

A Bayesian machine learning approach for online detection of railway wheel defects using track-side monitoring

Yi-Qing Ni^{ID} and Qiu-Hu Zhang

Abstract

Wheel condition assessment is of great significance to ensure the operation safety of trains and metro systems. This study is intended to develop a Bayesian probabilistic method for online and quantitative assessment of railway wheel conditions using track-side strain-monitoring data. The proposed method is a fully data-driven, nonparametric approach without the need of a physical model. To enable defect identification using only response measurement, the measured dynamic strain responses of rail tracks during the passage of trains are processed to elicit the normalized cumulative distribution function values representative of the effect of individual wheels, which in conjunction with the frequency points are used to formulate a probabilistic reference model in terms of sparse Bayesian learning. Through cleverly realizing sparsity by introducing hyper-parameters and their priors, the sparse Bayesian learning makes the resulting model to exempt from overfitting and generalize well on unseen data. Only the monitoring data in healthy state are needed in formulating the reference model. A novel Bayesian null hypothesis significance testing in terms of scale-invariant intrinsic Bayes factor, which does not suffer from the Jeffreys–Lindley paradox, is then pursued in the presence of new monitoring data collected from possibly defective wheel(s) to detect wheel defects and quantitatively assess wheel condition. The proposed method in fully Bayesian inference framework is verified by utilizing the real-world monitoring data acquired by a distributed fiber Bragg grating-based track-side monitoring system and comparing with the offline inspection results.

Keywords

Railway wheels, defect detection, track-side monitoring, sparse Bayesian learning, intrinsic Bayes factor, optical fiber sensors

Introduction

Passenger safety is the highest priority in mass transportation. This is especially true in modern high-speed rails in view of their mass transportation volume and fast speed. If a high-speed train runs in failure, it will result in a disastrous loss of mass lives and rail infrastructure. The current rail operation control systems do not have the functions of online detection of structural health and real-time response to potential structural failure, since in the current practice structural faults or damage are detected offline in depots or maintenance yards at scheduled time intervals. However, structural faults may occur during in-service operation; this issue is especially important for the high-speed trains that are in a high frequency of services. In this regard, development of effective online structural fault diagnosis methods is a core focus for preventing catastrophic failure as well as prolonging the service life of high-speed rails.

Wheel condition assessment is of great importance to ensure railway safety and to reduce the maintenance cost of railway infrastructure. Wheel defects because of wheel out-of-roundness, such as wheel flats, wheel shells, and wheel polygonization, can induce damage to both train and rail track, trimming down safety and ride comfort of in-service trains and increasing operation and maintenance costs for railway system.^{1–3} Early

Hong Kong Branch of Chinese National Engineering Research Center on Rail Transit Electrification and Automation, Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Corresponding author:

Yi-Qing Ni, Hong Kong Branch of Chinese National Engineering Research Center on Rail Transit Electrification and Automation, Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.
 Email: ceyqni@polyu.edu.hk

detection of wheel condition and timely re-profiling or replacement of defective wheels confer great benefits in railway safety and economy. During the past several decades, offline inspection and online monitoring techniques have been proposed for wheel condition assessment. Incipient offline inspection measures wheel profiles at workshop by contact-type measurement devices together with visual inspections.³ The inspection procedure is often costly in time and must be performed on a maintenance schedule. More time-effective ways are using noncontact-type measurement devices such as ultrasonic waves.^{4,5} Online monitoring techniques allow detection to be conducted in real time, aided by various sensing technologies such as piezoelectric accelerators,^{6–9} piezoelectric strain gauges,^{10–12} fiber Bragg grating (FBG) sensors,^{13–15} acoustic emission sensors,¹⁶ and laser sensors.^{17,18} The online monitoring techniques for rail system are categorized into onboard monitoring and track-side monitoring in accordance with the deployment of sensors. Onboard monitoring installs sensors on in-service trains, that is, particularly useful to monitor the deterioration of rail infrastructure such as rail track defects,^{19–21} rail fasteners,^{22,23} and rail irregularities.²⁴ The track-side monitoring^{6,7,10–18} installs sensors on tracks or surrounding areas, with the purpose of detecting the condition of in-service trains including wheel qualities.

Apart from various monitoring techniques, diagnosis and prognosis algorithms are at the core of research to realize precise wheel condition assessment. Belotti et al.⁷ explored high-frequency wavelet coefficient maxima from vertical accelerations of rails as an indicator of wheel flats. Jia and Dhanasekar⁸ investigated local wavelet energy average from vertical accelerations of bogies for identifying wheel flats. Based on wheel impact load detector, Stratman et al.¹⁰ proposed the maximum dynamic ratio (MDR) for wheel condition assessment. The MDR is defined as the ratio of maximum dynamic impact loads to static axle loads, and its value of 3 has been adopted as a threshold for flawed wheels in European Union.¹² In line with an FBG-based track-side monitoring system, Wei et al.¹³ defined a wheel condition index that is linearly proportional to the averaged strain alterations of rail bending but inversely proportional to train speed. Filograno et al.¹⁴ suggested that a 70% increase in strain energy of rail bending with respect to noise levels is a good measure of significant wheel defects. However, the aforementioned deterministic diagnostic algorithms are incapable of dealing with the uncertainties resulting from measurement of noise or error and randomness in wheel–rail interactions. Statistical models are deemed to achieve more reliable and persuasive diagnostic results.²⁵

Statistical approaches have recently been developed for wheel condition assessment. Skarlatos et al.⁶ attempted to establish a fuzzy-logic model for undamaged wheels by correlating maximum acceleration amplitudes with nominal train speeds and 1/3-octave bands from 80 Hz to 5 kHz. The statistical hypothesis test was conducted to investigate the probability of existence of wheel flats and the damage extents. Krummenacher et al.¹² proposed two automatic detection algorithms for wheel defects by means of machine learning methods. One algorithm employed support vector machine to learn and classify wavelet features extracted from the wheel impact load monitoring data, and the other automatically learned the original monitoring data and classified wheel conditions using deep artificial neural networks. The proposed methods could achieve at least 10% improvement in identifying wheel defects in comparison with the MDR, but defect extents were not identifiable. Liu and Ni¹⁵ assumed a Gaussian distribution for the normalized rail bending strain and employed the Chauvenet's criterion to signal wheel defects. Zhang et al.⁹ proposed a Bayesian dynamic linear approach for modeling ride quality evolution due to deteriorating wheel qualities and for probabilistic assessments of wheel condition with the use of onboard monitoring data of acceleration acquired from the running train.

This article aims to develop a Bayesian machine learning approach for online wheel condition detection using the track-side strain-monitoring data. The proposed method features the following merits: (a) it uses only the dynamic strain responses of tracks collected during the passage of trains; (b) it is fully data-driven and requires only the monitoring data collected in healthy state in formulating the reference model; (c) by means of sparse Bayesian learning (SBL), the built probabilistic reference model exempts from overfitting and bears favorable generalization ability due to sparsity embedded by SBL; and (d) the proposed method accounts for uncertainties arising from measurement and modeling errors. Toward the above, the Fourier amplitude spectra (FASs) of the measured track dynamic strain responses in healthy state are obtained to elicit normalized cumulative density functions (CDFs) that characterize the patterns of healthy wheels. The CDFs together with the corresponding frequency points are then used as outputs (response variables) and inputs (explanatory variables) to train a probabilistic reference model by means of SBL.^{26,27} There exist various sources of uncertainties in the measured track dynamic strain response data such as measurement noise and variability in the stochastic wheel–rail dynamics. The SBL allows the uncertainties arising from measurement and modeling errors to be accounted for in the model formulation. More

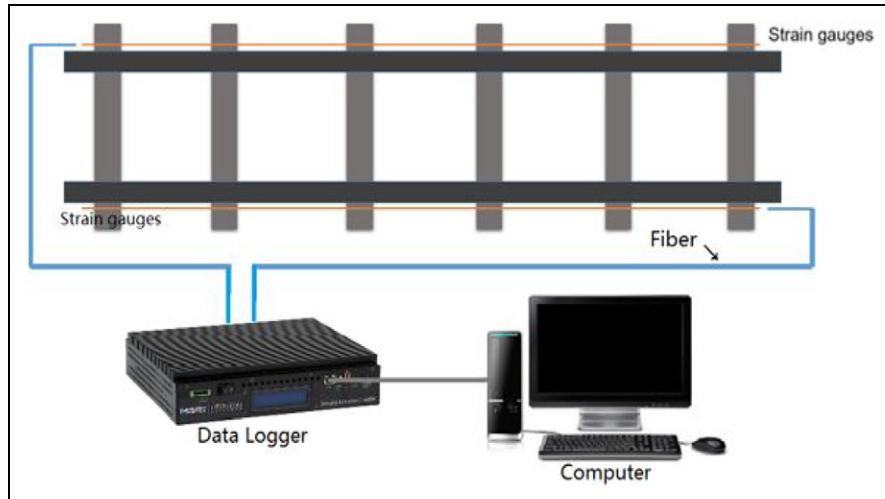


Figure 1. Track-side monitoring system using distributed FBG strain sensors.

importantly, through introducing hyper-parameters and sparsity-inducing priors, the SBL elicits a probabilistic regression model exempting from overfitting in terms of highly sparse representation. When new monitoring data on track dynamic strain responses coming from the effect of a possibly defective wheel are made available, the discrimination between the new measurements and the model predictions is evaluated in terms of an intrinsic Bayes factor (IBF) for defect detection and quantification. The IBF is derived through Bayesian null hypothesis significance testing (BNHST) which does not suffer from the Jeffreys–Lindley paradox.

