

Probabilistic analysis of crack growth in railway axles using a Gaussian process

Advances in Mechanical Engineering
2020, Vol. 12(9) 1–12
© The Author(s) 2020
DOI: 10.1177/1687814020936031
journals.sagepub.com/home/ade


Jiajie He¹, Yong Cui^{2,3} , Yusun Liu⁴ and Hui Wang¹

Abstract

To reduce maintenance costs, it is important to carry out probabilistic analyses on railway vehicle components. In this work, a data-driven approach based on a Gaussian process for regression is developed to determine the probability of axle failure caused by crack growth in railway axles. For complicated failure modes, it is difficult or even impossible to build a reliable analytical or simulation model before using an analytical approach. The main purpose of this work is to develop an algorithm to infer the distribution of crack growth from limited measured data without having to build an underlying model. The results of the case study show that the determined timing for the first inspection and the probability of failure coincide with the known results derived by analytically based approaches. The problems associated with modelling and calibration can be overcome by a data-driven approach. The developed Gaussian process model can serve as a complementary instrument to validate other analytically based approaches or numerical analyses. The model can also be applied to the probabilistic analyses of other railway components.

Keywords

Railway axle, reliability analysis, probability of failure, Gaussian process, regression

Date received: 30 January 2020; accepted: 28 May 2020

Handling Editor: Shahin Khoddam

Introduction

Preventive maintenance is a prevailing principle in the railway industry. Consequently, maintenance activities are scheduled according to the number of travelled kilometres or the operating time of a railway component. An optimal maintenance plan is desired to reduce maintenance costs. The engineering discipline of reliability, availability, maintainability and safety (RAMS) should be fully considered; specifically, reliability analysis is carried out to ensure that the resulting risk of a maintenance plan is below a tolerable level. Since many stochastic aspects are involved in railway operations and maintenance, probabilistic analysis is applied to assess reliability.

The service life of railway axles can be more than 30 years with a high number of loading cycles. During the whole life cycle, several inspection and maintenance activities should be planned. An example of defining

inspection intervals is illustrated in Figure 1. The probability of failure (PoF) is calculated during the service life. Before the PoF reaches the tolerable PoF, an inspection should be planned so that the PoF can be reduced again. At each inspection, an axle with a crack size that exceeds the detectable size will be identified. The axle will be repaired or put out of service. However, there are still many very small, undetectable cracks remaining on the axles. In the beginning, the

¹CRRC Qingdao Sifang Co., Ltd., Qingdao, China

²Institut für Eisenbahn- und Verkehrswesen, Universität Stuttgart, Stuttgart, Germany

³Hefei University, Hefei, China

⁴Robert Bosch GmbH, Gerlingen-Schillerhöhe, Germany

Corresponding author:

Yong Cui, Institut für Eisenbahn- und Verkehrswesen, Universität Stuttgart, Pfaffenwaldring 7, Stuttgart 70569, Germany.
Email: yong.cui@ievwwi.uni-stuttgart.de



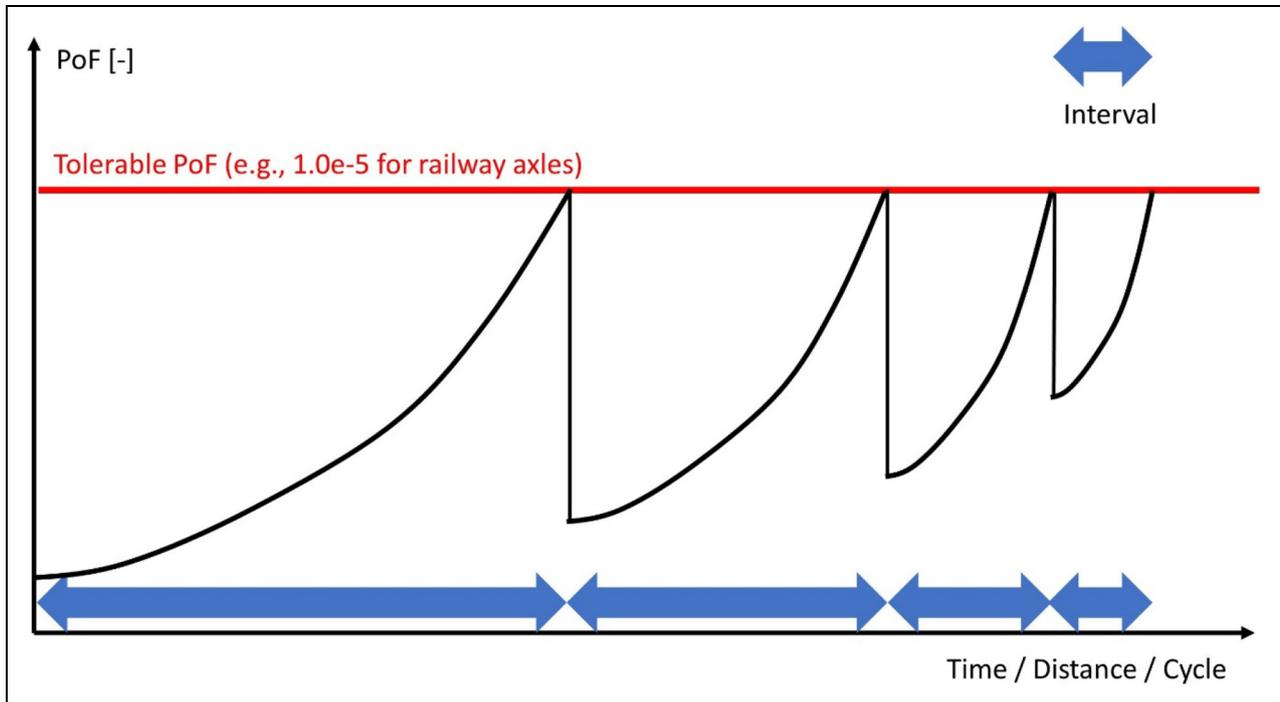


Figure 1. Definition of maintenance intervals based on the PoF.

growth rate of these small cracks is very slow. After a certain period of service, the accumulated loading effects will lead to a rapid increase in crack size. Therefore, the inspection interval will shorten over time. Traditional maintenance programmes are set in a fixed inspection interval. For example, an ultrasonic inspection should be carried out for a China Railway Rolling Corporation (CRRC) China Railway High-Speed (CRH)-2 high-speed train at each 90,000 km interval (approximately 3.5×10^7 rotation cycles). High maintenance costs will result from following a maintenance plan with overly short inspection intervals. However, if the inspection interval is too long, a crack may reach an intolerable depth before the next inspection, possibly leading to derailment. In this context, the objective of a probabilistic analysis is to find the maximum inspection interval that ensures that the PoF is under an acceptable level.

In the study by Zerbst et al.,¹ the mechanism of crack initiation in railway axles and the methods of determining a safe service life and damage tolerance are reviewed. Meanwhile, several research projects are being carried out in Europe, including EisenBahn FahrWerke (EBFW),² Bundesministerium für Bildung und Forschung (BMBF; German abbreviation for federal ministry of education and research) Project 19 P 0061 A–F 6,³ and wheelset integrated design and effective maintenance (WIDEM).⁴ In recent years, an investigation of crack growth and the optimisation of a maintenance plan for the Chinese railway high-speed (CRH) system has received great attention as well. This

work focuses on developing a probabilistic analysis of the crack growth in railway axles and uses railway axles for CRH trains as a case study.

