Kernel Density Estimation Based Gaussian and Non-Gaussian Random Vibration Data Induction for High-Speed Train Equipment

PENG WANG¹,²,³, HUA DENG¹,³, (Member, IEEE), YI MIN WANG², YUE LIU², AND YI ZHANG¹,³

¹School of Mechanical and Electrical Engineering, Central South University, Changsha 410083, China
²CRRC Zhuzhou Electric Locomotive Research Institute Company Ltd., Zhuzhou 412001, China
³State Key Laboratory of High Performance and Complex Manufacturing, Changsha 410083, China

Corresponding author: Yi Zhang (zhangyicsu@csu.edu.cn)

This work was supported by the National Key Technology Research and Development Program of the Ministry of Science and Technology of China under Grant 2015BAG14B01.

Abstract Because general statistics tolerance is not applicable to the induction of non-Gaussian vibration data and the methods for converting non-Gaussian data into Gaussian data are not always effective and can increase the estimation error, a novel kernel density estimation method in which induction is carried out on power spectral density data for the measured vibration of high-speed trains is proposed in this paper. First, data belonging to the same population of power spectral density are merged into the same feature sample. Then, the probability density function of all power spectral density values at the first frequency point is calculated through the kernel density estimation method, and the upper-limit estimate of all power spectral density values under the set quantile is obtained. This process is repeated, and the upper limit values of the power spectral density values at all frequency points can be obtained to convert the measured acceleration data to the acceleration power spectral density spectrum of the vibration test. Engineering examples are used to verify the proposed method. For the same Gaussian power spectral density data, the relative error between the root mean squares of the power spectral density spectrum obtained from induction by the kernel density estimation method and the statistics tolerance is 0.155% ~ 1.55%; for the non-Gaussian power spectral density data, the acceleration power spectral density spectrum of the non-Gaussian vibration can be obtained with the induction by the kernel density estimation method. The proposed kernel density estimation method satisfies the induction requirements for the measured Gaussian and non-Gaussian vibration data of high-speed trains with two different distributions, and its induction results have very good universality and estimation accuracy.

Index Terms Random vibration, data induction, kernel density estimation, non-Gaussian distribution, Gaussian distribution.

I. INTRODUCTION

When the operating speed of a high-speed train reaches 300 to 380 km/h, the vibration environment experienced by its on-board equipment also becomes harsher, and increasingly more fatigue failures, such as cracks in the components, the deformation and dropping of parts, and the loosening of fasteners, are triggered by the vibrations. Thus, the faults caused by vibration fatigue are an important factor affecting the safe operation of high-speed trains, and they pose a difficult problem that urgently needs to be solved to increase the service life of on-board equipment.

To ensure the fatigue durability and vibration environment adaptability of high-speed train equipment, manufacturers in various countries generally complete vibration testing for rail transit equipment according to IEC61373 [1]. The IEC61373 standard was formulated in 2010. At the time, only data from some European railways were used; the corresponding testing conditions were formulated according to the Gaussian data processing method, and the effect of non-Gaussian vibration on the onboard equipment was not taken into account. To cope with the continuous development
of vibration testing technology, IEC/TC9 MT 61373 convened a working group conference in December 2019 to clarify that the operating vibration data of trains in various countries will be collected under the premise of fully taking into account non-Gaussian data to reevaluate the harshness grades of vibration testing in IEC61373 [2], [3]. In addition, in the IEC 60068-2-64, which was published in 2019, guidelines for the non-Gaussian vibration test, including three methods for generating non-Gaussian random vibrations—amplitude adjustment, phase modification, and uneven phase—were added, but specific testing conditions were not given [4]. Therefore, non-Gaussian vibration has become a new direction for the development of vibration testing technology [5]–[8]. At the same time, scientific non-Gaussian vibration testing conditions still need to be formulated to ensure that this testing method is normative and operable. However, if vibration testing conditions have to be formulated according to the measured environmental data, an appropriate vibration data induction method, which can determine the success or failure of the vibration testing results, is crucial [2], [3], [10], [11].

Commonly used vibration data induction methods include the extremum enveloping method and the statistical tolerance method; these methods solve the induction problem of data that conform to a Gaussian distribution and are widely applied in engineering practice [9]–[14]. However, in the extremum envelope method, enveloping is carried out according to the maximum order of all data in an induction state without taking into account the data sample size and data distribution characteristics; therefore, the induced testing conditions are often overly harsh. The statistical tolerance method is a leap of vibration data induction from the extremum envelope method to a statistical method, but this method is only applicable to the processing of Gaussian vibration data and cannot satisfy the induction of non-Gaussian data [10], [11]. To solve the problem of the induction of measured non-Gaussian vibration data, the Johnson or Bootstrap method is often used. These two methods can convert non-Gaussian data into data that approximately obey a normal distribution, but their transformation process is not always effective; at the same time, more intermediary conversions will also increase the estimation error [15]–[18].

