Smart Structures and Systems, *Vol. 30, No. 3 (2022) 303-315*
https://doi.org/10.12989/sss.2022.30.3.303

Online railway wheel defect detection under varying running-speed conditions by multi-kernel relevance vector machine

Yuan-Hao Wei [1,2a], You-Wu Wang [1,2b] and Yi-Qing Ni*[1,2]

[1] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*
[2] *Hong Kong Branch of Chinese National Engineering Research Center on Rail Transit Electrification and Automation, Hong Kong*

(*Received June 28, 2022, Revised July 10, 2022, Accepted July 10, 2022*)

Abstract. The degradation of wheel tread may result in serious hazards in the railway operation system. Therefore, timely wheel defect diagnosis of in-service trains to avoid tragic events is of particular importance. The focus of this study is to develop a novel wheel defect detection approach based on the relevance vector machine (RVM) which enables online detection of potentially defective wheels with trackside monitoring data acquired under different running-speed conditions. With the dynamic strain responses collected by a trackside monitoring system, the cumulative Fourier amplitudes (CFA) characterizing the effect of individual wheels are extracted to formulate multiple probabilistic regression models (MPRMs) in terms of multi-kernel RVM, which accommodate both variables of vibration frequency and running speed. Compared with the general single-kernel RVM-based model, the proposed multi-kernel MPRM approach bears better local and global representation ability and generalization performance, which are prerequisite for reliable wheel defect detection by means of data acquired under different running-speed conditions. After formulating the MPRMs, we adopt a Bayesian null hypothesis indicator for wheel defect identification and quantification, and the proposed method is demonstrated by utilizing real-world monitoring data acquired by an FBG-based trackside monitoring system deployed on a high-speed trial railway. The results testify the validity of the proposed method for wheel defect detection under different running-speed conditions.

Keywords: model optimization; multi-kernel RVM; online detection; railway wheel defect; relevance vector machine (RVM); varying running speed

1. Introduction

The functional integrity of train wheels is critical for the safe operation of contact-type railway system. Troublesome events or even serious accidents can be caused by wheel defects, such as annoying noise (Petersson 2000, Wu and Thompson 2002, Kouroussis *et al*. 2014), abnormal vibration patterns (Morys 1999, Wu and Chi 2016), degradation of vehicle components (Dukkipati and Dong 1999, Barke and Chiu 2005), and even derailment (Johansson and Nielsen 2003, Jin *et al*. 2012). A variety of wheel defect detection strategies have been investigated in the past decades. Technically, these approaches can be divided into two branches: offline inspection and online monitoring. As a mature procedure of offline maintenance, the regular wheel re-profiling strategy is widely used due to its reliability and effectiveness (Pohl *et al*. 2004). With the development of novel techniques such as ultrasonic wave technique (Pau 2005, Wu *et al*. 2017), offline wheel condition assessment becomes more targeted and efficient, which can avoid the service-life reduction of wheelsets made by simply mileage-based re-profiling of all wheels in the overhauled train, eventually reducing the maintenance costs. Nevertheless, offline detection cannot reflect the health status of wheels in a real-time manner. Benefiting from the development of sensing technology, various kinds of sensors have come to the fore over the last two decades, such as laser sensors (Asplund *et al*. 2014), fiber Bragg grating (FBG) sensors (Wei *et al*. 2011, Filograno *et al*. 2013), acoustic emission sensors (Bollas *et al*. 2013), and piezoelectric strain gauges (Stratman *et al*. 2007, Milković *et al*. 2013, Krummenacher *et al*. 2017), which make real-time monitoring feasible. Online monitoring techniques have been extensively studied ever since. A comparison of wheel condition assessment techniques (Liu *et al*. 2018, 2019) shows that online methods have better monitoring efficiency than offline inspection strategies.

The diagnostic algorithms for online wheel health assessment can be generally grouped into two categories: deterministic ones and probabilistic ones. For deterministic algorithms, in most cases, a threshold needs to be empirically determined to distinguish the intact and deteriorated wheels. Using acceleration data, Belotti *et al*. (2006) developed a wavelet-based method to detect the bogie with damaged wheels. With the FBG technique, Wei *et al*. (2011) proposed a condition index (CI) for quantitative evaluation of wheel conditions. The structural health monitoring trends (SHMT) explored by Stratman *et al*. (2007) were used to represent the change of dynamic impact load with respect to time. A noticeable increase in the SHMT within a limited period indicates the presence of failed wheels. Filograno *et al*. (2013) designed a high-pass

*Corresponding author, Chair Professor,
E-mail: ceyqni@polyu.edu.hk
[a] M.Sc., Ph.D. Candidate
[b] Ph.D., Research Assistant Professor

ISSN: 1738-1584 (Print), 1738-1991 (Online)

filter to extract the wheel condition-related information. A considerable increase in the energy of the high-pass signal compared to the baseline indicates notable out-of-roundness of a wheel.

However, uncertainties inherent in the monitoring data, which can be caused by the environmental conditions or the stochastic interaction between wheels and rails, cannot be considered in the deterministic algorithms. In this connection, probabilistic algorithms have been increasingly investigated in the past decade, which are expected to yield more robust results (Jamshidi *et al*. 2016). For example, two techniques for wheel defect detection based on machine learning were developed by Krummenacher *et al*. (2017): the first one uses the wavelet transform to extract the features of vertical force measurements, and these features are further used to train a classifier using support vector machine; the second one realizes the automatic classification of different types of wheel defects through convolutional neural networks in the context of deep learning. Liu and Ni (2018) developed an FBG-based monitoring system to detect wheel tread defects. This approach introduced a 5-point quadratic polynomial to smooth every measured data segment of each wheel and Chauvet's criterion was used for anomaly detection of all processed segments individually. A Bayesian forecasting method was employed to identify outliers of acceleration-monitoring data captured from sensors equipped on the car body of a high-speed train (Zhang *et al*. 2018, Wang *et al*. 2020). The deterioration of wheels was indirectly reflected through the anomaly vibration of the car body, thereby the exact faulty wheel cannot be identified by this method. A fully Bayesian inference framework was exploited for online wheel condition monitoring by Ni and Zhang (2020). Enhanced by Bayesian sparse representation, a probabilistic regression model was formulated which achieved robust performance and successfully detected degraded wheels. However, the aforementioned methods fail to take the effect of train running-speed into account.

This study aims to formulate a data-driven, speed-tolerant, and response-only Bayesian-based wheel condition evaluation model in the context of relevance vector machine (RVM) (Tipping 2001). The frequency content of the obtained dynamic strain measurements and train running-speed are configured as the two-dimensional inputs to formulate a multiple probabilistic regression model (MPRM). Due to the time and cost constraints in field experiments, the training data are always sparse and discrete in containing diverse speed information compared to those under actual train operating conditions. However, the explanatory variable of speed in MPRM is still intended to be functionally continuous. In light of this, the MPRM should have satisfactory interpolation and extrapolation performance on the range of speed variables without support training data, which is in accord with the superiority of sparse representation embedded in the RVM algorithm. Considering that the radial basis function (RBF) is widely used in approximate high-precision solutions of various functions or models (Wong *et al*. 2002, Flyer and Wright 2009, Flyer *et al*. 2016), the Gaussian RBF is selected for kernel design of the RVM baseline, which facilitates the data fitting ability. However, the kernel width of RBF determines its range of action, which means the efficiency of an RBF will decrease with the increase of the Euclidean distance between its center and the prediction region. In other words, this property may weaken the prediction capacity of the MPRM. It was observed that the polynomial models with power function terms (PFTs) are adept at data prediction and generalization (Wei *et al*. 2018). Therefore, RBFs in conjunction with PFTs are adopted cooperatively as the kernel functions of the MPRM in this study to ameliorate the naive MPRM that employs RBFs only. A comparative study reveals that the optimized multi-kernel MPRM performs satisfactorily in both accuracy and generalization. Such a property endows the MPRM with favorable interpolation and extrapolation capabilities, which lays the foundation for the wheel defect detection task under various train running-speed conditions.

2. Basis of relevance vector machine and defect detection algorithm

2.1 RVM modeling

To account for the uncertainty and enhance the prediction/generalization capability of the formulated probabilistic regression model, sparse Bayesian learning and RVM (Tipping 2001) are cast in the modeling process. Marin and Robert (2007) have proved the advantage of Bayesian approaches in interpreting uncertainty. The RVM approach in the framework of sparse Bayesian regression is able to prevent overtraining/overfitting and obtain a sparse regression solution, thus avoiding a complex expression of the formulated model and greatly enhancing its prediction capability.

In the existence of noise, the nonlinear mapping between model input $\boldsymbol{x} \in \mathfrak{R}^{M\times D}$ and output $\boldsymbol{y} \in \mathfrak{R}^{M\times 1}$ can be expressed as

$$\boldsymbol{t} = \boldsymbol{y} + \boldsymbol{\epsilon} = \boldsymbol{\Psi w} + \boldsymbol{\epsilon} \tag{1}$$

where $\boldsymbol{t} \in \mathfrak{R}^{M\times 1}$ stands for target vector (measurements of the output), $\boldsymbol{w} = [w_1, w_2, \dots, w_N]^T$ represents the weight parameters; and $\boldsymbol{\epsilon} \in \mathfrak{R}^{M}$ is the additive noise vector with zero mean and variance of σ^2. In this study, the design matrix $\boldsymbol{\Psi} = [\boldsymbol{\Phi}; \boldsymbol{\rho}]$ contains two matrices consisting of various kernel functions. The matrix $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1(\boldsymbol{x}), \boldsymbol{\phi}_2(\boldsymbol{x}), \dots, \boldsymbol{\phi}_M(\boldsymbol{x})]$ consists of M vectors and each vector includes M elements, that is, $\boldsymbol{\phi}_i(\boldsymbol{x}) = [\phi(x_1, x_i), \phi(x_2, x_i), \dots, \phi(x_M, x_i)]^T$. The matrix $\boldsymbol{\rho}$ is made up of several PFTs, and the detailed setting of $\boldsymbol{\rho}$ is illustrated in Section 4.