As the distance and number of cycles increase over time, the inspection intervals will be shortened during the whole service life (see Figure 1). Instead of determining all the inspection intervals, this work calculates only the first inspection interval to evaluate the proposed method. The development of a maintenance plan during the whole service life will be investigated in future research.

In this article, the approaches for the probabilistic analysis of crack growth will be reviewed in section ‘Analytical, numerical and data-driven approaches for the probabilistic analysis of crack growth’. The constraints of analytically based approaches and numerical analyses are discussed. In section ‘Probabilistic analysis using a GP’, a data-driven approach based on a Gaussian process (GP) is proposed. The case study and the results are presented in section ‘Case study: regression with probabilistic layers’. The findings, the limitations of this work, and future directions are discussed in section ‘Conclusion and discussions’.

Analytical, numerical and data-driven approaches for the probabilistic analysis of crack growth

Martinez et al.⁵ summarised the approaches for analysing the crack growth in railway axles, including the analytically based approach, the numerical analysis and the

data-driven approach. These approaches can also be applied to probabilistic analysis. The main difference among these approaches is the mechanism for modelling the crack growth.

In the analytically based approach, the crack growth and propagation are modelled by using mathematical equations. In this work, Paris' law⁶ is applied to describe a power law relationship between the rate of crack growth and changes in the stress intensity factor (SIF). The SIF is derived from the stress, the crack depth and the geometry factor, depending on the loading and cracked body configuration. Paris' law and the calculation of the SIF are shown in equations (1)–(3)

$$K = \delta Y \sqrt{\pi a} \quad (1)$$

$$\Delta K = K_{max} - K_{min} \quad (2)$$

$$\frac{da}{dN} = C \Delta K^m \quad (3)$$

By integrating the differential equation (3), the crack growth will be solved for a certain number of cycles. In railway operations, the number of cycles can also be mapped to other dimensions, such as time or distance. In this article, a simplified implementation of Paris' law based on time is applied (see section 'Data preparation and calculation of the PoF based on the FORM/SORM') to generate sampled data and to verify the developed algorithm.

With the analytical model, the reliability of the system and the PoF can be estimated. In probabilistic analysis, multiple random variables and their joint distribution should be considered. In the study by Zhu et al.,⁷ the uncertainty quantification (UQ) is applied with multiplicative and additive methods, in order to identify and quantify the contribution of uncertainty from multiple types. Combined with classical statistics and data-driven approaches, the probabilistic model is updated through a Bayesian inference with experiment results. The first-order reliability method (FORM) and the second-order reliability method (SORM) have been widely used as an analytically based approach in recent decades.⁸ Given the analytical model, a limit state function can be defined to describe the boundary between the safe and failure states. By transforming the multiple correlated variables from physical space to a standardised uncorrelated space, the PoF can be calculated through a reliability index. Currently, the FORM and SORM have been implemented as standard features in many reliability analysis software programmes, for example, DAKOTA, OpenTURNS, SYSREL and PyRe. In this work, OpenTURNS is used to calculate the PoF based on an analytically based approach (see section 'Data preparation and calculation of the PoF based on the FORM/SORM').

The difficulty in using an analytically based approach increases with complicated crack geometries and different failure modes. Sometimes, it is even impossible to solve the partial differential equations in an analytical model. Supported by finite element method (FEM), numerical analysis is used to solve dynamic equations. The required inputs for numerical analysis are the component's geometry, initial cracks, properties of the materials, loading conditions, and other boundary and support conditions. During a simulation, the propagation of a crack and the other properties (e.g. stresses and SIF) can be obtained for further analysis.

Several commercial software programmes, for example, Zencrack and NASGRO, are available for fracture and fatigue analysis. These software programmes can be integrated with finite element software and software for reliability and uncertainty analysis. Software application and development are also very active in railway sectors. For example, the NASGRO model⁹ is used in the WIDEM project;¹⁰ the software INARA integrated with DAKOTA is developed for the probabilistic analysis in the EBFW-3 project,¹¹ and the programme ERWIN is developed to determine inspection intervals.³

Most research of crack growth and propagation on railway axles are concerned at individual cracks. It is difficult to investigate the characteristics of crack coalescence with an analytically based approach. In the study by Zhu et al.,¹² a probabilistic procedure for multiple surface fracture simulation is established. The behaviour of crack formation and propagation can be simulated with randomly generated parameters, including the positions and initial size of cracks, properties of the materials and fatigue crack growth rate. The failure modes, the probabilistic model of multiple surface cracking are further studied in Zhu et al.¹³ The remaining usage lives from initial cracks to critical cracks can be estimated from the probabilistic model.

The authors use the numerical approach to simulate crack growth for railway axles. The finite element model of a non-defective railway axle is built with the FE software Abaqus. The tool Zencrack for fracture mechanics simulations is used to simulate the progress of crack growth for a randomly generated initial crack. The progress of crack growth, including the size and the depth of the crack, stress and strain values, SIFs, J-integral values, Ct-integral values and T-stress values, can be read and evaluated from the simulation results (see Figure 2). Since the numerical analysis for the prediction of crack growth is very time-consuming, a method for predicting crack growth in railway axles with artificial neural network (ANN) is developed. The initial crack depth and crack growth can be predicted through the trained ANN, which coincides with the simulated FE analysis results.

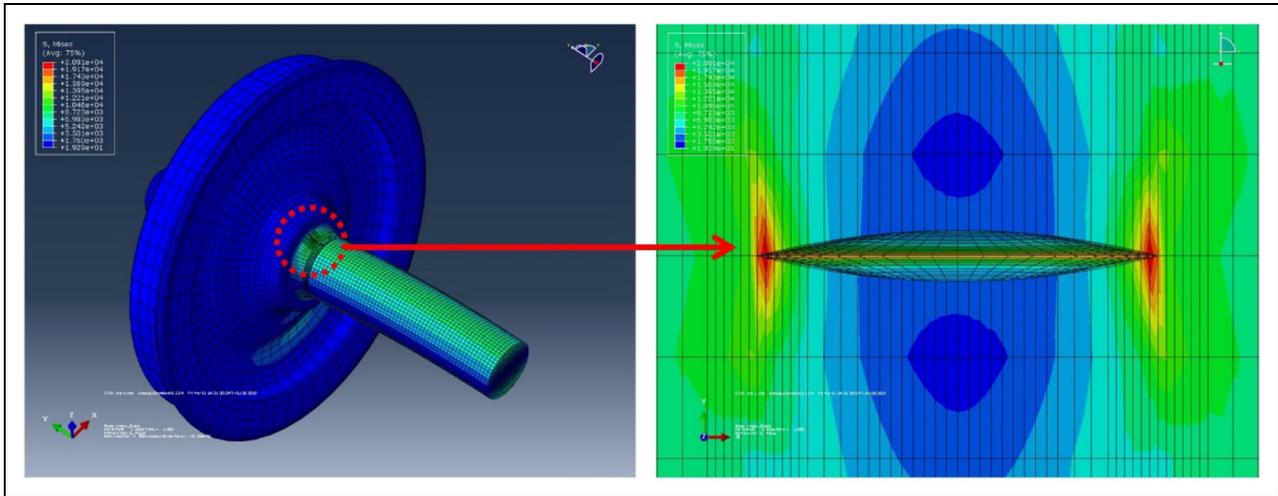


Figure 2. FEM analysis of crack growth in a railway axle.

