Reconstruction of an informative railway wheel defect signal from wheel–rail contact signals measured by multiple wayside sensors

Alemi, Alireza; Corman, Francesco; Pang, Yusong; Lodewijks, Gabriel

DOI
10.1177/0954409718784362

Publication date
2019

Document Version
Final published version

Published in
Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit

Citation (APA)
Alemi, A., Corman, F., Pang, Y., & Lodewijks, G. (2019). Reconstruction of an informative railway wheel defect signal from wheel–rail contact signals measured by multiple wayside sensors. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *233*(1), 49-62. https://doi.org/10.1177/0954409718784362

Original Article

Institution of MECHANICAL ENGINEERS
Journal of RAIL AND RAPID TRANSIT

Proc IMechE Part F:
J Rail and Rapid Transit
2019, Vol. 233(1) 49–62


Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0954409718784362
journals.sagepub.com/home/pif
SAGE

Reconstruction of an informative railway wheel defect signal from wheel–rail contact signals measured by multiple wayside sensors

Alireza Alemi[1], Francesco Corman[2], Yusong Pang[1] and Gabriel Lodewijks[3]

Abstract

Wheel impact load detectors are widespread railway systems used for measuring the wheel–rail contact force. They usually measure the rail strain and convert it to force in order to detect high impact forces and corresponding detrimental wheels. The measured strain signal can also be used to identify the defect type and its severity. The strain sensors have a limited effective zone that leads to partial observation from the wheels. Therefore, wheel impact load detectors exploit multiple sensors to collect samples from different portions of the wheels. The discrete measurement by multiple sensors provides the magnitude of the force; however, it does not provide the much richer variation pattern of the contact force signal. Therefore, this paper proposes a fusion method to associate the collected samples to their positions over the wheel circumferential coordinate. This process reconstructs an informative signal from the discrete samples collected by multiple sensors. To validate the proposed method, the multiple sensors have been simulated by an ad hoc multibody dynamic software (VI-Rail), and the outputs have been fed to the fusion model. The reconstructed signal represents the contact force and consequently the wheel defect. The obtained results demonstrate considerable similarity between the contact force and the reconstructed defect signal that can be used for further defect identification.

Keywords

Railway, wheel, defect, contact, condition monitoring, signal reconstruction

Date received: 11 December 2017; accepted: 28 April 2018

Introduction

Railway wheels are critical components and their maintenance is therefore a vital task. From the safety point of view, the defects of wheelsets are the main reasons for train accidents.[1] Wheel defects change the wheel–rail contact feature and sometimes generate a high impact force that is detrimental to the track and train. Unexpected wheel failures also reduce the availability of trains and cause delay in the transport services that reduces the reliability of the railway system. To make an effective and efficient maintenance plan, the condition of the wheels should be accurately measured or estimated. Hence, wheel condition monitoring has been and still is the subject of many studies.[2]

A wheel defect produces a contact force that is transferred to the track and vehicle. Therefore, the wheel condition can be indirectly estimated by measuring the wheel and rail responses such as strain, vibration, and acoustic. Installing sensors on every wheel is challenging due to the expense, implementation, and maintenance. For this reason, track-side measurement has been given more attention to measure the rail responses.

Wheel impact load detectors (WILDs) are common wayside wheel monitoring systems. The first generation of WILDs was introduced in 1983[3] and then rapidly became a widespread commercial system.[4] They measure the rail response such as strain[5] and vibration,[6] by a sensor or a set of sensors to estimate

[1]Faculty of Mechanical, Maritime and Material Engineering (3mE), Delft University of Technology, Delft, The Netherlands
[2]Institute for Transport planning and Systems, ETH Zurich, Zurich, Switzerland
[3]School of Aviation, The University of New South Wales, Sydney, Australia

Corresponding author:
Yusong Pang, Faculty of Mechanical, Maritime and Material Engineering (3mE), Delft University of Technology, Mekelweg 2, Delft 2628 CD, The Netherlands.
Email: Y.Pang@tudelft.nl

the condition of the in-service wheels. Different studies have attempted to interpret and use the data measured by WILDs to estimate the wheel condition.

Some wheel defects such as flats generate high-frequency components in the signals measured by sensors. Therefore, the defect can be detected by looking at high-pass filtered signal.[7] This method just detects the defect without any further information about its type and severity and can be used only if the defects generate signals containing high-frequency components. Therefore, long-wave defects such as periodic out of roundness (OOR) cannot be detected and identified using this method.

Another common criterion to quantify the wheel condition is the magnitude of the data acquired by a WILD. Stone et al.[8] and Nielsen and Johansson[9] investigated the wheel defects using the peak value criterion. They used the peak acceleration and the peak force collected by acceleration- and strain-based WILDs. The results showed a considerable fluctuation in the acceleration and force peaks especially when the trains had higher velocity and the wheels had more severe defects. Later,[10] Johansson and Nielsen investigated the effect of train velocity, the axle load, and the defect types on the peak forces measured. Their results presented an extreme variation even when the train velocity, the axle load, and the defects were kept constant in the repeated tests.

Partington[3] excluded the effect of axle load by means of two methods and defined two other criteria. First, force ratio that is calculated by dividing the peak force by the average force collected by multiple sensors. Second, dynamic force that is calculated by subtracting the peak force and the average force. In spite of excluding the effect of axle load, the train velocity is an out-of-control parameter that causes variation in the magnitude of the peak force, the force ratio, and the dynamic force. Johansson and Nielsen[10] related the fluctuation to the variation in track properties and to the random position of the defect with respect to the sensors. To measure more reliable data and to avoid this fluctuation, the track properties can be maintained similarly over the measurement station. Besides, WILDs can exploit multiple sensors to cover the whole wheel circumference. Asplund et al.[11] investigated the peak force, dynamic force, and force ratio criteria and finally concluded that they only detect the severe defects that greatly contribute to the contact force.

Another limitation is that these criteria fail to distinguish between the defect types. They classify the wheel into healthy and defective classes. The rate and mechanism of the wheel degradation are influenced by the defect type. Therefore, estimating the defect type is significant to provide a comprehensive estimate of the wheel condition. These criteria somehow show the existence of a defect but are unable to identify the defect type. In addition, a severe defect dominates the other defects of a wheel. Furthermore, the dynamic force and the force ratio of a wheel with multiple defects including a severe defect can be smaller than a wheel with a similar severe defect, because the average of the contact force for the first wheel is higher than the second one. Therefore, these criteria can lead to false interpretation. Another weakness of the current criteria is difficulty in detecting the minor defects such as spalling, periodic OOR, and small shelling at an early stage. As a result, developing an effective and reliable method for detecting and identifying the wheel defects is still an open issue.

The aim of this paper is to propose a method that can provide more information from the wheel defects to use for defect detection and identification. This paper presents the theoretical models of a data fusion process to reconstruct the defect signal from the discrete samples measured by multiple sensors such as WILDs. To achieve this purpose, the samples collected by multiple sensors are mapped over the circumferential coordinate to reconstruct a new informative signal. The reconstructed signal provides a pattern representing the wheel defect. As a result, the features of the reconstructed signal can be used for defect detection and identification.

The paper is organized as follows. The next section explains the configuration of the sensors and the corresponding issue of the partial observation. The space between the sensors causes a specific time lag between the signals measured. When the sensors have the same reference time, the time lag between the signals and the space lag between the sensors can be related to each other. Therefore, the ''Lag estimation between signals'' section estimates the time lags to determine the space relation between the collected data. Then, the ''Sampling methods'' section develops the sampling methods to determine the samples that should be selected from each signal as the output of each sensor. In the ''Data fusion and signal reconstruction'' section, the fusion method is developed to reconstruct a signal over the circumferential coordinate using the collected samples. Consequently, the ''Train velocity estimation'' section estimates the train velocity to define the sampling frequency of the collected data in the space domain. Finally, the last section validates the proposed method by reconstructing the informative signals from the data simulated by VI-Rail.[12]

Configuration of sensors and partial observation problem

A wheel–rail contact force represents the geometric pattern of the wheel. The generated contact force is transferred to the track and vehicle and can be measured on both sides by installing a sensor on the rail and wheel. When a sensor is installed on the wheel, it can move with the wheel to continuously measure the wheel response to the contact force. The sensor installed on the rail has three measurement zones with

respect to the wheel. First, the inactive zone in which the wheel is away from the sensor, producing a zero output. Second, the transient zone in which the wheel approaches or leaves the sensor, with increasing or decreasing in the sensor output. Third, the effective zone in which the wheel is on top of the sensor. The sensors collect data in these measurement zones but only the data from the effective zone are used here. When the effective zone is smaller than the wheel circumference, the sensor makes a partial observation and only senses a limited portion of the wheel. The position of the defect with respect to the effective zone of the sensor is out of control. Therefore, multiple sensors are commonly used to cover the wheel circumference.[5]

The outputs of the multiple sensors are usable when they sample in identical situations to have an identical transfer function. Therefore, the sensors should be mounted on a uniform track with a continuous structure. A schematic view of the sensors and the uniform track structure are illustrated in Figure 1(a). This structure consists of a continuous sleeper that provides identical transfer function for the sensors. Integration of the discrete samples collected by sensors gives the required contact signal over the circumferential coordinate.