Then the probability density function of the target vector $\boldsymbol{t}$ can be expressed as

$$P(\boldsymbol{t}|w, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} exp\left\{-\frac{1}{2\sigma^2} \parallel \boldsymbol{t} - \boldsymbol{\Psi w} \parallel^2\right\} \tag{2}$$

To constrain the complexity of the weight parameters $\boldsymbol{w}$, the RVM utilizes a zero-mean Gaussian prior over w_i's with the corresponding hyper-parameters α_i's

$$P(\boldsymbol{w}|\boldsymbol{\alpha}) = \prod_{i=1}^{N} N(0, \alpha_i^{-1}) \tag{3}$$

The posterior probability $P(\boldsymbol{w}, \boldsymbol{\alpha}, \sigma^2|\boldsymbol{t})$ of all the parameters can be decomposed as (Tipping 2001)

$$P(\boldsymbol{w}, \boldsymbol{\alpha}, \sigma^2|\boldsymbol{t}) = P(\boldsymbol{w}|\boldsymbol{t}, \boldsymbol{\alpha}, \sigma^2)P(\boldsymbol{\alpha}, \sigma^2|\boldsymbol{t}) \tag{4}$$

For brevity, σ^2 is replaced by β^{-1}, and the first item on the right-hand side of Eq. (4) is rewritten as

$$P(\boldsymbol{w}|\boldsymbol{t}, \boldsymbol{\alpha}, \beta) \sim N(\boldsymbol{m}, \boldsymbol{\Sigma}) \tag{5}$$

where the mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Sigma}$ are

$$\boldsymbol{m} = \beta \boldsymbol{\Sigma} \boldsymbol{\Psi}^{\boldsymbol{T}} \boldsymbol{t} \tag{6}$$

$$\boldsymbol{\Sigma} = (\mathbf{A} + \beta \boldsymbol{\Psi}^{\boldsymbol{T}} \boldsymbol{\Psi})^{-1} \tag{7}$$

where $\boldsymbol{A} = diag(\boldsymbol{\alpha})$.

To figure out the mean and covariance of the target vector $\boldsymbol{t}$, the values of the hyper-parameters $\boldsymbol{\alpha}$ and β should be iterated to maximize the log marginal likelihood (Tipping 2001)

$$ln P(\boldsymbol{t}|\boldsymbol{\alpha}, \beta)$$
$$\frac{N}{2} ln\beta - E(\boldsymbol{t}) - \frac{1}{2} ln|\boldsymbol{\Sigma}| - \frac{N}{2} ln(2\pi) + \frac{1}{2} \sum_{i=1}^{M} ln\alpha_i \tag{8}$$

Taking partial derivatives of Eq. (8) with respect to α_i's and β, we can obtain

$$\alpha_i = \frac{\gamma_i}{m_i^2} \tag{9}$$

$$\beta = \frac{N - \Sigma_i \gamma_i}{\| \boldsymbol{t} - \boldsymbol{\Psi} \boldsymbol{w} \|^2} \tag{10}$$

where $\gamma_i = 1 - \alpha_i \Sigma_{ii}$.

The initial mean $\boldsymbol{m}$ and covariance $\boldsymbol{\Sigma}$ in Eqs. (6)-(7) can be obtained by setting the hyper-parameters $\boldsymbol{\alpha}$ and β. Then, new $\boldsymbol{\alpha}$ and β can be obtained by maximizing the log marginal likelihood of Eq. (8), and the updated mean and covariance are determined thereafter. After several iteration processes, the final values of $\boldsymbol{m}$ and $\boldsymbol{\Sigma}$ can be attained once the criterion of convergence is satisfied.

The starting values of the hyper-parameters $\boldsymbol{\alpha}$ are extremely small, but with the process of iteration, most of α_i's will tend to infinity, which will result in the weight parameters as follows

$$\lim_{\alpha_i \to \infty} \boldsymbol{\Sigma} = \lim_{\alpha_i \to \infty} (\boldsymbol{A} + \beta \boldsymbol{\Psi}^{\boldsymbol{T}} \boldsymbol{\Psi})^{-1} = 0 \tag{11}$$

$$\Longrightarrow \lim_{\alpha_i \to \infty} \boldsymbol{m} = \lim_{\alpha_i \to \infty} \beta \boldsymbol{\Sigma} \boldsymbol{\Psi}^{\boldsymbol{T}} \boldsymbol{t} = 0 \tag{12}$$

Eqs. (11)-(12) illustrate that most weight parameters will reach the distribution $w_i \sim N(0,0)$, and thus the sparsity of the regression model is obtained automatically. The non-zero w_i parameters along with their corresponding input data are defined as "the relevance vectors".

The predictive distribution over $\boldsymbol{t}^*$ for new input $\boldsymbol{x}^*$ with N points is

$$P(\boldsymbol{t}^*|\boldsymbol{x}^*, \boldsymbol{\alpha}, \beta) = \int P(\boldsymbol{t}^*|\boldsymbol{w}, \beta) P(\boldsymbol{w}|\boldsymbol{t}, \boldsymbol{\alpha}, \beta) d\boldsymbol{w} = \mathcal{N}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \tag{13}$$

with

$$\boldsymbol{\mu}^* = \boldsymbol{\Psi}^* \boldsymbol{m} \tag{14}$$

$$\boldsymbol{\Sigma}^* = \beta^{-1} \boldsymbol{I}_N + \boldsymbol{\Psi}^* \boldsymbol{\Sigma} \boldsymbol{\Psi}^{*T} \tag{15}$$

where $\boldsymbol{\Psi}^*$ is the kernel function matrix of the new input $\boldsymbol{x}^*$. Tipping (2001) suggested ignoring the relatively small second term on the right-hand side of Eq. (15) so that the covariance matrix $\boldsymbol{\Sigma}^*$ could be simplified to a diagonal matrix with uniform elements σ_*^2, and then

$$\sigma_*^2 = \beta^{-1} \tag{16}$$

2.2 Bayesian approach for anomaly detection

To identify and quantify anomalies that deviate from the sparse Bayesian regression model, two hypotheses H_0 and H_1 are defined: the null hypothesis H_0 represents a statistical model in healthy status; while the hypothesis H_1 shifting the expectation of the null hypothesis H_0 with a pending h stands for a statistical model in anomaly status (Lipowsky *et al.* 2002). The statistical models H_0 and H_1 are $\mathcal{N}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ and $\mathcal{N}(\boldsymbol{\mu}^* + h, \boldsymbol{\Sigma}^*)$ respectively. Bayes factor (BF) (Marin and Robert 2007) defined as Eq. (17) can be employed as an indicator for the selection of H_0 and H_1 models on the basis of observation $\boldsymbol{y_p}$

$$BF = \frac{P(\boldsymbol{y_p}|H_1)}{P(\boldsymbol{y_p}|H_0)} = exp\left[\frac{1}{2}\left(\boldsymbol{\varepsilon_0^T} \boldsymbol{\Sigma^*}^{-1} \boldsymbol{\varepsilon_0} - \boldsymbol{\varepsilon_1^T} \boldsymbol{\Sigma^*} \boldsymbol{\varepsilon_1}\right)\right] \tag{17}$$

where $\boldsymbol{\varepsilon_0} = \boldsymbol{y_p} - \boldsymbol{\mu}^*$ and $\boldsymbol{\varepsilon_1} = \boldsymbol{y_p} - (\boldsymbol{\mu}^* + h)$; and $\boldsymbol{y_p}$ is the model output obtained from the RVM-based MPRM. In this study, an alternative expression that interprets BF on the log scale (Aitkin 1991, Kass and Raftery 1995) is adopted, which is

$$BF_{2ln} = 2ln\,(BF) = \boldsymbol{\varepsilon_0^T} \boldsymbol{\Sigma^*}^{-1} \boldsymbol{\varepsilon_0} - \boldsymbol{\varepsilon_1^T} \boldsymbol{\Sigma^*} \boldsymbol{\varepsilon_1} \tag{18}$$

With the intention of avoiding BF to be affected by the scale of data, a normalization step to make this factor scale-invariant is carried out

$$NBF_{2ln} = \frac{BF_{2ln}}{N} = \frac{1}{N}\left(\boldsymbol{\varepsilon_0^T} \boldsymbol{\Sigma^*}^{-1} \boldsymbol{\varepsilon_0} - \boldsymbol{\varepsilon_1^T} \boldsymbol{\Sigma^*} \boldsymbol{\varepsilon_1}\right) \tag{19}$$

where N is the number of elements in the output vector $\boldsymbol{y_p}$.

As demonstrated in Eq. (17), a BF smaller than one illustrates that the observed $\boldsymbol{y_p}$ is in accord with the hypothesis of the wheel in healthy status; otherwise, the data is in favor of the hypothesis that the wheel is defective. Correspondingly, according to Eq. (19), if NBF_{2ln} is

Table 1 Relationship between significance level ψ and shifting parameter h

ψ	61.7%	31.7%	13.36%	4.55%	1.24%	0.27%
h	$1\sigma_*$	$2\sigma_*$	$3\sigma_*$	$4\sigma_*$	$5\sigma_*$	$6\sigma_*$

negative, no damage is indicated; while when NBF_{2ln} is positive, damage is flagged, and a larger NBF_{2ln} warns a more serious defect. The choice of the shifting parameter h affects the risk of false-positive alarm in wheel defect detection, and the risk can be quantified by the significance level ψ (Zhang 2020)

$$\psi = 2\left[1 - \xi\left(\frac{h}{2\sigma_*}\right)\right] \tag{20}$$

where ξ is the cumulative distribution function of the standard normal distribution. The typical values of the significance level ψ and the corresponding shifting parameter h are given in Table 1.