The remainder of this article is organized as follows. Section “Feature extraction through track-side monitoring” describes the feature extraction from raw measurement data acquired by an FBG-based track-side monitoring system. Section “Model formulation by SBL” describes the formulation of a probabilistic reference model by means of SBL. Section “Bayesian hypothesis testing for wheel defect detection” delineates the detection of wheel defects and quantitative assessment of wheel condition by BNHST and scale-invariant IBF. Finally, section “Conclusion” gives the conclusions drawn.

Feature extraction through track-side monitoring

Track-side monitoring system

As illustrated in Figure 1, the FBG-based track-side monitoring system for this study consists of two arrays of FBG strain sensors deployed on two parallel tracks of a rail segment, two optical cables, a high-speed

optical interrogator, and a desktop or notebook computer. To facilitate the detection of minor wheel defects, the sensors are densely deployed along rail length but the instrumentation needs just to cover a range of rail slightly longer than the wheel perimeter. The FBG sensors are connected through optical cables to a high-speed optical interrogator which is controlled by a computer for data acquisition and processing. Both the interrogator and the computer can be located far away from the instrumented rail to facilitate remote monitoring.

Figure 2 shows the deployment of the track-side monitoring system on an in-service rail. In this implementation, each sensor array comprises 21 FBG gauges evenly spaced at 0.15 m intervals on rail foot of each single track, and the total instrumentation range reaches 3.0 m to enable the sensing of rolling action of the whole circumference of the wheel tread (the diameter of wheel is 0.92 m). Both the optical interrogator and the computer are operated in an auxiliary office which is about 120 m away from the monitoring area. The FBG sensors have been calibrated in laboratory before their installation to obtain strain sensitivity and temperature sensitivity, but in general the temperature compensation is not necessary as the time for online monitoring of the passage of a whole train just lasts for a few seconds, during which the environmental temperature does not change dramatically. The wheel-rail rolling friction may cause heating and cooling effects on the top surface of the rail, such that, the rail temperature varies to some extent, but such effects mainly influence the area of rail head and are less significant in rail foot.^{28,29} When both FBG strain gauges and temperature sensors are deployed, it is easy to compensate the effect of varying temperature by subtracting from

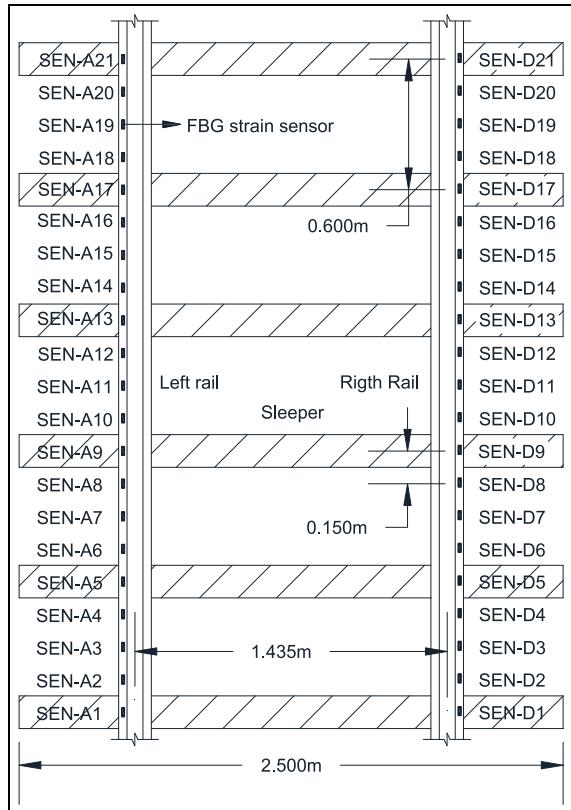


Figure 2. Deployment of FBG sensors.

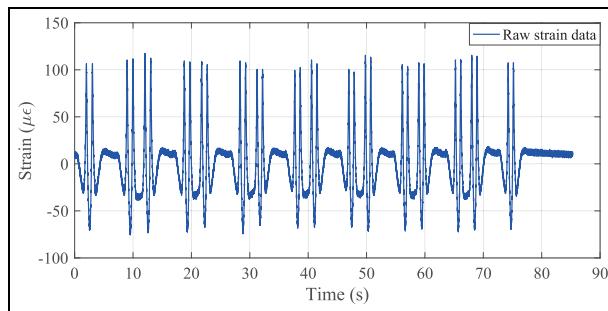


Figure 3. Rail foot strain recorded by SEN-D2.

the measured strain a thermal-induced ingredient which is equal to a constant coefficient times the difference between the measured instant temperature and the recorded temperature at the installation of the sensor.³⁰

Figure 3 illustrates the time history of longitudinal (bending) strain at rail foot monitored by the FBG sensor SEN-D2 (refer to Figure 2) when a typical eight-car passenger train passes through the instrumented rail at a nominal speed of 10 km/h. The sampling frequency f_s is set to be 5 kHz to ensure that the effect of minor wheel tread defect can be sensed. It is observed that the rail bending strain in the longitudinal direction varies

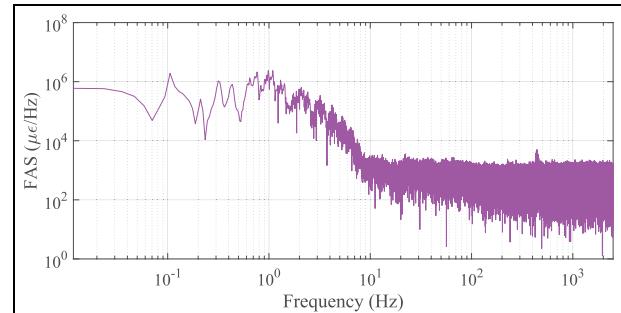


Figure 4. FAS of rail foot strain recorded by SEN-D2.

between -100 and $150 \mu\epsilon$, wherein the 32 peaks concur with the 32 wheels acting on each track. Figure 4 provides the corresponding FAS in a logarithmic scale and it is seen that the strain response is dominated by low-frequency components lower than 10 Hz when the train speed is 10 km/h. Although the FBG strain measurement depends on both loading of the train and running speed, the measured dynamic strains of rail bending can be classified into two parts: the low-frequency components and the high-frequency components separated by a cut-off frequency f_c .¹⁴ The low-frequency components are found to be primarily controlled by axle loads and wheel bases and thus they are often used for the dynamic weighing, axle counting, and identification of trains,²⁹ while the high-frequency components are mainly caused by wheel and rail surface roughness and measurement noises, and thus they are often used for wheel condition assessment.^{13,14} As specified later, the cut-off frequency depends only on the train speed. Thus, the classification of low-frequency components and high-frequency components is a relative concept, depending on the train speed.

Data pre-processing

To obtain the information relevant to wheel quality, Filograno et al.¹⁴ proposed an empirical formula to extract the wheel-sensitive response ingredients from rail strain-monitoring data, which is

$$f_c = kv \quad (1)$$

where f_c is the lower cut-off frequency, v is the train speed, and k is a proportional coefficient. The proportional coefficient k is set to be 1.0 Hz h/km after numerous experiments in line of different train running speeds. The response components in the frequency range higher than f_c are termed as detrended data, which can be extracted from raw measurement data using an ideal high-pass filter with a cut-off frequency f_c depending on train speed. The remaining part is the

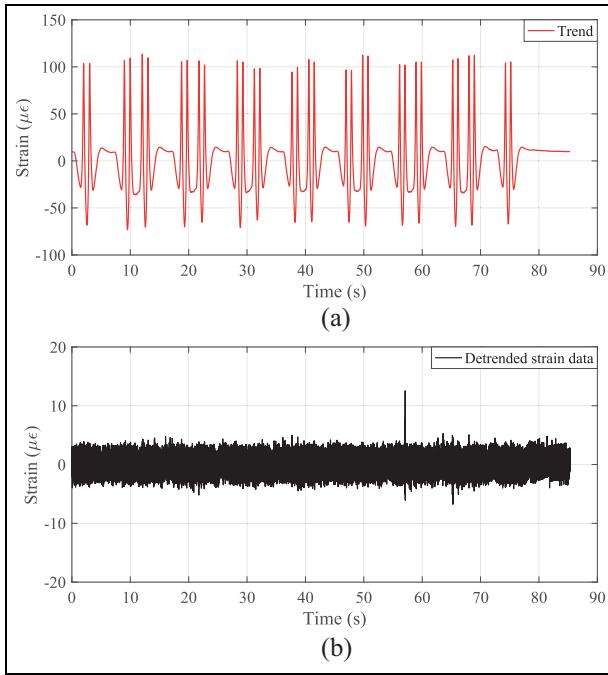


Figure 5. Rail strain recorded by SEN-D2: (a) trend response components containing frequencies lower than f_c and (b) detrended response components containing frequencies higher than f_c .

