Compilation Method of CNC Lathe Cutting Force Spectrum Based on Kernel Density Estimation of G-SCE

Shengxu Wang
School of Mechanical and Aerospace Engineering, Jilin University, Changchun, Jilin, 130022, China
Jialong He (hejl@jlu.edu.cn)
Jilin University
Guofa Li
School of Mechanical and Aerospace Engineering, Jilin University, Changchun, Jilin, 130022, China
Qingbo Hao
School of Mechanical and Aerospace Engineering, Jilin University, Changchun, Jilin, 130022, China
Hao Huang
School of Mechanical and Aerospace Engineering, Jilin University, Changchun, Jilin, 130022, China

Research Article

Keywords: cutting force spectrum, KDE, bandwidth, comprehensive evaluation

Posted Date: May 5th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-481931/v1

Version of Record: A version of this preprint was published at The International Journal of Advanced Manufacturing Technology on July 6th, 2021. See the published version at https://doi.org/10.1007/s00170-021-07541-1.

Compilation method of CNC lathe cutting force spectrum based on kernel density estimation of G-SCE

Shengxu Wang[1], Jialong He[1], Guofa Li[1], Qingbo Hao[2], Hao Huang[1]

Abstract: The cutting force spectrum of the CNC lathe is the basic data for the reliability design, reliability test and reliability evaluation of the CNC lathe and its components. Due to the complex and changeable turning conditions and different cutting processes, the cutting load presents multi-peak characteristic. In order to compile the CNC lathe cutting force spectrum accurately, a compilation method based on kernel density estimation (KDE) of goodness-smoothness comprehensive evaluation (G-SCE) is proposed. According to the measured dynamic cutting force, the KDE is used to establish the dynamic cutting force distribution of the CNC lathe. For the rule of thumb method (ROT) based on the mean integrated square error and the least squares cross-validation method (LCV) based on the integrated square error does not take into account the influence of different bandwidths on the goodness estimation and the smoothness of the estimated curve. The estimation accuracy test method based on multiple goodness-of-fit tests, and the smoothness test method based on the envelope curve are proposed. The entropy method is used to comprehensively calculate the estimated goodness index and the smoothness index to determine the optimal bandwidth. The case analysis shows that the method proposed can solve the problem of too large estimation error of parameter distribution for multimodal distribution. At the same time, it can better comprehensively evaluate the KDE under different bandwidths. In short, a new method of optimal bandwidth selection is proposed in the original method.

Keywords: cutting force spectrum; KDE, bandwidth; comprehensive evaluation

1 Introduction

The load spectrum is the image, table, matrix or other probability characteristic values obtained by counting the time history of the load of the product or its components. It shows the relationship between the load and the occurrence frequency or applied time [1]. Load spectrum can reflect the distribution law of random load, so it is widely used in reliability test of component structure [9,30], fatigue life prediction [5,8,10,31], reliability design [2,27], accelerated test [3] and other aspects. Compiling a scientific and reasonable load spectrum of CNC machine tools can provide basic data for the reliability design, reliability tests and reliability evaluation of key components [23,24].

The main steps of compiling the load spectrum include obtaining the load, counting the load and extrapolating the load [1,6,7], among which the extrapolation for load is the most important step. For the obtained cutting force data, the rain flow counting method is used for counting, and the counting result is the frequency of the mean and amplitude [11]. In the conventional process of building spectrum, the counting results are classified and combined to groups according to the set number, and the frequency statistical results of the mean and the amplitude of the cutting force are obtained. Finally, the statistical results are fitted to obtain the distribution function of the mean and amplitude. In this process, parametric estimation and non-parametric estimation can be used [1,2,4,]. Due to the diversity and complexity of load conditions, the load

✉ Jialong He
hejl@jlu.edu.cn
✉ Guofa Li
ligf@jlu.edu.cn

1. School of Mechanical and Aerospace Engineering, Jilin University, Changchun, Jilin, 130022, China.
2. Office of Academic Affairs, Aviation University of Air Force, Changchun 130000, China

exhibits multimodal characteristics, and the parametric extrapolation is also derived from the single distribution [1] to the mixed distribution [2,4]. However, there is often a large gap between the basic function assumption of the parametric model and the actual physical model [7,14]. Meanwhile, it is inaccurate to use the interval mean value to represent the load value in the interval [8] because the distribution of the load value in the interval is not necessarily uniform, which will produce larger errors. More importantly, the density distribution is greatly affected by the width of the sub-interval (that is, each histogram). If the same original data take different sub-interval ranges, the displayed results may be completely different [26]. Therefore, in this case, using parametric regression to model the load distribution is not accurate, and kernel density estimation (KDE) can solve this problem well.

Rosenblatt and Parzen [16] proposed the KDE method, one of the non-parametric estimation methods. Since the KDE does not use the prior knowledge of the data distribution [13,19] and does not attach any assumptions, it is a method to study the characteristics of the data distribution from the data sample itself [20]. So, the KDE is more comprehensively and accurately to reflect the distribution characteristics of the sample compared with the parametric estimation method. The establishment of density distribution function is to set a kernel function at each observation value, and take the bandwidth as the boundary. The kernel function within the boundary is the distribution probability of the observation value, and then the accumulation of the kernel function of the observations is the density distribution of the observations [28]. Compared with the parametric distribution method, the KDE does not need to group the original data, and it solves the influence of grouping on parametric estimation from the source [29]. Qin [14] used the KDE to estimate the probability distribution of wind speed in different regions, and believed that the kernel density method was superior to other parametric estimation methods. Siddharth [15] used the KDE to predict the multi-peak data of the smart meter, and it could better meet the prediction requirements.