Both the analytically based approach and numerical analysis require an accurate model to represent crack growth. The model should be calibrated by comparing the results from the model and the measured value from reality. The available measured data provide an option for the data-driven approach. With the data-driven approach, the measured data can be used directly to infer the behaviour of the system without having to build an analytical or simulation model. Due to the rapid development of machine learning and large data in recent years, the data-driven approach has been successfully applied in many applications. In the study by Cunha et al.,¹⁴ the methods for applying the back-mapping approach, the adaptive Gaussian method and the unscented filter method are compared with analytically based approaches and Monte Carlo methods. The results show that these data-driven approaches can achieve higher accuracy than the FORM and lead to a considerably smaller computation time than simulation-based approaches.

Essentially, there is a very strong assumption behind the analytical and numerical approaches: the existence and the mechanism of crack growth should be known before using probability analysis. If the eventuality of crack growth is not based only on assumptions, the problem can be overcome through a data-driven approach. Since prior knowledge of the system is no longer conceived as fixed, the algorithm is capable of learning and dramatically evolving. If the amount of data is limited, the data-driven approach can still draw conclusions from the available data. With increasing amounts of collected data, the results of the analysis can be continuously improved. Indeed, this is also the key idea behind Bayesian inference, which has been successfully used to solve stochastic problems,

including those addressed by probabilistic analysis and uncertainty analysis.

Inspired by the idea that increasingly more data can improve analysis, a data-driven approach with a GP is proposed in this work. With the model, the workflow for calculating the PoF can be divided into a training process and a prediction process. In section ‘Prediction of the PoF with a GP model’, an introduction to the regression model using a GP to predict the PoF is given. The procedure to train a GP model is described in section ‘Training a GP model’.

Probabilistic analysis using a GP

Prediction of the PoF with a GP model

In applying probabilistic analysis to crack growth, the available information is a set of measured data which consists of the measurement time $\mathbf{t} = \{t_i | i = 1, 2, \dots, n\}$ and the crack depth $\mathbf{a} = \{a_i | i = 1, 2, \dots, n\}$. With the given information, the objective of probabilistic analysis is to determine the distribution p of the crack depth at any time point. Hence, the PoF can be determined from p . However, it is difficult to calculate the underlying distribution p directly: either a closed mathematical form is not available, or the computational complexity is very high.¹⁵ Therefore, a regression model can be applied to learn from the measured data and predict the distribution of the crack depth.

With the regression model, distribution q is used to approximate the true distribution p . This approximated distribution q should be easily built with a known form. The GP model is a powerful tool that has been successfully applied in machine learning for regression and

classification.¹⁶ In this work, the tasks of crack growth analysis in the GP model involve training the model and predicting the approximated distribution q at any given time. The inputs of the GP model are the measured crack depth and measurement time, without having to model the internal variable including the load spectra and SIFs. The output of the prediction model is the distribution of the crack depth for a given time point t' .

A GP model is defined by a mean function $m(t)$ and a covariance function $k(t, t')$ with the parameter θ . In section ‘Training a GP model’, the definition of these functions and the process to fit the parameters will be explained in detail. A fitted GP model with trained parameters for mean and covariance functions generates random data such that any finite subset of the generated data follows a multivariate Gaussian distribution. The model creates the possibility of producing the distribution of the crack depth a' for any new time points t' with the given measured data a and t as

$$\begin{bmatrix} a \\ a' \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mu \\ \mu' \end{bmatrix}, \begin{bmatrix} \Sigma_{tt} & \Sigma_{tt'} \\ \Sigma_{t't} & \Sigma_{t't'} \end{bmatrix} \right) \quad (4)$$

with

$$\begin{aligned} \mu &= m(t), \mu' = m(t') \\ \Sigma_{tt} &= k(t, t), \Sigma_{t't'} = k(t', t') \\ \Sigma_{tt'} &= k(t, t'), \Sigma_{t't} = k(t', t) \end{aligned}$$

As defined in equation (4), the time points and the crack depth at each time point are the correlated variables in a GP. The vector of the mean and the covariance matrix in equation (4) can be calculated from the mean function and the kernel function, respectively. In addition, the noise in the measured data will also be considered; the noise will be added as noise variance σ_ε^2 to the covariance Σ_{tt} and $\Sigma_{t't'}$. The value of the noise variance will also be trained in the GP model.

To predict the distribution of the crack depth a' for new time points t' , the approximated posterior distribution $q(a'|a, t, t')$ can be calculated as

$$q(a'|a, t, t') \sim \mathcal{N}(\mu_{t'|t}, \Sigma_{t'|t}) \quad (5)$$

with

$$\begin{aligned} \mu_{t'|t} &= \mu' + \Sigma_{t't} \Sigma_{tt}^{-1} (a - \mu) \\ \Sigma_{t'|t} &= \Sigma_{t't'} - \Sigma_{t't} \Sigma_{tt}^{-1} \Sigma_{t't} \end{aligned}$$

From the posterior distribution $q(a'|a, t, t')$, the mean $\mu_{t'|t}$ and the standard deviation $\Sigma_{t'|t}$ at t' will be derived. A Gaussian distribution with the mean $\mu_{t'|t}$ and the standard deviation $\Sigma_{t'|t}$ will be constructed.

The PoF is then calculated from its cumulative distribution function (CDF) for the given critical depth. Hence, the distribution of the crack depth and the PoF at a certain time point can be predicted by the GP regression model.

Training a GP model

As described in section ‘Prediction of the PoF with a GP model’, a GP is defined by a mean function and a covariance function with the parameter θ . The task of training a GP model involves identifying suitable parameters for these functions. The mean function can be set as the mean value of the measured data a . It is also very common to set the mean function to 0 since the covariance function is able to capture and model the characteristics of the GP model.

For a probabilistic analysis of crack growth, the covariance is defined by a kernel function $k(t, t')$ between two time points t and t' . The kernel function describes the correlation between any two points. The key principle of setting a kernel function is that if t and t' are very close, the expected output distribution of the function should also be similar. The commonly used kernel functions for GPs include an exponential quadratic function, a rational quadratic function, a local periodic function, a linear function and combing kernels. In the study by Rasmussen and Williams¹⁶ and Murphy,¹⁷ the covariance and design of a kernel function are explained in detail. In this work, the exponential quadratic kernel is used

$$k(t, t') = \sigma^2 \cdot e^{-\frac{(t-t')^2}{2\ell^2}} \quad (6)$$

In equation (6), the value $t - t'$ represents the difference between two points. If the value is close to zero, this condition will lead to a maximum value of the kernel as σ^2 , thereby indicating that the two points are highly correlated. If the value $t - t'$ is very large, the correlation of the two points can be neglected with $k(t, t') \approx 0$.