response components lower than f_c , termed as trend. Figure 5(a) and (b) shows the trend and the detrended data, respectively, with a cut-off frequency of $f_c = 10$ Hz. By segmenting the strain response time history into datasets surrounding its peak values, we obtain 32 data segments (datasets), each lasting for around 0.3 s and consisting of about 1500 data points, in line with the effects of the 32 wheels. Figure 6 shows the 32 data segments (datasets) of the detrended strain response, and Figure 7 depicts the FASs of the 32 detrended data segments (datasets). Typically, an upper frequency limit f_u can be adopted to eliminate the frequency-domain response components that are mainly caused by random noises. The upper limit frequency f_u is the upper bound of the wheel defect-induced effect on the variation of rail bending strain in the frequency domain. It depends on the train speed and the possible maximum wheel defect. However, it is extremely difficult to explicitly determine such a relationship, partially because there is a lack of samples of the possible maximum wheel defect for field test but mainly because the wheel defect-induced impact on the rail varies nonlinearly with the train speed.^{1,2} In this case study, the average train speed in test is $v = 10$ km/h and the targeted maximum wheel defect length is $L_{\max} = 150$ mm. The upper limit frequency f_u is empirically set to be 300 Hz, which guarantees that the chosen frequency-domain

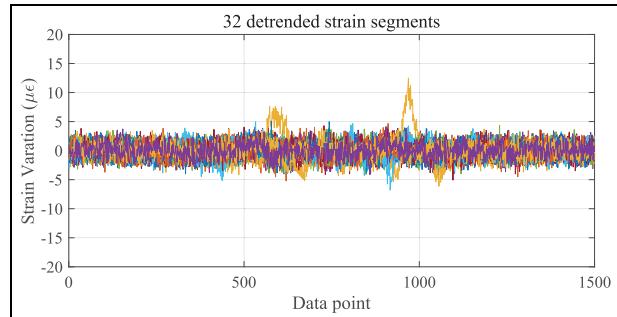


Figure 6. 32 detrended datasets extracted from rail strain recorded by SEN-D2.

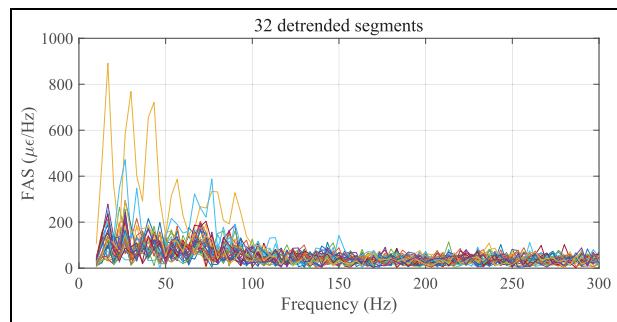


Figure 7. FASs of 32 detrended datasets recorded by SEN-D2.

response components can deliver the most information about wheel defect.

Feature extraction

After obtaining the FAS, its values $d(f)$ over the frequency band $[f_c, f_u]$ can be normalized by its integration d_I that is defined as

$$d_I = \int_{f_c}^{f_u} d(f) df \quad (2)$$

When a discrete Fourier transform (DFT) is employed, d_I can be approximated using the trapezoidal rule³¹ as

$$d_I \approx \sum_{i=1}^{N_1} \frac{d_i + d_{i+1}}{2} \Delta f \quad (3)$$

where Δf is the frequency resolution of DFT and N_1 is the number of frequency bins in $[f_c, f_u]$ given by

$$N_1 = \text{round} \left\{ \frac{f_u - f_c}{\Delta f} \right\} \quad (4)$$

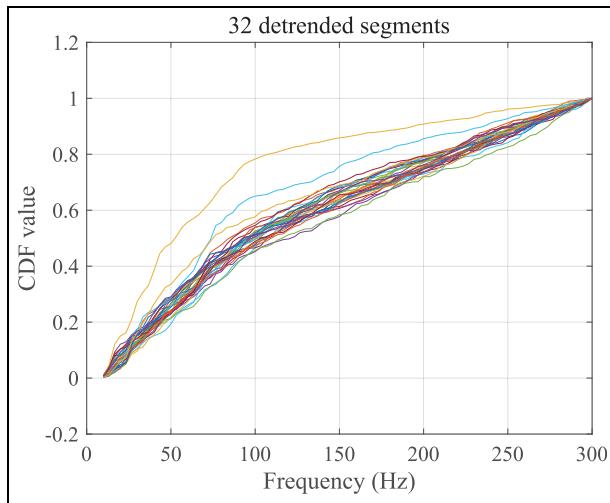


Figure 8. CDFs for normalized FASs of 32 detrended datasets recorded by SEN-D2.

where *round* is a mathematical operation for taking integer. As a result, the normalized FAS can be expressed in a probabilistic logic and its CDF on discrete frequency bins can be obtained as

$$y_i = \begin{cases} 0, & i=1 \\ \sum_{i=1}^{N_1-1} \frac{(d_i + d_{i+1})\Delta f}{2d_i}, & i \neq 1 \end{cases} \quad (5)$$

The values of CDF range between 0 and 1 in $[f_c, f_u]$. The CDFs for the normalized FASs of the 32 detrended data segments (datasets) are given in Figure 8 with $\Delta f = 3.3$ Hz and $N_1 = 88$. Later on, these CDFs will be used for training a probabilistic reference model by means of SBL, which will be served for wheel defect detection.

Model formulation by SBL

SBL

SBL^{26,27} is a nonparametric machine learning approach that shares characteristics in common with support vector machine,³² but produces probabilistic model outputs with dramatically few basis functions. Its ability of sparse representation and accurate prediction is primarily due to the Bayesian setting where uncertainty is taken into consideration and “inactive” basis terms can be automatically pruned through introducing hyperparameters in the prior distributions of weight parameters (sparsity-inducing priors).³³ As a result, the SBL exempts from the problem of overfitting which often occurs in classical least-squares and penalized least-squares. Due to the above merits, there has been an increasing interest in the application of SBL for

structural health monitoring (SHM).^{34–38} The basic theory of SBL for regression analysis is briefly introduced below. Given a dataset of input–output pairs $\{\mathbf{x}_i, y_i\}_{i=1}^N$ in which the outputs $\mathbf{y} = [y_1, \dots, y_N]^T$ are from a function or model $f(\mathbf{x})$ with additive noises $\boldsymbol{\varepsilon} = [\varepsilon_1, \dots, \varepsilon_N]^T$, the outputs can be expressed as

$$y_i = f(\mathbf{x}_i) + \varepsilon_i \quad (6)$$

A nonparametric approach for modeling $f(\mathbf{x}_i)$ is a linearly weighted sum of M basis (kernel) functions, given by

$$f(\mathbf{x}_i) = \mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i) = \sum_{i=1}^M w_i \phi_i(\mathbf{x}) \quad (7)$$

where $\boldsymbol{\phi}(\mathbf{x}) = [\phi_1(\mathbf{x}), \dots, \phi_M(\mathbf{x})]^T$ is a kernel vector and $\mathbf{w} = [w_1, \dots, w_M]^T$ is the associated weight vector. If the noise follows an independent and identically distributed Gaussian distribution with zero mean and variance σ^2 , the likelihood function for the outputs is

$$P(\mathbf{y}|\mathbf{w}, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^N}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - \Phi_1 \mathbf{w}\|^2\right) \quad (8)$$

where $\Phi_1 = [\boldsymbol{\phi}(\mathbf{x}_1), \dots, \boldsymbol{\phi}(\mathbf{x}_N)]^T$ is the design matrix for training the regression model. The maximum likelihood estimation of the unknown parameters \mathbf{w} and σ^2 from equation (8) may lead to severe overfitting and to avoid this, a common practice is to impose additional constraints on these unknowns to obtain a simpler model. In the Bayesian learning framework, this can be achieved by defining a set of explicit priors over these unknown parameters, which is also the key to SBL and available only in Bayesian setting. A popular choice of the prior for the weight vector \mathbf{w} is

$$P(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{i=1}^M P(w_i|\alpha_i) = \prod_{i=1}^M \frac{1}{\sqrt{2\pi/\alpha_i}} \exp\left(-\frac{1}{2} \alpha_i w_i^2\right) \quad (9)$$

where $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_M]^T$ is a hyper-parameter vector that controls the strength of the prior. A suitable hierarchical prior for $\boldsymbol{\alpha}$ is defined as

$$P(\boldsymbol{\alpha}) = \prod_{i=1}^M P(\alpha_i) = \prod_{i=1}^M \text{Gamma}(\alpha_i|a, b) \quad (10)$$

and the prior for the noise level σ^2 is defined as

$$P\left(\frac{1}{\sigma^2}\right) = P(\beta) = \text{Gamma}(\beta|c, d) \quad (11)$$

where $\beta \equiv 1/\sigma^2$ and

$$\text{Gamma}(\alpha|a, b) = \Gamma(a)^{-1} b^a \alpha^{a-1} e^{-b\alpha} \quad (12)$$

with the gamma function $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$. To make these priors non-informative, the parameters a, b, c , and d are fixed to small values. By Bayes' theorem, the joint posterior distribution for all the unknown parameters is given by

$$P(w, \alpha, \sigma^2 | y) = \frac{P(y|w, \alpha, \sigma^2)P(w, \alpha, \sigma^2)}{P(y)} \quad (13)$$