Compared with the kernel function, the bandwidth has more obvious influence on the accuracy of KDE [20,29]. Duong [17] proposed a method to determine the bandwidth of KDE by using the rule of thumb (ROT). The principle of ROT is to solve the optimal bandwidth with the goal of minimum mean integrated square error (MISE). At the same time, he also proposed the least squares cross-validation (LCV) method to determine the bandwidth, which is based on the minimum integrated square error (ISE) as the goal to solve the optimal bandwidth [18]. On this basis, Bayesian methods [13, 24], adaptive methods [21], diffusion methods [29] have been developed. Dias [21] modeled the river runoff with random data based on the KDE. Compared with the conventional method, it verified the effectiveness of the KDE. Bo HU [26] used the bandwidth determined by the ROT based on MISE to estimate the density of the wind speed in a specific area by KDE and established the probability density distribution of the wind speed. He also evaluated the sensitivity of each kernel function and confirmed that the Gaussian kernel function was optimal kernel function. Based on the above analysis, the ROT based on MISE and the LCV based on ISE are used as bandwidth determination methods, and the Gaussian kernel function is used as the basis function for KDE in this paper.

For different forms of data distribution, ROT and LCV have corresponding application conditions. It is necessary to determine which method is suitable to the corresponding data. Literature [14] proposed using Kolmogorov-Smirnov test (K-S test) and Chi-square test to determine the optimal KDE, but two or more tests may have opposite test results due to different test values. It is necessary to make a comprehensive evaluation of multiple test quantities. An evaluation method for KDE based on Goodness-Smoothness Comprehensive Evaluation (G-SCE) is proposed. It uses the K-S test, Chi-square test, Anderson-darling test (A-D test) and coefficient of determination to test the goodness of KDE. At the same time, considering that the smoothness of the KDE is also an important factor, the smoothness test of KDE based on the envelope curve is established. The entropy method is used to evaluate the weight of each goodness index and smoothness index and calculate the comprehensive evaluation value.

Based on the KDE of G-SCE, the method of establishing the dynamic cutting force spectrum of CNC machine tool is studied. KDE for rainflow counting results avoids the inaccuracy of probability distribution caused by grouping, and does not need to assume the probability distribution function (PDF). In section 2, the KDE is introduced and how to determine the optimal kernel bandwidth is described. In section 3, a comprehensive evaluation method for the goodness and smoothness of KDE is proposed. Section 4 introduces the acquisition of dynamic cutting force and results of KDE, and the results of the two bandwidth determination methods are compared by G-SCE values.

2 Theory and Method

2.1 Kernel Density Estimation

KDE uses observed random variables from an unknown distribution to estimate its density function and it does not require prior knowledge of the data distribution. It is a predictive estimation using only the sample data itself [25]. Assuming that the density of the random variable X is f, its value can be calculated by the following formula.

$$f(x)=\lim_{h\to 0}\frac{1}{2h}P(x-h<X<x+h) \quad (1)$$

where, h is the length of interval $[x\text{-}h, x+h]$, and $P(x\text{-}h<X<x+h)$ is the probability that the sample X in the interval. Therefore, the estimated density of $f(x)$ is:

$$f(x)=\frac{1}{2hN}n(x-h<X<x+h) \quad (2)$$

where, $n(x\text{-}h<X<x+h)$ is the number of $X_1, \ldots, X_j$ which is in the interval $[x\text{-}h, x+h]$.

Let $u=\frac{x-X_i}{h}$, when $|u|\le 1$, X_i is in the interval $[x\text{-}h, x+h]$, when $|u|>1$, X_i is not in the interval, so the weight function ω can be defined:

$$\omega(u)=\begin{cases}\frac{1}{2}, |u|\le 1\\ 0, |u|>1\end{cases} \quad (3)$$

Therefore, the estimated density of $f(x)$ at x can be expressed as:

$$f(x)=\frac{1}{Nh}\sum_{i=1}^{N}\omega(\frac{x-X_i}{h}) \quad (4)$$

Since the definition of the above formula is discontinuous at the boundary of each estimation interval, this will cause the loss of some probability distribution information. In order to achieve a continuous estimation of the probability density, Silverman [25] proposed kernel function $K(x)$ to replace the weight function, then the kernel density of $f(x)$ at x is estimated as:

$$f(x)=\frac{1}{Nh}\sum_{i=1}^{N}K(\frac{x-X_i}{h}) \quad (5)$$

where, h is the bandwidth, N is the number of samples.

In KDE, the kernel function $K(x)$ is a weight function, and its shape controls the utilization degree of x when estimating the $f(x)$. There are many types of kernel functions, and commonly used are Uniform, Triangle, Epanechnikov, Gaussian, Quartic and Triweight [25]. Take the Gaussian kernel function as an example to introduce the principle of KDE. For samples with unknown distribution, there is a kernel density function at each sample, and the probability density value at data x can be determined by formula (5). Specific examples are shown below. There is a sample set [0.9, 0.6, 2.9, 2.7, 2.8, 2]. As shown by the red dotted line in Fig.1, at each sample point, there is a Gaussian kernel function, result of KDE is the blue line in Fig.1.

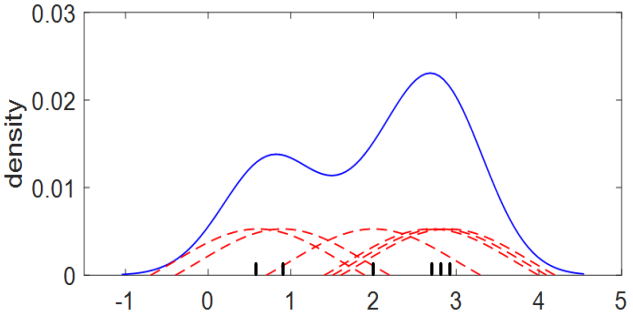

Fig.1 Gaussian function to estimate kernel density

At present, there is no recognized optimal kernel function type. When the bandwidth calculation method is determined, the kernel function type has little effect on KDE and the effect can be ignored compared with that of bandwidth [20].

