The 2017 World Congress on
Advances in Structural Engineering and Mechanics (ASEM17)
28 August - 1 September, 2017, Ilsan(Seoul), Korea

Condition Assessment of In-service High-speed Trains Based on Bayesian Inference and Time Series Analysis

*Lin-Hao Zhang[1)] Yi-Qing Ni[2)] You-Wu Wang[3)] and Chi Xu[4)]

[1), 2)] *National Rail Transit Electrification and Automation Engineering Technology Research Center (Hong Kong Branch)*
[2), 3), 4)] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*
[1)] *zhanglinhao92@gmail.com*

ABSTRACT

Ride quality is an important role in the evaluation of the operational performance of high-speed trains. In this research, on-board monitoring data has been collected by a data acquisition system deployed on an in-service high-speed train for condition assessment. Firstly, the measured original vibration signals of car body floor were processed by Sperling operator and the obtained values of ride quality index were used to represent the train's operation conditions. Next, a novel technique based on combining Bayesian forecasting and time series analysis is adopted for assessing the obtained ride quality index sequence aimed at evaluating the operation condition of high-speed trains. In detail, Bayesian forecasting enables the calculation of conditional probabilities, together with time series analysis, the one-step forecasting probability density functions (PDFs) are available before the next observation. The change detection is carried out by comparing the current model and the alternative model (whose mean value is shifted by a prescribed offset) to determine which one can fit the actual observation better. When the comparison result implies that the alternative model performances better, then a potential change has been detected. Furthermore, to determine if the current observation is a single outlier or the beginning of the trend of significant change, a particular logic called Bayes factor is developed. The significant change represents that there is a great variation of operation performance of high-speed trains occurred. In this paper, the introduced method is employed for condition assessment of high-speed train by using online data acquired under different wheel quality conditions.

[1)] Research assistant
[2)] Professor
[3)] Ph.D. candidate
[4)] Ph.D. student

1. INTRODUCTION

High-speed rail (HSR) has currently received large popularity because of providing comfortable and convenient travel to people. Running at and above 200 km/h speed and operating in adequate traffic density, HSR attracts a lot of consumers in comparison with traditional transportations. To keep up with the requirements for high quality of service and safety of operations, it is meaningful and necessary to conduct condition assessment of high-speed train (HST). During routine operations, HST as a moving system running at such high speeds always produces various dynamic responses. To be specific, wheel-rail interaction is the major inducement of vehicle vibrations, which contains track irregularity (e.g., rail corrugation, worn wheel and rail profiles, track imperfections, etc.) (Remennikov 2008). In particular, some investigations indicate that vibrations at low-frequency in the range of 5-10 Hz are mainly induced by wheel-rail contact point bouncing at two sides of abrasion concave (Huang 2013). Indeed, the low-frequency contributes a lot in producing uncomfortable feelings of passengers (Kim 2003). One of the key finding of our previously research work is that the vibrations at low-frequency range can transfer into car body (Wang 2016). Naturally, different wheel conditions of the vehicle can influence the ride comfort of passengers. Hence, ride comfort as an indicator is employed to reflect operation and service quality of HST in this paper. A vibration-based index in terms of Sperling index (Nakagawa 2010) describing ride quality using for data preprocessing and further analysis.

Dynamic responses of trains are closely related to route states, but only few research for ride comfort evaluation use on-board monitoring data (Kim 2004; Ni 2015). In this investigation, an on-board data acquisition system has been installed on an in-service high-speed train. By equipped with accelerometers and interrogators, the system has measured dynamic performance of car body during routine operations. Particularly, vibrations in two operation conditions corresponding to wheels before and after lathing process has been recorded. The values of ride quality index obtained from the calculation of Sperling operator every 100 seconds using measured online monitoring data. The achieved ride quality index sequence representing the running state of HST is assessed by a novel technique which is based on the application of dynamic linear models (DLMs) and Bayesian forecasting. This detection algorithm has already been evaluated in the field of gas turbine monitoring (Lipowsky 2010). In this research, the introduced method in the context of dynamic linear model and Bayesian forecasting is demonstrated for operation condition assessment of HST by using the monitoring data measured under different wheel quality conditions. Specific logics including Bayes factor and cumulative Bayes factor are formulated to help determine whether the current observation is an outlier or the beginning of a significant change. The detection results are beneficial in terms of monitoring running conditions of HSTs and making wheel maintenance planning.