Typically, it is difficult to compute the joint posterior $P(w, \alpha, \sigma^2 | y)$ analytically from equation (13). Alternatively, the joint posterior $P(w, \alpha, \sigma^2 | y)$ can be decomposed into

$$P(w, \alpha, \sigma^2 | y) = P(w|y, \alpha, \sigma^2)P(\alpha, \sigma^2 | y) \quad (14)$$

where $P(\alpha, \sigma^2 | y)$ can be easily calculated as the weight vector w can be integrated out analytically as follows³⁹

$$\begin{aligned} P(\alpha, \sigma^2 | y) &\propto P(y|\alpha, \sigma^2) = P(y|w, \sigma^2)P(w|\alpha)dw \\ &= \mathcal{N}(\mu_1, \Sigma_1) \\ &= \frac{1}{\sqrt{(2\pi)^N |\Sigma_1|}} \\ &\exp\left[-\frac{1}{2}(y - \mu_1)^T \Sigma_1^{-1} (y - \mu_1)\right] \end{aligned} \quad (15)$$

where

$$\mu_1 = \mathbf{0}_N \quad (16)$$

$$\Sigma_1 = \sigma^2 \mathbf{I}_N + \Phi_1 \mathbf{A}^{-1} \Phi_1^T \quad (17)$$

with $\mathbf{A} = \text{diag}(\alpha_1, \dots, \alpha_M)$. Most-plausible point estimators α_{MP} and σ_{MP}^2 can be derived through maximization of equation (15) with respect to α and σ^2 . In Bayesian inference, the maximization of the marginal likelihood is known as a type-II maximum likelihood method. After deriving α_{MP} and σ_{MP}^2 , the posterior distribution over the weights is given by

$$\begin{aligned} P(w|y, \alpha, \sigma^2) &= \frac{P(y|w, \sigma^2)P(w|\alpha)}{P(y|\alpha, \sigma^2)} \\ &= \frac{1}{\sqrt{(2\pi)^{N+1} |\Sigma_2|}} \\ &\exp\left[-\frac{1}{2}(w - \mu_2)^T \Sigma_2^{-1} (w - \mu_2)\right] \end{aligned} \quad (18)$$

where the posterior covariance and mean vector for the weights are, respectively, given as

$$\Sigma_2 = (\sigma^{-2} \Phi_1^T \Phi_1 + \mathbf{A})^{-1} \quad (19)$$

$$\mu_2 = \sigma^{-2} \Sigma_2 \Phi_1^T y \quad (20)$$

Given the most-plausible point estimators α_{MP} and σ_{MP}^2 , predictions can be made on new test points $\mathbf{x}_* = [\mathbf{x}_{*,1}, \dots, \mathbf{x}_{*,S}]^T$ to obtain model outputs $\mathbf{y}_* = [y_{*,1}, \dots, y_{*,S}]^T$, in terms of the predictive distribution

$$P(y_* | y, \alpha_{MP}, \sigma_{MP}^2) \approx \int P(y_* | w, \sigma_{MP}^2) P(w | y, \alpha_{MP}, \sigma_{MP}^2) dw \quad (21)$$

where S is the number of the new test points. Since the terms in the aforementioned integration are both Gaussian, it turns out that

$$\begin{aligned} P(y_* | y, \alpha_{MP}, \sigma_{MP}^2) &\approx \mathcal{N}(\mu_3, \Sigma_3) \\ &= \frac{1}{\sqrt{(2\pi)^S |\Sigma_3|}} \\ &\exp\left[-\frac{1}{2}(y_* - \mu_3)^T \Sigma_3^{-1} (y_* - \mu_3)\right] \end{aligned} \quad (22)$$

with

$$\mu_3 = \Phi_2 \mu_2 \quad (23)$$

$$\Sigma_3 = \sigma_{MP}^2 \mathbf{I}_S + \Phi_2 \Sigma_2 \Phi_2^T \quad (24)$$

where $\Phi_2 = [\phi_i(\mathbf{x}_{*,1}), \dots, \phi_i(\mathbf{x}_{*,S})]^T$ is the prediction matrix, and μ_3 and Σ_3 are the posterior mean and covariance of outputs, respectively. The predictive covariance comprises the sum of two uncertain components: the estimated noise level on the data and that due to the uncertainty in the prediction of the weights. In implementation by iteration,²⁷ the maximum marginal likelihood method (type-II maximum likelihood method) makes a point approximation of α and σ^2 by maximizing $P(y|\alpha, \sigma^2)$ at each iteration step. The iterative updating of the hyper-parameters α and the noise level σ^2 typically leads many α_i 's to diverge toward infinity, resulting in the posterior of the corresponding weights highly peaked at zero. It means that the related basis functions $\phi_i(x)$ are irrelevant (or "inactive") and can be pruned from the model expressed in equation (7). Thus, the SBL offers an automatic regularization approach to make sparsity come out through pruning irrelevant ("inactive") basis functions in the iteration process.

Implementation for model learning

In general, monitoring data at the model training stage is lopsided: it is prodigal in healthy state, but niggard (even null) in defective state. Wheel defects are diverse in type and extent, and monitoring data from defective wheels of each type can be very limited. By contrast, data for healthy wheels are often abundant. In the worst case where undamaged data are also not available, it is necessary to resort to a precise and validated physical model such as a high-fidelity wheel–rail interaction model and the method is no longer free of model. This study is intended to develop a nonparametric reference model using only the monitoring data from healthy wheels in the training phase, with which defective wheels can be identified in the testing phase. To help account for uncertainties arising from different sources, this model is established in a probabilistic framework by means of SBL. In this section, the monitoring data acquired by the sensor SEN-D2 as described in the preceding section are taken as an example to illustrate the model training by SBL and this process can be easily applied to other sensors.

The monitoring data of rail foot strain response acquired by the sensor SEN-D2 contain the information about all 32 wheels (refer to Figure 3). The sequence of data is then separated and filtered to obtain the detrended datasets (Figure 6), their FASs (Figure 7), and normalized CDFs (Figure 8) stemming from individual wheels. The probabilistic model is trained in terms of SBL using the normalized CDFs which are elicited from the data collected by the sensor SEN-D2 during multiple trips of the train with all the wheels being in healthy state. Because of good adaptability, the Gaussian kernels are employed in this study as basis functions, given by

$$\phi_i(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\gamma^2}\right) \quad (25)$$

where γ is the Gaussian kernel width. In the SBL framework developed by Tipping,^{26,27} the kernel width needs to be predefined. In this study, a random initialization strategy is utilized to explore the optimal kernel width γ and three indices are employed to investigate the performance of model trained using different kernel widths. The first index is the root mean square residual (RMSR), defined as

$$\text{RMSR} = \sqrt{\sum_{i=1}^{N_2} \sum_{j=1}^{N_1} \frac{(y_{ij} - \mu_{3j})^2}{N_1 N_2}} \quad (26)$$

where N_1 is the number of frequency bins for each observed CDF, N_2 is the total number of CDFs used,

y_{ij} is the observed value of the i th CDF on the j th frequency bin, and μ_{3j} is the expectation value predicted by the SBL model. The second index is the mean standardized log loss (MSLL), given by

$$\begin{aligned} \text{MSLL} &= -\frac{1}{N_1 N_2} \log P(\mathbf{y}) \\ &= \frac{1}{2N_1 N_2} \sum_{i=1}^{N_2} \sum_{j=1}^{N_1} \\ &\quad \left[\log(2\pi\Sigma_3(n, n) + \Sigma_3^{-1}(n, n)(y_{ij} - \mu_{3j})^2) \right] \end{aligned} \quad (27)$$

where $\Sigma_3(n, n)$ is the n th diagonal element of the covariance matrix Σ_3 with $n=j+(i-1)\times N_1$. The third index is the sparse ratio κ , defined as

$$\kappa = \frac{N_{RV}}{N_1 N_2} \times 100\% \quad (28)$$

where N_{RV} is the number of non-zero weights in the trained model.