2.2 Determination of Bandwidth

Bandwidth is a key parameter that affects the accuracy of KDE. It reflects the overall smoothness of the kernel density curve and the proportion of observed

data points in the formation of the kernel density curve. Different bandwidths have different estimation results as shown in Fig.2. When the bandwidth is unreasonable, it will bring a poor effect to the KDE. The final density curve is too smooth or too steep. Therefore, it is very necessary to use a suitable method to select the bandwidth.

In order to evaluate the approximation effect of the KDE, evaluation features should be introduced to evaluate the degree of difference between the estimated PDF and the true PDF. Commonly used evaluation features are MISE and ISE.

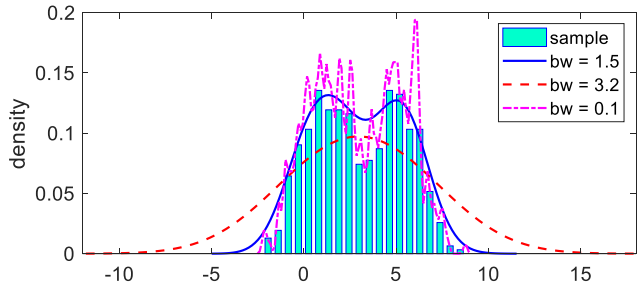

Fig.2 KDE results under different bandwidths

(1) The expression of MISE is as follows:

$$MISE=E\left[\int\left\{\overline{f(x)}-f(x)\right\}^2 dx\right] \tag{6}$$

The (ROT) [17] is introduced to derive the bandwidth.

$$h_{ROT}=\left[\frac{\frac{1}{2\sqrt{\pi}}}{\sigma^{-5}\frac{3}{8\sqrt{\pi}}}\right]^{\frac{1}{5}} n^{-\frac{1}{5}}\approx 1.06\hat{\sigma}_x n^{-\frac{1}{5}} \tag{7}$$

where, $\hat{\sigma}_x$ is the standard deviation of the sample, $\overline{X}$ is the mean of sample, and n is the number of samples.

However, the method may lead to over-smooth for multimodal density functions. Therefore, Silverman [25] proposed an improved ROT method to improve the estimation accuracy of complex density functions.

$$h_{ROT}=1.06\times\min\left\{\hat{\sigma}_x,\frac{\hat{R}}{1.34}\right\}n^{-\frac{1}{5}} \tag{8}$$

where, $\hat{R}$ is the sample interquartile range.

(2) The calculation of ISE is shown in the following formula:

$$ISE=\int\left[\hat{f}_n(x)-f(x)\right]^2 dx=\int[\hat{f}_n(x)]^2 dx-2\int\hat{f}_n(x)f(x)dx+\int[f(x)]^2 dx \tag{9}$$

The LCV is introduced to calculate the bandwidth.

$$h_{LCV}=\arg\min_h\left[\int[\hat{f}_n(x)]^2 dx-2\int\hat{f}_n(x)f(x)dx\right] \tag{10}$$

There is currently no specific selection basis to determine which method can be applied to a certain condition, especially in the field of dynamic cutting force spectrum. The bandwidths obtained by different methods have different values, which have a greater impact on the effect of KDE. Therefore, it is necessary to use appropriate methods to evaluate the effects of KDE obtained by different methods.

3 Goodness-Smoothness Comprehensive Evaluation (G-SCE)

Different bandwidths will affect the accuracy of KDE and the smoothness of the KDE curve. Literature [7,8] shows that the dynamic cutting force distribution is uneven, showing a multi-peak distribution, which leads to different estimation effects at different positions with different bandwidths. A goodness-smoothness comprehensive evaluation method based on entropy method is proposed to evaluate the estimation results for cutting force data. As shown in Fig.3, after bandwidths are calculated by ROT and LCV, smoothness inspection method based on envelope curve and goodness test based on four indexes are combined to evaluate the estimation results of different bandwidth comprehensively by entropy method. Finally, the optimal bandwidth suitable for dynamic cutting force is determined when the comprehensive index F is the smallest.

3.1 Goodness test for KDE

The KDE is obtained when the overall distribution is unknown，the K-S test, A-D test, Chi-square test and coefficient of determination can solve this problem well [27].

The K-S test defines the statistic D based on the largest vertical difference between the observation cumulative PDF $F(xi)$ and the empirical cumulative

PDF *G(xi)*.

$$D=\max_{1\le i\le N}\left|(F(x_i)-G(x_i))\right| \quad (11)$$

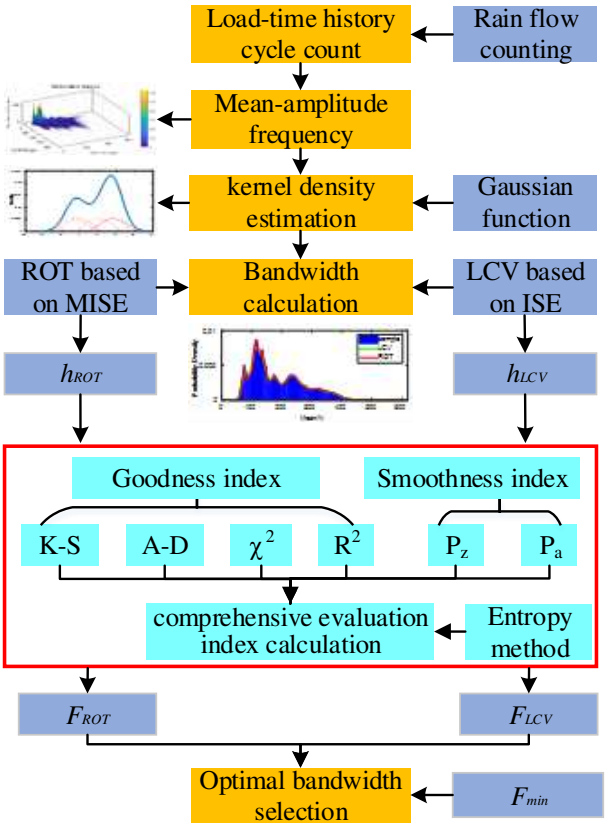

Fig.3 flow chart of G-SCE

The A-D test measures the integral of the squared distance between the empirical cumulative PDF and the observation cumulative PDF. The A-D test quantity statistic A^2 is

$$A^2=n\int_{-\infty}^{+\infty}\frac{\{G(x_i)-F(x_i)\}^2}{F(x_i)(1-F(x_i))}dF(x) \quad (12)$$