2. MEASUREMENT SETUP

In this research, the data acquisition system has been employed on an in-service high-speed train. As showing in Fig. 1, the system is composed of sensors, interrogator, and laptops which are respectively used to collect acceleration data of car body floor and simultaneously conduct recoding work, purposing for obtaining ride quality values and then evaluating the operation condition of high-speed trains. This on-board monitoring work has been lasted for more than one month. During the test, the train operated at its normal speed.

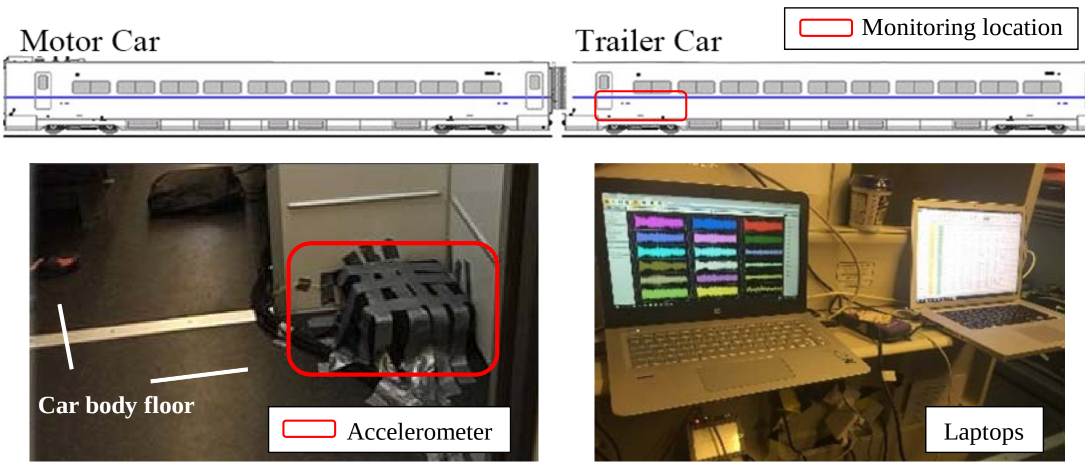

Fig. 1 The data acquisition system deployed on an in-service high-speed train

This online monitoring system has constantly acquired acceleration signals on train operations. In detail, the accelerometer is mounted on the car body floor of a trailer car; it recorded the vibration at two directions (i.e., lateral and vertical). The acceleration data is stored by the system with the sampling rate of 5000 Hz for analysis. In this research, the vibration data is used to reveal the dynamic performance of car body. Then, pre-processing for the acquired vibration data is conducted by Sperling operator in order to obtain the ride quality index sequence and represent the operation condition of the high-speed train for continuous time. Meanwhile, to assess the effectiveness and reliability of the adopted novel technique which combines two technique (i.e., Bayesian forecasting and time series analysis), the calculation values of ride quality index in two operation conditions which corresponds to the acceleration data measured before and after wheel lathing have been employed.

3. METHODOLOGY

3.1 Sperling index

Sperling index is a specific indicator used to connect subjective comfort feelings of majority passengers with the objective physical variables of a running vehicle (Nakagawa 2010). Based on this statement, Sperling index is capable of reflecting the condition of the running vehicles. By acquiring acceleration data on car body floor, the

Sperling index calculation is based on Eq. (1) (Zhou 2009), and the smaller value obtained, then the higher ride quality achieves.

$$W = \left(W_1^{10} + W_2^{10} + \cdots + W_n^{10}\right)^{0.1} = \left(\sum_{i=1}^{n} W_i^{10}\right)^{0.1} \tag{1}$$

where *n* is the number of frequency contents $\{f_1, f_2, f_3, \ldots, f_n\}$ of the measured acceleration in time history after fast Fourier transform (FFT). W_i stands for Sperling index value of each contributed component, and determined by:

$$W_{\mathrm{i}} = 7.08\left(\frac{A_i}{f_i} F\left(f_i\right)\right)^{0.1} \tag{2}$$

where A_i records the vibration acceleration at frequency f_i which unit is Hz, and $F\left(f_i\right)$ is the frequency modification factor which as a piecewise-linear function of f_i referred to (Standards-GB5599 1985). Sperling operator processes the measured signals every 100 seconds and converts them into ride quality index symbolizing operation conditions of the monitored HST in each time cycle.