3. Monitoring system and features of measured strain data

3.1 FBG-based wheel condition monitoring system

An FBG-strain gauge array system along with its data acquisition instruments is illustrated in Fig. 1. In this system, two FBG arrays are deployed along the foot of rail tracks on both sides. In recognizing that the circumference of the train wheels in field experiments is about 2.9 m, the length of the sensor array is set to be 3 m to capture the complete strain response during the passage of each wheel (rolling). Each array contains 21 FBG strain sensors evenly distributed on the rail foot with 0.15 m intervals, and such a dense arrangement can characterize abundant wheel defect detection performance (Liu and Ni 2018). This system also includes a high-performance optical interrogator along with a computer that controls the entire data acquisition procedure, and makes the sampling rate in data acquisition as high as 5000 Hz.

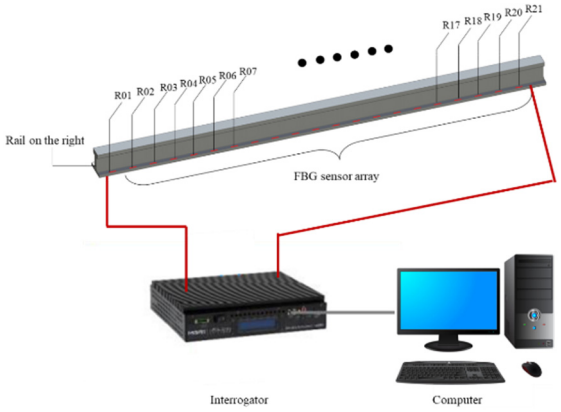

Fig. 1 FBG-based monitoring system

3.2 Features of measured strain data

It is known from previous studies that the low-frequency components of rail dynamic strain responses are mainly caused by wheelbases and axle loads (Zhang 2020), while the high-frequency ingredients contain more wheel-rail surface roughness relevant information (Wei *et al*. 2011, Filograno *et al*. 2013). Filograno *et al*. (2013) suggested dividing the response measurements into high-frequency portions and low-frequency portions with a demarcation f_d to extract defect-related information from the data, and f_d is defined as

$$f_d = kv \tag{21}$$

where k is the proportional coefficient which is set as 1.0 Hz h/km in this study based on the previous research (Ni and Zhang 2020), and v is the running speed of the train. It has been revealed that utilizing the extracted high-frequency components can prevent the signal from being notably affected by loading conditions (Johansson and Nielsen 2003, Uzzal *et al*. 2008).

The rail strain data collected under different train speeds of 10 km/h, 20 km/h, 30 km/h, 40 km/h, and 50 km/h are preliminarily filtered according to Eq. (21). Fig. 2(a) shows a typical raw strain signal collected by the 10th FBG sensor on the right side rail (shorthand for R10) when an 8-car electric multiple unit (EMU) train passes through the instrumented rail at a speed of 10 km/h. This time-series record contains 32 peaks which represent the strain responses induced by the corresponding 32 wheels on one side of the train (4 wheels for each car). Each deployed sensor as shown in Fig. 1 records its individual strain time history, and a collection of the data acquired from all sensors will be used to offer a comprehensive assessment of wheel condition. Fig. 2(b) illustrates the high-pass filtered component of the signal in Fig. 2(a) with frequencies higher than 10 Hz. The filtered data around the 32 peaks located in Fig. 2(a) are selected and segmented, each segment encompassing 1500 data points (which is about 0.3s when the speed is 10 km/h); the length of the data segments is scaled according to the running speed.

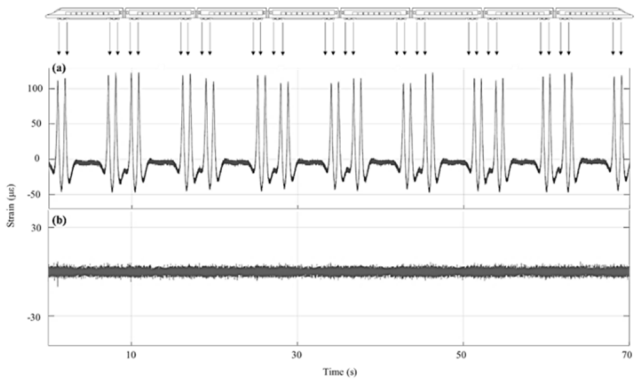

Fig. 2 Measured strain response of R10: (a) raw data; (b) high-pass filtered data

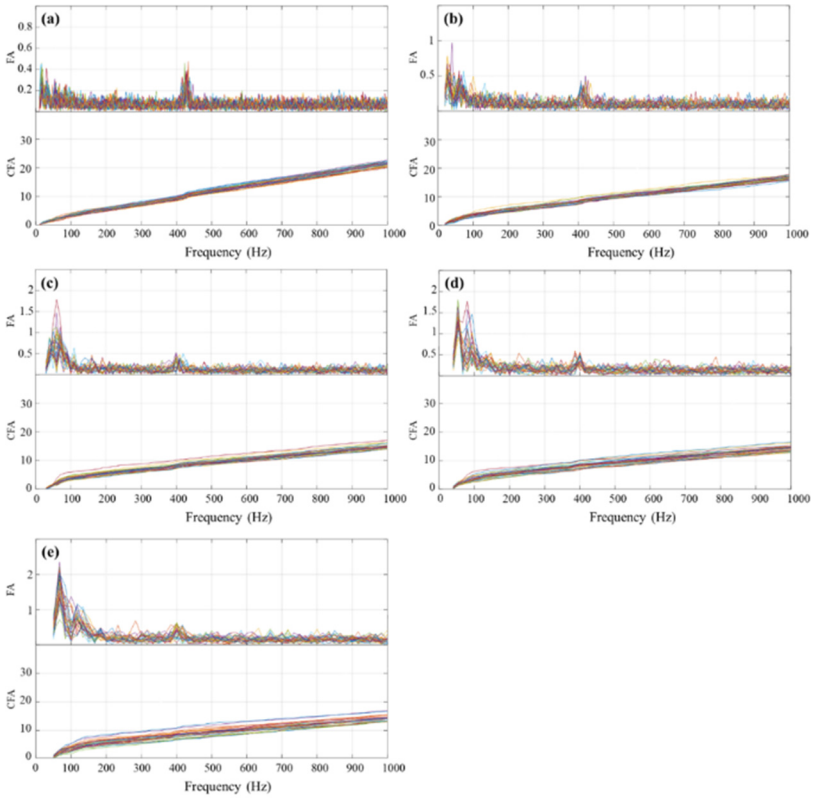

Fig. 3 FA and CFA of strain response segments under different running speeds: (a) 10 km/h; (b) 20 km/h; (c) 30 km/h; (d) 40 km/h; (e) 50 km/h

Fig. 3 depicts the Fourier amplitudes (FA) of the extracted strain response segments along with their corresponding cumulative Fourier amplitudes (CFA) under various running-speed conditions acquired by the sensor R10. The lower limits of the concerned frequency bands are determined by Eq. (21) according to the speed condition of each measured strain data and the upper limit is empirically fixed at 1000 Hz to ensure sufficient information contained in the frequency band. It can be seen that the FAs in Fig. 3 are fluctuating, and such variation is not conducive to modeling. To make the results more consistent and smoother, we introduce CFA by defining

$$y_n = \sum_{i=1}^{n} x_i \tag{22}$$

where x_i represents the value of Fourier amplitude at each frequency point and y_n is the corresponding converted CFA. All the strain dynamic response data collected by the monitoring system are pre-processed in the preceding way. Subsequently, the obtained CFAs will be used for the formulation of MPRM.

4. RVM-based MPRM formulation for measured strain data

4.1 Bayesian approach for anomaly detection

Since the frequency and train speed are considered as two input variables of the RVM-based MPRM, the dimension $\boldsymbol{D}$ of $\boldsymbol{x}$ is 2. The two-dimensional $\boldsymbol{\phi}_i(\boldsymbol{x})$ in the design matrix $\boldsymbol{\Phi}$ is given as

$$\begin{aligned}\boldsymbol{\phi}_i(\boldsymbol{x}) &= \boldsymbol{\phi}\left(\begin{bmatrix} x_i^{D1} \\ x_i^{D2} \end{bmatrix}\right) \\ &= exp\left(\frac{\|x^{D1} - x_i^{D1}\|^2 + \|x^{D2} - x_i^{D2}\|^2}{-\gamma^2}\right)\end{aligned} \tag{23}$$

where x^{D1} represents the variable of speed and x^{D2} represents the variable of frequency; and γ is the kernel width, which is a parameter affecting the accuracy and sparsity of the model, that needs to be predetermined. It is observed that with the increase of the kernel width, the number of the relevance vectors gradually decreases, and the model is transferred from overfitting to underfitting (Ben-Hur *et al.* 2008). Therefore, determination of the kernel width is crucial for the MPRM establishment.

To determine an appropriate kernel width γ, two indicators, i.e., root-mean-square error (RMSE) and redundancy rate (RR), are evaluated. The RMSE or root-mean-square deviation is a general index measuring the difference between model-predicted results and observations, which can compare the prognosis abilities of different models of a prescribed dataset (Hyndman and Koehler 2006, Willmott and Matsuura 2006, Pontius *et al.* 2008). The RMSE is defined as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2} \tag{24}$$

where y_i is the CFA values and t_i is the corresponding MPRM output; N is the number of training data points. The indicator RR is defined as the ratio between the number of relevant vectors and the training data. This index reflects the model fitting effect of a formulated model to prevent the over-fitting or under-fitting in a roundabout way. On the premise of ensuring accuracy, the RR index is desired to be smaller to increase the smoothness and generalization of the model. The RR is given by

$$RR = \frac{N_w}{N} \times 100\% \tag{25}$$

where N_w is the number of the relevance vectors. The smaller the RR index, the less the number of relevance vectors needed, and the better the sparse performance of the model.