The exponential quadratic kernel function is stationary; this stationary state only depends on the relative position of two time points t and t' . With the increase in time, the mean value and the variance of the crack depth will also increase. To represent the dependency of the absolute position, an additional scale factor for the variance is introduced to multiply the mean value of the crack depth.

The parameters σ and ℓ , the scale factor of the variance and the noise variance σ_ε (see section ‘Prediction of the PoF with a GP model’) will be learned during the training process. The general term θ is used to refer to all the trainable parameters. During a training process, the given measured data will be batched as \tilde{t} and \tilde{a}

in many iterations. To derive the approximated distribution q for the batched data $\tilde{\mathbf{r}}$ and $\tilde{\mathbf{a}}$, the parameters θ should be trained to maximise the marginal likelihood $q(\tilde{\mathbf{a}}|\tilde{\mathbf{r}}, \theta)$

$$\theta^* = \operatorname{argmax}_{\theta} q(\tilde{\mathbf{a}}|\tilde{\mathbf{r}}, \theta) = \operatorname{argmin}_{\theta} [-\log(q(\tilde{\mathbf{a}}|\tilde{\mathbf{r}}, \theta))] \quad (7)$$

$$\begin{aligned} \log(q(\tilde{\mathbf{a}}|\tilde{\mathbf{r}}, \theta)) &= -\frac{1}{2}(\tilde{\mathbf{a}} - \mu_{\theta})^{\top} \Sigma_{\theta}^{-1}(\tilde{\mathbf{a}} - \mu_{\theta}) \\ &\quad -\frac{1}{2} \log|\Sigma_{\theta}| - \frac{d}{2} \log 2\pi \end{aligned} \quad (8)$$

In equation (7), the optimal parameters θ^* are derived by minimising the negative log-likelihood loss. This minimisation is equivalent to maximising the marginal likelihood since the log function is a monotone transformation. Minimising the negative log-likelihood enables a product calculation to be converted into a summation calculation. The potential for underflow can also be avoided. The negative log-likelihood is calculated in equation (8); the negative log-likelihood can be minimised through a gradient-based approach.

Many software packages implement a GP model for training and regression. In section ‘Case study: regression with probabilistic layers’, a case study of the regression of the crack growth process and a comparison of the GP model with the analytical approach will be presented.

Case study: regression with probabilistic layers

Data preparation and calculation of the PoF based on the FORM/SORM

Before starting the inference by the GP model, the data for training and validation should be prepared. Theoretically, the ‘true’ value of the PoF and the distribution of the crack depth are unknowable. To test and validate the results obtained with the GP model, a simplified analytical model is built. With the analytical model, the PoF can be calculated in advance by the FORM and SORM (see section ‘Analytical, numerical and data-driven approaches for the probabilistic analysis of crack growth’). Then, a set of data for the crack depth at different time points can be randomly generated from the analytical model. These data will be used as if they were measured in reality. In this section, the term ‘measured data’ refers to the data generated from the analytical model. The PoF estimated from the GP model will be compared with the analytical results, which are calculated with the FORM and SORM.

The log-log plot of the real crack growth rate is illustrated in Figure 3(a). Some algorithms, for example, the NASGRO equation,⁸ are used to model the relation between the crack growth rate and the SIF range. As

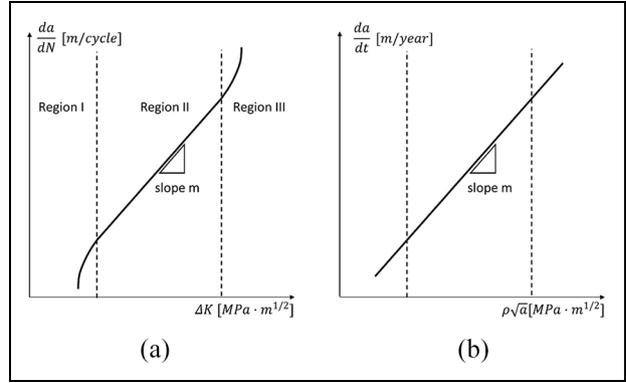


Figure 3. Illustration of the simplified closed-form solution in a log-log plot: (a) real crack growth rate and (b) simplified crack growth rate.

described in section ‘Prediction of the PoF with a GP model’, the inputs of the GP model are the measured crack depth and measurement time, and it is not required to know the information about the SIF range. To compare the results between the analytical approach and data-driven approach, reliability analysis with the FORM and SORM should be carried out with the same input information. Therefore, a closed-form solution for the crack depth depending on time is required. However, it is not possible to have such a closed mathematical formula for realistic crack growth. Hence, the crack growth rate is calculated through a simplified model. For testing and verification, Paris’ law is rewritten in a simplified way, as shown in equation (9). The solution is presented in equation (10), which is the solved form of the differential equation given in equation (9). The simplified linear function on the log-log plot is shown in Figure 3(b). The intension of this simplified model is not to calculate the real crack growth rate. Instead, it is used to provide the FORM and SORM in a closed form for reliability analysis. Meanwhile, the GP model can estimate the PoF with the data generated by the same model

$$\frac{da}{dt} = C \cdot (\rho\sqrt{a})^m \quad (9)$$

$$a = \left[C \cdot \frac{(2-m)}{2} \cdot \rho^m \cdot t + a_0^{(2-m)/2} \right]^{2/(2-m)} \quad (10)$$

The critical depth of the crack is defined as $a_{critical} = 0.05$ m. Random variables will be introduced to imitate the stochastic behaviour of crack growth. The parameters ρ , C and a_0 are random values, following a normal distribution with the given mean and standard deviation. According to British Standards Institution¹⁸ and Beretta and Carboni,¹⁹ the parameter m is deterministic in most reliability analyses. The value of m and the mean value of C are taken from the

Table 1. The PoF results with the FORM, SORM and GP model.

Year	PoF with the SORM	PoF with the FORM	PoF with the GP regression model
1	5.565e-53	2.151e-06	1.110e-16
2	1.216e-23	2.150e-06	7.506e-11
3	3.102e-13	2.150e-06	2.450e-09
4	3.660e-08	2.150e-06	1.147e-07
5 (>10 ⁻⁵ !)	2.519e-05	2.716e-05	4.136e-05
6	1.285e-03	1.364e-03	1.974e-03
7	1.508e-02	1.577e-02	1.444e-02
8	7.267e-02	7.499e-02	6.826e-02
9	1.995e-01	2.033e-01	2.461e-01
10	3.802e-01	3.829e-01	4.338e-01
Average error	–	0.096%	1.059%

SORM: second-order reliability method; FORM: first-order reliability method; GP: Gaussian process.