3.2 Time series analysis

Dynamic linear model The dynamic linear models (DLMs) are regarded as a special case of a general state space model which is commonly used in time series analysis. In DLM, a system equation also called state equation is used to describe state parameter varying in time series. Together with an observation equation which presents the relationship between the measurement and the unknow state parameters, then the dynamic changes can be demonstrated (Petris 2009).

Observation equation: $$Y_t = F_t^T \theta_t + \upsilon_t, \qquad \upsilon_t \sim N(0, V_t) \tag{4}$$

System equation: $$\theta_t = G_t \theta_{t-1} + \omega_t, \qquad \omega_t \sim N(0, W_t) \tag{5}$$

where F_t and G_t denote the vector of known parameters (of order $p \times 1$) and known constant (of order $p \times p$), respectively. While θ_t is a $p \times 1$ vector of unknow parameters; $\{V_t\}$ and $\{W_t\}$ are two independent sequences of independent Gaussian random vectors with mean zero and known variance, respectively; Y_t is the observation series at time t. In this case, a second-order DLM has been employed to model the dynamic process. To be specific, the basic ideal of describing a time series is breaking the observed measurement down into its mathematical elements including value and gradient (Lipowsky 2010). In this study, the values of ride quality index calculated by Sperling operator above serve as the measurement instead of using original signals directly in DLM. Therefore, the following matrices are applied in this DLM:

$$\theta = \begin{pmatrix} \mu \\ \beta \end{pmatrix} \qquad F = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \qquad W = \begin{pmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix} \qquad V = \sigma^2 \tag{6}$$

where μ and β respectively represent the guess value of parameter value and parameter gradient; ω_1 and ω_2 are the variances in value and gradient, respectively. the measurement variance is provided by the square of its standard deviation σ.

3.3 Bayesian forecasting

Bayes' theorem enables the calculation of conditional probabilities. Based on this statement, Bayesian inference is applied to forecast and predict by the use of PDFs. To develop the synergy in combination with time series analysis, the following notation is applied:

$$P(\theta_t | D_t) \sim N(m_t, C_t), \qquad D_t = Y_1, Y_2, \ldots, Y_t \tag{7}$$

where $\{Y_1, Y_2, \ldots, Y_t\}$ stands for the observations. The equation presents the PDF of state parameter θ at time t based on the measurements. This PDF is normally distributed with a mean of m and a variance of C. If the state at time t has been given, the posterior distribution of the state parameter and the predictions for one-step forecast at time $t+1$ obtains from the following equations (Petris 2009):

Poster distribution of state parameters: $$P(\theta_{t+1} | D_t) \sim N(a_{t+1}, R_{t+1}) \tag{8}$$

One-step forecasting distribution: $$P(Y_{t+1} | D_t) \sim N(f_{t+1}, Q_{t+1}) \tag{9}$$

where means and variances above can be calculated by the following equations:

Mean: $$a_{t+1} = G_{t+1} \cdot m_t, \qquad f_{t+1} = F_{t+1}^T \cdot a_{t+1} \tag{10}$$

Variance: $$R_{t+1} = G_{t+1} C_t G_{t+1}^T + W_{t+1}, \qquad Q_{t+1} = F_{t+1}^T R_{t+1} F_{t+1} + V_{t+1} \tag{11}$$

In this inference process, if we obtain the measurement at time $t+1$, the corrections are based on the Eq. (7) and the distribution of the state parameters is updated to $P(\theta_{t+1} | D_{t+1}) \sim N(m_{t+1}, C_{t+1})$ where the mean and variance are given by:

Mean: $$m_{t+1} = a_{t+1} + A_{t+1} e_{t+1} \tag{12}$$

Variance: $$C_{t+1} = R_{t+1} - A_{t+1} A_{t+1}^T Q_{t+1} \tag{13}$$

where $e_{t+1} = Y_{t+1} - f_{t+1}$ and $A_{t+1} = R_{t+1} F_{t+1} / Q_{t+1}$. In the next sections, a special logic called Bayes factor has been employed in use of identifying single outliers and significant changes.