The CFAs of the filtered and segmented data acquired by R10 under various train speeds are utilized to formulate the MPRM based on the RVM algorithm. A series of γ varying from 0.1 to 300 along with its corresponding RMSE and RR are shown in Fig. 4 for the determination of a suitable kernel width. It can be seen that the RMSE fluctuates and gradually rises for the kernel width γ ranging between 0.1 and 110. There is a notable increase in RMSE when the kernel width changes from $\gamma = 110$ to $\gamma = 120$ and then the trend stabilizes. The RR displays a considerable decline when the kernel width changes from $\gamma = 0.1$ to $\gamma = 60$, and the subsequent drop of the RR is relatively mild. Considering the tendencies of the above two indicators comprehensively, it is most cost-effective when the kernel width is set to be $\gamma = 110$. In order to examine the extrapolation capability of the formulated MPRM model, the monitoring data at the running speed of 50 km/h are excluded from the MPRM formulation. About 1% (34 terms) of the total training data points are deduced as the relevance vectors through the RVM algorithm. According to Eqs. (1) and (23), the 34 non-zero weights along with the associated kernels of the MPRM are

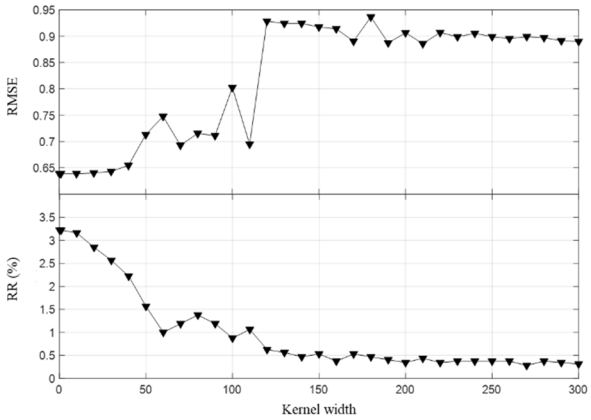

Fig. 4 Kernel width against RMSE and RR

$$\begin{aligned}\boldsymbol{w} = [&2.31, -27.12, 57.13, 934.68, -1418.31, 1379.72, 320.02, 1384.24, -1881.42, 2177.84,\\ &1825.62, 467.76, -40.33, 75.44, -528.56, -3189.41, -1685.66, -68.27, 31.59, -2175.96,\\ &-1977.22, 16.42, -110.02, 214.85, 517.79, 2286.87, -2468.73, 1991.88, 945.63, -950.73,\\ &1073.52, 856.44, 294.15, -123.76]^{\boldsymbol{T}}\end{aligned}$$

$$\begin{aligned}\Psi_{S0} = \Big[&\phi\left(\begin{bmatrix}10\\10\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\80\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\320\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\400\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\400\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\480\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\480\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\720\end{bmatrix}\right),\\ &\phi\left(\begin{bmatrix}10\\760\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\800\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\880\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\960\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\1000\end{bmatrix}\right), \phi\left(\begin{bmatrix}20\\40\end{bmatrix}\right), \phi\left(\begin{bmatrix}20\\360\end{bmatrix}\right), \phi\left(\begin{bmatrix}20\\840\end{bmatrix}\right),\\ &\phi\left(\begin{bmatrix}20\\840\end{bmatrix}\right), \phi\left(\begin{bmatrix}30\\30\end{bmatrix}\right), \phi\left(\begin{bmatrix}30\\240\end{bmatrix}\right), \phi\left(\begin{bmatrix}30\\520\end{bmatrix}\right), \phi\left(\begin{bmatrix}30\\680\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\80\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\280\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\320\end{bmatrix}\right),\\ &\phi\left(\begin{bmatrix}40\\480\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\560\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\600\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\640\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\720\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\760\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\800\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\880\end{bmatrix}\right),\\ &\phi\left(\begin{bmatrix}40\\960\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\1000\end{bmatrix}\right)\Big]\end{aligned}$$

where x^{D1} is the input variable of speed dimension and x^{D2} is the input variable of frequency dimension. For instance, $\boldsymbol{\phi}\left(\begin{bmatrix}10\\80\end{bmatrix}\right)$ indicates that a derived relevance vector has its speed of 10 km/h and frequency of 80 Hz, and the associated weight is 2.31 from the above expressions. For model uncertainty, the unified variance of this MPRM is $\sigma_*^2 = \beta^{-1} = 0.49$ according to Eq. (16). In the formulation of this MPRM, we intentionally exclude the data collected at the speed of 50 km/h with the intent to verify the extrapolation performance of the formulated MPRM. This model is shown in Fig. 5. It is seen that while the predicted results (extrapolation) of the model at 50 km/h are less accurate than those at other speeds, the model is capable of extrapolation. Fig. 6 shows a reference MPRM model which is trained by using the data covering all 5 available speeds from 10 km/h to 50 km/h. It is seen that the reference model fits the data better than the previous extrapolation model. Also, an MPRM model is trained without using the data collected at the speed of 40 km/h with the intent to verify the interpolation performance of the formulated MPRM. This model is shown in Fig. 7. Figs. 8 and 9 illustrate comparisons between the reference model and the extrapolation/interpolation models. It is apparent that the interpolation performance is superior to the extrapolation performance. For the sake of convenience in expression, in this study the reference model refers to the MPRM model trained using data under the speed conditions of 10-50 km/h, while the extrapolation model is trained without using data at 50 km/h, and the interpolation model is trained without using data at 40 km/h.

4.2 Model optimization

The design matrix of the RVM kernels can be implanted

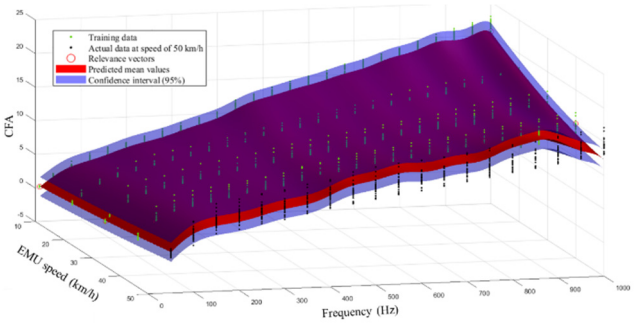

Fig. 5 Extrapolation case: MPRM formulated without using the data collected at 50km/h speed

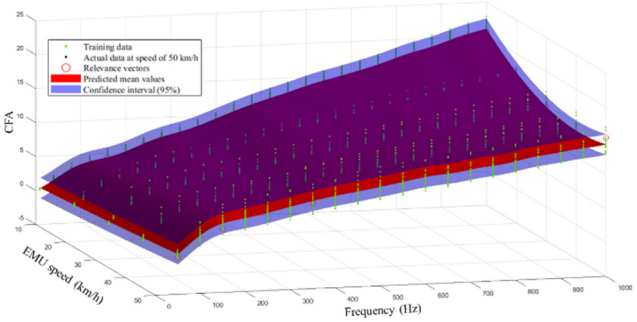

Fig. 6 Reference model using data at all available speeds

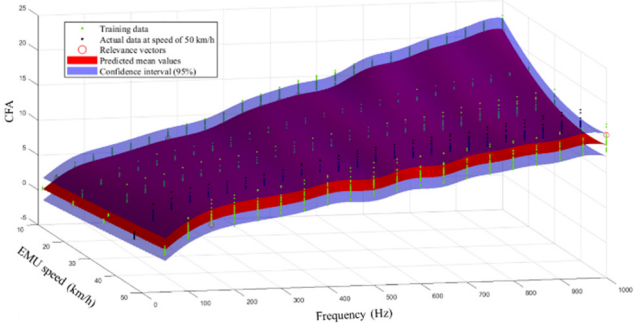

Fig. 7 Interpolation case: MPRM formulated without using the data collected at 40km/h speed

flexibly, which makes it possible to enhance the generalization capability of the MPRM model. In this section, the PFTs appending to the previous RBFs' design matrix are introduced for the sake of model optimization. Benefiting from the embedding of sparse characteristics, the RVM-based model can conclude the indispensable items in PFTs automatically while expunging unnecessary elements (whose weight coefficients are nearly equal to zero). From the preceding section, it is observed that the expression of the MPRM is not complicated. Therefore, the exponentials of PFTs are empirically set as [0:0.1:5] containing 51 terms, which are deemed sufficient. Then two novel design matrix scenarios are defined

$$\boldsymbol{\Psi}_{S1} = [\boldsymbol{\Phi}, \boldsymbol{\rho}_{D1}] = [\boldsymbol{\phi}_1(\boldsymbol{x}), \ldots, \boldsymbol{\phi}_M(\boldsymbol{x}), \boldsymbol{x}_{D1}^0, \ldots, \boldsymbol{x}_{D1}^5] \quad (26)$$

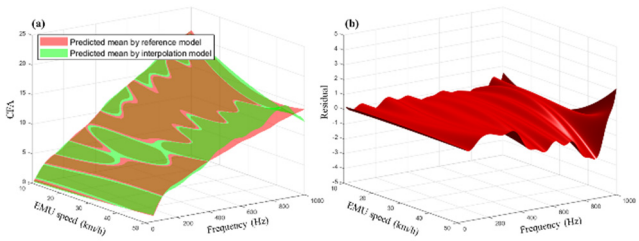

Fig. 8 (a) Comparison of the predicted mean values between the extrapolation model and reference model; (b) residuals between the means

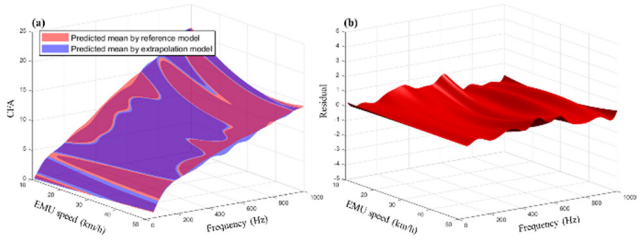

Fig. 9 (a) Comparison of the predicted mean values between the interpolation model and reference model; (b) residual between the means

$$\begin{aligned}\boldsymbol{\Psi}_{S2} &= [\boldsymbol{\Phi}, \boldsymbol{\rho}_{D1}, \boldsymbol{\rho}_{D2}] \\ &= [\boldsymbol{\phi}_1(\boldsymbol{x}), \ldots, \boldsymbol{\phi}_M(\boldsymbol{x}), \boldsymbol{x}_{D1}^0, \ldots, \boldsymbol{x}_{D1}^5, \boldsymbol{x}_{D2}^0, \ldots, \boldsymbol{x}_{D2}^5]\end{aligned} \quad (27)$$

where scenario 1 (S1) attaches PFTs to the speed variable only, while scenario 2 (S2) appends PFTs to both speed and frequency variables. For S2, the weight parameters of the two input dimensions in PFTs are independent since the trends of the model with respect to the two variables (speed and frequency) are anisotropic. Three scenarios including the original case (set as S0) where only RBF is employed in kernel design are further explored and compared.