Then, SBL model is trained by successively increasing the kernel width γ from 1 to 100. Figure 9(a) to (d) shows four SBL models in conjunction with their predicted expectations, and the associated 95% confidence intervals when the kernel width $\gamma = 1$ are equal to 1, 26, 41, and 100, respectively. Figures 10–12, respectively, show the variation of RMSR, MSLL, and the sparse ratio κ against the kernel width γ . From Figure 9(a), it can be seen that the trained model is relatively complicated when the kernel width is too small and as a result, 79 relevance vectors (basis functions) are needed to represent the CDFs. Figure 9(b) and (c) provides two better alternative models, which are smoother and sparser with only six and four relevance vectors used, respectively. When the kernel width continues to increase, the probabilistic model learned by SBL will lose expressive ability as evidenced in Figure 9(d), where only three relevance vectors remain in the model but they are insufficient to characterize the CDFs. As shown in Figures 10 and 11, both RMSR and MSLL increase in general with the kernel width γ . Within the kernel width range considered, there exist two different local optimal kernel widths, which are $\gamma=26$ and 41 for RMSR and $\gamma=25$ and 40 for MSLL, in relation to different explanations about the complexity of the probabilistic model. By contrast, the sparsity ratio κ almost deceases continuously with the kernel width γ as shown in Figure 12. As a compromise between expressive ability and sparseness, the kernel width $\gamma=40$ is considered in this study as the optimal value to construct the probabilistic model that is fairly simple and favorably consistent with the likelihood of the training data. There exist only four relevance vectors

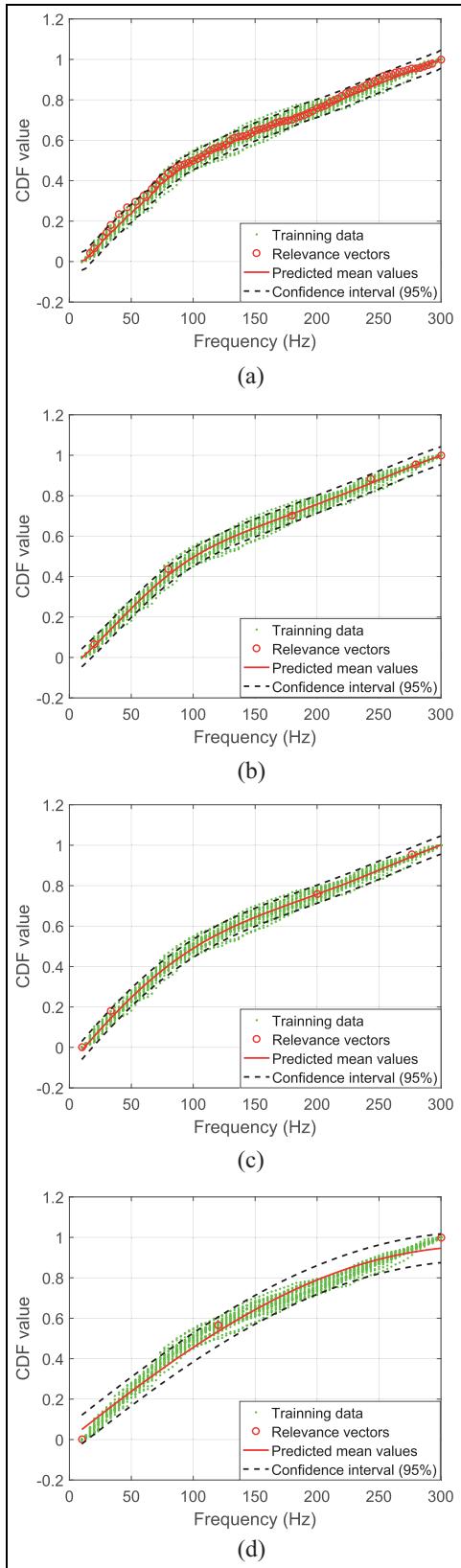


Figure 9. Model training with different kernel widths: (a) $\gamma = 1$, (b) $\gamma = 26$, (c) $\gamma = 41$, and (d) $\gamma = 100$.

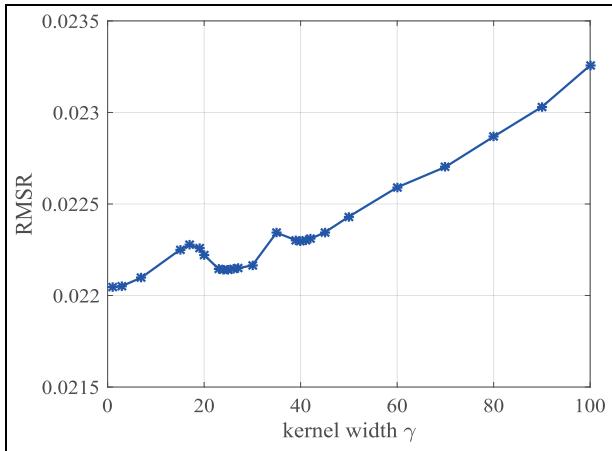


Figure 10. RMSR against kernel width γ .

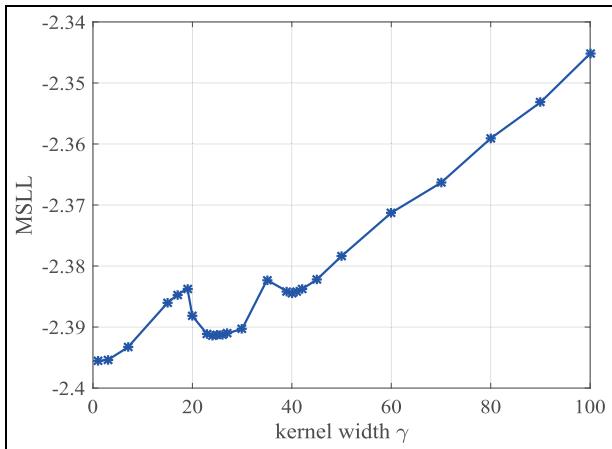


Figure 11. MSLL against kernel width γ .

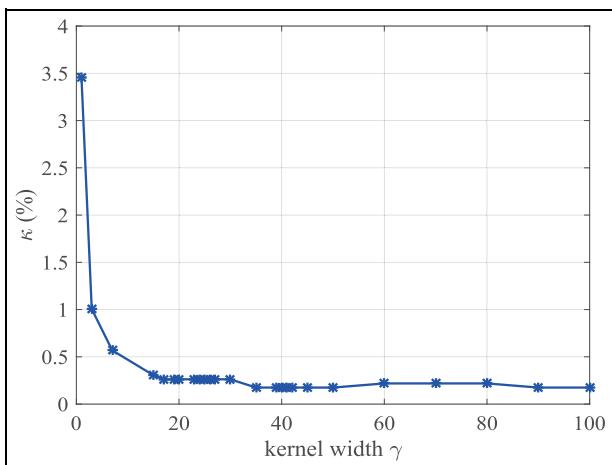


Figure 12. Sparsity ratio κ against kernel width γ .

Table 1. Active weights and associated kernel functions.

Weight number	Weight distribution	Associated kernel function
w_1	$\mathcal{N}(-5.35, 0.05^2)$	$\exp\left(-\frac{\ x-1\ ^2}{2 \times 40^2}\right)$
w_{209}	$\mathcal{N}(5.65, 0.048^2)$	$\exp\left(-\frac{\ x-9\ ^2}{2 \times 40^2}\right)$
w_{1769}	$\mathcal{N}(-1.73, 0.020^2)$	$\exp\left(-\frac{\ x-69\ ^2}{2 \times 40^2}\right)$
w_{2263}	$\mathcal{N}(2.24, 0.016^2)$	$\exp\left(-\frac{\ x-88\ ^2}{2 \times 40^2}\right)$

Table 2. Performance comparison of SBL and BGL models.

Performance index	SBL	BGL
RMSR	0.0223	0.0222
MSLL	-2.3844	-2.3835
κ	0.17%	3.85%

SBL: sparse Bayesian learning; BGL: Bayesian generalized linear; RMSR: root mean square residual; MSLL: mean standardized log loss.

(basis functions) in the probabilistic model when $\gamma=40$, justifying a sparse representation of the built model. Table 1 shows the four relevance vectors including the distributions of active (non-zero) weights and the associated kernel functions, where the frequency band of interest $x_f=[10 \text{ Hz}, 300 \text{ Hz}]$ has been scaled to $x=[1, 88]$ using a linear transform $x=3x_f/10 - 2$ to facilitate computational accuracy and efficiency.

To validate the benefit of the SBL framework, a comparative study is conducted using the non-sparse Bayesian generalized linear (BGL)⁴⁰ model to learn the same training data. Table 2 provides a comparison between the SBL and the BGL models in terms of three performance indices. It is seen that while the learning performance in terms of RMSR and MSLL is comparable between the SBL and the BGL models, the SBL model utilizes dramatically fewer basis functions (much lower sparsity ratio), giving rise to a much simpler probabilistic model with stronger prediction ability.

In this study, the SBL model is formulated using the data collected under the nominal running speed of 10 km/h. This is due to the fact that the track-side monitoring system is often installed at a location immediately before trains arrive at a railway station or terminal, for the ease of operation, management, and maintenance. As a result, it allows both low-speed trains and high-speed trains to pass the monitoring area at a fixed speed. Certainly, the SBL model can be adaptive to a wide range of running speeds by incorporating an extra variable (train speed) in the basis functions in equation (25) if enough data covering different train speeds are available.

The formulated model is deemed to be robust to loading conditions as the loading conditions, such as fully loaded and non-loaded trains, do not significantly influence the wheel defect-incurred impact on the railway track, which has been validated by both numerical modeling⁴¹ and field test.¹ Environmental conditions, such as wet track, ice, debris, and extremely temperature, may affect the pattern of wheel–rail interaction. To ensure the reliability of defect detection results by the proposed method, it is preferable to collect the monitoring data in the testing phase under the similar environmental conditions as to obtain the training data.

The model does not evolve with time because it is formulated to represent the initial defect-free state of railway wheels. If new monitoring data collected later are confirmed from healthy state as well, the model can be refined using the newly collected data as explained in the following section. In principle, the model does not need to be updated over time. Instead, the model can be utilized to investigate the deterioration of wheel condition over time.