With the kernel estimation method [19]–[22], there is no need to assume in advance that the data obey a certain specific standard parameter distribution, and this method can directly fit the Probability Density Function (PDF) of the data sample. Due to this unique advantage, this method has attained wide application for Gaussian or non-Gaussian data processing in many fields, such as engineering, medicine, finance, and the Internet [23]–[27]. However, according to the literature, there is still no reporting on the application of the kernel estimation method in vibration data induction. In this paper, the fact that the kernel density estimation method does not need to assume the sample distribution in advance is utilized, and an induction method for the Power Spectral Density (PDF) of the measured vibration based on density estimation is proposed to thereby obtain the upper-limit estimate of the PDF data under a set confidence level, which satisfies the induction requirements for the measured Gaussian and non-Gaussian vibration data of high-speed trains. Through engineering examples, the induction method is verified to have very good universality and estimation accuracy.

II. PROBLEM DESCRIPTION

The measured vibration data of high-speed trains are time-domain acceleration data in which abnormal data such as idling-speed operation and transient impact while hooking up are included. To this end, time-domain data should be tested before data induction to reject measurement data that cannot be effectively modified, and then the measured time-domain data are tested for stationarity, periodicity, ergodicity, and normality. Then, the commonly used Cooley-Tukey method [28] is used to obtain the PDF of the time-domain acceleration data.

In the JGB/Z 126 [10], MIL-STD-810G [12], and NASA-hdbk-7005 [13] standards, it is suggested that the statistical tolerance method be used to carry out induction on the vibration data that conform to a stationary Gaussian distribution, and the induction method is given.

Set all channel measured acceleration data PDF as \( \tilde{G}_{k(i,j)}(i = 1, 2, \ldots, L_1; j = 1, 2, \ldots, M_1; k = 1, 2, \ldots, N_1) \), where \( L_1 \) is the number of channels, \( M_1 \) is the sample capacity, and \( N_1 \) is the number of spectral lines; set the PDF of the same data channel within the scope of the frequency band being analyzed to obey a \( \chi^2 \) distribution. Calculate the root mean square value \( RMS(i,j) \) of \( \tilde{G}_{k(i,j)} \); then when the statistical degree of freedom is greater than 45, \( RMS(i,j) \) approximately obeys normal distribution.

The statistical tolerance induction method mainly includes the following two steps [10], [11]:

A. STEP 1: MERGING CHANNELS TO FORM THE FEATURE SAMPLE

1) PARAMETER ESTIMATION

Estimate the mean value \( \overline{X}_{i} \) and the variance \( S_{i}^2 \) of \( RMS(i,j) \):

\[
\begin{align*}
\overline{X}_{i} &= \frac{1}{M_1} \sum_{j=1}^{M_1} RMS(i,j) \\
S_{i}^2 &= \frac{1}{M_1 - 1} \sum_{j=1}^{M_1} [RMS(i,j) - \overline{X}_{i}] 
\end{align*}
\]

(1)

2) HYPOTHESIS TESTING

Calculate the statistics \( F_n(i,m) \) and \( t_n(i,m) \) for the mean value \( \overline{X}_{i} \) and the variance \( S_{i}^2 \):

\[
\begin{align*}
F_n(i,m) &= \frac{S_{i}^2}{S_{m}^2} \\
t_n(i,m) &= \frac{\overline{X}_{i} - \overline{X}_{m}}{\sqrt{(S_{i}^2 + S_{m}^2)/M_1}}
\end{align*}
\]

(2)

If the PDF values of channels \( i \) and \( m \) belong to the same population, then \( F_n(i,m) \) obeys the F distribution at the
degree of freedom \((M_1 - 1, M_1 - 1)\), and \(t_{(i, m)}\) obeys the central \(t\) distribution at the degree of freedom \(2(M_1 - 1)\).

Under a given confidence level \((1 - \alpha)\), if (3) holds, then channels \(i\) and \(m\) are merged into the same population.

\[
\begin{align*}
F_{(M_1 - 1, M_1 - 1); \alpha / 2} & \leq F_{n(i, m)} \leq F_{(M_1 - 1, M_1 - 1); \{1 - \alpha / 2\}} \\
|t_{n(i, m)}| & \leq t_2(M_1 - 1; \{1 - \alpha / 2\})
\end{align*}
\]  

(3)

Merge the PSD belonging to the same population to form the feature sample \(G_k(p, q)\) \((p = 1, 2, \ldots, P_1; q = 1, 2, \ldots, Q_p)\), where \(P_1\) is the number of feature samples, and \(Q_p\) is the feature sample capacity.