The Chi-square test is the degree of deviation between the actual observation value of the statistical sample and the theoretical value. The chi-square test is

$$\chi^2=\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i} \quad (13)$$

where O_i is the observed frequency for group i and E_i is the expected frequency for group i.

The coefficient of determination R^2 is the ratio of the regression sum of squares to the total deviation of squares.

$$R^2=1-\frac{S_{re}}{S_t} \quad (14)$$

where, S_{re} is the residual sum of squares, $S_{re}=\sum_{i=1}^{n}(y_i-f_i)^2$, S_t is the total sum of squares, $S_t=\sum_{i=1}^{n}(y_i-\bar{y})^2$. y_i is the sample observation value, f_i is the model prediction value, and $\bar{y}$ is the average observation value.

3.2 Smoothness test for KDE

The envelopes of the peak and valley points of the estimated curve are drawn respectively, and the difference between the coordinate values of the upper and lower envelopes is defined as the local smoothness of the curve. The smaller the value, the smoother the curve. The calculation formula of the local smoothness index P_z is

$$P_z=\max\left|Pu_i-Pl_i\right| \quad (15)$$

where, Pu_i is the coordinate value of upper envelope, and Pl_i is the coordinate value of lower envelope.

Define the area of the part included in the upper and lower envelopes as the envelope area, P_a, whose expression is:

$$P_a=\int_{x_{\min}}^{x_{\max}}f_u(x)dx-\int_{x_{\min}}^{x_{\max}}f_l(x)dx \quad (16)$$

where, $f_u(x)$ is the expression of the upper envelope, $f_l(x)$ is the expression of the lower envelope.

3.3 Comprehensive evaluation method based on entropy method

The entropy method believes that different criteria vary with the goal when determining an optimal goal. The value of a criterion changes greatly when the target value changes, indicating that this criterion has a greater impact in the final decision-making and occupies a larger weight value. When the target value changes, the value of a criterion is basically no change, it means that this criterion has little influence in the final decision-making, and its influence can be less considered. Therefore, the entropy method can be used to evaluate the weight of each criterion in the final decision.

The value of the test corresponding to different

bandwidths is also different. Suppose the test value of *j-th* test (total *n*) corresponding to *i-th* bandwidth (total *m*) is x_{ij}, and the evaluation matrix is *X*.

$$X=\begin{pmatrix} a_{11} & \mathrm{K} & a_{1n} \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ a_{m1} & \mathrm{L} & a_{mn} \end{pmatrix} \quad (17)$$

The normalization method is used to obtain the standardized fitting test value, and the entropy value E_j of each test can be calculated by formula (18):

$$E_j=-k\sum_{i=1}^{m}\hat{x}_{ij}\ln\hat{x}_{ij} \quad (18)$$

where, $k=\dfrac{1}{\ln m}$.

The entropy weight of each test can be calculated by formula (19).

$$w_j=\frac{1-E_j}{\sum_{j=1}^{n}(1-E_j)} \quad (19)$$

After obtaining the weight vector $w=(w_1, w_2,...,w_n)$, the value of comprehensive test can be calculated by formula (20).

$$F=\sum\theta\hat{x}_{ij}\times w \quad (20)$$

where, θ is the coefficient of the evaluation index, for the benefit index (the larger the index value, the better the effect), θ = -1, for the cost index (the smaller the index value, the better the effect), θ =1. The smaller the final value *F* of comprehensive test index, the better the effect of KDE.

4 Case study

4.1 Cutting force data achievement

Stainless steel (1Cr18Ni9Ti) was chosen as the cutting material and the coated carbide insert (CNMG120408) was selected for the test. The cutting parameters are shown in Table.1. 36 sets of process parameters were carried out.

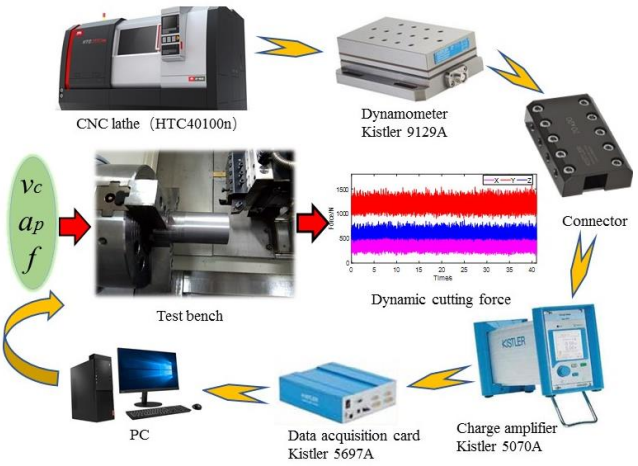

Fig.4 Test bench for cutting force measurement

The dynamic turning force measurement system includes Kistler multi-component dynamometer (9129A), charge amplifier (5070A), data acquisition card (5697A) and computer, as shown in Fig.4. The dynamometer is fixed on the tool holder of the CNC lathe (HTC 40100n) and the tool bar (MCLNL2525M12) was fixed on the dynamometer through the connector.