Bayesian hypothesis testing for wheel defect detection

A variety of diagnostic criteria are available for damage or fault identification and quantification. In recognizing the shortcomings of the commonly used distance-based diagnostic methods, statistical hypothesis tests have gained growing interest in SHM applications. For example, Bayesian point null hypothesis testing (BPNHT) has been attempted for damage or fault identification and quantification in terms of Bayes factor.^{40,42,43} It is more robust than the distance-based diagnostic methods in that its resulting risk is averaged over the priors for unknown parameters in the hypotheses. However, the BPNHT is sensitive to the priors of the unknown parameters, giving rise to the so-called Jeffreys–Lindley paradox. In this study, we introduce a novel damage diagnostic logic in terms of BNHST and IBF, which does not suffer from the Jeffreys–Lindley paradox and the effect of sample size (data scale).

BNHST

In the previous section, a probabilistic model has been formulated by SBL using monitoring data in the state of healthy wheels. This model characterizes the stochastic CDFs in healthy state. Thus, a null hypothesis H_0 can be defined for healthy wheel using the built probabilistic model. Because of the lack of monitoring data from defective wheels in the model training phase, we are unable to elicit an alternative hypothesis representative of defective wheels by using the training data.

Instead, the alternative hypothesis H_1 for defective wheel is made here by shifting the expectations $\boldsymbol{\mu}_3$ with small values of \mathbf{h} .⁴⁴ Thus, referring to equation (22), the two hypotheses are given by

$$H_0 : \mathbf{y}_u \sim \mathcal{N}(\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3) \text{ for healthy wheel} \quad (29)$$

$$H_1 : \mathbf{y}_u \sim \mathcal{N}(\boldsymbol{\mu}_3 + \mathbf{h}, \boldsymbol{\Sigma}_3) \text{ for defective wheel} \quad (30)$$

where $\mathbf{y}_u = [y_{u,1}, \dots, y_{u,S}]^T$ is the new CDFs from wheels (healthy or defective) to be assessed, and $\boldsymbol{\mu}_3$ and $\boldsymbol{\Sigma}_3$ are the predicted expectations and the covariance as given in equations (23) and (24), respectively. The same covariance matrix is presumed in equations (29) and (30) for the computational efficiency. More importantly, it enables us to focus on the identification of the systematic change in defect-sensitive features (the means of the stochastic CDFs) caused by wheel defects. \mathbf{h} is related to the false-positive diagnostic risk, denoted as ϑ , which is referred to as the probability of healthy wheels being falsely diagnosed as defective. There is a one-to-one relationship between \mathbf{h} and ϑ , given by

$$\vartheta = 2 \left\{ 1 - \Phi \left[\frac{1}{2S} \left(\boldsymbol{\Sigma}_3^{-1/2} \mathbf{h} \right)^T \times \mathbf{1}_S \right] \right\} \quad (31)$$

where Φ is the cumulative distribution function of the standard normal random variable, and $\mathbf{1}_S$ is a unit column vector of size S . The Bayes factor is defined as the ratio of the likelihood probabilities of the two hypotheses, given by

$$BF_{01} = \frac{P(\mathbf{y}_u | H_0)}{P(\mathbf{y}_u | H_1)} = \exp \left[\frac{1}{2} \left(\boldsymbol{\varepsilon}_1^T \boldsymbol{\Sigma}_3^{-1} \boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_0^T \boldsymbol{\Sigma}_3^{-1} \boldsymbol{\varepsilon}_0 \right) \right] \quad (32)$$

where $\boldsymbol{\varepsilon}_0 = \mathbf{y}_u - \boldsymbol{\mu}_3$ and $\boldsymbol{\varepsilon}_1 = \mathbf{y}_u - (\boldsymbol{\mu}_3 + \mathbf{h})$. Since BF_{01} is always non-negative, it can be converted to the logarithmic scale for convenience of comparison among a large range of values. It is also useful to consider twice the natural logarithm of the Bayes factor that is on the same scale as the familiar deviance and likelihood ratio test statistics,⁴⁵ given by

$$2 \ln(BF_{01}) = \boldsymbol{\varepsilon}_1^T \boldsymbol{\Sigma}_3^{-1} \boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_0^T \boldsymbol{\Sigma}_3^{-1} \boldsymbol{\varepsilon}_0 \quad (33)$$

Kass and Raftery⁴⁶ suggested interpreting $2 \ln(BF_{01})$ between 0 and 2 as weak evidence in favor of H_0 , between 2 and 6 as positive evidence, between 6 and 10 as strong evidence, and $2 \ln(BF_{01}) > 10$ as very strong evidence. Negative Bayes factor of the same magnitude is said to favor the alternative hypothesis H_1 by the same amount. However, the Bayes factor defined earlier relies heavily on sample size such that different Bayes factor thresholds should be used when the sample size is highly different. To overcome this drawback, a novel Bayes factor, termed as intrinsic Bayes factor,

is proposed in this study to eliminate the effect of sample size (data scale), which is defined as follows

$$IBF_{01} = \sqrt[S]{BF_{01}} = \sqrt[S]{\exp \left[\frac{1}{2} \left(\boldsymbol{\varepsilon}_1^T \boldsymbol{\Sigma}_3^{-1} \boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_0^T \boldsymbol{\Sigma}_3^{-1} \boldsymbol{\varepsilon}_0 \right) \right]} \quad (34)$$

where S is the size of \mathbf{y}_u . Twice the natural logarithm of IBF_{01} yields

$$2 \ln(IBF_{01}) = \frac{1}{S} \left(\boldsymbol{\varepsilon}_1^T \boldsymbol{\Sigma}_3^{-1} \boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_0^T \boldsymbol{\Sigma}_3^{-1} \boldsymbol{\varepsilon}_0 \right) \quad (35)$$

As identical Gaussian kernels are used in the model trained by SBL, the predicted uncertainties can be approximated by $\boldsymbol{\Sigma}_3 \approx \sigma_{MP}^2 \mathbf{I}_S$.²⁶ Thus, the IBF_{01} and its twice the natural logarithm reduce to

$$IBF_{01} = \sqrt[S]{\exp \left[\frac{1}{2\sigma_{MP}^2} (\boldsymbol{\varepsilon}_1^T \boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_0^T \boldsymbol{\varepsilon}_0) \right]} \quad (36)$$

$$2 \ln(IBF_{01}) = \frac{1}{S\sigma_{MP}^2} (\boldsymbol{\varepsilon}_1^T \boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_0^T \boldsymbol{\varepsilon}_0) \quad (37)$$

Accordingly, the false-positive diagnostic risk ϑ is approximately given by

$$\vartheta = 2 \left[1 - \Phi \left(\frac{\mathbf{h}^T \times \mathbf{1}_S}{2S\sigma_{MP}} \right) \right] \quad (38)$$

When new monitoring data from potentially defective wheel(s) are made available, the predicted expectations $\boldsymbol{\mu}_3$ and covariance $\boldsymbol{\Sigma}_3$ can be obtained from the built reference model. Then, the IBF_{01} or its twice the natural logarithm is calculated to infer the existence of defective wheel(s) and to assess the degree of wheel defect(s). Using the foregoing classification criterion on discrimination strength (more detailed classification intervals are available in Kass and Raftery⁴⁶) to quantitatively assess the extent of defect is relatively subjective before physical calibration.

If the new monitoring data are confirmed from healthy state, the newly collected data can be used to update or refine the current SBL model by taking the joint posterior distribution $P(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2 | \mathbf{y})$ of the unknown parameters in the current model as a data-based prior for the updated model. The model accuracy will be enhanced since the model posterior uncertainty $\boldsymbol{\Sigma}_3$ in equation (24) can be reduced by increasing the training data. As mentioned before, the formulation of the reference model does not use monitoring data in defective states. However, when monitoring data in line with confirmed defects are available, the classification intervals can be calibrated and pertinently refined to provide more precise mapping between the range of IBF and the extent of wheel defect.

In comparison with the new defect-sensitive features characterized by the SBL model, the raw defect-sensitive features cannot be directly adopted to perform Bayesian hypothesis significance testing in the presence of new monitoring data because their uncertainties are not quantified. Moreover, the raw and the new features have different feature inputs and thus they are not directly comparable.

Diagnostic results of wheel defects using a single sensor

A blind test was later conducted by replacing some healthy wheels by defective wheels and running the train equipped with defective wheels on the rail instrumented with the track-side monitoring system. It provides in-situ monitoring data to verify the proposed wheel defect detection method. Moreover, after the blind test, the suspected defective wheels have been delivered to a workshop for offline wheel radius deviation measurement. As such, a comparison between the online diagnostic results by the proposed method and the offline inspection results can be made. It is worth noting that the monitoring data collected by a single sensor might be unable to capture the defect-relevant information in case the minor defective tread (e.g. a small flat) did not roll over the rail section deployed with the sensor; it would result in a false negative if using only the data from the single sensor. When using the monitoring data from all the deployed sensors, more reliable defect detection results would be obtained because the effect of minor defective tread must be sensed by at least one sensor if the sensors are densely deployed along a rail segment longer than the wheel perimeter. In the following section, both the wheel defect detection results using the monitoring data from a single sensor (SEN-D2) and from all the sensors are presented.