Fig. 10 illustrates the variation of RMSE along with the corresponding number of terms including RBFs and PFTs (if any) with respect to kernel width under scenarios S0, S1 and S2 using 10-40 km/h training data. Since the definition of RR is the ratio between the number of relevance vectors to the training data, considering that the PFTs are additional elements that cannot be counted into the total number of training data and in order to make the comparison easier, the total number of terms in each condition is considered as a more intuitive measure of the model redundancy. For the indicators RMSE and the number of terms, it is found that the scenarios S1 and S2 exhibit superiority over S0 in that the precision is improved significantly at no cost to model complexity, especially when the kernel width is larger than 100. With the priority of model accuracy, the appropriate kernel width is 120 for S1 and 260 for S2, respectively. Fig. 11 provides a composition about the number of terms in the design matrix. It is seen that the number of PFTs is stable regardless of the kernel width, which is different from the decreasing trend in the case of RBF. The non-zero weight parameters along with their design matrix in scenarios S1 and S2 are

$$\boldsymbol{w}_{S1} = [-2523.02, -2.267, 2549.81, -1801.45, -19.79, -10.42, 1.35, 10.64, 14.08, 5537.93, 1090.88, -4919.03, 3757.69, -440.63, -1407.60, 499.73, -266.59, 122.10, 9.86, -17.17, 69.17, -44.16, 0.02, 2.02 \times 10^{-3}, -3.88 \times 10^{-9}]^T$$

$$\boldsymbol{\Psi}_{S1} = \phi\left(\begin{bmatrix}10\\10\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\120\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\160\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\200\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\360\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\520\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\720\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\920\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\1000\end{bmatrix}\right), \phi\left(\begin{bmatrix}20\\20\end{bmatrix}\right), \phi\left(\begin{bmatrix}20\\240\end{bmatrix}\right), \phi\left(\begin{bmatrix}30\\40\end{bmatrix}\right), \phi\left(\begin{bmatrix}30\\80\end{bmatrix}\right), \phi\left(\begin{bmatrix}30\\280\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\120\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\160\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\200\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\320\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\520\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\960\end{bmatrix}\right), x_{D1}^{0}, x_{D1}^{0.1}, x_{D1}^{1.4}, x_{D1}^{1.7}, x_{D1}^{5}\Big]$$

$$\boldsymbol{w}_{S2} = [128.80, 135.48, -2.57, -58.54, -34.77, -312.92, 129.65, -70.49, 493.12, -784.51, 70.10, -29.11, 0.068, -1.25 \times 10^{-08}, 0.0023, -5.40 \times 10^{-13}]^T$$

$$\boldsymbol{\Psi}_{S2} = \Big[\phi\left(\begin{bmatrix}10\\10\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\120\end{bmatrix}\right), \phi\left(\begin{bmatrix}10\\600\end{bmatrix}\right), \phi\left(\begin{bmatrix}20\\520\end{bmatrix}\right), \phi\left(\begin{bmatrix}30\\800\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\160\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\560\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\960\end{bmatrix}\right), \phi\left(\begin{bmatrix}40\\1000\end{bmatrix}\right), x_{D1}^{0}, x_{D1}^{0.1}, x_{D1}^{1.2}, x_{D1}^{5.0}, x_{D1}^{1.9}, x_{D1}^{5.0}\Big]$$

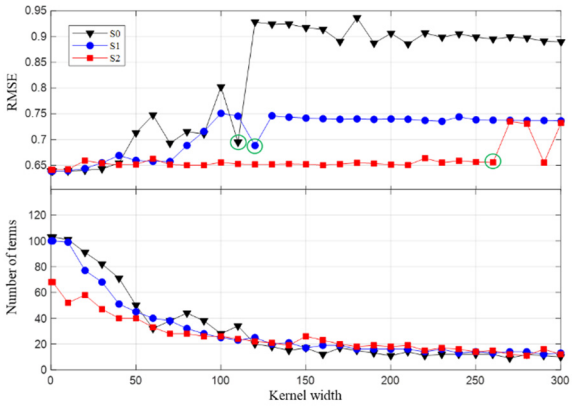

Fig. 10 Kernel width against RMSE and RR under scenarios S0, S1 and S2

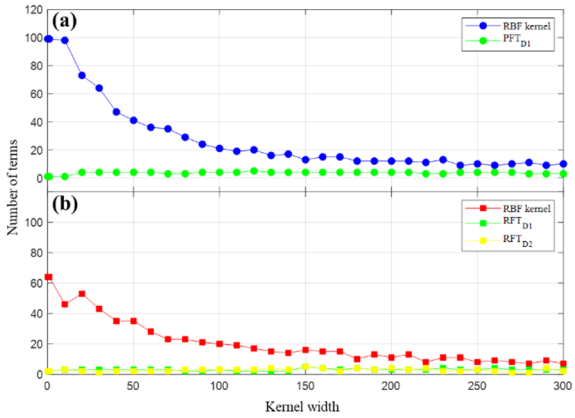

Fig. 11 Number of RBF kernels and PFTs against kernel width under (a) S1; (b) S2

Fig. 12 compares the RMSE and the number of terms among scenarios S0, S1 and S2. The comparison suggests that S2 achieves better performance than S0 and S1 with the fewest relevance vectors (15 terms) and the smallest RMSE value. Figs. 13 and 14 provide a comparison among the extrapolation, interpolation, and reference models in S1. It is seen that after incorporating PFTs to the speed variable, the S1 model significantly improves the extrapolation performance, but its improvement to the interpolation performance is not significant. Figs. 15 and 16 provide a comparison among the extrapolation, interpolation, and reference models in S2. It is observed that both extrapolation performance and interpolation performance are largely enhanced after incorporating PFTs to both speed and frequency variables. Besides, the absolute values of residuals between the interpolation/extrapolation models and the reference model under these scenarios are further used for quantitative comparison. Two metrics are adopted for this purpose: (a) the average of the absolute residuals (AAR); and (b) the maximum of the absolute residuals (MAR). The results are shown in Fig. 17, which justify again that the S2 model outperforms the other two and it can significantly improve both interpolation performance and extrapolation performance.

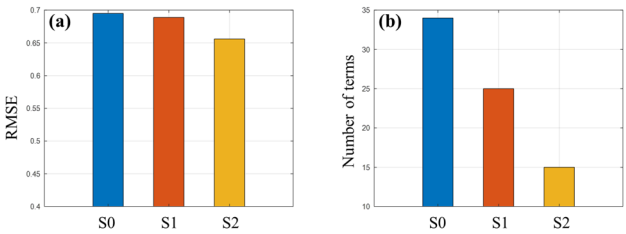

Fig. 12 Comparison among scenarios S0, S1 and S2: (a) RMSE; (b) Number of terms

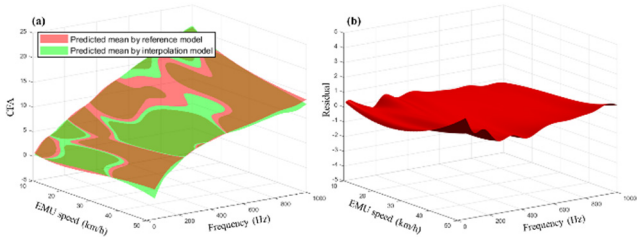

Fig. 13 (a) Comparison of the predicted mean values between the S1 extrapolation model and reference model; (b) residual between the means

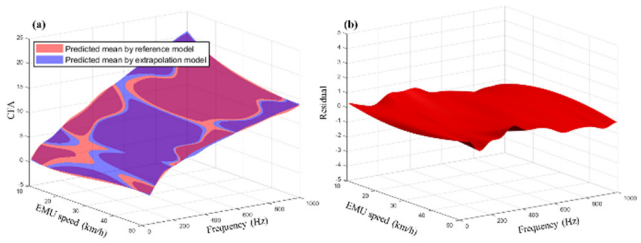

Fig. 14 (a) Comparison of the predicted mean values between the S1 interpolation model and reference model; (b) residual between the means

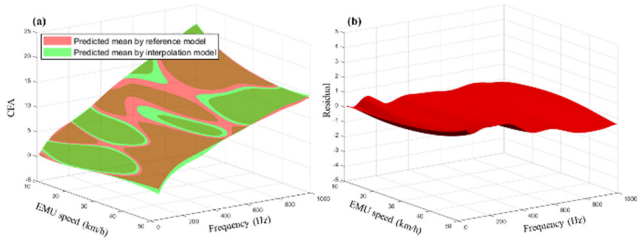

Fig. 15 (a) Comparison of the predicted mean values between the S2 extrapolation model and reference model; (b) residual between the means