3.4 Detection algorithm

Outlier detection This detection logic is based on the calculation of Bayes factors aimed at identifying potential outliers. Bayesian forecasting enables generation of PDF for the next observation as aforementioned. In this process, the detection is carried out by comparing the measurement with current model (forecast distribution for current time, marked as Model 0) and the alternative model (achieves by shifting a prescribed offset of mean value h of the current model, marked as Model 1). The Bayes factor obtains from the ratio of two models above:

$$H_t = \frac{PDF \quad value \quad of \quad Model1}{PDF \quad value \quad of \quad Model0} \tag{14}$$

Since the PDFs of both model are normally distributed, then the Bayes factor can be specified:

$$H_t = \exp\left(\frac{\pm 2h \cdot (Y_t - f_t) - h^2}{2Q_T^2}\right) \tag{15}$$

where H_t serves as a criterion in indicating of which model can better fit the measurement. In the case of $H > 1$, it implies that the better fitness is achieved from alternative model instead of the current one, then a potential outlier has been detected. On the other hand, the plus and minus sign in the Eq. (13) are examined in parallel for detecting positive and negative outliers, respectively (Lipowsky 2010). In the view of Jeffreys (1961), he suggested that the outlier detection use a threshold of $H_{\min} = 10$. However, the shift value of h should be determined according to the required confidence level. In this study, $h = 1.645\sigma_t$ at 90% confidence interval has been set. After that, an uncertainty limit (*ucl*) can be obtained by the following equation:

$$ucl = \frac{\ln(H_{\min})}{h}\sigma_t^2 + \frac{h}{2} \tag{16}$$

In this situation ($h = 1.645\sigma_t$ and $H_{\min} = 10$), the uncertainty limit is equal to $2.22\sigma_t$ which means an observation will be detected as an outlier when its deviation from the mean value of current model is larger than *ucl*.

Change detection To estimate and distinguish between the beginning of change and a single outlier, the aforementioned Bayes factor has been extended as presenting in the below:

$$H_t(k) = \prod_{t-k+1}^{t} H_t, \qquad k = 1,2,\ldots,l_{\max} \tag{17}$$

where $l_{\max}$ denotes the maximum number of Bayes factors considered. The maximum cumulative Bayes factor is calculated by the following equation:

$$L_t = H_t(l_t) = \max(H_t(k)), \qquad 1 \le l_t \le l_{\max} \tag{18}$$

where l_t is referred to as the run length. It can be calculated recursively by $l_t = 1 + l_{t-1}$ in which $t > 1$ ($l_t = 1$ at $t = 1$), and run length threshold of l_t is $l_{\min} = 4$ recommended by Pole (1994). In the procedure of change detection, a notification of change at time t when $L_t > H_{\min}$ is marked as time of notification ($TON = t$); the information of time when a change occurred is hold and recorded as time of occurrence ($TOC = t - l_t + 1$).

In summary, the procedure of detection consists of two steps. The first step is to calculate the single Bayes factor and judge whether an observation is an outlier, and the criterion for catching an outlier is presented as $H_t > H_{\min}$ and $H_{t-1} < H_{\min}$. If not, the procedure runs to the change detection. In this step, the cumulative Bayes factor L_t and run length l_t are calculated. In contrast to the outlier detection, the significant change can be triggered by the following two reasons (Lipowsky 2010): (i) the occurrence of two consecutive Bayes factors $H_t > H_{\min}$, which is equivalent to $L_t > H^2_{\min}$; (ii) the concurrence of $L_t > H_{\min}$ and $l_t > l_{\min}$. Once a change is detected, the retrospective adjustment will be conducted by resetting the time to the TOC and by setting the mean value of Model 0 to the measurement ($f_t = Y_t$); following by model adjustment starting at the TOC.