**B. STEP 2: CALCULATION OF THE UPPER TOLERANCE LIMIT**

1) **PARAMETER ESTIMATION**

Carry out the transformation on the obtained feature sample according to (4) to obtain the sample \(x_k(p, q)\) that approximately obeys a normal distribution:

\[
x_k(p, q) = \sqrt{G_k(p, q)} \quad (k = 1, 2, \ldots, N_1; \quad p = 1, 2, \ldots, P_1; \quad q = 1, 2, \ldots, Q_p)
\]  

(4)

Estimate the mean value \(\bar{x}_k(p)\) and the variance \(S_k^2(p)\) of the sample \(x_k(p, q)\):

\[
\begin{align*}
\bar{x}_k(p) &= \frac{1}{Q_p} \sum_{q=1}^{Q_p} x_k(p, q) \\
S_k^2(p) &= \frac{1}{Q_p - 1} \sum_{q=1}^{Q_p} [x_k(p, q) - \bar{x}_k(p)]^2
\end{align*}
\]  

(5)

2) **CALCULATION OF THE UPPER TOLERANCE LIMIT COEFFICIENT**

Under a given confidence level \((1 - \alpha)\) and quantile \(\beta\), calculate the upper tolerance limit coefficient as follows:

\[
F = \frac{t_{Q_p - 1; \{1 - \alpha / 2\}}}{\sqrt{Q_p}} + Z_\beta \sqrt{\frac{Q_p - 1}{\chi_{Q_p - 1; \alpha}^2}}
\]  

(6)

3) **ESTIMATION OF THE UPPER TOLERANCE LIMIT OF THE PSD**

The upper tolerance limit estimation for the \(p\)th feature sample is

\[
G_k(p) = [\bar{x}_k(p) + F \cdot S_k(p)]^2
\]  

(7)

The aforementioned statistical tolerance method is applicable to the processing of Gaussian vibration data but cannot be used for the induction of non-Gaussian data. To solve the induction problem of measured non-Gaussian vibration data, the Johnson or Bootstrap method is used [11], [12], [21]. Before Step 2 above, it is necessary to first convert non-Gaussian data into data that approximately obey a normal distribution by using the Johnson or Bootstrap method. Step 2 is then carried out. However, the method is complex as many iterations are needed to find the optimal conversion function, resulting in that the overall conversion efficiency is low and the calculation cost is large. Because the conversion will increase the estimation error and is not always effective, new methods need to be studied and proposed to realize the induction of non-Gaussian data. This is the motivation for proposing the kernel density estimation induction method to carry out induction on Gaussian and non-Gaussian data in this paper.

**III. NEW DATA INDUCTION METHOD BASED ON KERNEL DENSITY ESTIMATION**

Generally, when carrying out induction on data, it is necessary to know that the data obey a certain specific standard parameter distribution. For Gaussian data, there are standard induction methods, but the induction of non-Gaussian data is a greater challenge. Because the kernel density estimation method does not require the inducted data to obey a certain specific standard parameter distribution, this method has the unique advantage of directly fitting the probability density function. In this section, the application of the kernel density estimation method to the induction of Gaussian and non-Gaussian data is explored.

**A. PROCEDURES OF THE NEW INDUCTION METHOD**

Set the acceleration PSD for the \(j\)th test of the \(i\)th channel in the vibration test as \(G_{k(i, j)}\) \((i = 1, 2 \ldots L; j = 1, 2 \ldots M, \ k = 1, 2 \ldots N)\), where \(L\) is the number of test channels; \(M\) is the number of samples for a certain test channel, which is the number of PSD values for a certain channel; and \(N\) is the number of spectral lines. Therefore, \(G_{k(i, j)}\) represents the PSD value at the \(k\)th frequency point of the PSD for the \(i\)th test channel and the data induction method is as follows:

1) Carry out hypothesis testing on the data of all channels (a total of \(L\)) and merge the PSD samples belonging to the same population as the same feature sample to thereby obtain \(P\) merged feature samples.

2) Assuming that \(S(S \leq M)\) portions of the PSD for the \(i\)th \((i \leq L)\) channel are obtained, then SPDs values for each portion of the PSD at the \(k\)th \((k \leq N)\) frequency point can be obtained. Now, carry out kernel density estimation on the SPDs values that satisfy an arbitrary distribution to obtain the probability density function \(f_k(x)\) for the \(S\) PSD values at the \(k\)th frequency point, and then obtain the upper-limit estimate \(x_\beta\) of the SPD values under the set quantile \(\beta\).

3) Repeat the first step of the process to obtain the upper limit values of the PSD at all frequency points (a total of \(N\)).

4) Carry out the calculations for all the channels in the same feature sample (a total of \(P\)) according to the methods in the second and third steps. Then \(P\) measured spectra \(G(L)\) of random vibrations under the quantile \(\beta\) can be obtained.

The data induction flow is shown in Figure 1.
Start

Merge $L$ channels to form $P$ feature samples

Import $M$ portions of PSD data from channel $i$

Import $M$ PSDs at frequency point $k$

Carry out kernel density estimation on $M$ PSD values at frequency point $k$, obtain $x_p$

Obtain the $x_p$ of all ($N$) frequency points

Obtain $P$ portions of PSD ($G(L)$)

End

FIGURE 1. Data induction flowchart.