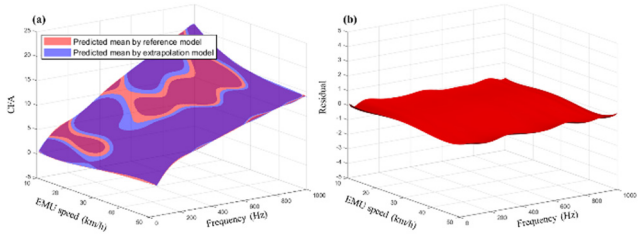

Fig. 16 (a) Comparison of the predicted mean values between the S2 interpolation model and reference model; (b) residual between the means

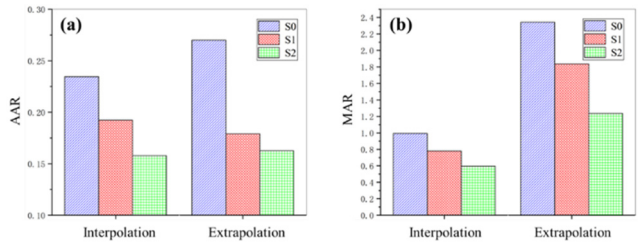

Fig. 17 Quantitative comparison among scenarios S0, S1 and S2: (a) AAR; (b) MAR

From the above results, the multi-kernel RVM-based MPRM with refinement of PFTs is perceived to have the following advantages:

- The accuracy of the formulated model is improved;
- Both interpolation performance and extrapolation performance are enhanced, which are essential for MPRM to work under a wide range of speed conditions with only sparse training data;
- The number of terms in the optimized MPRM model (S2 with 15 terms) is significantly reduced in comparison with the original model (S0 with 34 terms).

5. Diagnosis of wheel defects

5.1 Single-sensor diagnosis performance

As aforementioned, the formulation of MPRM utilizes only the dynamic strain response measurements of a rail segment generated by the impact of passing wheels in healthy state, and any undetermined data that deviates from the trained model will be viewed as a symptom of the occurrence of suspicious defects. Making use of the hypothesis testing strategy delineated in Section 2.2, the possible defective wheels can be identified with their damage severity being quantified. In the present study, the offset h is set as $6 \times \beta^{-\frac{1}{2}}$ resulting in a 0.27% probability of misidentifying an intact wheel as damaged. One may choose different values of h according to specific risk mitigation requirements, e.g., a compromise between the risks of false-positive and false-negative.

Fig. 18(a) depicts the NBF_{2ln} values derived from the reference model including all available speed conditions of sensor R10. These testing strain responses are excited by a running EMU train with several deteriorated wheels. The 1st right wheel is diagnosed as defective under all 5 various speed statuses, whose NBF_{2ln} values are all over 50. Besides, the NBF_{2ln} values of the 4th and 29th right wheels also warn defects only when train speed is 50km/h while neither of them over 5. It is claimed that a NBF_{2ln} exceeding 10 shows significantly compelling evidence against the hypothesis H_0 (Kass and Raftery 1995), which means the 1st wheel has a very high possibility of damage while the evidence of the 4th and 29th wheels is not strong.

The actual running speed of a train may be different from those used in MPRM formulation, thereby the feasibility of the formulated MPRM in interpolation/ extrapolation needs to be verified. Figs. 18(b) and (c) provide a comparison of the NBF_{2ln} values obtained from the reference model in Fig. 18(a) and from the interpolation/extrapolation models given in the previous section: Fig. 18(b) shows the NBF_{2ln} values obtained from the interpolation and reference models when the train speed is 40 km/h, while Fig. 18(c) shows the difference of NBF_{2ln} values between the extrapolation and reference models at the speed of 50 km/h. It is found that the extrapolation model identifies 3 suspected damaged wheels and also the reference model does. However, the difference of NBF_{2ln} values between the extrapolation and the reference models is more striking than that between the interpolation and the reference models, which indicates that the performance of interpolation is more reliable. After verification of the interpolation/extrapolation performance, the raw data can be directly inputted into the model to accomplish the wheel condition assessment within a range rather than a specific value of the speed variable, which is essential for practical application of the proposed method. If no damage is flagged, the detected data can further be used to update the existing model to enhance its performance. In

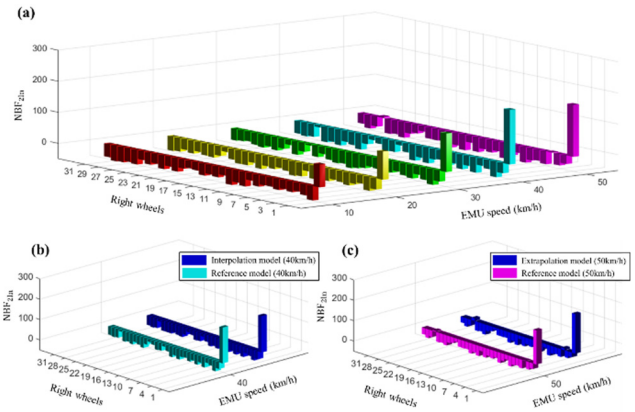

Fig. 18 Diagnosis of right wheels using data from R10: (a) NBF_{2ln}'s of the reference model; (b) comparison of NBF_{2ln}'s between the interpolation and reference models at speed of 40 km/h; (c) comparison of NBF_{2ln}'s between the extrapolation and reference models at speed of 50 km/h

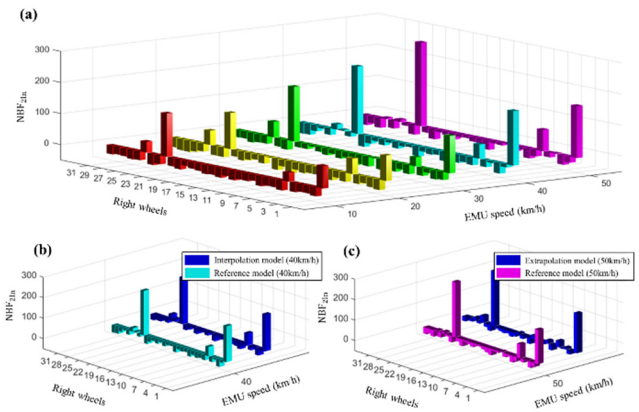

Fig. 19 Diagnosis of right wheels using data from all available sensors: (a) NBF_{2ln}'s of the reference model; (b) comparison of NBF_{2ln}'s between the interpolation and reference models at speed of 40 km/h; (c) comparison of NBF_{2ln}'s between the extrapolation and reference models at speed of 50 km/h

view of the fact that faulty wheels are rare among all wheels in a train (Liu and Ni 2018), the sparse representation property of the RVM baseline can make the regression model robust and immune to interference from the case that data with unobservable defect-relevant signals are fed into the model by mistake. This is because of the merit that the regression model will not be substantially affected by a small amount of abnormal data.

5.2 Multi-sensor diagnosis performance

Since an individual sensor may not be sufficient to detect all potential defective wheels, we further conduct the diagnosis by simultaneously using the data from all the deployed sensors. The integrative diagnosis result for each wheel is defined as the largest NBF_{2ln} derived from MPRMs of individual sensors to maximize the diagnostic efficacy

$$NBF_{2ln} = max\{NBF_{2ln}{}^{1}, NBF_{2ln}{}^{2}, \dots, NBF_{2ln}{}^{S}\} \quad (28)$$

where S herein refers to the number of FBG sensors installed on each side of the rail segment, which is equal to 21 in this study. Fig. 19(a) shows the overall condition diagnosis results of wheels on the right side obtained by the reference models including all available speeds. This figure indicates 4 potential defective wheels with positive NBF_{2ln} values under all speed conditions, that is, the 1st, 6th, 24th, and 27th wheels. Compared with the diagnosis results using only single sensor R10, the integrative results reveal three more suspicious damaged wheels. In addition, there are four inconspicuous positive NBF_{2ln} values smaller than 5 only when the speed is higher than 40 km/h. Figs. 19(b) and (c) show the interpolation and extrapolation results compared with the corresponding results of the reference models. It is found that the results of interpolation models match those of the reference models better; while several additional marked defective wheels appear in the extrapolation models

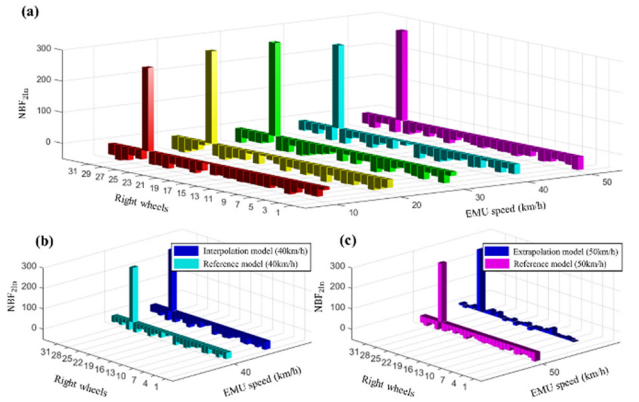

Fig. 20 Diagnosis of left wheels using data from all available sensors: (a) NBF_{2ln}'s of the reference model; (b) comparison of NBF_{2ln}'s between the interpolation and reference models at speed of 40 km/h; (c) comparison of NBF_{2ln}'s between the extrapolation and reference models at speed of 50 km/h

compared to the results of the reference models. A similar situation occurs for wheels on the left side, where only the 27th wheel presents relatively large NBF_{2ln} values in the diagnosis results of both reference and interpolation models, shown in Figs. 20(a) and (b). However, almost every wheel flags tiny damage in the extrapolation scenario, which is illustrated in Fig. 20(c). According to the discussions given in Section 3.4, the performance of extrapolation is worse than that of interpolation, so the diagnosis results from the extrapolation model need to be treated with caution.