In the implementation of the proposed method, the shift parameters are set to be $\mathbf{h} = 5\sigma_{MP} \times \mathbf{1}_S$ at the price of the diagnostic risk of 1.24% that healthy wheels are falsely diagnosed as defective, where $\mathbf{1}_S$ is a unit column vector of size S . The choice of \mathbf{h} reflects the understanding of professionals regarding the presumed discrepancy in the features of wheels with and without defect. Its choice is essentially a balance between the false-positive and false-negative error rates. If the values of \mathbf{h} are smaller, a higher rate of false alarm of wheel defect occurs, while larger values of \mathbf{h} correspond to a higher rate of missing alarm. For different application domains, one may use customized values of \mathbf{h} , according to the total loss caused by the two types of errors. The obtained IBFs for assessing the condition of all right wheels when using the monitoring data from

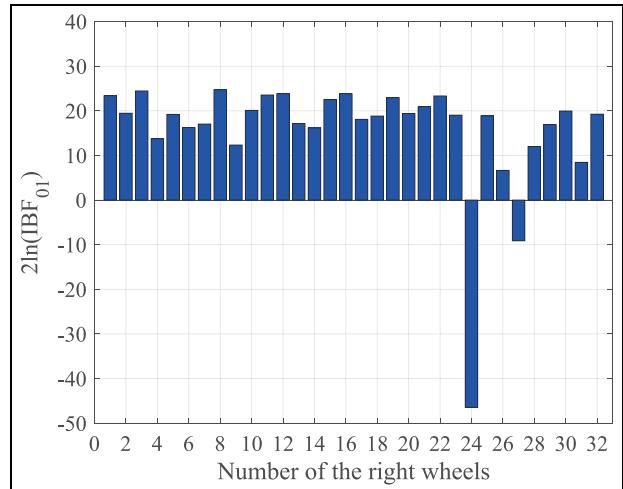


Figure 13. Condition assessment of right wheels using monitoring data from SEN-D2 deployed on right rail track.

the sensor SEN-D2 deployed on the right railway track are shown in Figure 13. It is observed that the IBF for the 24th right wheel is less than -10 , which advocates very strongly that it is heavily defective. The IBF for the 27th right wheel is between -10 and -6 , providing evidence that it is moderately defective. By contrast, the IBFs for other right wheels are all positive and therefore, these wheels are diagnosed as undefective.

Diagnostic results of wheel defects by integrating all sensors

The wheel defect detection is then pursued using the monitoring data from all sensors deployed on one side of the rail. As we are interested in finding defective wheels, the smallest IBF is used in the assessment, which is given by

$$2\ln(IBF_{01}) = \min\{2\ln(IBF_{01}^1), \dots, 2\ln(IBF_{01}^{N_S})\} \quad (39)$$

where N_S is the number of sensors deployed on each side of the rail. For the track-side monitoring system described in section “Feature Extraction through Track-side Monitoring,” $N_S = 21$.

Figure 14 provides the diagnostic results on the condition of the right wheels using the monitoring data from all 21 sensors deployed on the right rail track. It can be seen that the IBFs associated with the 1st, 6th, 24th, 27th, and 31st right wheels are negative, suggesting that the five wheels are potentially defective with different degrees. The 24th wheel is most heavily defective, while the 31st wheel is most weakly defective. With IBFs being positive, the other right wheels are diagnosed as healthy. By comparing Figure 14 with Figure 13, it is found that using only the monitoring data from the sensor SEN-D2 fails to identify the defects on the

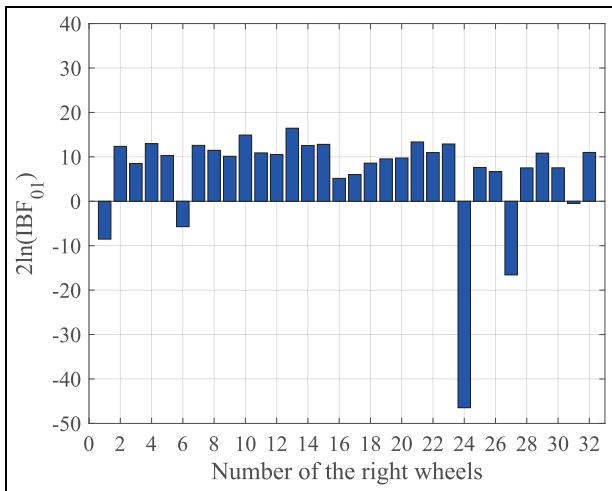


Figure 14. Condition assessment of right wheels using monitoring data from all sensors deployed on right rail track.

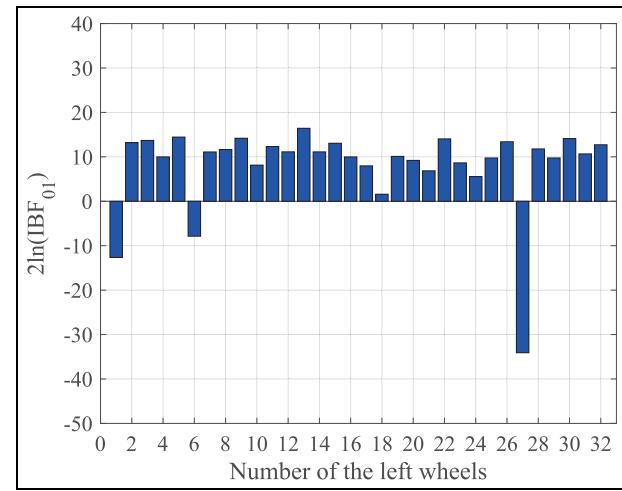


Figure 15. Condition assessment of left wheels using monitoring data from all sensors deployed on left rail track.

1st, 6th, and 31st right wheels, while the results from the two settings indicate defects on the 24th and 27th wheels.

Figure 15 illustrates the diagnostic results on the condition of the left wheels when using the monitoring data from all 21 sensors deployed on the left rail track. It is observed that the IBFs for the 1st, 6th, and 27th wheels are negative, indicating that the three wheels are defective with the 27th wheel being most heavily defective. The other left wheels are diagnosed as healthy. The values of IBFs provide a quantitative measure to assess the degree of wheel defects. Smaller values of IBF indicate in general worse wheel conditions.

To validate the diagnostic results by the proposed online method, offline inspection on the suspected defective wheels was conducted afterwards in a workshop. The offline wheel radius deviation measurement indicates that the 1st, 6th, 24th, and 27th right wheels, and the 1st, 6th, and 27th left wheels are indeed defective, with large flats found on the 24th (36.0 mm in length), 27th right wheel (26.9 mm), and the 27th left wheel (34.6 mm). Although the 31st right wheel is diagnosed as weakly defective by the proposed method, it is in fact healthy. This warns us of the diagnostic risk imposed by the proposed method. Overall, the diagnostic results by the proposed online method are in good agreement with the offline measurement results.

Conclusion

A Bayesian machine learning approach for online detection of wheel defects and condition has been

proposed in this study. With the aid of an FBG-based track-side monitoring system, CDFs for the normalized FASs of rail foot strain responses in the healthy state of wheels are extracted as characteristic features to train a probabilistic reference model by SBL. In the Bayesian probabilistic framework, the formulated model can account for uncertainties arising from the monitoring data (e.g. measurement noise and variability in the stochastic wheel–rail dynamics) and modeling error, and the SBL enables the model to perform well in either characterizing the training data or predicting unseen data. Because only a few basis functions are involved in the model, its computational efficiency is quite competitive especially on prediction, enabling fast diagnosis in wheel condition assessment. The diagnostic logic in terms of BNHST and scale-invariant IBF allows the unsupervised defect detection to be executed in a fully probabilistic inference context, ranging from the probabilistic model development in the training phase to the wheel defect diagnosis in the testing phase.

The proposed method is verified using the in-situ monitoring data acquired by a track-side monitoring system during the passage of a train with all wheels being healthy and with some wheels being defective, respectively, and is validated through comparing the diagnostic results obtained by the proposed online method and by the offline wheel radius deviation measurement. It turns out the following findings: (a) when using CDFs as characteristic features, a sparse representation of the probabilistic model containing only a few basis functions (four Gaussian kernels in the case of the optimal kernel width $\gamma=40$) can be obtained to

favorably characterize the stochastic CDFs; (b) the proposed method may give rise to false-negative diagnostic results when using only the monitoring data from a single sensor, but more accurate diagnostic results can be obtained using the monitoring data from all the deployed sensors; and (c) when using the monitoring data from all the sensors, the diagnostic results on the wheel defects and condition by the proposed online method are in good agreement with the offline inspection results. It should be mentioned that the choice of the shifting values \mathbf{h} affects the evaluated value of IBF (or its twice the natural logarithm) and therefore has an influence on defect detection results. The value of \mathbf{h} is directly related to the significance level (diagnostic risk). In this study, \mathbf{h} is set to be $5\sigma_{MP} \times 1_s$ at the price of the diagnostic risk of 1.24% that healthy wheels are falsely diagnosed as defective, and a false-positive “defect” on the 31st right wheel is falsely alarmed (albeit it is identified to be a very weakly defective according to its IBF). When offline radius deviation measurement results on wheels of different defect extents are available, a more appropriate choice of the shifting values \mathbf{h} can be made. Similarly, when the correspondence between the evaluated values of IBF and the real defect extents of defective wheels detected by offline inspection is established, the generic classification intervals defined from a mathematical logic can be refined to be more precise and specific for defect degree classification of the monitored wheels based on the evaluated value of IBF.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work described in this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (grant no. PolyU 152014/18E). The authors would also like to appreciate the funding support by the Innovation and Technology Commission of Hong Kong SAR Government to the Hong Kong Branch of Chinese National Rail Transit Electrification and Automation Engineering Technology Research Center (grant no. K-BBY1).