Note: The bold value indicates that at this timepoint the PoF exceeded the acceptable threshold (1e-5) for the first time.

material of carbon steel S38C, which is used for the railway axles of a type of CRH train produced by CRRC Qingdao Sifang Co., Ltd. Since the standard deviation of C is not available, it is assumed to be 10% of the mean value according to the value provided in WIDEM.⁴

The greatest challenge is to determine the distribution of the stress and geometry factor in the simplified model. In a fatigue crack growth model, the stress will be either modelled by the load spectra or from a known distribution of stress amplitudes. However, the real load spectra and stress distribution cannot be directly used in the simplified model. In region I shown in Figure 3(a), the crack growth rate increases very slowly with a small ΔK near the threshold. The crack growth will be too fast if the simplified linear relationship in the log-log plot is applied to region I. The SORM and FORM cannot deal with a complicated piecewise function for the crack growth rate. In addition, the value Y is highly dependent on the geometry of the crack. The differential equation is difficult to solve with a purely analytical approach. Usually, numerical analysis should be applied. For this reason, a simplified parameter ρ combined with the stress and geometry factor is introduced. The distribution of ρ is chosen to generate a set of ‘fictive’ PoFs through the FORM and SORM, which fit the empirical data in service. Depending on the operating programmes with different operators, the range of the number of rotation cycles per year for the investigated axle can be from 4.6×10^7 (approximately 120,000 km/year) to 1.5×10^8 (approximately 400,000 km/year). Taking the lower value of the range (4.6×10^7 cycles/year), the PoF will reach 10^{-5} after 5 years of service, which corresponds to the PoF calculated in Table D 7.1.8.9 in the study by Maierhofer et al.⁹

The difficulties in modelling crack growth for reliability analysis also show the advantages of the GP

model. With traditional data fitting, many variables (e.g. the load spectra and threshold SIF range) and the assumptions about their distributions should be known in advance. With the developed GP model, only the measured crack depth and measurement time are required. The developed GP model can estimate the PoF without building an underlying analytical or simulation model. The results can be used to validate other analytical and simulation models. Due to the different behaviours of crack growth, many regression models need to separately fit the data for the near-threshold region and linear region.¹⁶ The regression of the GP model can be consistently performed within one GP.

Using the software OpenTURNS,²⁰ the PoF is separately calculated with the FORM and SORM (see Table 1 in section ‘Regression with probabilistic layers in TensorFlow probability’). Since the results calculated from the SORM are more accurate than those calculated from the FORM, the SORM results are used as a benchmark. Meanwhile, 400 pairs of measured data $y = \{(t_i, a_i) | i = 1, 2, \dots, 400\}$ are randomly generated by the random parameters ρ C and a_0 (see Figure 5 in section ‘Regression with probabilistic layers in TensorFlow probability’). These data will be used in the GP model for training and regression.

Regression with probabilistic layers in TensorFlow probability

In this work, the GP model is implemented with TensorFlow probability (TFP), a library for probabilistic reasoning and statistical analysis. TFP enables the integration of a probabilistic model with deep learning on high-performance hardware (tensor processing unit (TPU), graphics processing unit (GPU)).²¹ In TFP, various distributions are provided as low-level building blocks, and different probability layers, an inference algorithm, and optimisers are implemented as high-

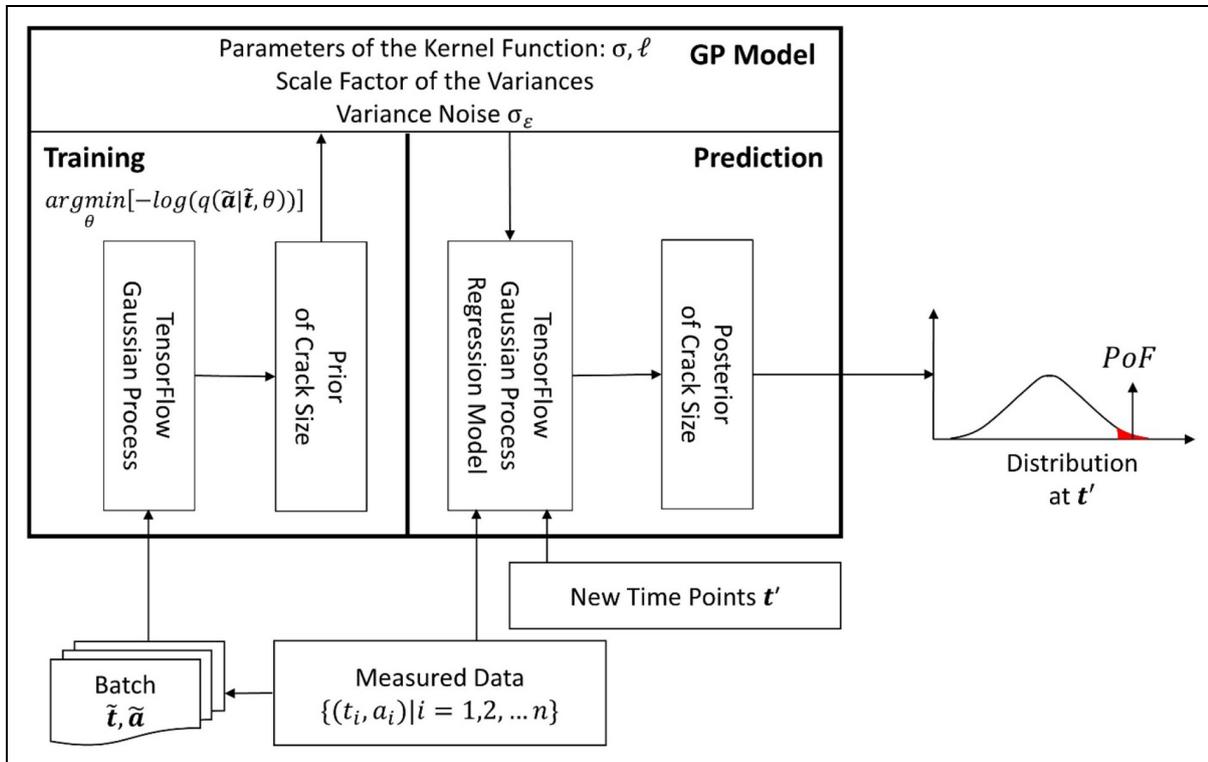


Figure 4. The structure of the GP model with TensorFlow probability.

level constructs. The structure of the GP model for probabilistic regression and reliability analysis is shown in Figure 4.

The GP model is trained through a predefined ‘GP’ class, which is an encapsulated high-level application programming interface (API). The model accepts the batched training data as input and outputs the inferred marginal distribution of the GP as the prior. Before training, the parameters of the kernel function, the scale factor of the variance and the noise variance are initialised. With the fed batched data, the parameters of the model are iteratively trained to minimise the negative log-likelihood (see equations (7) and (8)).

The distribution of the crack depth for new time points can be predicted through the class ‘GP Regression Model’, which outputs a posterior from the given measured data, taking the trained parameters in the GP model as the prior. The failure of the probability will be calculated by calling the CDF on the output distributions of the crack depth, with the condition that the crack depth is larger than the critical depth.