Tracks with continuous sleepers are not common. In addition, creating a uniform track structure needs a dramatic change in the rail and sleeper structure. Therefore, a typical rail–sleeper structure is considered. The typical rail–sleeper structure (Figure 1(b)) causes dissimilar rail responses in different points along the rail. In this case, the outputs of the sensors have to be calibrated with respect to the sensor position in the longitudinal direction. To avoid this complexity, a symmetric structure of the sensors can be used. To configure this structure, the sensors should be mounted on the positions with an identical situation as displayed in Figure 1(c) and (d). By assuming a healthy track without any irregularities, this configuration assures that every measurement refers to a comparable rail and sleeper condition, and the only variable is the wheel condition.

Figure 2 shows the results of a field measurement[13] presenting the passage of four wheels by variation in the signals with four peaks. In this example, the third wheel had a 60 mm flat with 1 mm depth, and the other wheels were healthy. Figure 2(a) shows the strain signal converted to the contact force. This signal was measured by the strain sensors mounted on a sleeper bay. Figure 2(b) shows the rail bending moment above the sleeper. The variation in this signal is not as sharp as the signal in Figure 2(a) but clearly shows the passage of four wheels and the existence of the flat in the third wheel in 2.98 s. In Figure 2(a) and (b), the third wheel faces the sensors with the flat part. Figure 2(c) shows the same wheels measured by a sensor mounted above another sleeper with a distance from the prior sensor used in Figure 2(b).

The defect of the third wheel obviously influenced the measured signal and generated a specific pattern with a downward and an upward deflection in the defective area. Regardless of the sensor type, the pattern of the defective area of the third wheel can be seen in both Figure 2(a) and (b). This pattern was also

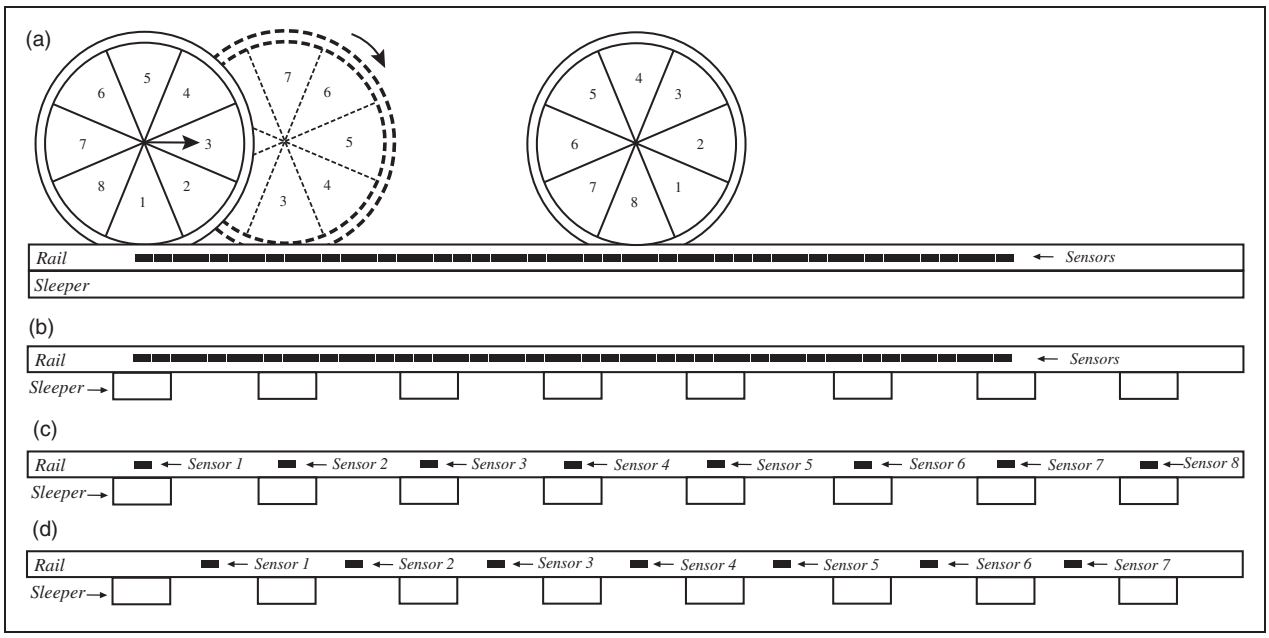

Figure 1. The configuration of the wheel, rail, sleepers, and sensors for (a) a uniform track structure with joined sensors, (b) the typical rail–sleeper structure with joined sensors, (c) the typical structure with discrete sensors on the sleepers, and (d) the typical structure with discrete sensors between the sleepers.

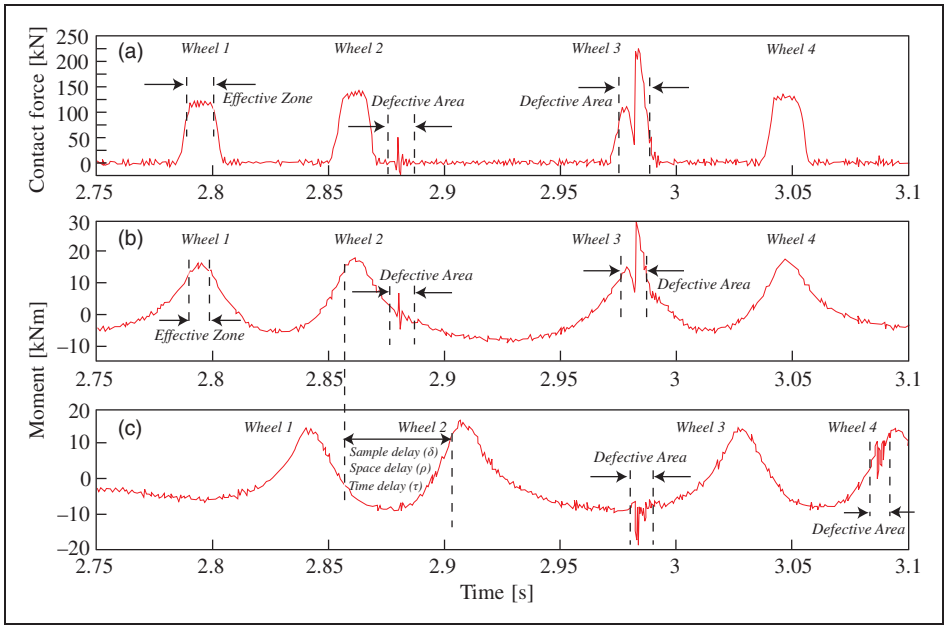

Figure 2. (a) The vertical wheel–rail contact forces measured in a sleeper bay, (b) rail bending moment above the sleeper, and (c) rail bending moment above another sleeper with a distance from the prior sensor.[13]

sensed near the second wheel in 2.88 s in the inactive zone due to the previous turn of the wheel. In Figure 2(c), the defective area of the third wheel was not sensed by the effective zone and appeared in the inactive zone in 2.98 s, and in the transient zones in 3.09 s. The effective zone in Figure 2(b) reaches 30 kNm at the maximum for the third wheel, and the signal in Figure 2(c) reaches 15 kNm. By selecting only the magnitude of the signal as the representative output of the sensor, the pattern of the signal related to the wheel flat is neglected.

To reconstruct a signal over the circumferential coordinate, two sampling frequencies should be considered carefully. First, the sensor sampling frequency (f_t) that is defined in the time domain. For example, 10 kHz sampling frequency means the sensor collects 10,000 samples per each second. Second, the space sampling frequency (f_s) that is defined in the space domain and determines the sampling frequency in the unit of space. Increasing the train velocity increases the distance between the collected samples and decreases the space sampling frequency. For example, when a wheel is moving with 10 m/s velocity and a sensor is sampling with 10 kHz time domain frequency, the distance between the samples collected by the sensor is 1 mm in the space domain and the space sampling frequency is 1000 samples/m. When the wheel is moving with 40 m/s velocity, the same sensor with the same sampling frequency will collect samples with 4 mm distance in the space domain that means the space sampling frequency is 250 samples/m. To have a signal over the circumferential coordinate, the sensors should sample from the wheel to the extent that the signal can be reconstructed using the data sampled. In Figure 1(c) and (d), the distance between the sensors leads to discrete sampling from the wheel circumference. Therefore, in spite of the sufficient sensor sampling frequency (f_t), it is not possible to reconstruct a signal from the samples collected in this way.