**B. FORMATION OF THE FEATURE SAMPLE BY MERGING CHANNELS**

Before carrying out kernel density estimation, the PSD samples belonging to the same population should be merged into the same feature sample. When the data approximately obey a Gaussian distribution, the sample parameter testing method can be used to carry out hypothesis testing. When they do not obey a Gaussian distribution, then nonparametric statistics are used to determine if the distribution functions of the two populations are equal. There are many nonparametric test methods including the runs test, the signed rank sum test, and the Kolmogorov–Smirnov (K-S) test [29], [30]. For the K-S test, a distance between the empirical distribution functions of two samples is defined first by the K-S statistic and then the null distribution of this statistic is calculated under the null hypothesis that the samples come from the same distribution [29]. Among these test methods, the runs test and the K-S test are relatively complex, and the signed rank sum test is relatively simple and efficient in data utilization. Therefore, the signed rank sum test method is used in this study.

Let there be PSD samples of the two channels $i$ and $j$. The sample capacities of the two channels are both $n$. Calculate and obtain the $RMS$ value of each PSD as $RMS_i(x)$, $RMS_j(y)(x, y = 1, 2, \ldots , n)$, and then carry out a sign rank sum test on the two $RMS$ samples, $RMS_i$, $RMS_j$; if the test result accepts $H_0$, then the PSD data of the $i$ and $j$ channels are deemed to be from the same population and are merged into the same feature sample.

**C. KERNEL DENSITY ESTIMATION**

1) **BASIC CONCEPT OF KERNEL DENSITY ESTIMATION**

Set the distribution function of variable $X$ as $F(x)$. Its probability density function is

$$f(x) = F'(x) = \lim_{h \to 0} \frac{F(x + h) - F(x - h)}{2h}$$ (8)

Because one estimate of the distribution function is an empirical distribution function, its definition is

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x)$$ (9)

In (9), $I$ is an indicator function.

Substituting (8) into (9), the estimation of the probability density function is obtained:

$$f_n(x) = \frac{F_n(x + h) - F_n(x - h)}{2h}$$ (10)

From (10), one can obtain

$$2nhf_n(x) = \sum_{i=1}^{n} I(x - h \leq X_i \leq x + h)$$ (11)

Equation (11) obeys the binomial distribution $B(n, F(x + h) - F(x - h))$.

In (11), when $h \to 0$, the mean value $E[f_n(x)]$ and the variance $D[2nhf_n(x)]$ are respectively given as

$$E[f_n(x)] = \frac{1}{2h} [F(x + h) - F(x - h)] \rightarrow f(x)$$ (12)

$$D[2nhf_n(x)] = n[1 - (F(x + h) - F(x - h))] \times [F(x + h) - F(x - h)] \rightarrow 0$$ (13)

When $h \to 0$ and $nh \to \infty$, then

$$D[2nhf_n(x)] = \frac{1}{4nh^2} [1 - (F(x + h) - F(x - h))] \times [F(x + h) - F(x - h)] \rightarrow 0$$ (14)

Based on (13) and (14), the estimator $f_n(x)$ is the consistent estimation of $f(x)$, and the variance tends to 0; thus, $f_n(x)$ is an estimation of $f(x)$.

Set $k(u) = \frac{1}{2} I(|u| \leq 1)$ and substitute $k(u)$ into (10) to obtain

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} I = \frac{1}{n} \sum_{i=1}^{n} k\left(\frac{X_i - x}{h}\right)(-1 \leq \frac{X_i - x}{h} \leq 1)$$ (15)
2) KERNEL FUNCTION SELECTION

Commonly used kernel functions include the uniform kernel function, the triangular kernel function, Epanechnikov kernel functions, and the Gaussian kernel function. The evaluation of these kernel functions can be achieved by comparing their PDFs of the same PSD data, which are shown in Figure 2. This figure shows that different kernel functions have little influence on kernel density estimation, which is consistent with the existing study that the selection of different kernel functions has little effect on the estimation results [20]. However, a slight difference exists. It can be seen from Figure 2 that, Gaussian kernel and Epanechnikov kernel functions are smoother for PDF, a triangle kernel function is the second, and uniform kernel function is the worst. For this reason, the most commonly used Gaussian kernel function is selected in this study.

![PDF Influence of Different Kernel Functions](image)

**FIGURE 2.** Influence of different kernel functions on PDF.

The Gaussian kernel function $k(u)$ can be expressed as

$$k(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2) \quad (16)$$

According to the definition of the probability density function, $\int_{-\infty}^{+\infty} k(u)du = 1$. When selecting the kernel function shown in (16), one has

$$\int_{-\infty}^{+\infty} f_n(x)dx = \frac{1}{nh} \sum_{i=1}^{n} \int_{-\infty}^{+\infty} k\left(\frac{X_i - x}{h}\right)dx$$

$$= \frac{1}{nh} \sum_{i=1}^{n} \left\{ \int_{-\infty}^{+\infty} k(u)du \right\} = 1 \quad (17)$$

From (17), the kernel estimation $f_n(x)$ of $f(x)$ is the probability density function.

As the proposed kernel density estimation method is suitable for continuous random variables, it can be used for all variables satisfying continuous distribution. Moreover, if a random variable satisfies the single peak unbiased distribution, its PDF calculated by kernel density estimation will be more accurate.