In Figure 5, the predicted posterior probability distribution from the GP regression model is shown. The blue points are the ‘measured’ data randomly generated by equation (10) (see section ‘Data preparation and calculation of the PoF based on the FORM/SORM’) for time t . Another set of 400 randomly generated time points t' is used to predict the posterior at these time points. The red line represents the predicted mean of

the crack depth, and the light grey area is the interval of $\pm 0.1\sigma$. The mean and the variance of the crack depth also increase over time.

The PoF can be further predicted from year 1 to year 10. The time $t' = \{1, 2, \dots, 10\}$ is given to the GP regression model to predict the posterior of $a' = \{a'_i | i = 1, 2, \dots, 10\}$. The resulting distribution is used to calculate the CDF for $a'_i \geq a_{critical}$.

In this work, the results calculated by the SORM are used as an analytical benchmark. In Figure 6, the predicted PoF (the red points) is compared with the results of the SORM (the blue line). The predicted PoF can approximate the trend of the SORM results. It is difficult to see the difference before year 7 due to the very small value of the PoF. The results are further compared with the log-PoF in Figure 7 between the prediction results (the red line), the FORM results (the orange line) and the SORM results (the blue line). From year 1 to year 4, the prediction results based on the GP model is closer to the results of the SORM than FORM.

The detailed results of the PoF with the GP regression model and analytical methods are presented in Table 1. Both results indicate that an inspection should be planned at the end of the fourth year, since the PoF will exceed the threshold of the PoF (10^{-5}) at the fifth year. Compared to the PoF calculated by the SORM, the average error of the PoF with the GP regression model is 1.059%, and the average error of the PoF with the FORM is 0.096%.

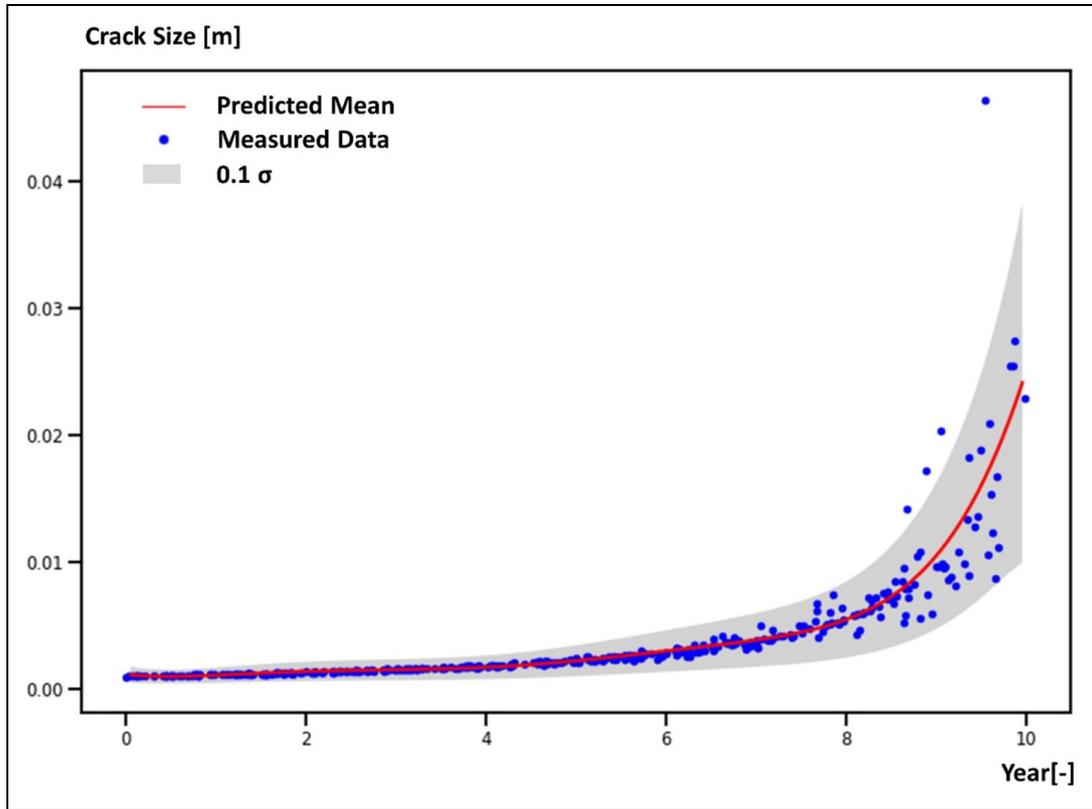


Figure 5. Prediction of the posterior from the GP regression model.

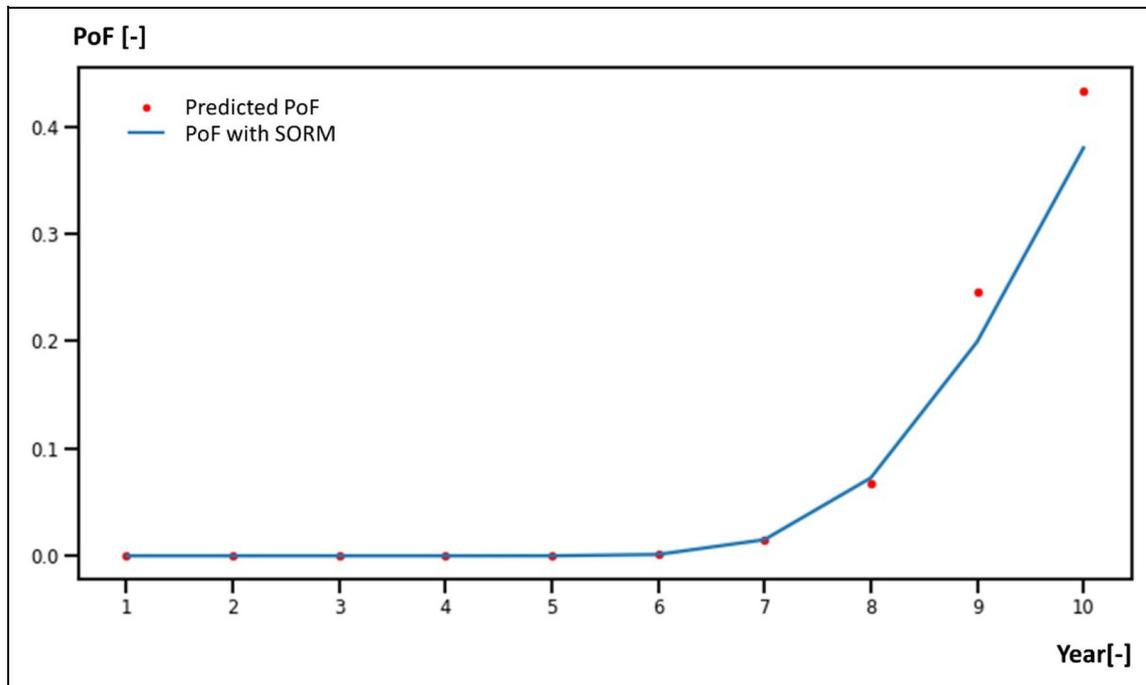


Figure 6. Predicted PoF compared to the results from the SORM.