According to the Nyquist sampling criterion, reconstructing the actual signal is perfectly possible when the sampling frequency (f_s) is at least twice the highest frequency contained in the signal ($f_{\max}$); otherwise, it leads to aliasing[14]

$$f_s > 2f_{\max} \tag{1}$$

In accordance with the sensor configuration, the sleeper interval is a determining factor that defines the sensor intervals. In fact, only a limited number of samples from the wheel circumference can be collected on every wheel revolution. This sampling method leads to signal distortion (aliasing). The space sampling frequency is definitely far from the Nyquist frequency and therefore presents a new challenge for the sampling in the space domain.

Lag estimation between signals

The patterns of the rail bending moment signal in Figure 2(b) and (c) are generally similar except only having a delay and some variations due to the wheel defect. The delay can be presented in three different ways: time delay (τ), space delay (ρ), and sample delay (δ). The time delay indicates the wheel travel time between two sensors (time dimension, s). The space delay indicates the spatial turn of the wheel with respect to the prior sensor, which is equal to the sensor intervals (space dimension, m). Finally, the

sample delay shows the number of samples in the second signal that lagged behind the first signal (a number without dimension).

In this research, the signals presented in Figure 2(b) and (c) are modeled in the time domain as follows

$$z_1(t) = w(t) + g_1(t) + n_1(t) \tag{2a}$$

$$g_1(t) = w(t) \cdot g(t) \tag{2b}$$

$$\begin{aligned} &\{t | t \in \mathbb{R}, 0 \leqslant t \leqslant T\} \\ &z_2(t) = w(t-\tau) + g_2(t) + n_2(t) \end{aligned} \tag{3a}$$

$$\begin{aligned} &g_2(t) = w(t-\tau) \cdot g(t) \\ &\{t | t \in \mathbb{R}, 0 \leqslant t \leqslant T\} \end{aligned} \tag{3b}$$

$z_1(t)$ and $z_2(t)$ are the signals measured by two consecutive sensors in the time domain. $w(t)$ is the signal generated by the wheel movement and contains low-frequency components, that we call it the wheel signal. This signal is a function of track and vehicle dynamics as the fundamental parameters, in addition to the axle load and the train velocity as the operational parameters. Due to the sensor distance, and the wheel movement, the wheel signal $w(t)$ shifts over time and space. τ is the time delay between the signals $z_1(t)$ and $z_2(t)$, and the $n_1(t)$ and $n_2(t)$ are the uncorrelated noises. The signals are defined in the closed interval between zero reference time and T that is the measurement time. The time interval between each sample is $1/f_t$ second and the time delay between two signals is τ second.

$g(t)$ is the signal generated by the wheel defect and is a function of the defect geometry. The defect signal $g(t)$ is a periodic signal that is repeated on every wheel revolution. The sensors have a limited effective zone; therefore, they observe a limited portion of the defect signal. The wheel signal $w(t)$ operates as a window function that has almost a zero value outside the effective zone. Therefore, the product of the wheel signal $w(t)$ and the defect signal $g(t)$ generates a partial view of the defect signal. $g_1(t)$ is the partial view of the defect signal measured by the first sensor. This signal superimposes on the $w(t)$ and mostly contains high-frequency components. As a result, $g_1(t)$ is also a function of the wheel signal. $g_2(t)$ is the partial view of the defect signal measured by the second sensor and is superimposed on $w(t-\tau)$. Figure 3 illustrates a schematic view of the wheel signal $w(t)$, the defect signal $g(t)$, the windowed defect signal $g_1(t)$, and the measured signal $z_1(t)$. Bear in mind that, this paper aims to reconstruct the defect signal $g(t)$, from the measured signals $z(t)$.

The measured signals can be also presented in the space domain as

$$z_1(x) = w(x) + g_1(x) + n_1(x) \tag{4a}$$

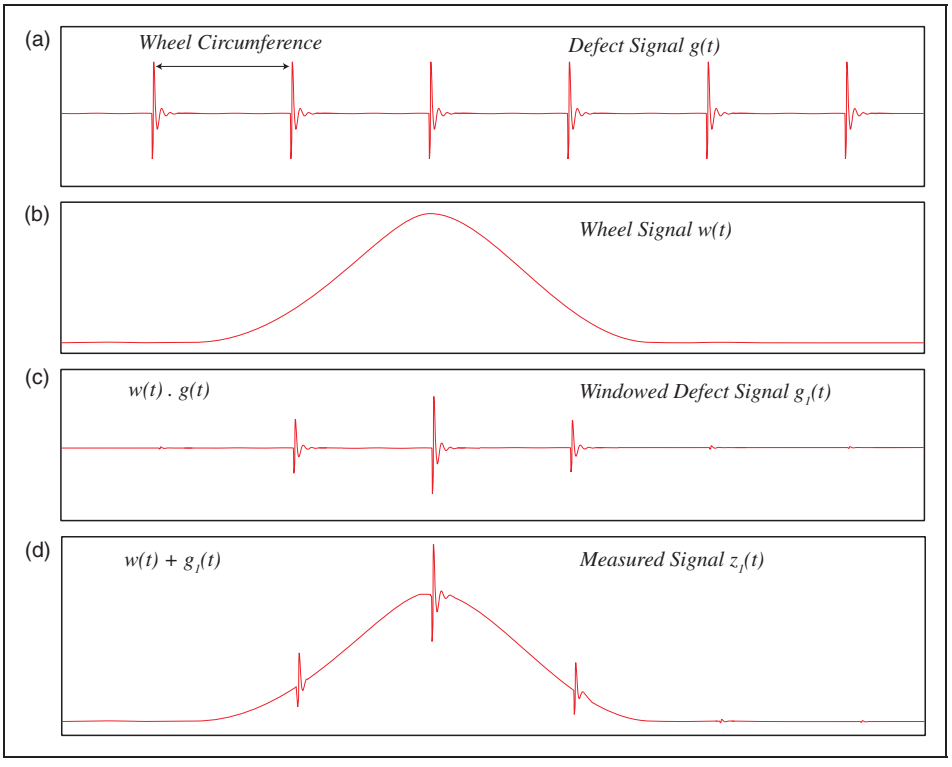

Figure 3. The schematic view of (a) the defect signal $g(t)$, (b) the wheel signal $w(t)$, (c) the windowed defect signal $g_1(t)$, and (d) the measured signal $z_1(t)$.

$$g_1(x) = w(x) \cdot g(x) \quad \{x | x \in \mathbb{R} : 0 \leqslant x \leqslant X\} \tag{4b}$$

$$z_2(x) = w(x - \rho) + g_2(x) + n_2(x) \tag{5a}$$

$$g_2(x) = w(x - \rho) \cdot g(x) \quad \{x | x \in \mathbb{R} : 0 \leqslant x \leqslant X\} \tag{5b}$$

$z_1(x)$, $z_2(x)$, $w(x)$, $g(x)$, $g_1(x)$, $g_2(x)$, $n_1(x)$, and $n_2(x)$ are the signals in the space domain. The signals are defined in the closed interval between zero reference place and X that is the length passed by the wheel over the sensors. The space interval between each sample is $1/f_s$ meter and the space delay between two signals is ρ meter.

The measured signals can be also presented without dimension as

$$z_1(i) = w(i) + g_1(i) + n_1(i) \tag{6a}$$

$$g_1(i) = w(i) \cdot g(i) \quad \{i | i \in \mathbb{Z}^+ : 0 \leqslant i \leqslant I\} \tag{6b}$$

$$z_2(i) = w(i - \delta) + g_2(i) + n_2(i) \tag{7a}$$

$$g_2(i) = w(i - \delta) \cdot g(i) \quad \{i | i \in \mathbb{Z}^+ : 0 \leqslant i \leqslant I\} \tag{7b}$$

The dimensionless signals are defined in the closed interval between the first sample and I that is the length of signal. In this case, the delay between two signals is δ samples.