3) BANDWIDTH SELECTION

Bandwidth selection will have a greater effect on the results. If the bandwidth is too large, then the density curve is too smooth, and the results are not accurate enough; if the bandwidth is too small, then the shape of the curve is complicated. Therefore, selecting a reasonable bandwidth is very important.

If the Gaussian kernel function shown in (16) is selected, according to the rule of thumb [20], the optimal bandwidth can be obtained as

$$h_{opt} = 1.06 \min\{\hat{\sigma}, \frac{R}{1.34}\} n^{-\frac{1}{5}} \quad (18)$$

In (18), $\hat{\sigma}$ is the standard deviation of the sample, and $R$ is the interquartile range.

Substituting (17) and (18) into (15), the probability density function can be obtained as

$$f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^2} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} k\left(\frac{X_i - x}{h}\right)(-1 \leq \frac{X_i - x}{h} \leq 1)$$

$$= \frac{1}{2.657n} \sum_{i=1}^{n} n^{1/5} \min\{\hat{\sigma}, \frac{R}{1.34}\} \exp\left(-\frac{1}{2}\left(\frac{X_i - x}{h}\right)^2\right) \quad (19)$$

4) QUANTILE SELECTION

Set the probability density function of variable $X$ as $f(x)$. The distribution function is $F(x)$; if $x_\beta$ is satisfied, then

$$P(X \leq x_\beta) = F(x_\beta) = \int_{-\infty}^{x_\beta} f(x)dx = \beta \quad (20)$$

In this case, $x_\beta$ is the quantile that satisfies $P(X \leq x_\beta) = \beta$, and $x_\beta$ is the upper-limit estimate of the $S$ PSDs under set quantile $\beta$.

In summary, in using the kernel density estimation method, there is no need to assume that the data being inducted obeys a certain specific standard parameter distribution. Therefore, the kernel density estimation method is applicable to the induction of non-Gaussian and Gaussian data.

IV. ENGINEERING EXAMPLES

The PSD of the measured time-domain acceleration data of high-speed trains is calculated according to the Cooley-Tukey method, and the sign rank sum testing method is applied to merge all PSD values into $P$ same feature samples. Then, the PSD data of a certain channel is extracted from the same feature samples to carry out induction.

Let the PSD data extracted from the $i$th channel be recorded as $G_k(i,j)$ ($i$ is the channel number, $j = 1, 2, \ldots, 51$, and $k$ is the number of spectral lines). The $i$th channel has a total of 51 portions of the PSD spectra, and the frequencies of all PSD values are from 2 to 349.6Hz. The frequency resolution is 0.4Hz, and each portion of the PSD is dispersed into 8,740 frequency points.

A. GAUSSIAN VIBRATION DATA INDUCTION

1) GAUSSIAN MEASURED DATA

The measured vibration data from Channel No. 2 in this example are shown in Figure 3. A total of 51 PSDs are placed
at the same coordinates as shown in Figure 4. The fifty-one PSDs were extracted from 51 sets of measured vibration data of channel No. 2 by the Cooley-Tukey method. These PSDs will be used to induce the measured spectrum. Now, arbitrarily select data at the frequency point \( f = 2.8 \) Hz (see Table 1) as an example to induce the PSD upper limit \( x_{\beta} \) of the frequency point. The Anderson-Darling normality test is used to carry out a normality test on the data in Table 1 and the corresponding quantile-quantile plot is shown in Figure 5. The middle straight line of Figure 5 is an indicator of the normal distribution. If the data well obey the normal distribution, the calculated statistics plotted in Figure 5 will be closely distributed along the straight line. The upper and lower sidelines of Figure 5 represent the lower and upper limits of the 95% confidence levels for each percentile, respectively. According to the normality test, \( p = 0.595 \) (> 0.05) is obtained, which means that the data approximately obey a Gaussian distribution (see Figure 5).

2) INDUCTION BY THE STATISTICAL TOLERANCE METHOD

Because the data conform to a Gaussian distribution, the statistical tolerance method in Section II can be used. First, the mean value \( m = 0.00004288 \) and standard deviation \( \sigma = 0.000005244 \) of the data in Table 1 are calculated. Then, the upper tolerance limit coefficients of the acceptance test and the qualification test are calculated to be \( F_{90/50} = 1.3087 \) and \( F_{90/90} = 1.6600 \), respectively. Here, \( F_{90/50} \) is the tolerance upper limit coefficient with 0.9 quantile and 50% confidence level, and \( F_{90/90} \) is the tolerance upper limit coefficient with 0.9 quantile and 90% confidence level. As a result, the calculated upper tolerance limits are \( X_{90/50} = 5.03196684237743e-05 \) and \( X_{90/90} = 5.15838893824322e-05 \), respectively. Here, \( X_{90/90} \) is the upper limit of PSDs induced by the statistical tolerance method with 0.9 quantile and 90% confidence level, and \( X_{90/50} \) is the upper limit of PSDs induced by the statistical tolerance method with 0.9 quantile and 50% confidence level. The PSDs with \( F_{90/50} \) and \( F_{90/90} \) obtained by induction are shown in Figure 6.