Table.1 Process parameters for cutting force measurement test

No.	v_c/(m/min)	a_p/mm	f/(mm/r)	No.	v_c/(m/min)	a_p/mm	f/(mm/r)
1	100	1.2	0.35	19	70	0.6	0.2
2	100	1.2	0.3	20	70	0.6	0.1
3	100	1.2	0.2	21	40	1.2	0.35
4	100	1.2	0.1	22	40	1.2	0.3
5	70	0.2	0.35	23	40	1.2	0.2
6	70	0.2	0.2	24	40	1.2	0.1
7	70	0.2	0.3	25	100	0.2	0.35
8	70	0.2	0.1	26	100	0.2	0.3
9	40	0.6	0.35	27	100	0.2	0.2
10	40	0.6	0.3	28	100	0.2	0.1
11	40	0.6	0.2	29	70	1.2	0.35
12	40	0.6	0.1	30	70	1.2	0.3
13	100	0.6	0.35	31	70	1.2	0.2
14	100	0.6	0.3	32	70	1.2	0.1
15	100	0.6	0.2	33	40	0.2	0.35
16	100	0.6	0.1	34	40	0.2	0.2
17	70	0.6	0.35	35	40	0.2	0.3
18	70	0.6	0.3	36	40	0.2	0.1

4.2 Data processing

The rainflow counting method is widely used in engineering, especially the fatigue life analysis of

engineering parts [11]. It expresses the measured load history in the form of discrete load cycles. 36 sets of measured dynamic cutting force data were recombined according to different components (Fig.5(a)) and counted by rainflow counting method. The counting result is the "From-to" cycle of the cutting force components in the three directions and the mean and amplitude are used to describe the cycle. The two-dimensional rainflow matrix of X direction is shown in Fig.5(b).

As shown in Fig.5, the rainflow counting result is the frequency of amplitude value and mean value. In the previous studies, the statistical results are divided into predetermined load intervals. In the subsequent statistical analysis, there are two problems as mentioned in Introduction, one is that the mean value of the interval representing the load leads the parameter regression analysis inaccurate because the distribution of the load in the interval is not necessarily uniform. The other is that the density function is greatly affected by the width of the sub-interval (that is, each histogram). If the same original data takes different sub-interval ranges, the displayed results may be completely different, and the statistical results will vary with the changing interval.

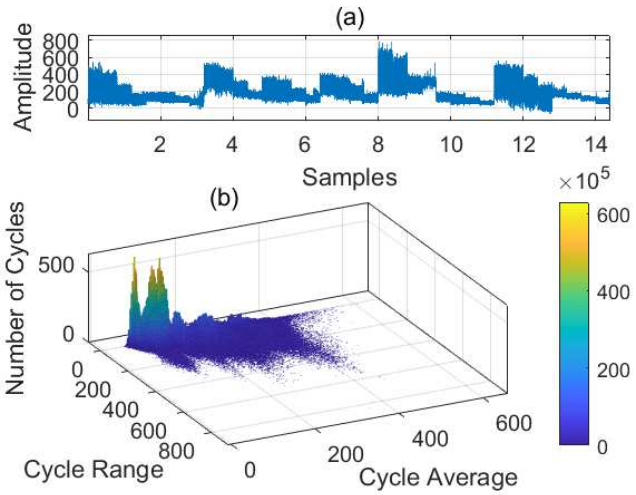

Fig.5 Dynamic cutting force in X direction(a) and its rainflow counting result(b).

4.3 Estimation of the probability density

When a single parametric distribution or a mixed parametric distribution is used for fitting, the two-dimensional rainflow counting matrix needs to be grouped, for example, into a 32*32 matrix or a 64*64 matrix. However, the KDE is different from parametric distribution estimation that the results of rainflow counting donot need to be grouped, which reduces the impact caused by grouping. Instead, it uses the mean and amplitude counting results after rainflow counting directly. Based on evaluation results of the G-SCE, the ROT and LCV methods for calculating the bandwidth of the kernel density function are judged to determine the optimal KDE of the cutting force distribution.

Based on the above method, the KDE is performed on the rainflow counting results of the dynamic cutting force in different directions, and the results obtained are shown in Fig.6-Fig.9. The results of the goodness test and smoothness test of KDE for dynamic cutting force in X direction are shown in Table.2.

Rainflow counting is performed on the dynamic cutting force in X direction, and 427,589 mean-amplitude cycles are obtained. After dividing the cycles into specific intervals, the two-dimensional matrix of the average-amplitude frequency is shown in Fig.5(b). It can be found that the distribution is not uniform and there are multiple peaks at different locations. The difference between the peak and the valley is quite large. It is difficult to use the parametric estimation to fit them.

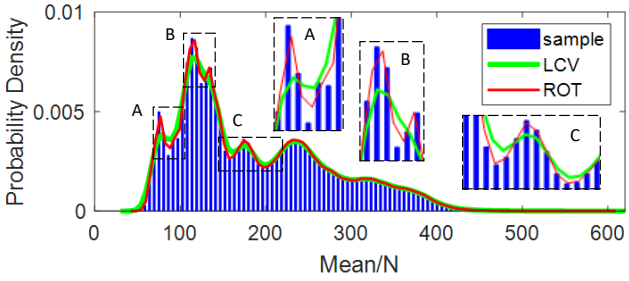

Fig.6 Probability density Estimation of the mean of dynamic cutting force in X direction

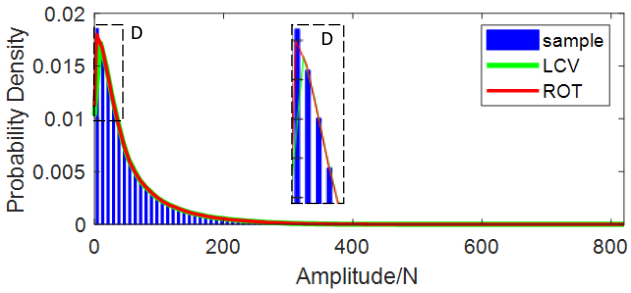

Fig.7 Probability density estimation of amplitude of dynamic cutting force in X direction