The delay between two signals such as $z_1(i)$ and $z_2(i)$ displayed in Figure 2 can be estimated by looking for the maximum cross-correlation between the signals.[15] The cross-correlation function can be calculated as follows

$$R_{z_1 z_2}(\gamma) = \sum_{i=1}^{I-\gamma} z_1(i + \gamma) z_2(i), \quad \gamma = 0, 1, 2, \ldots \tag{8}$$

The cross-correlation between the signals $R_{z_1 z_2}(\gamma)$ involves shifting one of the signals and summing the multiplication of the two signals. Therefore, the cross-correlation is a function of the lag between the signals (γ). The lag γ that maximizes the cross-correlation value presents the sample delay δ

$$\delta = \arg\max_{\gamma} R_{z_1 z_2}(\gamma) \tag{9}$$

The space delay (ρ) is equal to the space distance between two consecutive sensors that is a known value, but the time delay (τ), which is the time difference between the signals, should be estimated. The time delay (τ) can be calculated using the time interval between each sample ($1/f_t$) and the sample delay δ as follows

$$\tau = \frac{1}{f_t} \times \delta \tag{10}$$

where τ is the time delay between two signals, f_t is the sampling frequency of the sensors in the time domain, and δ is the sample delay between two signals.

Sampling methods

The multiple sensors (M sensors) start sampling at the same time with an identical sampling frequency f_t. Therefore, each sensor measures $I = T \times f_t$ samples over T second. As a result, M sensors measure M signals that have equal length (I samples). These signals include the samples from inactive, transient, and effective zones and generate a dataset as follows

$$Z_{m,i} = \begin{pmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,I-1} & z_{1,I} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,I-1} & z_{2,I} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ z_{M-1,1} & z_{M-1,2} & \cdots & z_{M-1,I-1} & z_{M-1,I} \\ z_{M,1} & z_{M,2} & \cdots & z_{M,I-1} & z_{M,I} \end{pmatrix} \quad \{i | i \in \mathbb{Z}^+, 0 \leqslant i \leqslant I\} \quad \{m | m \in \mathbb{Z}^+, 0 \leqslant m \leqslant M\} \tag{11}$$

Figure 4 illustrates a schematic view of the configuration of the proposed sensors, to explain the measurement zones and the required parameters. Figure 4(a) demonstrates the configuration of the wheel, rail, and the sensors that measure the rail response at different places. Figure 4(b) shows a schematic pattern of a defect signal $g(t)$. Figure 4(c) shows the inactive, transient, and effective zones of the first sensor. In Figure 4(d), the multiple sensors measure the rail response at different places. Each sensor makes a partial observation. The sensors provide different outputs in their effective zone due to the defect signal. Every sensor collects multiple samples in the effective zone that is coming from a specific portion of the wheel circumference. These samples are the combination of the wheel signal $w(t)$ and the defect signal $g(t)$. The number of samples collected in the effective zone is identical in every sensor if the train passes the sensors with a constant velocity and the sensors sample with an identical sampling frequency f_t. In Figure 4(d), the sensors collect N samples in their effective zone.

The effective zone is between the increasing and decreasing transient zones. In this zone, the first derivative of the signal is almost zero. The location of the effective zone can be determined using a low-pass filtered signal of the measured signal. The local maximum of the low-pass filtered signal shows the

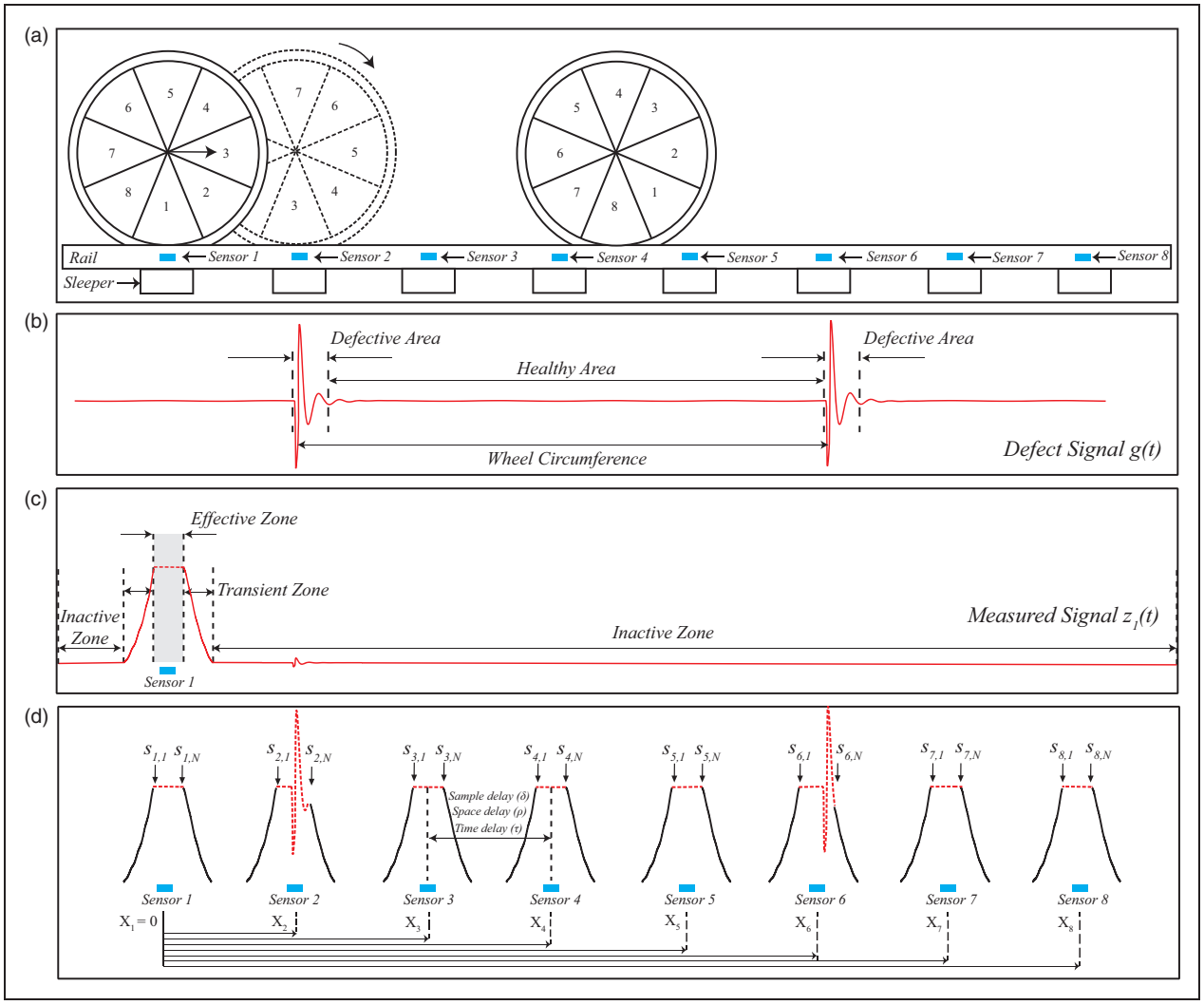

Figure 4. (a) The configuration of the wheel, rail, sleepers, and sensors; (b) the defect signal; (c) inactive, transient, and effective zones of a sensor; and (d) the multiple sensors that collect multiple samples in their effective zone.

middle point of the effective zone. The length of the effective zone depends on the physical property of the sensor and the sensor position. By knowing the middle point and the length of the effective zone, the beginning point is determined.

The signals have similar patterns but with a δ delay. Therefore, the corresponding points of the signals in two consecutive sensors have the following relation

$$z_{1,i} \mapsto z_{2,(i+\delta)} \tag{12}$$

It means that the sample i in the signal z_1 measured by the first sensor maps to sample $i+\delta$ in the second signal z_2. For example, when the ith sample of the signal in Figure 2(b) is the representative sample of the second wheel, the $(i+\delta)$th sample will be the corresponding sample of the second wheel in Figure 2(c). In general, when the sensors have equal space delay (ρ), and the wheel moves with the constant velocity, the relation between the corresponding points of the first signal to any other signal (measured by the sensor m) will be as follows

$$z_{1,i} \mapsto z_{m,(i+(m-1)\times\delta)} \tag{13}$$

We use the samples of the effective zone. Therefore, when the multiple sensors (M sensors) collect multiple samples (N samples) from the passage of a wheel, we can generate the following dataset from the collected samples

$$S_{m,n} = \begin{pmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,N-1} & s_{1,N} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,N-1} & s_{2,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s_{M-1,1} & s_{M-1,2} & \cdots & s_{M-1,N-1} & s_{M-1,N} \\ s_{M,1} & s_{M,2} & \cdots & s_{M,N-1} & s_{M,N} \end{pmatrix} \tag{14}$$

$$\{n | n \in \mathbb{Z}^+, 1 \leqslant n \leqslant N\}$$
$$\{m | m \in \mathbb{Z}^+, 1 \leqslant m \leqslant M\}$$
$$S_{m,n} \subseteq Z_{m,i}$$

In this dataset, each row presents the samples of the effective zone collected by each sensor. The space distance between the samples of each column (distance between the identical samples collected by two consecutive sensors, e.g. $s_{1,1}$ and $s_{2,1}$) is equal to the space distance between the sensors (ρ) that is a known value. Therefore, the space distance of the sensors defines the space distance between the samples of each column.