To verify the developed GP model, it is very common to compare the predicted PoF with analytical results. In this work, the analytic results are calculated

from a simplified model that is easy to solve. If the prediction from the GP model coincides with the known results, the model and the results can be interpreted as

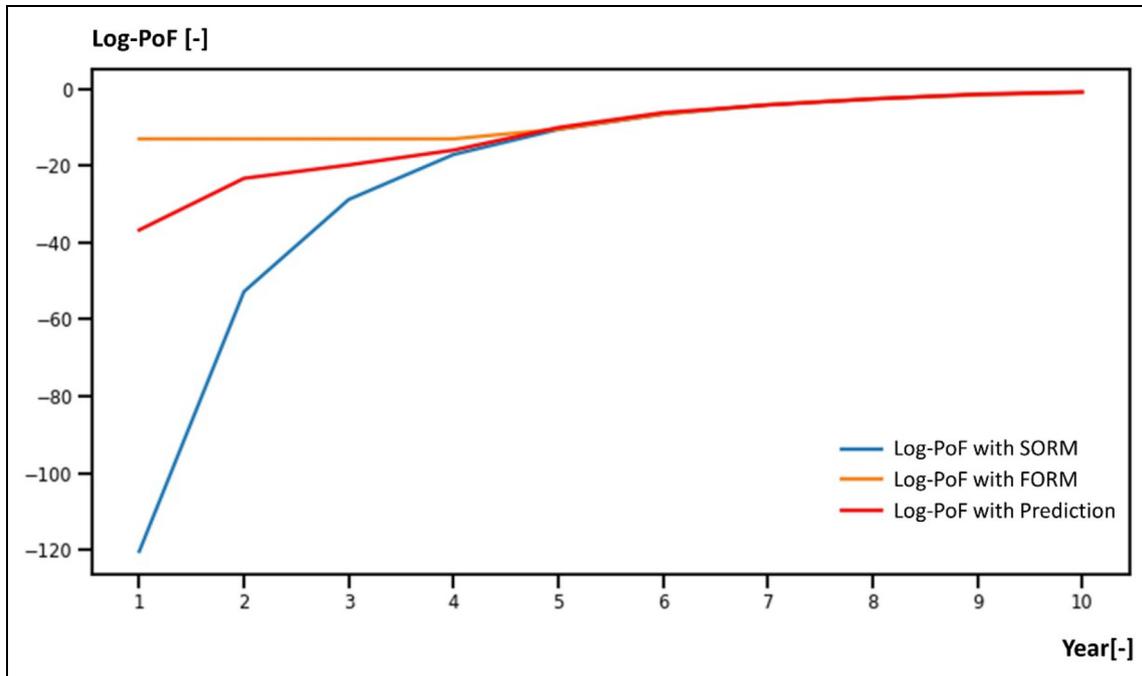


Figure 7. Comparison of the log-PoF between the prediction, SORM and FORM results.

‘correct’. However, there is no clear definition of the credibility of the results. For example, the compared results in Table 1 show a large divergence in the magnitude of the PoF during the first 4 years. In this work, the timing that reaches the threshold of the PoF (10^{-5}) is applied to justify the results, since it is critical for scheduling maintenance activities. For this reason, the results of the GP model are considered credible.

The deviation between the GP regression model and the analytical results becomes significant after year 8. The deviations are caused by the high variance and insufficient information provided by the measured data. This issue should be deeply investigated for developing a whole maintenance plan after the first inspection (e.g. after the fifth year in this case study).

Conclusion and discussions

Findings in this work

In this work, a data-driven approach based on a GP for regression to determine the PoF caused by crack growth in railway axles is developed. For complicated failure modes, it is difficult or even impossible to build a reliable analytical or simulation model before using an analytical approach. The main purpose of this work is to develop an algorithm to infer the distribution of the crack growth from a limited amount of measured data without having to build an underlying model.

After the GP model is trained with the measured data, the distribution of the crack depth can be

predicted for any time point. The PoF will be calculated accordingly. Hence, the necessary maintenance activity can be scheduled to ensure that the PoF does not exceed a predefined threshold. The results of the case study show that the determined timing for the first inspection, as well as the PoF, coincides with the known results derived by analytically based approaches. The problems associated with modelling and calibration can be overcome by the data-driven approach. The developed GP model can also serve as a complementary instrument to validate other analytically based approaches or numerical analyses.

Perspectives for further research

The fracture properties of the railway axle are dominantly determined by the crack growth rate and load spectra. Region II (see Figure 3) was already confirmed to fit well through a comparison with the GP model. However, region I and region III should be deeply investigated. Especially, in the early stages of crack growth, it is necessary to determine how the errors in region I affect the predicted results. A 1:1 test bench to the test crack growth with bending and rotating stress was built in CRRC Sifang. Experiments for small-scale specimens and full-scale axles are planned. In the study by Beretta and Carboni,²² the load interactions under constant and variable amplitudes for different scales are compared. This shows that the effect of the load interaction for full-scale axles is not negligible. The load spectrum should be fully considered in the prediction model.

The effect of parameters on the prediction results needs to be checked. A comparison of the contribution of the parameters on the prediction results will help to understand which parameters are essential for improving the prediction results. The parameters that should be investigated include those for the crack growth model (e.g. the empirical parameters C and m , the SIF, the stress ratio and the geometry factor), the linear material properties (e.g. the elastic modulus and Poisson's ratio), the residual stresses, the numerical errors and the dimensional tolerances.

As the development and application of the GP are still very active, the issues of tuning the hyper-parameters and model configuration should be addressed. Current development of the GP model is expected so that highly reliable and stable results can be delivered with less manual effort. The accuracy of the model should be further improved, especially for time points with a high variance. Meanwhile, a deep investigation related to the challenge of determining how much available data should be used is required. On one hand, less data will limit the generalisation of the inferred results; on the other hand, more data collected from actual situations increases the computation complexity. In recent years, the variational Gaussian process (VGP) was developed.²³ It has been proven that the computational complexity of the algorithm can be up to $\mathcal{O}(N^3)$. In the study by Titsias,²⁴ a method of using a so-called inducing (pseudo) point instead of the full measured data was developed to reduce the computational effort. This method is also integrated into the optimiser of the TensorFlow VGP layer. The optimiser can find the optimum locations of the inducing points to minimise the negative evidence lower bound (ELBO).¹⁵ Although the method of inducing points can balance the performance and accuracy, the issue of setting hyper-parameters (e.g. the number of inducing points) should still be investigated and justified. Since the ELBO can only reach a local optimum, the hyper-parameters in the model should be tuned through several rounds of searches. In addition, the choice of the kernel is still questionable. A simplified model with fewer configuration requirements for TFP is expected in the future.

Further research and development will involve developing a maintenance plan for all rounds of inspections in the whole service life. In this work, the probability of crack initiation and the threshold SIF range are not considered with the simplified analytical model. The mechanism of generating an initial crack, the probability of detecting a crack and other stochastic aspects, including the stress and loading during operations, will be studied. The data-driven approach will be examined and validated by other analytically based approaches and numerical analysis. Furthermore, the inference method can be applied to analyse the daily collected

data for online detection and diagnosis so that condition-based maintenance for railway axles can be achieved.