4. ANALYSIS AND DISCUSSION

Conducting the condition assessment of HST requires a set of stable and effective data, the on-board data acquisition system has been employed on an in-service HST for a couple of days during the routine operations. The measured vibration signals of the car body floor are processed and analysed in this section based on the novel technique introduced above. Firstly, with the help of Sperling operator, the original collected data has been converted into ride quality index sequence reflecting the operation conditions of HST every 100 seconds. In the assessing procedure, a second-order DLM has been employed for modeling the dynamic process. Then, Bayesian forecasting provides the prediction of dynamics (i.e., value and gradient). Two special logics are developed and carried out for outlier detection and change detection. During monitoring work, vibration data of car body under different wheel quality conditions (corresponding to before and after lathing the wheels) has been recorded. In this study, we use this introduced novel method to identify the variation of operation performance which is induced by wheels imperfectly rounded or deterioration. The analysis is conducted in two cases by using the monitoring data acquired from different intervals of the route, which aims at examining the results independent or not.

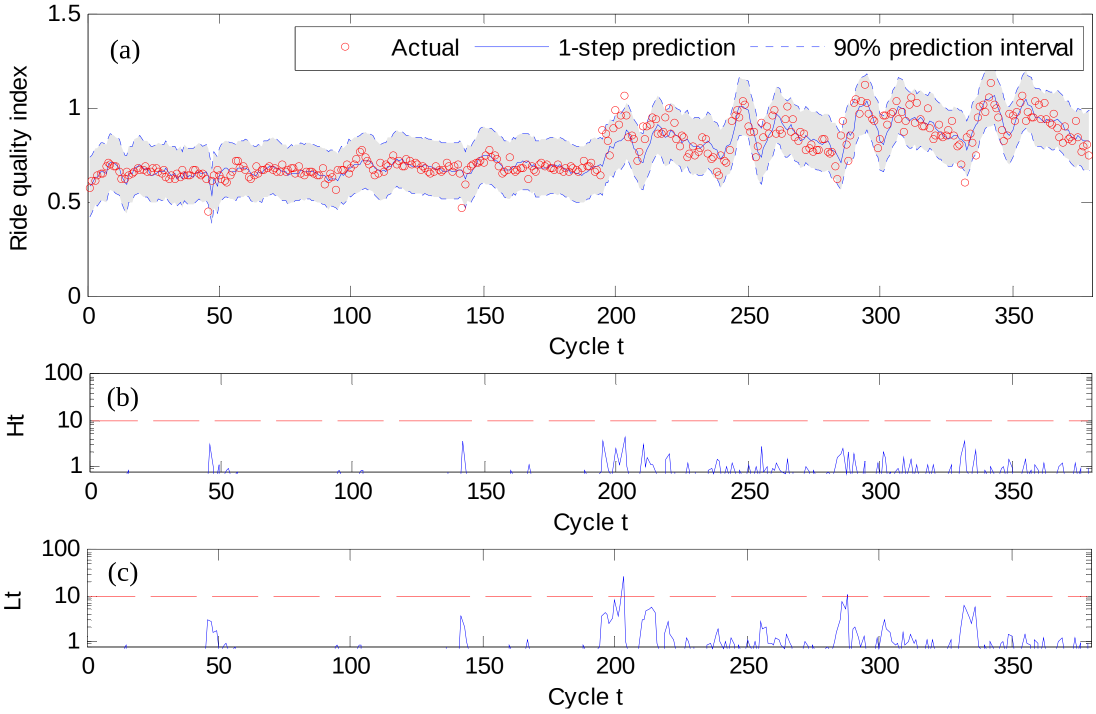

Fig. 2 Ride quality index for outlier and change detection in case 1.