Table.2 KDE and test results of dynamic cutting force in X direction

X direction		Bandwidth	Goodness index				Smoothness index		G-SCE
			χ^2	D	R^2	A^2	P_z	P_a	F
Mean	LCV	6.720	10.2045	0.1215	0.7484	10.6872	0.0871	0.0561	0.1889
	ROT	2.624	1.1456	0.0587	0.8932	1.2434	0.0279	0.1254	0.0590
Amplitude	LCV	2.706	51.6785	0.1633	0.6869	52.0638	0.0101	0.1294	0.3781
	ROT	1.860	44.8621	0.1263	0.7567	45.0816	0.0464	0.1017	0.3697
Weight			0.3554	0.0503	0.0042	0.3482	0.2025	0.0394	

Table.3 KDE and test results of dynamic cutting force in Y direction

Y-direction		Bandwidth	Goodness index				Smoothness index		G-SCE
			χ^2	D	R^2	A^2	P_z	P_a	F
Mean	LCV	23.022	18.7076	0.1541	0.4293	19.5979	0.0349	0.1487	0.1441
	ROT	10.682	3.3240	0.0870	0.7041	3.5632	0.1774	0.4752	0.2116
Amplitude	LCV	3.834	40.7891	0.1193	0.8434	41.1571	0.0325	0.1576	0.2388
	ROT	2.173	38.2639	0.0799	0.9007	38.4312	0.2607	0.1621	0.3702
Weight			0.2321	0.0356	0.0354	0.2235	0.3225	0.1509	

Table.4 KDE and test results of dynamic cutting force in Z direction

Z-direction		Bandwidth	Goodness index				Smoothness index		G-SCE
			χ^2	D	R^2	A^2	P_z	P_a	F
Mean	LCV	12.457	20.4923	0.2225	0.1738	21.6829	0.0115	0.1234	0.1064
	ROT	6.563	4.4986	0.1493	0.4850	4.8927	0.0213	0.1722	0.0600
Amplitude	LCV	3.598	31.6233	0.1055	0.8605	32.0481	0.0946	0.1952	0.2276
	ROT	2.861	29.0629	0.0879	0.8845	29.3621	0.3204	0.1295	0.4810
Weight			0.1405	0.0589	0.1251	0.1328	0.5259	0.0167	

The method proposed in this paper is used to establish the dynamic cutting force distribution of CNC lathe. The LCV method and ROT are used to calculate the bandwidth, and the results of the KDE determined by the two bandwidths are tested by the goodness test and the smoothness test. The results in Table.2 show that the estimation accuracy of the kernel density estimation determined by the ROT is higher than that of the LCV method, regardless of the mean or the amplitude. The smoothness of the KDE curve determined by the LCV method is indistinguishable from the ROT. Finally, the weights of tests are calculated and the result of comprehensive estimation is that ROT is more suitable for the distribution estimation of the mean and amplitude.

The curves of KDE are shown in Fig.6 and Fig.7. Fig.6 is the estimation result of the mean distribution and Fig.7 is the estimation result of the amplitude distribution. The green line is the KDE curve obtained using the LCV method, the red line is the KDE curve obtained using the ROT, and the histogram is the mean-frequency result of the cycles obtained from rainflow counting which is divided into 100 intervals. It can be seen from A, B, and C in Fig.6 that when the mean value is small, the peak value of the probability distribution is large, and the difference between the peak value and the valley value is large, which has a bad influence on the estimation result, whether it is LCV method or the ROT. The estimation results in the first half are not completely consistent with reality, which also verifies that the same bandwidth will have different effects on the KDE for different data. But comparing the two methods, the ROT can better reflect the real situation. Similarly, for the amplitude, its distribution is a unimodal light-tailed

distribution. The result of KDE is shown in Fig.7 and the ROT is better than the LCV method.

Similarly, data processing and rain flow counts are performed on the acquired dynamic cutting forces in Y direction and Z direction. The rainflow counting result of dynamic cutting force in Y direction has 469,330 cycles, and 471551 cycles in Z direction. It is similar with the performance in X direction, the distribution of the mean of the cutting force in Y and Z directions is not uniform. There are multiple frequency peaks at different locations and the difference between the peak and the valley is quite large. It is difficult to use the parametric estimation to fit the mean.

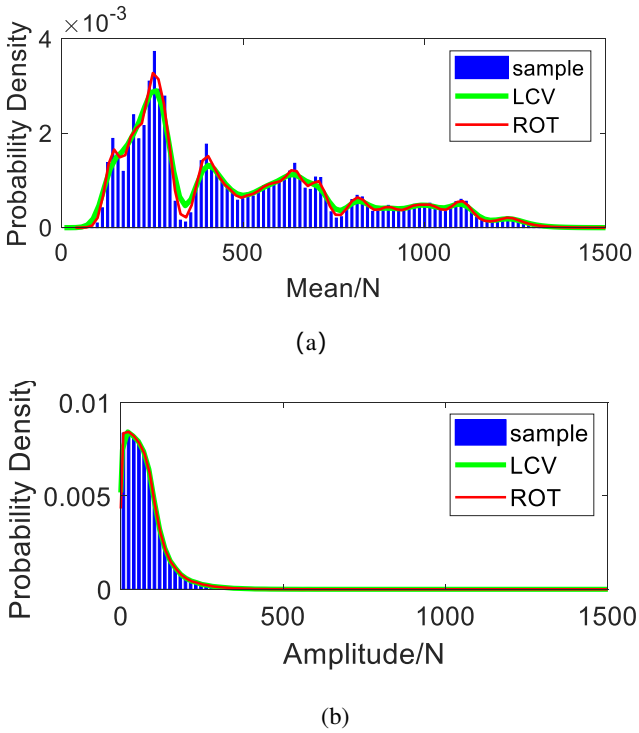

Fig.8 KDE of mean(a) and amplitude(b) of dynamic Cutting Force in Y Direction.