Data fusion and signal reconstruction

The sensors collect a few samples on every wheel revolution (as presented in Figure 4). The sampling frequency in the space domain f_s obviously violates the Nyquist criterion by subsampling lower than the fundamental frequency of the signal in the space domain. To respond to the Nyquist sampling challenge, the nature of the defect signal gives a hint. The defect signal is a periodic signal that is replicated in every wheel revolution. The distances between the main peaks indicate the wheel circumference that is the fundamental period of the signal. The samples selected from different sensors can be mapped over the circumferential coordinate using the wheel circumference and the sensors' configuration. Figure 5 presents a schematic illustration of the fusion process, in which X_m is the space position of the sensors, L_w is the wheel circumference length, and Y_m is the corresponding position of the sensors over the circumferential coordinate.

In Figure 5, sensors 1–5 sample from the first revolution, and the sensors 6–8 sample from the second revolution. Y_m determines the position of the sensors 6–8 over the wheel circumference. The samples collected by sensors 6–8 fill the gaps between the sensors 1 and 5 and improve the quality of the signal. By extending the sampling procedure to the other turns, more samples from different portions of the wheel are collected to fill the missing data. When the sample/cycle ratio is not an integer quantity, other replications of the wheel revolution collect supplementary samples. Instead, this method will sample multiple times the same points when the circumference is 3000 mm (954.9 mm diameter) with 600 mm sensor interval. In this case increasing the number of sensors is not useful for collecting the missing data. By bearing in mind the range of the wheel diameter between 840 and 920 mm,[16] and assuming the 600 mm sensor interval, the number of sampling from the wheel revolution will be 4.39–4.81 times per wheel revolution. Hence, increasing the number of sensors improves the signal quality.

Data fusion for single sampling method

In this subsection, only a single sample is used as the output of each sensor, which is called single sampling method (SSM). By selecting the sample $s_{1,1}$ as the output of the first sensor for the wheel, the sample $s_{2,1}$ will be the output of the second sensor for the

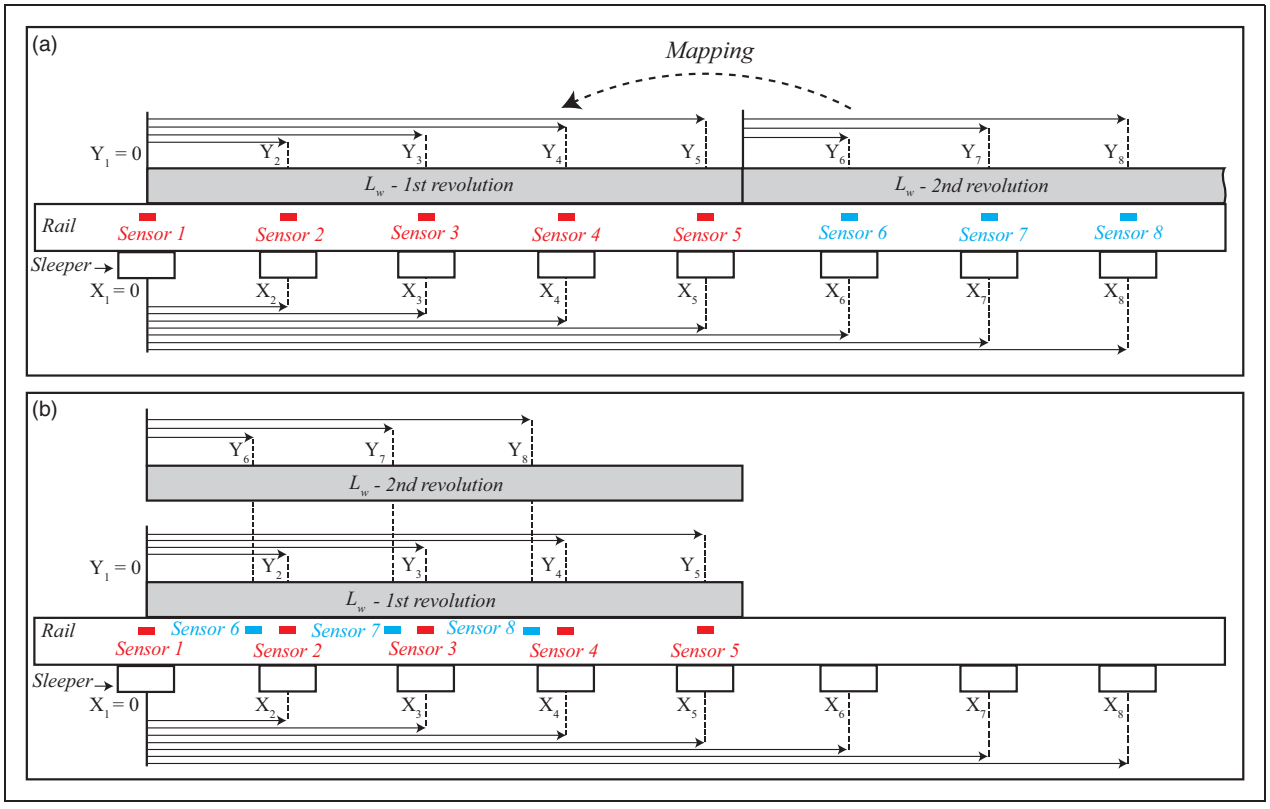

Figure 5. The illustration of the fusion process.

wheel that measures another point of the wheel with ρ distance in the space domain. As a result, a set of samples as the output of different sensors for the wheel are acquired as follows

$$S_{m,1} = \begin{bmatrix} s_{1,1} \\ s_{2,1} \\ \vdots \\ s_{M-1,1} \\ s_{M,1} \end{bmatrix} \quad \{m|m \in \mathbb{Z}^+, 1 \leqslant m \leqslant M\} \tag{15}$$

The samples of the subdataset ($S_{m,1}$) can be fused over the circumferential coordinate to generate a signal for the wheel using the following equation

$$Y_{m,1} = X_m - \left(L_w \times \left\lfloor \frac{X_m}{L_w} \right\rfloor\right) \quad \{m|m \in \mathbb{Z}^+, 1 \leqslant m \leqslant M\} \tag{16}$$

where $Y_{m,1}$ is the corresponding position of the samples over the circumferential coordinate, X_m is the space position of the sensors, L_w is the wheel circumference length, and $\lfloor\rfloor$ is the round operator toward the nearest integer less than or equal to the element. The remainder after division of the sensor position by the circumference length determines the sensor position on the circumferential coordinate. A new signal (ψ_s) is generated using the magnitude ($S_{m,1}$) and the position ($Y_{m,1}$) of the samples as follows

$$\psi_s = [Y_{m,1}, S_{m,1}] \tag{17}$$

The signal (ψ_s) reconstructed by SSM has M samples over the circumferential coordinate.

Data fusion for multiple sampling method (MSM)

In this subsection, MSM uses all the data collected in the effective zone. To do this the space distance between the samples of each row (λ) should be estimated. For example, the space distance between $s_{1,1}$ and $s_{1,2}$ is required. When the first sample of a sensor is positioned over the circumferential coordinate, the other samples collected by the sensor in the effective zone have the following positions

$$Y_{m,n} = Y_{m,1} + ((n-1) \times \lambda) \quad \{n|n \in \mathbb{Z}^+, 1 \leqslant n \leqslant N\} \quad \{m|m \in \mathbb{Z}^+, 1 \leqslant m \leqslant M\} \tag{18}$$

Then the multiple samples (N samples) measured by the sensors are positioned using the space distance between the samples (λ). As a result, the reconstructed signal (ψ_s) is generated using the magnitude ($S_{m,n}$) and the position ($Y_{m,n}$) of the samples

$$\psi_s = [Y_{m,n}, S_{m,n}] \tag{19}$$

The MSM reconstructs the signal (ψ_s) by $M \times N$ samples. Intuitively, these samples are not uniformly distributed over the circumferential coordinate.

