uncertainties of heterogeneous data. Risk Analysis, 2015; 35(4):624-641.
11. Faghih-Roohi S, Ong YS, Asian S, Zhang AN. Dynamic conditional value-at-risk model for routing and scheduling of hazardous material transportation networks. Annals of Operations Research, 2016; 247(2):715-734.
12. Siddiqui A, Verma M. An expected consequence approach to route choice in the maritime transportation of crude oil. Risk Analysis, 2013; 33(11):2041-2055.
13. Paltrinieri N, Bonvicini S, Spadoni G, Cozzani V. Cost-benefit analysis of passive fire protections in road LPG transportation. Risk Analysis, 2012; 32(2):200-219.
14. Wu DD, Chen SH, Olson DL. Business intelligence in risk management: some recent progresses. Information Sciences, 2014; 256:1-7.
15. Wu DD, Xie K, Gang C, Ping G. A risk analysis model in concurrent engineering product development. Risk Analysis, 2010; 30(9):1440-1453.
16. Fumeo E, Oneto L, Anguita D. Condition based maintenance in railway transportation systems based on big data streaming analysis. Procedia Computer Science, 2015; 53:437-446.
17. Zoeteman A, Dollevoet RPB, Li Z. Dutch research results on wheel/rail interface management: 2001-2013 and beyond. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2014; 228(6):642-651.
18. Liu X, Barkan C, Saat M. Analysis of derailments by accident cause: evaluating railroad track upgrades to reduce transportation risk. Transportation Research Record: Journal of the Transportation Research Board, 2011; (2261):178-185.
19. Liu X, Saat M, Barkan C. Analysis of causes of major train derailment and their effect on accident rates. Transportation Research Record: Journal of the Transportation Research Board, 2012; 15(2289):154-163.
20. Sandström J, Ekberg A. Predicting crack growth and risks of rail breaks due to wheel flat impacts in heavy haul operations. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2009; 223(2):153-161.
21. Molodova M, Li Z, Núñez A, Dollevoet R. Automatic detection of squats in railway infrastructure. IEEE Transactions on Intelligent Transportation Systems, 2014; 15(5):1980-1990.
22. Fan Y, Dixon S, Edwards RS, Jian X. Ultrasonic surface wave propagation and interaction with surface defects on rail track head. NDT & E International, 2007; 40(6):471-477.
23. Song Z, Yamada T, Shitara H, Takemura Y. Detection of damage and crack in railhead by using eddy current testing. Journal of Electromagnetic Analysis and Applications, 2011; 3:546-550.
24. Mariani S, Nguyen T, Phillips RR, Kijanka P, di Scalea FL, Staszewski W, Fateh M, Carr G. Noncontact ultrasonic guided wave inspection of rails. Structural Health Monitoring, 2013; 12(5-6): 539-548.
25. Li Y, Trinh H, Haas N, Otto C, Pankanti S. Rail component detection, optimization, and assessment for automatic rail track inspection. IEEE Transactions on Intelligent Transportation Systems, 2014; 15(2):760-770.
26. Li Q, Ren S. A visual detection system for rail surface defects. IEEE Transactions on Applications and Reviews: Systems, Man, and Cybernetics, Part C, 2012; 42(6):1531-1542.
27. Li Q, Ren S. A real-time visual inspection system for discrete surface defects of rail heads. IEEE Transactions on Instrumentation and Measurement, 2012; 61(8):2189-2199.
28. Skakun S, Kussul N, Shelestov A, Kussul O. Flood hazard and flood risk assessment using a time series of satellite images: a case study in Namibia. Risk Analysis, 2014; 34(8):1521-1537.
29. Sharkey J, Scarfe L, Santeramo I, Garcia-Finana M, Park BK, Poptani H, Wilm B, Taylor A, Murray P. Imaging technologies for monitoring the safety, efficacy and mechanisms of action of cell-based regenerative medicine therapies in models of kidney disease. European Journal of Pharmacology, 2016; 790:74-82.
30. Singh V. Higher pollution episode detection using image classification techniques. Environmental Modeling & Assessment, 2015; 21(5):591–601.
31. Moradiamin M, Memari A, Samadzadehaghdam N, Kermani S, Talebi A. Computer aided detection and classification of acute lymphoblastic leukemia cell subtypes based on microscopic image analysis. Microscopy Research and Technique, 2016; 79(10):908-916.
32. Javaid M, Javid M, Rehman MZU, Shah SIA. A novel approach to CAD system for the detection of lung nodules in CT images. Computer Methods and Programs in Biomedicine, 2016; 135:125-139.
33. Jamshidi A, Núñez A, Dollevoet R, Li Z. Robust and Predictive Fuzzy Key Performance Indicators for Condition-Based Treatment of Squats in Railway Infrastructures. Journal of Infrastructure Systems, 2017; 04017006.
34. Jamshidi A, Faghih-Roohi S, Núñez A, Babuska R,

De Schutter B, Dollevoet R, Li Z. Probabilistic defect-based risk assessment approach for rail failures in railway infrastructure. Proceedings of the 14th IFAC Symposium on Control in Transportation Systems (CTS), 2016; Istanbul, Turkey.
35. Xu D, Zhao W. Reliability prediction using multivariate degradation data. Proceedings of the IEEE in Reliability and Maintainability Symposium. 2005; Alexandria, VA, USA.
36. Li L, Wang J, Leung H, Jiang C. Assessment of catastrophic risk using Bayesian network constructed from domain knowledge and spatial data. Risk Analysis, 2010; 30(7):1157-1175.
37. Droguett EL, Mosleh A. Bayesian methodology for model uncertainty using model performance data. Risk Analysis, 2008; 28(5):1457-1476.
38. Kazemi R, Mosleh A. Improving default risk prediction using Bayesian model uncertainty techniques. Risk Analysis, 2012; 32(11):1888-1900.
39. Faghih-Roohi S, Xie M, Ng KM. Accident risk assessment in marine transportation via Markov modelling and Markov chain Monte Carlo simulation. Ocean Engineering, 2014; 91:363-370.
40. Marques R, Bouville C, Ribardiere M, Santos LP, Bouatouch K. A spherical Gaussian framework for Bayesian Monte Carlo rendering of glossy surfaces. IEEE Transactions on Visualization and Computer Graphics, 2013; 19(10):1619-1632.
41. Gilks WR. Markov chain Monte Carlo. John Wiley & Sons, Ltd; 2005.
42. Neal RM. Slice sampling. Annals of Statistics, 2003:705-741.
43. Faghih-Roohi S, Hajizadeh S, Núñez A, Babuska R, De Schutter B. Deep convolutional neural networks for detection of rail surface defects. Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN), 2016; Vancouver, Canada.
44. Krizhevsky A, Sutskever I, Hinton GE. Image net classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 2012; 1097-1105.
45. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015; 521(7553):436-444.
46. Sainath TN, Kingsbury B, Saon G, Soltau H, Mohamed AR, Dahl G, Ramabhadran B. Deep convolutional neural networks for large-scale speech tasks. Neural Networks, 2015; 64:39-48.
47. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 2014; 15(1):1929-1958.
48. Bengio Y. Learning deep architectures for AI. Foundations and Trends in Machine Learning, 2009; 2(1):1-127.
49. Li Z, Dollevoet R, Molodova M, Zhao X. Squat growth–some observations and the validation of numerical predictions. Wear, 2011; 271(1-2):148–157.
50. Bassett Jr G, Koenker R. Asymptotic theory of least absolute error regression. Journal of the American Statistical Association, 1978; 73(363):618-622.