5.3 Offline inspection

To verify the proposed online wheel defect detection method, the offline wheel condition diagnosis results through the radius deviation technique were obtained, as

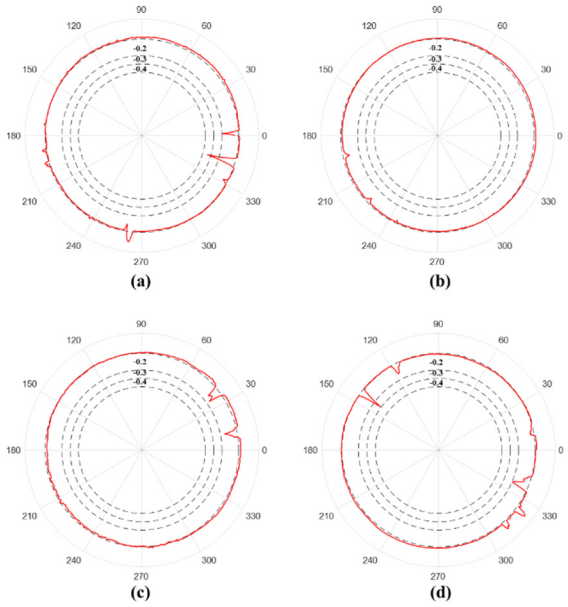

Fig. 21 Offline inspection results of four verified defective wheels: (a) the 1st right-side wheel; (b) the 6th right-side wheel; (c) the 24th right-side wheel; (d) the 27th left-side wheel

Table 2 Details of wheel defects through offline inspection

Wheelset	Side	Defect Qty	Depth of wheel defects (mm)			
1	Right	3	0.055	0.342	0.200	-
6	Right	2	0.062	0.056	-	-
24	Right	2	0.162	0.151	-	-
27	Left	4	0.292	0.126	0.044	0.196

illustrated in Fig. 21. It is found that there are four wheels with real defects, which are the 1st, 6th, and 24th wheels on the right side and the 27th wheel on the left side, and the details of the detected wheel defects are given in Table 2. The results of offline inspection indicate that the online diagnostic results by the extrapolation model are relatively inaccurate compared with the interpolation and reference models. It can be seen from Fig. 19 that four wheels on the right side have been identified as damaged under all speed conditions: the 1st, 6th, 24th and 27th wheels, but the 27th wheel is confirmed intact in the offline inspection. In Fig. 20, the diagnosis results of the reference and interpolation models only flag one defective wheel on the left side, i.e., the 27th wheel, which is consistent with the offline inspection. The fairly large NBF_{2ln} values reveal a serious defect of the 27th wheel which can be observed in Fig. 21(d) that the left-side 27th wheel is indeed the most degraded. It can be concluded that the 27th wheel on the right side is falsely alarmed due to the influence of the opposite left-side wheel, that is, a severely damaged wheel can affect the dynamic pattern of its opposite wheel (Liu and Ni 2018). Apart from the above wheels, the NBF_{2ln} values of right-side wheels No. 4, 11, 26, and 29 are also positive when the speed is over 40 km/h, while they are determined to be healthy in offline inspection. It is worth

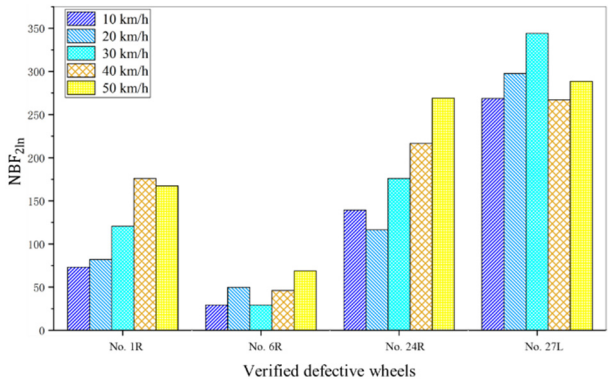

Fig. 22 NBF_{2ln}'s derived from the reference models on four verified defective wheels

mentioning that the NBF_{2ln} values of the above false-positive wheels are all less than 10. For a clearer illustration, the NBF_{2ln} values derived from the reference models of the four actually defective wheels are extracted and plotted in Fig. 22. It can be seen that the NBF_{2ln} values of the 1st, 24th and 27th wheels are all over 50 while that of the 6th wheel is much lower, which is in line with the damage severity detected by offline inspection that the 1st, 24th, and 27th wheels are seriously deteriorative and the degradation to the 6th wheel is relatively minor. Even though the depth of defect on the 6th wheel is only about 0.06 mm, its NBF_{2ln} values are over 20 in all available speed cases, which provides sufficient evidence for damage identification. It is easy to find that the offline inspection results match well with those of the reference/interpolation models, but differ from the results of the extrapolation model, which should be treated with caution in practical application of the proposed method.

5.4 Discussion

The feasibility of the online wheel defect identification method proposed in this study has been verified through comparison with the offline radius deviation inspection results. The following issues are worthy of further discussions and explorations for better applications of the proposed method:

Considering that the length of the sensor array just needs to cover the circumference of the target wheel (3m in this study), it is better to keep the speed of the passing train to be the same as one of the speeds used in the MPRM formulation. If it is difficult, we can alternatively control the speed of the train within the range of running speeds covered by the monitoring data used for formulating the MPRM model, since the interpolated data can still obtain satisfactory identification results. In practical applications, it is preferred to avoid using the monitoring data collected under the running speeds exceeding the range covered by the MPRM model. It is also noteworthy that the dynamic pattern of an intact wheel may be affected by its opposite defective wheel, causing it to be falsely alarmed.

Through the verification of the online wheel defect identification results by offline inspection, it is suggested that main concern be given to the wheels with their

identified NBF_{2ln} values greater than 10, and this value was also recommended by Kass and Raftery (1995) to strongly reject the null hypothesis (the hypothesis that the wheel is healthy).

6. Conclusions

By utilizing dynamic strain measurements obtained by a trackside monitoring system, the RVM-based MPRM procedure is proposed for online wheel defect identification in this study. The train speed as a variable is integrated into the MPRM, enabling it to evaluate the wheel health status under various running-speed conditions. The multi-kernel design matrix of RVM is formulated to include both RBFs and PFTs, which endows the MPRM with high accuracy and robustness in model prediction. In addition to RBFs, the model incorporating PFTs in both speed and frequency variables performs optimally in terms of interpolation and extrapolation. The model where PFTs are introduced only in the speed variable performs second, and the model using only RBFs performs the worst. Meanwhile, the model incorporating PFTs in both speed and frequency variables has the least number of relevance vectors among the three models, which indicates it is the most concise.

The case study illustrates that diagnosis of faulty wheels using a single sensor may result in omissions. The multi-sensor diagnosis strategy can avoid this shortcoming. The study shows that the proposed online wheel condition identification results are in good agreement with those of offline inspection. Through the adoption of NBF_{2ln}, the severity of wheel damage can be estimated. However, it should be noted that if the monitoring data to be used for online diagnosis are collected under the running speeds which exceed the speed range covered by the formulated MPRM model, the diagnosis results might be inaccurate. In this case, the monitoring data collected under the higher running speeds in healthy state of the wheels should be incorporated to reformulate and refine the MPRM model.

Acknowledgments

The work described in this paper is supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (SAR), China (Grant No. PolyU 152014/18E), and a grant from the Hong Kong, Macao, Taiwan Science and Technology Innovation Cooperation Key Project of Sichuan Province, China (Grant No. 2020YFH0178). The authors would also like to appreciate the funding support by the Innovation and Technology Commission of the Hong Kong SAR Government to the Hong Kong Branch of National Rail Transit Electrification and Automation Engineering Technology Research Center (Grant No. K-BBY1).

Yuan-Hao Wei: https://orcid.org/0000-0001-9439-0780
You-Wu Wang: https://orcid.org/0000-0003-2293-4712
Yi-Qing Ni: https://orcid.org/0000-0003-1527-7777

References

Aitkin, M. (1991), "Posterior Bayes factors", *J. Royal Statist. Soc.: Series B (Methodological)*, **53**(1), 111-128. https://doi.org/10.1111/j.2517-6161.1991.tb01812.x

Asplund, M., Famurewa, S. and Rantatalo, M. (2014), "Condition monitoring and e-maintenance solution of railway wheels", *J. Quality Maint. Eng.*, **20**(3), 216-232. https://doi.org/10.1108/JQME-05-2014-0027

Barke, D.W. and Chiu, W.K. (2005), "A review of the effects of out-of-round wheels on track and vehicle components", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, **219**(3), 151-175. https://doi.org/10.1243/095440905X8853

Belotti, V., Crenna, F., Michelini, R.C. and Rossi, G.B. (2006), "Wheel-flat diagnostic tool via wavelet transform", *Mech. Syst. Signal Process.*, **20**(8), 1953-1966. https://doi.org/10.1016/j.ymssp.2005.12.012

Ben-Hur, A., Ong, C.S., Sonnenburg, S., Schölkopf, B. and Rätsch, G. (2008), "Support vector machines and kernels for computational biology", *PLoS Computat. Biol.*, **4**(10), e1000173. https://doi.org/10.1371/journal.pcbi.1000173

Bollas, K., Papasalouros, D., Kourousis, D. and Anastasopoulos, A. (2013), "Acoustic emission monitoring of wheel sets on moving trains", *Constr. Build. Mater.*, **48**, 1266-1272. https://doi.org/10.1016/j.conbuildmat.2013.02.013

Dukkipati, R.V. and Dong, R. (1999), "Impact loads due to wheel flats and shells", *Vehicle Syst. Dyn.*, **31**(1), 1-22. https://doi.org/10.1076/vesd.31.1.1.2097