The sampling frequency of a sensor determines the time interval between the samples collected by the sensors. By considering the constant sampling frequency in the time domain, the train velocity determines the space frequency (space distance) of the samples collected by the sensor. The space interval between the samples can be defined using the space delay ρ and the samples delay δ as follows

$$\lambda = \frac{\rho}{\delta} \tag{20}$$

This relation can be rewritten based on the train velocity V and the sensor sampling frequency f_t as the influential factors as

$$\lambda = \frac{V}{f_t} \tag{21}$$

The space distance between the samples (λ) determines the space resolution of the measurement in the space domain. For example, when a sensor is sensing by 10 kHz sampling frequency, for a train with 10 m/s velocity, the space distance between the samples is 1 mm. In addition, the sensors have a limited effective zone. Therefore, the number of samples that can be used as the outputs of the sensors is determined by the space distance between the samples (λ) and the length of the effective zone (L_e) as presented below

$$N = \frac{L_e}{\lambda} \tag{22}$$

To determine the space distance between the samples (λ), the space delay (ρ) and the samples delay (δ) can be directly used as presented in equation (20). Moreover, the train velocity can be indirectly used in equation (21) that is estimated in the next section.

Estimation of train velocity

Filograno et al.[17] and Tam et al.[18] estimated the train velocity using the passage time between two axles. To find the axle distance, they counted the axle numbers and compared with the known information of different trains to identify the train type and the matching axle distance. This method uses only one sensor but relies on the other information about the trains that should be provided from other sources. Here, we estimate the train velocity using the multiple sensors that do not require identifying the train type. The velocity

is the space passed over the unit of time. Accordingly, the train velocity can be estimated using the space delay and the time delay as

$$V = \frac{\rho}{\delta} \times f_t \tag{23}$$

where ρ and f_t are the known values and the sample delay (δ) is estimated by the cross-correlation in equations (8) and (9).

Validation test

To assess the fusion model, a validation test is designed. The purpose of the data fusion process is to reconstruct the wheel defect signal $g(t)$ from the measured signals $z(t)$. If the reconstructed signals represent the features of the wheel defects, the fusion model has fulfilled its intended purpose. The validation test contains two steps: data generation and data fusion.

The general overview of the validation test and the detailed flowchart of the process are presented in Figure 6. The data generation step uses VI-Rail to model the interaction of a rail and defective wheels. VI-Rail is a commercial multibody dynamics software that has been built upon MSC Adams. In the first step, the defect model generates the defect on the wheel. Then, using the required parameters, VI-Rail models the wheel–rail interaction and generates some outputs. The data generated by VI-Rail is exported to MATLAB as the input of the data fusion process. In Figure 6(b), the position of the wheel (effective zone of the sensor) is estimated using a low-pass filter. The delay between the signals and the train velocity is used to select the required samples and make the dataset $S_{m,n}$. In Figure 6(c), the data collected in dataset are fused to reconstruct a new signal, using the wheel circumference and the sensor configuration.

VI-Rail models the interaction of the track and vehicle by considering their subsystems such as car body, sleepers, rail pads, wheelsets, primary and secondary suspensions, dampers, and antiroll bars. The simulations are carried out for a passenger vehicle based on the Manchester Benchmarks.[19] The assembly model consists of a vehicle and a flexible track. The vehicle is one wagon composed of a car body and two bogies each having four S1002 wheels.

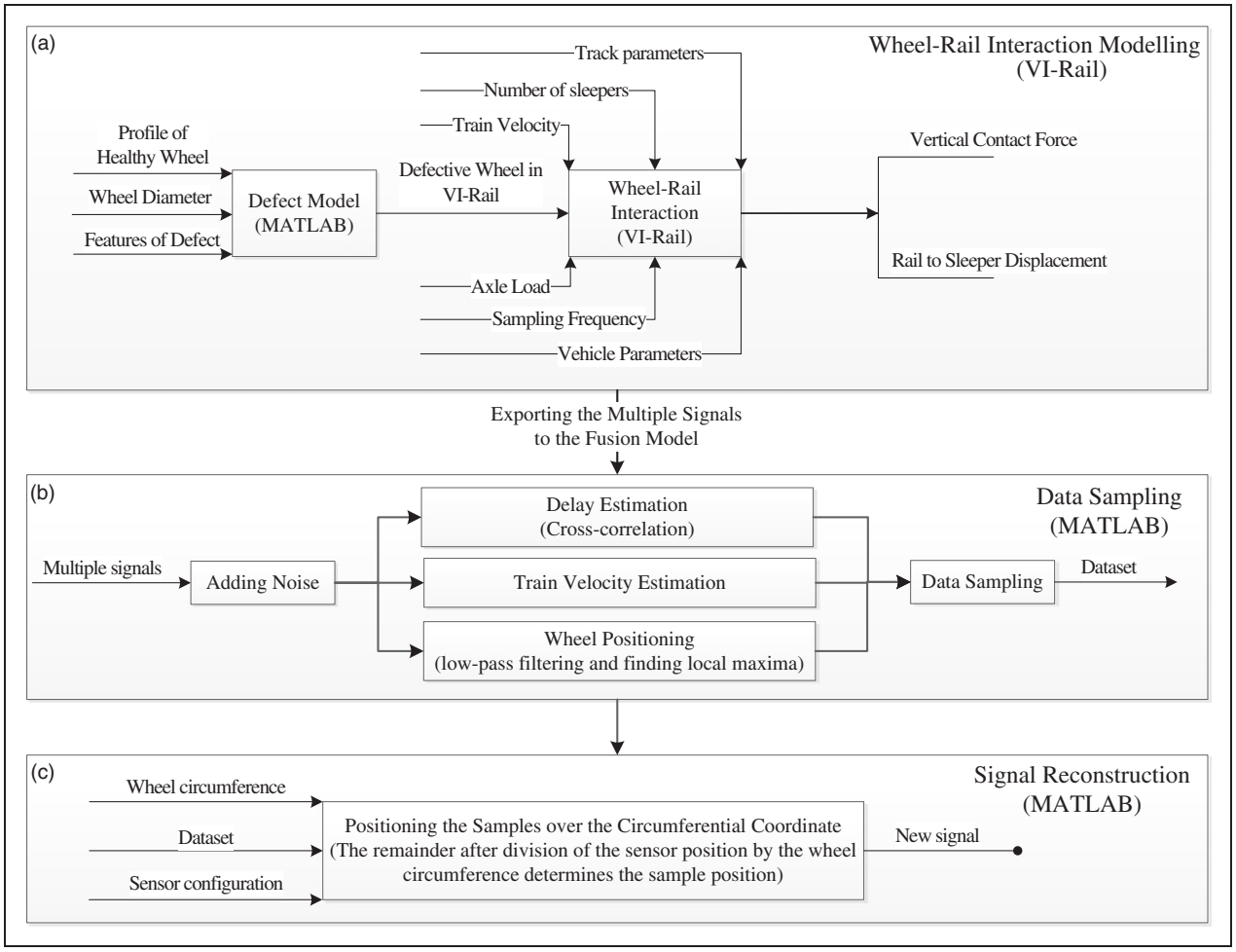

Figure 6. The process of the validation test.

The flexible track contains a straight UIC60 rail. In this model the rail mass and the inertia properties are concentrated on each rail sleeper. The detailed explanation of the track and vehicle structure falls outside the scope of this article.

Nielsen and Johansson[9] classified and reviewed the wheel defects and discussed the reasons of their development. Wheel flats are the severe defects that cause high impact forces. According to the Swedish criterion the wheels with 40–60 mm flat length should be reprofiled as detected during visual inspections.[10] Hence, this research considers two flats with 40 and 60 mm length (0.4 and 1 mm depth, respectively). The flats are on the nominal contact region. Wheel flats generate defect signals containing high-frequency components while the wheels with periodic OOR generate contact force with low frequency variation. The OOR covers the entire wheel profile and the circumference. Therefore, to model an OOR, the wheel diameter is varied according to the defect shape. This research considers a third-order out-of-round wheel that has three harmonics around the wheel circumference with 0.3 mm amplitude.

Output of the data generation process

VI-Rail provides a range of outputs such as the contact force, rail and sleeper acceleration, and rail and sleeper displacement. The primary desired output is the rail strain that is used in practice, but VI-Rail cannot provide the rail strain signal. By considering the rail as a transducer, the contact force signal is transformed into the rail response such as strain, acceleration, and displacement. In this research due to lack of strain signal, the vertical rail to sleeper displacement is used as the output of the data generation process. Every sleeper is considered as a sensor that measures the rail to sleeper displacement signal. The sleepers have a discrete and periodic configuration like the sensors' configuration.