The bandwidths of KDE for the mean and amplitude are calculated by LCV method and ROT, and results are shown in Table.3 and Table.4. The curves of KDE are shown in Fig.8 and Fig.9. It can be seen from Fig.8(a) and Fig.9(a) that the distributions of mean are multimodal, and when the value is small, the peak and the difference between the peak and the valley are larger, and the phenomenon has a bad influence on the estimation. The mean of the cutting force in Z direction is more concentrated in the place where the value is small, and its change trend is more obvious. While the mean of the cutting force in Y direction is more concentrated in the middle of the distribution, and for the amplitude, its distribution is a unimodal light-tailed distribution. Compared with the X direction, the amplitudes of cutting force in Y and Z directions are more balanced at the front end of the image, and the trend changes more slowly. After using the G-SCE method and calculating the value of F, the results of goodness test and smoothness test for KDE of the dynamic cutting force in Y direction and Z direction are shown in Table.3 and Table.4. It can be seen that the LCV method is better to determine the bandwidth of the mean and amplitude, and the bandwidth is 23.022 and 3.834 respectively. Similarly, for the dynamic cutting force in Z direction, the ROT is better than LCV method for determining the bandwidth of mean, and its value is 6.563. But the LCV method is better for amplitude and the bandwidth is 3.598.

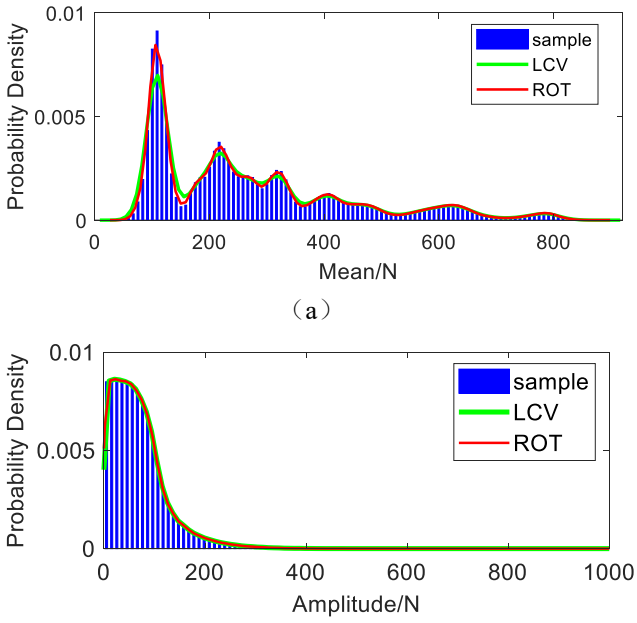

Fig.9 KDE of mean(a) and amplitude(b) of dynamic cutting force in Z Direction

5 Conclusion

Aiming at the problem that the multimodal distribution of dynamic cutting force of CNC lathe cannot be estimated well with the conventional distribution modeling method, a modeling method base

on KDE is proposed. As the ROT method based on MISE and the LCV method based on ISE do not fully consider the impact of bandwidth on the accuracy and smoothness of KDE when determining the bandwidth. The G-SCE method is proposed. The Gaussian function is selected as the kernel function, and the ROT and LCV methods are used as the bandwidth selection method. The K-S test, A-D test, chi-square test and coefficient of determination are chosen as the test indexes to evaluate the goodness and the smoothness test method based on the envelope curve is established. The entropy method judged the goodness index and the smoothness index comprehensively at last.

Rainflow counting is performed on the measured dynamic cutting force of 36 sets. The results show that the mean of dynamic cutting force presents a random multimodal distribution, and a single parameter distribution cannot meet the statistical requirements. The probability density distribution is estimated using the KDE based on G-SCE, and the results show that whether the ROT or the LCV method, the KDE can describe the multi-peak distribution characteristics of the cutting force well, especially at locations with small peak changes.

Case analysis shows that the G-SCE method used to estimate the goodness and smoothness can well solve the difference performance of ROT and LCV methods under different sample data. The results show that in X direction, the KDE determined by the ROT can estimates better for the mean and amplitude of the observed dynamic cutting force, and its bandwidth is 2.624 and 1.860 respectively. For the dynamic cutting force in Y direction, the LCV method better estimates its mean and amplitude, with bandwidths of 23.022 and 3.834 respectively. For the dynamic force in Z direction, the ROT performs better in the mean with a bandwidth of 6.563, but the LCV method performs better in terms of amplitude with a bandwidth of 3.598.

Declarations

Funding This work was supported by the National Natural Science Foundation of China (51905209); Free Exploration Key Project of Natural Science Foundation of Jilin Province Science and Technology Development Plan, China (2020122332JC); Science and Technology Research Project of Education Department of Jilin Province, China (Grant No. 42180); and Program for JLU Science and Technology Innovative Research Team (JLUSTIRT).

Conflicts of interest/Competing interests

Not applicable

Availability of data and material All the data presented and/or analyzed in this study are available upon request to the corresponding author.

Authors' contributions Shengxu Wang: conceptualization, methodology, literature study and model validation, and writing, reviewing and editing.

Guofa Li: methodology, investigation, writing, original draft and visualization.

Jialong He: supervision, investigation, writing, reviewing and editing, visualization.

Qingbo Hao: reviewing and editing.

Hao Huang: investigation and editing.

All authors read and approved the final manuscript.