3) INDUCTION BY THE KERNEL DENSITY ESTIMATION METHOD

According to the kernel density estimation method in Section III, first calculate the PDF of the data in Table 1 using (19), as shown in Figure 7. Then, according to (20) and under the condition of the selected
quantile $\beta = 0.9$, the $PSD$ upper limit value $X_{90} = 5.0319684237743e-05$ under $\beta = 0.9$ can be calculated from (16). The induced $PSD$ is shown in Figure 8, which is similar to that obtained by using the statistical tolerance method as shown in Figure 6.

**B. NON-GAUSSIAN VIBRATION DATA INDUCTION**

1) NON-GAUSSIAN MEASURED DATA

The measured vibration data from Channel No. 17 in this example are shown in Figure 9. A total of 51 $PSDs$ are placed at the same coordinates as shown in Figure 10. The fifty-one $PSDs$ were extracted from 51 sets of measured vibration data of channel No. 17 by the Cooley-Tukey method.

Now, arbitrarily select data at the frequency point $f = 2.8$Hz (see Table 2) as an example to induce the $PSD$ upper limit $X_\beta$ of the frequency point. The Anderson-Darling normality test is used to carry out a normality test on the data in Table 2 and the corresponding quantile-quantile plot is shown as Figure 11. As the calculated statistics plotted in Figure 11 do not distribute closely along the straight line, the data do not obey the normal distribution. Moreover, $p < 0.05$ is obtained according to the normality test, which indeed means that the data do not obey a Gaussian distribution (see Figure 11).
2) INDUCTION BY THE KERNEL DENSITY ESTIMATION METHOD

As the data do not conform to a Gaussian distribution, the statistical tolerance method in Section II cannot be used. According to the proposed kernel density estimation method, first, calculate the PDF of the data in Table 1 through (19) as shown in Figure 12. Then, according to (20) and under the condition of the selected quantile $\beta = 0.9$, the PSD upper limit value $X_{90} = 5.49122571451624e-04$ is calculated. Here, $X_{90}$ is the upper limit of PSD induced by the proposed kernel density estimation method. The induced PSD is shown in Figure 13.

| TABLE 2. Fifty-one PSD values when $f = 2.8$Hz. |
|-----------------------------------------------|
| $PSD$ Values                                 |
|------------------------------------------------|
| 0.00607271 | 0.00617236 | 0.00552427 | 0.00713118 |
| 0.0053854  | 0.00577637 | 0.00589008 | 0.00606482 |
| 0.0075423  | 0.00505927 | 0.00554827 | 0.00555551 |
| 0.00899663 | 0.0060237  | 0.00612467 | 0.00607755 |
| 0.00736606 | 0.00589859 | 0.00542922 | 0.00695092 |
| 0.00908079 | 0.0063368  | 0.00680075 | 0.00632229 |
| 0.00705665 | 0.00678042 | 0.0062797  | 0.00661976 |
| 0.00951175 | 0.00612819 | 0.00609492 | 0.00719168 |
| 0.00625654 | 0.00651644 | 0.00669668 | 0.00764524 |
| 0.00602248 | 0.0062307  | 0.00684711 | 0.00941626 |
| 0.00616346 | 0.00734889 | 0.00810037 | 0.00813828 |
| 0.00731429 | 0.00590717 | 0.00717807 | 0.00685731 |
| 0.00584607 | 0.00782444 | 0.00864349 | |

FIGURE 12. The probability density function of the data in Table 3.

C. DISCUSSION OF THE INDUCTION RESULTS

First, the accuracy of the induced PSD for Gaussian data by using the proposed kernel density estimation method is discussed. For Gaussian data, the PSD obtained by using the proposed kernel density estimation method shown in Figure 8 looks similar to that obtained by using the statistical tolerance method shown in Figure 6. The upper limit values of the PSDs shown in Figures 6 and 8 for the same Gaussian data are obtained first by using the statistical tolerance method and the kernel density estimation method respectively, and then the RMS values of the PSDs are calculated, as shown in Table 3. Based on Table 3, the relative errors of $X_{90}$ with $X_{90/50}$ and $X_{90/90}$ are 0.155% and 1.55%, respectively. The results show that the PSD obtained by using the proposed method for Gaussian vibration data is very close to that obtained by using the statistical tolerance method, demonstrating that the induction accuracy of the proposed method is high for Gaussian vibration data.

Second, the effectiveness of the proposed kernel density estimation method on the induction of non-Gaussian data is discussed. For the induction of non-Gaussian vibration data, the existing Johnson or Bootstrap method is complex and inefficient because many iterations are needed to find the optimal conversion function. However, the proposed kernel density estimation method can be used directly for the induction of non-Gaussian vibration data. According to the proposed kernel density estimation method, the induced PSD for the non-Gaussian vibration signal is shown in Figure 13. To demonstrate the reasonability of the proposed induction method, the induced PSD shown in Figure 13 with kurtosis value 7 is inputted to a closed-loop vibration test system that consists of a VR9500 non-Gaussian vibration controller made by the Vibration Research Corporation (USA). The random vibration data can then be obtained from the closed-loop vibration test system as shown in Figure 14. It can be seen from Figure 14 that a non-Gaussian random vibration that is related to the inputted PSD and the kurtosis value is replicated.