Acknowledgements

During this work, Ullrich Martin organised the cooperation between the project partners. Fabian Hantsch provided advice for solving the differential equations for crack growth. The FE model for analysing crack growth in a railway axle was created by Euiyoul Kim. Laura Martinez summarised the literature in the field. This work is supported by Universität Stuttgart, Chinesisch-Deutsches Forschungs- und Entwicklungszentrum für Bahn- und Verkehrstechnik Stuttgart and Hefei University.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This work was financially supported by the CRRC Qingdao Sifang Co., Ltd. (grant number SF/GY-Liang-2017-253).

ORCID iD

Yong Cui  <https://orcid.org/0000-0003-0330-1637>

References

- Zerbst U, Beretta S, Köhler G, et al. Safe life and damage tolerance aspects of railway axles – a review. *Eng Fract Mech* 2013; 98: 214–271.
- EBFW. Project: Eisenbahnfahrwerke (EBFW), 2013, www.ebfw3.net (accessed 21 July 2019).
- Traupe M, Meinen H and Zenner H. Sichere und wirtschaftliche Auslegung von Eisenbahnfahrwerken. *Abschlussbericht Bmbf-projekt* 2004; 19: 0061.
- WIDEM. Project: wheel-set integrated design and effective maintenance, 2007, www.widem.org (accessed 21 July 2019).
- Martinez L, Cui Y and Martin U. Restlebensdauer rissgeschädigter Radsatzwellen. *ETR- Eisenbahntechnische Rundschau* 2019; 6: 54–60.
- Paris P and Erdogan F. A critical analysis of crack propagation laws. *J Basic Eng* 1963; 85: 528–533.
- Zhu SP, Huang HZ, Peng W, et al. Probabilistic physics of failure-based framework for fatigue life prediction of aircraft gas turbine discs under uncertainty. *Reliab Eng Syst Safe* 2016; 146: 1–12.
- Madsen HO, Krenk S and Lind NC. *Methods of structural safety*. New York: Courier Corporation, 2006.
- Maierhofer J, Pippan R and Gänser HP. Modified NAS-GRO equation for physically short cracks. *Int J Fatigue* 2014; 59: 200–207.

10. Wei L. D7.1.7: procedure to define NDT periodicity as function of vehicle service profile; D7.1.8: results of probability analysis: probabilistic model. In: *European project 'Wheelset integrated design and effective maintenance' (WIDEM)*, 2018, pp. 62–69.
11. Gänser HP, Kunter KH, Deisl A, et al. Forschungsprojekt Eisenbahnlaufwerke 3 – Ergebnisse. 45. Tagung, “Moderne Schienenfahrzeuge”, 15.-17.04.2019 TU Graz.
12. Zhu SP, Foletti S and Beretta S. Evaluation of size effect on strain-controlled fatigue behavior of a quench and tempered rotor steel: experimental and numerical study. *Mater Sci Eng: A* 2018; 735: 423–435.
13. Zhu SP, Hao YZ and Liao D. Probabilistic modelling and simulation of multiple surface crack propagation and coalescence. *Appl Math Model* 2020; 78: 383–398.
14. Cunha A, Caetano E and Ribeiro P. Computation of probabilistic crack analysis. In: *Proceedings of the 9th international conference on structural dynamics, EURO-DYN*, Porto, 30 June–2 July 2014.
15. Blei DM, Kucukelbir A and McAuliffe JD. Variational inference: a review for statisticians. *J Am Stat Assoc* 2017; 112: 859–877.
16. Rasmussen CE and Williams CK. *Gaussian processes for machine learning*. Cambridge: MIT Press, 2006.
17. Murphy KP. *Machine learning: a probabilistic perspective*. Cambridge: MIT Press.
18. BS 7910:2013. Guide to methods for assessing the acceptability of flaws in metallic structures.
19. Beretta S and Carboni M. Experiments and stochastic model for propagation lifetime of railway axles. *Eng Fract Mech* 2006; 73: 2627–2641.
20. OpenTURNS. Software documentation: openturns 1.13 Documentation, 2005, [Openturns.github.io/openturns/latest/contents.html](https://openturns.github.io/openturns/latest/contents.html) (accessed 21 July 2019).
21. TensorFlow. Software. TensorFlow probability, 2019, www.tensorflow.org/probability (accessed 21 July 2019).
22. Beretta S and Carboni M. Variable amplitude fatigue crack growth in a mild steel for railway axles: experiments and predictive models. *Eng Fract Mech* 2011; 78: 848–862.
23. Tran D, Ranganath R and Blei DM. The variational Gaussian process, 2015, <https://arxiv.org/abs/1511.06499>
24. Titsias M. Variational learning of inducing variables in sparse Gaussian processes. In: *Proceedings of the Twelfth International Conference on Artificial intelligence and statistics, Proceedings of Machine Learning Research (PMLR)*, Hilton Clearwater Beach Resort, Clearwater Beach, FL, USA, 16–18 April 2009, pp. 567–574.

Appendix

Notation

a	measured data of the crack depth $a = \{a_i i = 1, 2, \dots, n\}$ at measurement time t	\tilde{a}	batched observed crack depth $\{a_i i = 1, 2, \dots, d\}$
a	crack depth, the dependent variable (m)	a_0	initial crack depth, a random value with a mean of $1.00e-3$ and standard deviation of $1.00e-4$ for the case study (m)
		C	empirical random parameter with a mean of $5.48e-15$ and standard deviation of $5.48e-16$ for the case study (–)
		d	number of batched training data
		$k(t, t')$	kernel function to define covariance
		K	stress intensity factor (SIF) ($\text{MPa m}^{1/2}$)
		K_{max}, K_{min}	maximum and minimum value of the SIF ($\text{MPa m}^{1/2}$)
		l	length scale of the kernel
		m	constant parameters with a value of 3.53 for the case study (–)
		$m(t)$	mean function
		N	number of cycles (–)
		$q(\tilde{a} \tilde{l}, \theta)$	marginal likelihood
		t	measurement time $\mathbf{t} = \{t_i i = 1, 2, \dots, n\}$
		t, t'	time point
		\tilde{t}	batched time points $\{t_i i = 1, 2, \dots, d\}$
		Y	geometry factor (–)
		δ	stress (MPa)
		ΔK	range of the SIF ($\text{MPa m}^{1/2}$)
		θ	parameters defined in a GP model
		θ^*	optimal parameters to approximate the distribution p
		μ, μ'	vector of the mean calculated from the mean function
		$\mu_\theta, \Sigma_\theta$	calculated mean vector and covariance matrix from the current value of θ
		ρ	simplified parameter combined with the stress and geometry factor Y (see equation (1)), a random value with a mean of $2.70e + 4$ and standard deviation of $2.35e + 4$ for the case study (MPa)
		σ	amplitude that controls the maximum value of the kernel
		σ_ϵ^2	noise variance for the noises in the measured data
		$\Sigma_u, \Sigma_{t'}, \Sigma_{t''}, \Sigma_{t''t'}$	matrix of covariance calculated from the kernel function