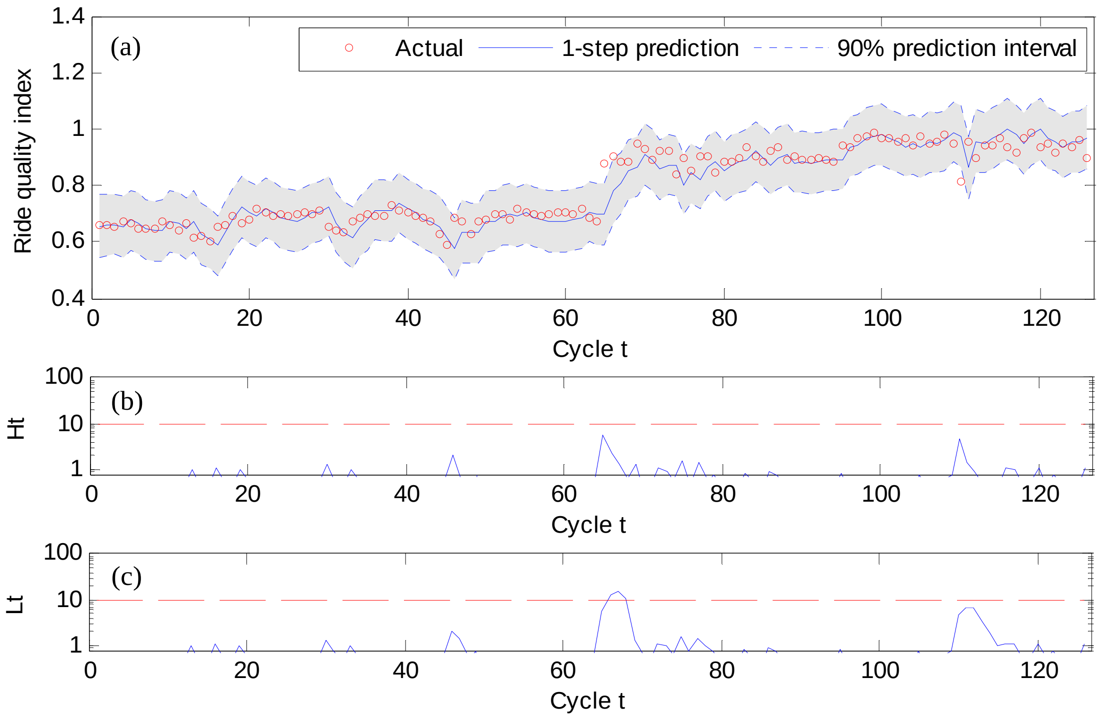

Fig. 3 Ride quality index for outlier and change detection in case 2.

The application of the introduced novel technique to condition assessment of HST is illustrated in Fig. 2. and Fig. 3. To be specific, the example data involved in two cases is extracted from the measurement which is collected on the car body floor when the train ran through the same interval of the route before and after lathing the wheels. The shift parameter was set to $h = 1.645\sigma$ resulting in an uncertainty limit of $ucl = 2.22\sigma$. The upper diagram of Fig. 2(a) and Fig. 3(a) comprises the actual values (red point) which obtain from original vibration signals by Sperling operator at an equivalent cycle time, the results of 1-step forecasting (guess for mean value) by Bayes' theorem described in section 3.3 are presented in blue line, and the 90% prediction interval is denoted by gray shadows. The Bayes factor and accumulative Bayes factor are used to check for outlier detection and significant change detection, respectively. The calculation results are illustrated at the bottom of each figure. It can be observed that, although some observations falling outside of the 90% prediction interval, the calculation values of Bayes factor are blow to the threshold of $H_{\min} = 10$ in the overall process of two cases, which indicates that there no outlier has been detected as shown in Fig. 2(b) and Fig 3(b). On the other hand, Fig. 2(c) and Fig. 3(c) show the cumulative Bayes factor for change detection. It can be seen that the calculation values exceed the warning line (the threshold of $L_{\min} = 10$) at $t = 201$ (i.e., $TON_1 = 201$) and $t = 67$ (i.e., $TON_2 = 67$) of two cases, respectively. Meanwhile, the corresponding run length of two cases is $l_t = 5$ and $l_t = 2$. Then, the real time of occurrences of two cases are at $t = 196$ (i.e., $TOC_1 = 196$) and $t = 65$ (i.e., $TOC_2 = 65$), which is well mapping the actual time point of lathing the wheels. After that, the model adjustment is carried out by the algorithm as shown in Fig. 2(a) and Fig. 3(a).