Figure 7 displays the typical rail to sleeper displacement signals for two consecutive sleepers while the first wheel has 40 mm flat. In this signal, the peaks corresponding to the passage of the wheels were close to each other and generated two big peaks containing two smaller peaks. These signals show the variation in the vertical rail to sleeper position sensed in one side of the sleeper. They have four peaks representing the passage of a wagon with four wheels in that side of track. The wheel flat produced the defect signal containing high-frequency components. The defect signal is superimposed on the displacement signals. Figure 7(b) shows a delay due to the distance between the sleepers.

Results of the data fusion process

Figure 8 presents the samples collected by 59 sleepers (sensors) for a wheel with a 40 mm flat using the MSM. In this figure, the samples represent all samples measured by the effective zones collected in the dataset $S_{m,n}$. As is clear from Figure 8, the collected samples provide a limited piece of information.

Figure 9 makes a comparison between the signals reconstructed by the SSM and MSM based on equations (16) and (18). In this example, the samples collected in Figure 8 are fused to reconstruct new informative signals. Figure 9(a) shows the contact force that is provided by VI-Rail. The contact force is transferred to the wheel and rail and makes the dynamic response of the wheel and rail. In this example, the rail to sleeper displacement signal is used as

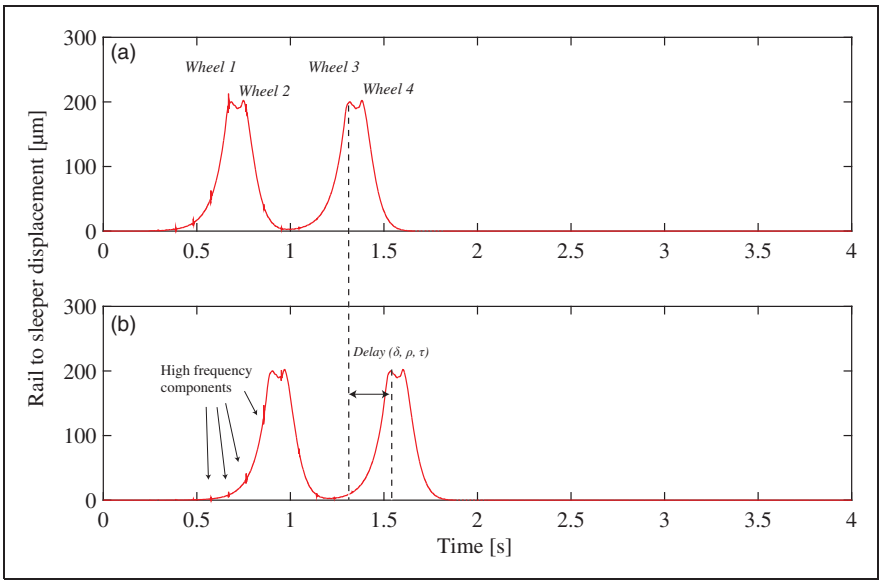

Figure 7. (a) The rail to sleeper displacement signal for the passage of four wheels while the first wheel is defective. This signal is considered as the measured signal $z(t)$ and (b) The rail to sleeper displacement signal for the consecutive sleeper as the second sensor.

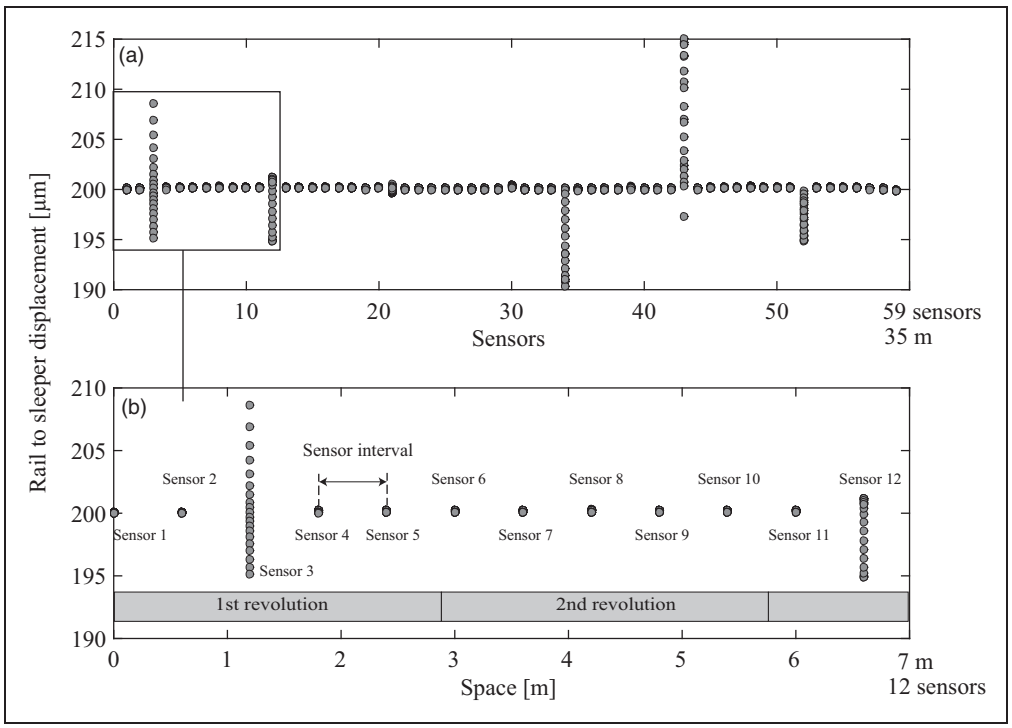

Figure 8. (a) The simulated data sampled by 59 sensors using the MSM and (b) the magnified view of the plot (a).

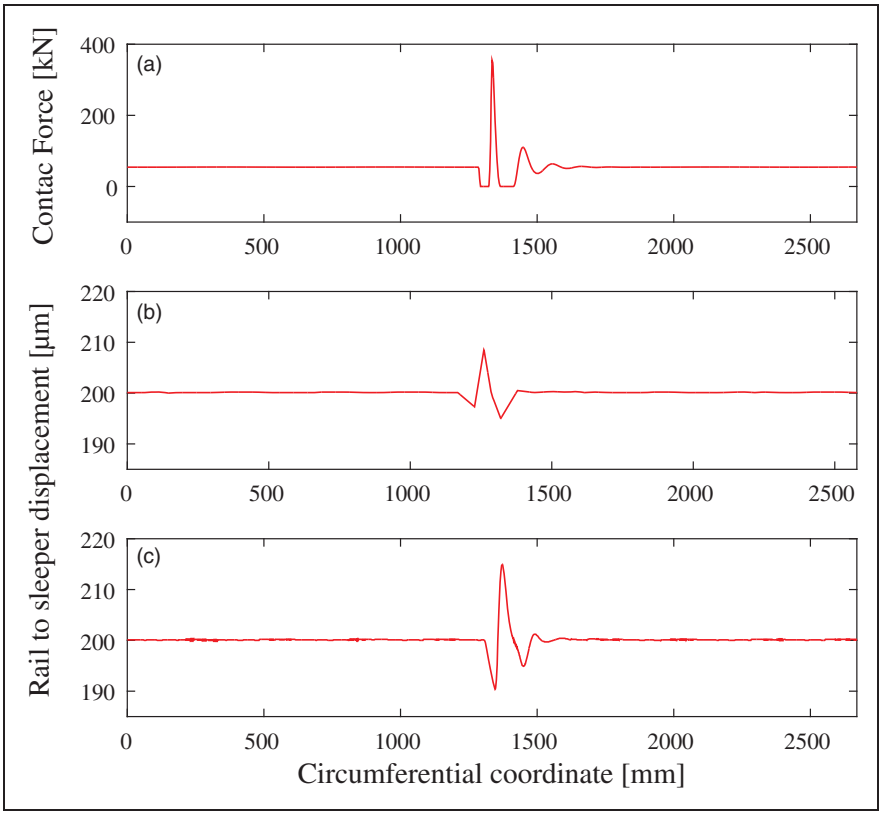

Figure 9. (a) The simulation result of a wheel–rail contact force for a wheel with 40 mm flat and 30 m/s velocity. The signals reconstructed from the rail to sleeper displacement signal collected by 59 sensors using (b) the SSM and (c) the MSM.

the measured signals $z(t)$. In Figure 9(b), the SSM used 59 sensors to collect data and reconstruct a signal. In Figure 9(c), the MSM used the same sensors but exploited more samples. In Figure 9(b), the first downward of the defective area was not sensed completely, while in Figure 9(c), the MSM overcame this problem. The reconstructed signals represent the contact force signal that is a function of the wheel defect.