According to the above discussion, several results can be obtained as follows. The proposed kernel density estimation method can be used directly for the induction of both...
Gaussian and non-Gaussian random vibration data satisfying continuous distribution, avoiding the normality tests on the measured data, and decreasing the calculation cost. Especially for the induction of non-Gaussian random vibration data, it is not necessary to first convert non-Gaussian data into data that approximately obey a normal distribution. Thus, the complexity of data induction and the error of the induced PSD can decrease greatly because many iterations are not needed to find the optimal conversion function. As a result, the proposed method can be widely used for the induction of Gaussian and non-Gaussian random vibration data.

V. CONCLUSION
A novel method for carrying out induction on the power spectral density data by applying the kernel density estimation is proposed for the measured vibration of high-speed trains. First, the power spectral density data belonging to the same population are merged into the same feature sample, and the probability density function of all power spectral density data at the first frequency point is calculated through the kernel density estimation method. Then, the upper-limit estimates of all power spectral density values under the set quantile are obtained, and the upper limit of the power spectral density values at all frequency points can be calculated by repeating this process to convert the measured acceleration vibration data to the acceleration power spectral density spectrum of the vibration test.

Finally, two engineering examples are used to verify the proposed data induction method for Gaussian and non-Gaussian data. For the same Gaussian distribution data, the relative error between the root means square values of the power spectral density spectrum obtained from the induction using the kernel density estimation method, and the ordinary statistics tolerance is 0.155% ~ 1.55%. For the non-Gaussian distribution data, the kernel density estimation method is used to carry out the induction and an acceleration power spectral density spectrum is obtained for a non-Gaussian vibration test. The proposed method satisfies the induction requirements for the measured Gaussian and non-Gaussian vibration power spectral density data of high-speed trains, which have different distributions, and its induction results have very good universality and estimation accuracy.

REFERENCES
[1] Railway Applications—Rolling Stock Equipment—Shock and Vibration Tests, IEC Standard 61373, 2010.
[2] N. NEVEUX, IEC/TC 9/ADG 18-MT 61373 Proposal for the Revision of IEC 61373-Railway applications-Rolling Equipment-Equipment and Vibration Tests, MT61373 Kick Off Meeting, Standard IEC/TC 9 MT 61373 Working Group, Dec. 2019.
[3] N. Vincent, “Technical support for the revision of EN61373 standard,” VibraTec, Kuala Lumpur, Malaysia, Tech. Rep., Mar. 2018, pp. 1–6.
[4] Environmental Testing—Part 2-64: Tests—Test Fh: Vibration, Broadband Random and Guidance, IEC Standard 60068-2-64, 2019.
[5] B. Colin, “Characterization of random Gaussian and non-Gaussian stress processes in terms of vibration fatigue,” Mech. Ind., vol. 18, no. 8, pp. 1–12, Nov. 2017.
[6] R. Zheng, H. Chen, D. Vandenpiete, and Z. Luo, “Multi-exciters stationary Gaussian random vibration test with time domain randomization,” Mech. Syst. Signal Process., vol. 122, pp. 103–116, May 2019.
[7] R. H. Zheng, H. H. Chen, and X. D. He, “Probability distributions control for multi-input multi-output stationary non-Gaussian random vibration test,” J. Vib. Control, vol. 24, no. 5, pp. 5201–5210, 2017.
[8] Q. Zheng, F. Wang, B. Ai, and Z. Zhong, “Multicarrier downlink transmission for high-speed railway in non-Gaussian noise channels,” IEEE Access, vol. 6, pp. 52607–52615, Oct. 2018.
[9] Measurement Method of Environmental Parameters for Electric and Electronic Products, Vibration Data Processing and the Induction Method, GB Standard GB/T 10593.3-1990, 1990.
[10] The Inductive Methods for Environmental Measured Data of Vibration and Shock, GJB Standard GJBZ 126-1999, 1999.
[11] H. J. Yuan, M. Luo, and T. M. Jiang, “Study on inductive method for environmental measured data of random,” Acta. Aeronautica Et. Astronautica Sinica, vol. 28, pp. 115–117, Jan. 2007.
[12] Environmental Engineering Considerations and Laboratory Tests, Standard 810G, USA Dept. Defense, Arlington, VA, USA, 2008.
[13] Dynamic Environmental Criteria, Standard NASA-HDBK-7005, 2001.
[14] A. Coulon, B. Nèlain, and N. Vincent, “Interest of equivalent damage methods for railway equipment qualification to vibrations,” Procedia Eng., vol. 133, pp. 714–725, Dec. 2015.
[15] J. F. Shapiro and S. S. Shapiro, “The Johnson system: Selection and parameter estimation,” Technometrics, vol. 22, no. 2, pp. 239–246, May 1980.
[16] B. Efрон, “Bootstrap methods: Another look at the Jackknife,” Ann. Stat., vol. 7, no. 1, pp. 1–26, Jan. 1979.
[17] L. Wasserman, All of Nonparametric Statistics. New York, NY, USA: Springer, 2006.
[18] M. Luo and T. M. Jiang, “Induction for random vibration environmental measured data using Johnson system distributions,” J. Syst. Eng. Electron., vol. 12, pp. 2176–2179, 2017.
[19] E. Parzen, “On estimation of a probability density function and mode,” Ann. Math. Statist., vol. 33, no. 3, pp. 1065–1076, Sep. 1962.
[20] M. Carbon, L. T. Tran, and B. Wu, “Kernel density estimation for random fields,” Stat. Probabil. Lett., vol. 36, pp. 115–125, Dec. 1997.
[21] B. Qin and F. Xiao, “A non-parametric method to determine basic probability assignment based on kernel density estimation,” IEEE Access, vol. 6, pp. 73509–73519, Dec. 2018.
[22] Q. Xiao and S. Zhou, “Simulating non-Gaussian stationary stochastic processes by translation model,” IEEE Access, vol. 7, pp. 34555–34569, Apr. 2019.
[23] E. Richard, “Application of kernel smoothing to estimate the spatiotemporal variation in risk of STEC O157 in England,” Spatial Statiotemporal Epidemiol., vol. 32, pp. 1–10, Feb. 2020.
[24] P. T. Roy, L. Jofre, J.-C. Jouhaud, and B. Cuenot, “Versatile sequential sampling algorithm using kernel density estimation,” Eur. J. Oper. Res., vol. 284, no. 1, pp. 201–211, Jul. 2020.
[25] J. D. Parente, “Kernel density estimates for sepsis classification,” Stat. Probabil. Lett., vol. 161, pp. 1–9, Jun. 2020.
[26] H. D. Dau, T. Lalo, and R. Servien, “Exact asymptotic limit for kernel estimation of regression level sets,” Stat. Probabil. Lett., vol. 161, pp. 1–9, Jun. 2020.
[27] M. Ahsan, “Robust adaptive multivariate Hotelling’s T2 control chart based on kernel density estimation for intrusion detection system,” Expert Syst. Appl., vol. 145, pp. 1–31, May 2020.
[28] C. Li and J. Guan, “Code acquisition method of GPS signals based on WFTA, PFA and cooley-Tukey FFT,” in Proc. IEEE Int. Conf. Intell. Control, Auton. Detection High-End Equip., Beijing, China, Jul. 2012, pp. 131–134.