5. CONCLUSION REMARKS

In this work, an on-board data acquisition system was deployed on an in-service high-speed train to collect vibration signals of car body during routine operations. A vibration-based operator based on Sperling index has been employed for pre-processing the measured original signals, and the obtained ride quality index sequence is used to represent the running condition of trains for continuoues time. Then, a novel technique combining Bayesian forecasting and time series analysis is used to conduct condition assessment. The special logics included in this method have been carried out for outlier detection and change detection. This technical is experimentally demonstrated by using the monitoring data acquired under different wheel quality conditions (corresponding to before and after lathing wheels). The analysis results in two case studies indicate that this method has successfully applied in identifying the variations of running condition. At the same time, it can successfully detect the time when the operation state changes. In addition, this method enables providing 1-step forecasting with prediction intervals, which is not only valuable for monitoring and assessing the operation condition of high-speed trains, but also favorable to design maintenance plan of wheels.

ACKNOWLEDGEMENT

The work described in this paper was supported by the National Natural Science Foundation of China (Grant No. 11602210). The authors also would like to appreciate the funding support by the Innovation and Technology Commission of Hong Kong SAR Government to the Hong Kong Branch of National Rail Transit Electrification and Automation Engineering Technology Research Center (Project No: K-BBY1).

REFERENCES

China National Bureau of Standards. (1985), "GB5599: Railway vehicles specification for evaluation of dynamic performance and accreditation test", Beijing, China.
Huang, Z. W., Cui, D. B., Du, X., and Jin, X. S. (2013), "Influence of deviated wear of wheel on performance of high-speed train running on straight tracks", *Journal of the China Railway Society*., **35**(2), 14-20.
Jeffreys, H. (1961). *Theory of probability*. Clarendon Press, Oxford.
Kim, Y. G., Kwon, H. B., Kim, S. W., Kim, C. K., and Kim, T. W. (2003), "Correlation of ride comfort evaluation methods for railway vehicles", Proceedings of the Institution of Mechanical Engineers, Part F: *Journal of Rail and Rapid Transit*, **217**(2), 73-88.
Kim, Y. G., Kim, S. W., Park, C. K., Kim, K. H., and Park, T. W. (2004), "Analysis on the characteristics of the ride comfort for high speed trains on the high-speed line/conventional line", *Transactions of the Korean Society for Noise and Vibration Engineering*., **14**(10), 999-1006.
Lipowsky, H., Staudacher, S., Bauer, M., and Schmidt, K. J. (2010), "Application of Bayesian forecasting to change detection and prognosis of gas turbine performance", *Journal of Engineering for Gas Turbines and Power*., **132**(3), 031602.
Nakagawa, C., Shimamune, R., Watanabe, K., and Suzuki, E. (2010), "Fundamental study on the effect of high frequency vibration on ride comfort", *Journal of Mechanical Systems for Transportation and Logistics*., **3**(1), 287-293.
Ni, Y. Q., Liu, X. Z., Zhao, W. Z., and Liang, S. L. (2015), "Outlier detection in sensor-assisted online ride comfort assessment of high-speed trains", *Structural Health Monitoring 2015*.
Petris, G., Petrone, S., and Campagnoli, P. (2009), "Dynamic linear models", *Dynamic Linear Models with R* (2009): 31-84.
Pole, A., West, M., and Harrison, J. (1994), "Applied Bayesian forecasting and time series analysis", *CRC*., press.
Remennikov, A. M., and Kaewunruen, S. (2008), "A review of loading conditions for railway track structures due to train and track vertical interaction", *Structural Control and Health Monitoring*., **15**(2), 207-234.
Wang, X, Ni, Y. Q, Zhang, L. H, and Sun, Q. (2016), "Understanding of dynamic interaction of an in-service high-speed train via on-board monitoring", *1st International Workshop on Structural Health Monitoring for Railway System*, Qingdao, Shandong Province. (Invited for Presentation).
Zhou, J., Goodall, R., Ren, L., and Zhang, H. (2009), "Influences of car body vertical flexibility on ride quality of passenger railway vehicles", *Proceedings of the Institution*

of Mechanical Engineers., Part F: Journal of Rail and Rapid Transit, **223**(5), 461-471.