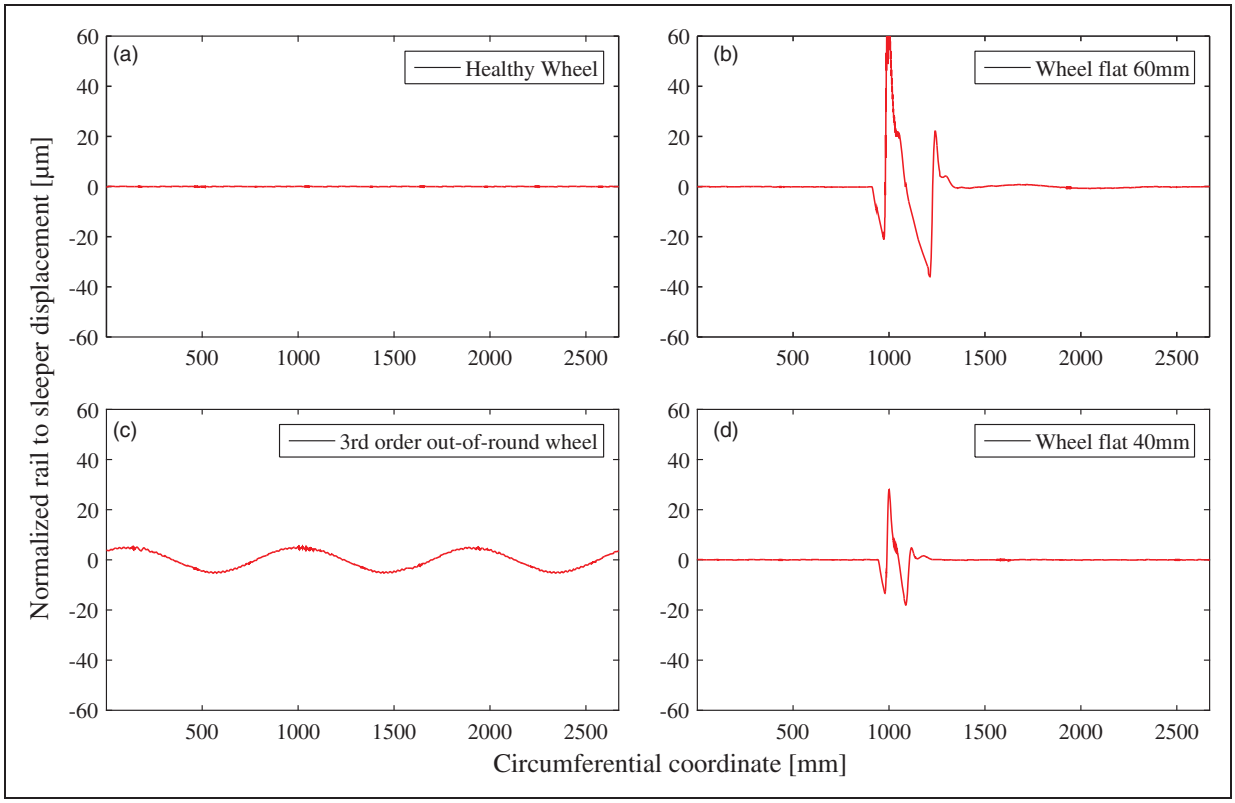

Figure 10. The signals reconstructed for the wheels with different defects: (a) healthy wheel, (b) 60 mm wheel flat, (c) third-order out-of-round wheel, and (d) 40 mm wheel flat. The signals have been normalized by subtracting the average of the signals.

Figure 10 presents other examples of the signals reconstructed by the proposed method for a healthy wheel, 40 and 60 mm flats, and third-order out-of-round wheel. The difference between the signals is due to the wheel condition. These signals adequately represent the features of the wheel defects. For example, the reconstructed signal in Figure 10(c) shows a sinusoidal wave with three periods covering the wheel circumference that is accurately representing the third-order out-of-round wheel defect. The defect signals in Figure 10(b) and (d) have similarity due to the type of defects and have differences due to their severity. As a result, defect signals can be reconstructed to use for defect detection and identification.

Conclusion

The magnitude of the contact force contains a limited piece of information about the wheel defect. Therefore, this paper proposed a fusion method to reconstruct a signal containing the pattern of the contact force that is a function of the wheel defect. To achieve this purpose, this paper has developed the required fusion method and described the theoretical relations between the samples collected by multiple sensors such as WILDs. The results of the validation test showed that the defect signals reconstructed by the proposed method completely represented the features of the wheel defects used in the data generation step. Therefore, the proposed method opens up the possibility of detecting and identifying the defects including the minor and long-wave defects at an early stage.

The fusion process is influenced by several parameters such as number of sensors, length of the effective zone, and wheel circumference as the fundamental period of the defect signal. The effect of the influential parameters on the reconstructed signals should be investigated further.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was partly supported by the Ministry of Science, Research and Technology of the Islamic Republic of Iran (MSRT).

ORCID iD

Alireza Alemi http://orcid.org/0000-0002-4693-0022

References

1. Chong SY, Lee JR and Shin HJ. A review of health and operation monitoring technologies for trains. *Smart Struct Syst* 2010; 6: 1079–1105.

2. Alemi A, Corman F and Lodewijks G. Condition monitoring approaches for the detection of railway wheel defects. *Proc IMechE, Part F: J Rail and Rapid Transit* 2017; 231: 961–981.
3. Partington W. Wheel impact load monitoring. *Proc ICE Transp* 1993; 100: 243–245.
4. Brickle B, Morgan R, Smith E, et al. *Identification of existing and new technologies for wheelset condition monitoring*. London, United Kingdom: Technical report, Rail Safety and Standards Board (RSSB), 2008.
5. Stratman B, Liu Y and Mahadevan S. Structural health monitoring of railroad wheels using wheel impact load detectors. *J Fail Anal Prev* 2007; 7: 218–225.
6. Belotti V, Crenna F, Michelini RC, et al. Wheel-flat diagnostic tool via wavelet transform. *Mech Syst Signal Process* 2006; 20: 1953–1966.
7. Filograno ML, Corredera P, Rodriguez-Plaza M, et al. Wheel flat detection in high-speed railway systems using fiber Bragg gratings. *IEEE Sens J* 2013; 13: 4808–4816.
8. Stone D, Kalay S and Tajaddini A. Statistical behaviour of wheel impact load detectors to various wheel defects. In: *International wheelset congress*, Sydney: National conference publication (Institution of Engineers, Australia), NSW, August 1991, pp.9–13.
9. Nielsen JCO and Johansson A. Out-of-round railway wheels – a literature survey. *Proc IMechE, Part F: J Rail and Rapid Transit* 2000; 214: 79–91.
10. Johansson A and Nielsen JO. Out-of-round railway wheels – wheel-rail contact forces and track response derived from field tests and numerical simulations. *Proc IMechE, Part F: J Rail and Rapid Transit* 2003; 217: 135–146.
11. Asplund M, Palo M, Famurewa S, et al. A study of railway wheel profile parameters used as indicators of an increased risk of wheel defects. *Proc IMechE, Part F: J Rail and Rapid Transit* 2016; 230: 323–334.
12. VI-Grade Engineering Software & Services. *VI-Rail 16.0 documentation*. Technical report, VI-grade GmbH, Marburg, Germany, 2014.
13. Nielsen J. Out-of-round railway wheels. In: R. Lewis and U. Olofsson *Wheelrail interface handbook*. Woodhead Publishing: Cambridge, Elsevier, 2009, pp.245–279. License number: 4322550888936.
14. Jerri A. The Shannon sampling theorem – its various extensions and applications: a tutorial review. *Proc IEEE* 1977; 65: 1565–1596.
15. Knapp C and Carter G. The generalized correlation method for estimation of time delay. *IEEE Trans Acoust Speech Signal Process* 1976; 24: 320–327.
16. UIC Leaflet. *UIC 510-2 OR, trailing stock: wheel and wheel sets conditions concerning the use of wheels of various diameters*, Paris, France: International Union of Railways. Technical Report, May 2004.
17. Filograno ML, Corredera Guillen P, Rodriguez-Barrios A, et al. Real-time monitoring of railway traffic using fiber Bragg grating sensors. *IEEE Sens J* 2012; 12: 85–92.
18. Tam HY, Liu SY, Guan BO, et al. Fiber Bragg grating sensors for structural and railway applications. In: Rao YJ, Kwon OY and Peng GD (eds) *Proceedings of SPIE – the International Society for Optical Engineering*, Bellingham, USA: SPIE Publications, vol. 5634, pp.85–97.
19. Iwnicki S. Manchester benchmarks for rail vehicle simulation. *Veh Syst Dyn* 1998; 30: 295–313.