ORCID iD

Yi-Qing Ni  <https://orcid.org/0000-0003-1527-7777>

References

1. Nielsen JCO and Johansson A. Out-of-round railway wheels: a literature survey. *Proc IMechE, Part F: J Rail and Rapid Transit* 2000; 214(2): 79–91.
2. Johansson A and Nielsen JCO. Out-of-round railway wheels-wheel-rail contact forces and track response derived from field tests and numerical simulations. *Proc IMechE, Part F: J Rail and Rapid Transit* 2003; 217(2): 135–146.
3. Johansson A. Out-of-round railway wheels-assessment of wheel tread irregularities in train traffic. *J Sound Vib* 2006; 293(3–5): 795–806.
4. Pohl R, Erhard A, Montag HJ, et al. NDT techniques for railroad wheel and gauge corner inspection. *NDT & E Int* 2004; 37(2): 89–94.
5. Pau M. Ultrasonic waves for effective assessment of wheel-rail contact anomalies. *Proc IMechE, Part F: J Rail and Rapid Transit* 2005; 219(2): 79–90.
6. Skarlatos D, Karakasis K and Trochidis A. Railway wheel fault diagnosis using a fuzzy-logic method. *Appl Acoust* 2004; 65(10): 951–966.
7. Belotti V, Crenna F, Michelini RC and Rossiet al. Wheel-flat diagnostic tool via wavelet transform. *Mech Syst Signal Pr* 2006; 20(8): 1953–1966.
8. Jia S and Dhanasekar M. Detection of rail wheel flats using wavelet approaches. *Struct Health Monit* 2007; 6(2): 121–131.
9. Zhang LH, Wang YW, Ni YQ, et al. Online condition assessment of high-speed trains based on Bayesian forecasting approach and time series analysis. *Smart Struct Syst* 2018; 21(5): 705–713.
10. Stratman S, Liu Y and Mahadevan S. Structural health monitoring of railroad wheels using wheel impact load detectors. *J Fail Anal Prev* 2007; 7(3): 218–225.
11. Milković D, Simić G, Jakovljević Ž, et al. Wayside system for wheel-rail contact forces measurements. *Measurement* 2013; 46(9): 3308–3318.
12. Krummenacher G, Ong CS, Koller S, et al. Wheel defect detection with machine learning. *IEEE Trans Intel Transp Syst* 2018; 19(4): 1176–1187.
13. Wei C, Xin Q, Chung WH, et al. Real-time train wheel condition monitoring by fiber Bragg grating sensors. *Int J Distrib Sens Netw* 2012; 8(1): 409048.
14. Filograno ML, Corredora P, Rodriguez-Plaza M, et al. Wheel flat detection in high-speed railway systems using fiber Bragg gratings. *IEEE Sens J* 2013; 13(12): 4808–4816.
15. Liu XZ and Ni YQ. Wheel tread defect detection for high-speed trains using FBG-based online monitoring techniques. *Smart Struct Syst* 2018; 21(5): 687–694.
16. Bolas K, Papasalouros D, Kourousis D, et al. Acoustic emission monitoring of wheel sets on moving trains. *Construct Build Mater* 2013; 48: 1266–1272.
17. Lagnebäck R. *Evaluation of wayside condition monitoring technologies for condition-based maintenance of railway vehicles*. Doctoral Dissertation, Luleå University of Technology, Luleå, 2007.
18. Asplund M, Famurewa S and Rantatalo M. Condition monitoring and e-maintenance solution of railway wheels. *J Qual Maint Eng* 2014; 20(3): 216–232.
19. Resendiz E, Hart JM and Ahuja N. Automated visual inspection of railroad tracks. *IEEE Trans Intel Transp Syst* 2013; 14(2): 751–760.
20. Li Y, Trinh H, Haas N, et al. Rail component detection, optimization, and assessment for automatic rail track inspection. *IEEE Trans Intel Transp Syst* 2014; 15(2): 760–770.

21. Gibert X, Patel VM and Chellappa R. Deep multitask learning for railway track inspection. *IEEE Trans Intel Transp Syst* 2017; 18(1): 153–164.
22. Aytekin C, Rezaeitabar Y, Dogru S, et al. Railway fastener inspection by real-time machine vision. *IEEE Trans Syst Man Cybernet: Part A Syst Human* 2015; 45(7): 1101–1107.
23. Chen J, Liu Z, Wang H, et al. Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network. *IEEE Trans Instrum Meas* 2018; 67(2): 257–269.
24. Westeon PF, Ling CS, Roberts C, et al. Monitoring vertical track irregularity from in-service railway vehicles. *Proc IMechE, Part F: J Rail and Rapid Transit* 2007; 221(1): 75–88.
25. Jamshidi A, Roohi SF, Núñez A, et al. Probabilistic defect-based risk assessment approach for rail failures in railway infrastructure 2016; 49(3): 73–77.
26. Tipping ME. Sparse Bayesian learning and the relevance vector machine. *J Mach Learn Res* 2001; 1: 211–244.
27. Tipping ME and Faul AC. Fast marginal likelihood maximisation for sparse Bayesian models. In: *Proceedings 9th international workshop on artificial intelligence and statistics*, Key West, Florida, USA, 3–6 January 2003.
28. Jing L and Han L. Further study on the wheel–rail impact response induced by a single wheel flat: the coupling effect of strain rate and thermal stress. *Veh Syst Dynam* 2017; 55(12): 1946–1972.
29. Filograno ML, Guillén PC, Rodríguez -Barrios A, et al. Real-time monitoring of railway traffic using fiber Bragg grating sensors. *IEEE Sens J* 2012; 12(1): 85–92.
30. Wang YW, Ni YQ and Wang X. Real-time defect detection of high-speed train wheels by using Bayesian forecasting and dynamic model. *Mech Syst Signal Pr* 2020; 139: 106654.
31. Stoer J and Bulirsch R. *Introduction to numerical analysis*. 3rd ed. New York: Springer, 2002.
32. Cortes C and Vapnik V. Support-vector networks. *Machine Learn* 1995; 20(3): 273–297.
33. Tipping ME. Bayesian inference: an introduction to principles and practice in machine learning. In: Bousquet O, Von Luxburg U and Ratsch G (eds) *Advanced lectures on machine learning*. Berlin: Springer, 2006.
34. Wu Q, Zhang YD, Amin MG, et al. Structural health monitoring exploiting MIMO ultrasonic sensing and group sparse Bayesian learning. In: *Proceedings of the conference on signals, systems and computers*, Pacific Grove, CA, 2–5 November 2014. New York: IEEE.
35. Huang Y and Beck JL. Hierarchical sparse Bayesian learning for structural health monitoring with incomplete modal data. *Int J Uncertain Quant* 2015; 5(2): 139–169.
36. Huang Y, Beck JL and Li H. Hierarchical sparse Bayesian learning for structural damage detection: theory, computation and application. *Struct Safet* 2017; 64: 37–53.
37. Wu B, Huang Y, Chen X, et al. Guided-wave signal processing by the sparse Bayesian learning approach employing Gabor pulse model. *Struct Health Monit* 2017; 16(3): 347–362.
38. Wu B, Li H and Huang Y. Sparse recovery of multiple dispersive guided-wave modes for defect localization using a Bayesian approach. *Struct Health Monit* 2019; 18(4): 1235–1252.
39. Quinonero-Candela J. *Learning with uncertainty: Gaussian processes and relevance vector machines*. PhD Thesis, Technical University of Denmark, Lyngby, 2004.
40. Wang J, Liu XZ and Ni YQ. A Bayesian probabilistic approach for acoustic emission-based rail condition assessment. *Comput Aided Civil Infrastruct Eng* 2018; 33(1): 21–34.
41. Uzzal RUA, Ahmed W and Rakheja S. Dynamic analysis of railway vehicle-track interactions due to wheel flat with a pitch-plane vehicle model. *J Mech Eng* 2008; 39(2): 86–94.
42. Jiang X and Mahadevan S. Bayesian probabilistic inference for nonparametric damage detection of structures. *J Eng Mech* 2008; 134(10): 820–831.
43. Sankararaman S and Mahadevan S. Bayesian methodology for diagnosis uncertainty quantification and health monitoring. *Struct Control Health Monit* 2013; 20(1): 88–106.
44. Lipowsky H, Staudacher S, Bauer M, et al. Application of Bayesian forecasting to change detection and prognosis of gas turbine performance. *J Eng Gas Turbine Power* 2010; 132(3): 031602.
45. Jeffreys H. *The theory of probability*. Oxford: Oxford University Press, 1961.
46. Kass RE and Raftery AE. Bayes factors. *J Am Stat Assoc* 1995; 90(430): 773–795.