[29] M. Lovric, “R. H. C. Lopes. Kolmogorov-Smirnov Test,” in International Encyclopedia of Statistical Science. Berlin, Germany: Springer, 2011.

[30] Y. Q. Li, Nonparametric Statistical Analysis. Chengdu, China: Southwest Jiaotong Univ. Press, (in Chinese), 2010.

PENG WANG received the B.E. and M.E. degrees in mechanical engineering. He is currently pursuing the Ph.D. degree in mechanical and electronic engineering from Central South University, China. He is also a Senior Engineer with CRRC Zhuzhou Electric Locomotive Research Institute Company Ltd., and an Expert with IEC/TC9. His research interests include environmental and reliability test technology of rail transit products.

HUA DENG (Member, IEEE) received the B.E. degree in electrical engineering from the Nanjing Aeronautical Institute, Nanjing, China, in 1983, the M.E. degree in electrical engineering from Northwestern Polytechnical University, Xi’an, China, in 1988, and the Ph.D. degree in manufacturing engineering from the City University of Hong Kong, Kowloon, Hong Kong, in 2005. He is currently a Professor with the School of Mechanical and Electrical Engineering, Central South University, Changsha, China. His research interests include intelligent control and learning, robotic dynamics and control, biomechatronics, and modeling and control of complex distributed parameter systems.

YI MIN WANG was born in Changzhou, Jiangsu, China, in 1963. He received the B.E. degree in electric locomotive from Southwest Jiaotong University, in 1986. He is currently a Professor with CRRC Zhuzhou Electric Locomotive Research Institute Company Ltd. His research interests include design of rail transit power supply products, electromagnetic compatibility technology, and environment and reliability test technology.

YUE LIU was born in Hanzhong, Shaanxi, China, in 1996. He received the B.E. degree in mechanical engineering from Chongqing University, China. He has been an Assistant Engineer with CRRC Zhuzhou Electric Locomotive Research Institute Company Ltd. His main research interest includes vibration and impact test technology of rail transit products.

YI ZHANG received the B.E. and Ph.D. degrees from the School of Mechanical and Electrical Engineering, Central South University, China. He is currently a Postdoctoral Fellow with the School of Mechanical and Electrical Engineering, Central South University. He is also a Lecturer with the College of Mechanical and Electrical Engineering, Changsha University, China. His current research interests include mechanism design and system modeling and control.

***