Filograno, M.L., Guillén, P.C., Rodríguez-Barrios, A., Martín-López, S., Rodríguez-Plaza, M., Andrés-Alguacil, Á. and González-Herráez, M. (2011), "Real-time monitoring of railway traffic using fiber Bragg grating sensors", *IEEE Sensors J.*, **12**(1), 85-92. https://doi.org/10.1109/JSEN.2011.2135848

Filograno, M.L., Corredera, P., Rodriguez-Plaza, M., Andres-Alguacil, A. and Gonzalez-Herraez, M. (2013), "Wheel flat detection in high-speed railway systems using fiber Bragg gratings", *IEEE Sensors J.*, **13**(12), 4808-4816. https://doi.org/10.1109/JSEN.2013.2274008

Flyer, N. and Wright, G.B. (2009), "A radial basis function method for the shallow water equations on a sphere", *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **465**(2106), 1949-1976. https://doi.org/10.1098/rspa.2009.0033

Flyer, N., Barnett, G.A. and Wicker, L.J. (2016), "Enhancing finite differences with radial basis functions: experiments on the Navier–Stokes equations", *J. Computat. Phys.*, **316**, 39-62. https://doi.org/10.1016/j.jcp.2016.02.078

Hyndman, R.J. and Koehler, A.B. (2006), "Another look at measures of forecast accuracy", *Int. J. Forecast.*, **22**(4), 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Jamshidi, A., Roohi, S.F., Núñez, A., Babuska, R., De Schutter, B., Dollevoet, R. and Li, Z. (2016), "Probabilistic defect-based risk assessment approach for rail failures in railway infrastructure", *IFAC-PapersOnLine*, **49**(3), 73-77. https://doi.org/10.1016/j.ifacol.2016.07.013

Jin, X., Wu, L., Fang, J., Zhong, S. and Ling, L. (2012), "An investigation into the mechanism of the polygonal wear of metro train wheels and its effect on the dynamic behaviour of a wheel/rail system", *Vehicle Syst. Dyn.*, **50**(12), 1817-1834. https://doi.org/10.1080/00423114.2012.695022

Johansson, A. and Nielsen, J.C. (2003), "Out-of-round railway wheels—wheel-rail contact forces and track response derived from field tests and numerical simulations", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, **217**(2), 135-146. https://doi.org/10.1243/095440903765762878

Kass, R.E. and Raftery, A.E. (1995), "Bayes factors", In: *J. Am. Statist. Assoc.*, **90**(430), 773-795. https://doi.org/10.2307/2291091

Kouroussis, G., Connolly, D.P. and Verlinden, O. (2014), "Railway-induced ground vibrations–a review of vehicle effects", *Int. J. Rail Transport.*, **2**(2), 69-110. https://doi.org/10.1080/23248378.2014.897791

Krummenacher, G., Ong, C.S., Koller, S., Kobayashi, S. and Buhmann, J.M. (2017), "Wheel defect detection with machine learning", *IEEE Transact. Intell. Transport. Syst.*, **19**(4), 1176-1187. https://doi.org/10.1109/TITS.2017.2720721

Lipowsky, H., Staudacher, S., Bauer, M. and Schmidt, K.J. (2010), "Application of Bayesian forecasting to change detection and prognosis of gas turbine performance", *J. Eng. Gas Turbines Power*, **132**(3). https://doi.org/10.1115/1.3159367

Liu, X.Z. and Ni, Y.Q. (2018), "Wheel tread defect detection for high-speed trains using FBG-based online monitoring techniques", *Smart Struct. Syst.*, *Int. J.*, **21**(5), 687-694. https://doi.org/10.12989/sss.2018.21.5.687

Liu, X.Z., Ni, Y.Q. and Zhou, L. (2018), "Condition-based maintenance of high-speed railway vehicle wheels through trackside monitoring", *Proceedings of the Second International Workshop on Structural Health Monitoring for Railway System*, Qingdao, China, October.

Liu, X.Z., Xu, C. and Ni, Y.Q. (2019), "Wayside detection of wheel minor defects in high-speed trains by a Bayesian blind source separation method", *Sensors*, **19**(18), 3981. https://doi.org/10.3390/s19183981

Marin, J.M. and Robert, C. (2007), *Bayesian Core*: *A Practical Approach to Computational Bayesian Statistics*, Springer, New York, NY, USA.

Milković, D., Simić, G., Jakovljević, Ž., Tanasković, J. and Lučanin, V. (2013), "Wayside system for wheel–rail contact forces measurements", *Measurement*, **46**(9), 3308-3318. https://doi.org/10.1016/j.measurement.2013.06.017

Morys, B. (1999), "Enlargement of out-of-round wheel profiles on high speed trains", *J. Sound Vib.*, **227**(5), 965-978. https://doi.org/10.1006/jsvi.1999.2055

Ni, Y.Q. and Zhang, Q.H. (2020), "A Bayesian machine learning approach for online detection of railway wheel defects using track-side monitoring", *Struct. Health Monitor.*, **20**(4), 1536-1550. https://doi.org/10.1177/1475921720921772

Pau, M. (2005), "Ultrasonic waves for effective assessment of wheel-rail contact anomalies", *Proceedings of the Institution of Mechanical Engineers*, *Part F*: *Journal of Rail and Rapid Transit*, **219**(2), 79-90. https://doi.org/10.1243/095440905X8808

Petersson, M. (2000), "Noise-related roughness of railway wheel treads-full-scale testing of brake blocks", *Proceedings of the Institution of Mechanical Engineers*, *Part F*: *Journal of Rail and Rapid Transit*, **214**(2), 63-77. https://doi.org/10.1243/0954409001531342

Pohl, R., Erhard, A., Montag, H.J., Thomas, H.M. and Wüstenberg, H. (2004), "NDT techniques for railroad wheel and gauge corner inspection", *NDT & e Int.*, **37**(2), 89-94. https://doi.org/10.1016/j.ndteint.2003.06.001

Pontius, R.G., Thontteh, O. and Chen, H. (2008), "Components of information for multiple resolution comparison between maps that share a real variable", *Environ. Ecol. Statist.*, **15**(2), 111-142. https://doi.org/10.1007/s10651-007-0043-y

Stratman, B., Liu, Y. and Mahadevan, S. (2007), "Structural health monitoring of railroad wheels using wheel impact load detectors", *J. Fail. Anal. Prevent.*, **7**(3), 218-225. https://doi.org/10.1007/s11668-007-9043-3

Tipping, M.E. (2001), "Sparse Bayesian learning and the relevance vector machine", *J. Mach. Learn. Res.*, **1**(Jun), 211-244. https://doi.org/10.1162/15324430152748236

Uzzal, R.U.A., Ahmed, W. and Rakheja, S. (2008), "Dynamic analysis of railway vehicle-track interactions due to wheel flat with a pitch-plane vehicle model", *J. Mech. Eng.*, **39**(2), 86-94. https://doi.org/10.3329/jme.v39i2.1851

Wang, Y.W., Ni, Y.Q. and Wang, X. (2020), "Real-time defect detection of high-speed train wheels by using Bayesian forecasting and dynamic model", *Mech. Syst. Signal Process.*, **139**, 106654. https://doi.org/10.1016/j.ymssp.2020.106654

Wang, S.M., Jiang, G.F., Ni, Y.Q., Lu, Y., Lin, G.B., Pan, H.L., Xu, J.Q. and Hao, S. (2022), "Multiple damage detection of maglev rail joints using time-frequency spectrogram and convolutional neural network", *Smart Struct. Syst.*, *Int. J.*, **29**(4), 625-640. https://doi.org/10.12989/sss.2022.29.4.625

Wei, C., Xin, Q., Chung, W.H., Liu, S.Y., Tam, H.Y. and Ho, S.L. (2011), "Real-time train wheel condition monitoring by fiber Bragg grating sensors", *Int. J. Distrib. Sensor Networks*, **8**(1), 409048. https://doi.org/10.1155/2012/409048

Wei, B., Xie, N. and Hu, A. (2018), "Optimal solution for novel grey polynomial prediction model", *Appl. Mathe. Modell.*, **62**, 717-727. https://doi.org/10.1016/j.apm.2018.06.035

Willmott, C.J. and Matsuura, K. (2006), "On the use of dimensioned measures of error to evaluate the performance of spatial interpolators", *Int. J. Geograph. Inform. Sci.*, **20**(1), 89-102. https://doi.org/10.1080/13658810500286976

Wong, S.M., Hon, Y.C. and Golberg, M.A. (2002), "Compactly supported radial basis functions for shallow water equations", *Appl. Mathe. Computat.*, **127**(1), 79-101. https://doi.org/10.1016/S0096-3003(01)00006-6

Wu, X. and Chi, M. (2016), "Study on stress states of a wheelset axle due to a defective wheel", *J. Mech. Sci. Technol.gy*, **30**(11), 4845-4857. https://doi.org/10.1007/s12206-016-1003-y

Wu, T.X. and Thompson, D.J. (2002), "A hybrid model for the noise generation due to railway wheel flats", *J. Sound Vib.*, **251**(1), 115-139. https://doi.org/10.1006/jsvi.2001.3980

Wu, Y., Du, X., Zhang, H.J., Wen, Z.F. and Jin, X.S. (2017), "Experimental analysis of the mechanism of high-order polygonal wear of wheels of a high-speed train", *J. Zhejiang Univ.-Sci. A*, **18**(8), 579-592. https://doi.org/10.1631/jzus.A1600741

Zhang, Q. (2020), "Sparse Bayesian learning approach for damage detection in a population of nominally identical structures", Ph.D. Dissertation; The Hong Kong Polytechnic University, Hong Kong.

Zhang, L.H., Wang, Y.W., Ni, Y.Q. and Lai, S.K. (2018), "Online condition assessment of high-speed trains based on Bayesian forecasting approach and time series analysis", *Smart Struct. Syst.*, *Int. J.*, **21**(5), 705-713. https://doi.org/10.12989/sss.2018.21.5.705