References

1. He J, Wang S, Li G, et al. Compilation of NC lathe dynamic cutting force spectrum based on two-dimensional mixture models[J]. The International Journal of Advanced Manufacturing Technology, 2018, 98(1-4): 251-262. https://doi.org/10.1007/s00170-018-2067-x.
2. Letot C, Serra R, Dossevi M, et al. Cutting tools reliability and residual life prediction from degradation indicators in turning process[J]. International Journal of Advanced Manufacturing Technology, 2016, 86(1-4):495-506. https://doi.org/10.1007/s00170-015-8158-z
3. He X, Li T, Li Y, et al. Developing an accelerated flight load spectrum based on the nz - N curves of a fleet[J]. International Journal of Fatigue, 2018, 117(DEC.): 246-256. https://doi.org/10.1016/j.ijfatigue.2018.08.005
4. Geng S , Liu X , Yang X , et al. Load spectrum for automotive wheels hub based on mixed probability distribution model[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2019, 2

33(14): 3707-3720. https://doi.org/10.1177/0954407019832433

5. Meddour, Ikhlas, Yallese, et al. Prediction of surface roughness and cutting forces using RSM, ANN, and NSGA-II in finish turning of AISI 4140 hardened steel with mixed ceramic tool[J]. The International Journal of Advanced Manufacturing Technology, 2018, 97(5): 1931-1949. https://doi.org/10.1007/s00170-018-2026-6
6. Zhang Z, Qi Y, Cheng Q, et al. Machining accuracy reliability during the peripheral milling process of thin-walled components[J]. Robotics & Computer Integrated Manufacturing, 2019, 59(OCT.):222-234. https://doi.org/10.1016/j.rcim.2019.04.002
7. Wang Y, Jia Y, Qiu J, et al. Load spectra of CNC machine tools[J]. Quality & Reliability Engineering International, 2015, 16(3): 229-234. https://doi.org/10.1002/1099-1638(200005/06)16:3<229::AID-QRE314>3.0.CO;2-U
8. Li G, Wang S, He J, et al. Compilation of load spectrum of machining center spindle and application in fatigue life prediction[J]. Journal of Mechanical science and Technology, 2019, 33(4): 1603-1613. https://doi.org/10.1007/s12206-019-0312-3
9. Elsayed E A. Overview of Reliability Testing[J]. IEEE Transactions on Reliability, 2012, 61(2): 282-291. https://doi.org/10.1109/TR.2012.2194190
10. Wang S, Liu X, Jiang C, et al. Prediction and evaluation of fatigue life for mechanical components considering anelasticity‐based load spectrum[J]. Fatigue & Fracture of Engineering Materials & Structures, 2021,44(1): 129-140. https://doi.org/10.1111/ffe.13340
11. P, Heuler, H, et al. Generation and use of standardized load spectra and load-time histories[J]. International Journal of Fatigue, 2005, 27(8): 974-990. https://doi.org/10.1016/j.ijfatigue.2004.09.012
12. Lu Y, Bi W, Zhang X, et al. Calculation method of dynamic loads spectrum and effects on fatigue damage of a full-scale carbody for high-speed trains[J]. Vehicle System Dynamics, 2019, 58(7): 1037-1056. https://doi.org/10.1080/00423114.2019.1605080
13. Mark, J, Brewer. A Bayesian model for local smoothing in kernel density estimation[J]. Statistics & Computing, 2000, 10(4): 299-309. https://doi.org/10.1023/A:1008925425102
14. Qin Z, Li W, Xiong X. Estimating wind speed probability distribution using kernel density method[J]. Electric Power Systems Research, 2011, 81(12):2139-2146. https://doi.org/10.1016/j.epsr.2011.08.009
15. Siddharth, Arora, James. Forecasting electricity smart meter data using conditional kernel density estimation[J]. Omega, 2016, 59: 47-59. https://doi.org/10.1016/j.omega.2014.08.008
16. Gao G. A Parzen-Window-Kernel-Based CFAR Algorithm for Ship Detection in SAR Images[J]. Geoence & Remote Sensing Letters IEEE, 2011, 8(3): 557-561. https://doi.org/10.1109/LGRS.2010.2090492
17. DUONG, Tarn; HAZELTON, Martin. Plug-in bandwidth matrices for bivariate kernel density estimation[J]. Journal of Nonparametric Statistics, 2003, 15(1): 17-30. https://doi.org/10.1080/10485250306039
18. Duong T, Hazelton M L. Cross-validation Bandwidth Matrices for Multivariate Kernel Density Estimation[J]. Scandinavian Journal of Stats, 2005, 32(3):485-506. https://doi.org/10.1111/j.1467-9469.2005.00445.x
19. BOTEV, Zdravko I., et al. Kernel density estimation via diffusion[J]. The annals of Statistics, 2010, 38(5): 2916-2957. https://doi.org/10.1214/10-AOS799
20. Yin X F, Hao Z F. Adaptive Kernel Density Estimation using Beta Kernel[C], International Conference on Machine Learning & Cybernetics. IEEE, 2007:3293-3297.
21. Yin F, Zoubir A M. Robust positioning in NLOS environments using nonparametric adaptive kernel density estimation[C], IEEE International Conference on Acoustics. IEEE, 2012, 3517-3520.
22. Kaya B, Oysu C, Ertunc H M. Force-torque ba

sed on-line tool wear estimation system for CN C milling of Inconel 718 using neural networks [J]. Advances in Engineering Software, 2011, 4 2(3):76-84. https://doi.org/10.1016/j.advengsoft.2 010.12.002
23. Vadgeri S S, Patil S R, T. Chavan S. Static and Fatigue Analysis of Lathe spindle for Maximum Cutting Force[J]. Materials Today Proceedings, 2018, 5(2): 4438-4444.
24. Zhang X, King M L, Hyndman R J. A Bayesia n approach to bandwidth selection for multivari ate kernel density estimation[J]. Computational Statistics & Data Analysis, 2006, 50(11): 3009-3031. https://doi.org/10.1016/j.csda.2005.06.019
25. Wglarczyk S. Kernel density estimation and its application[J]. ITM Web of Conferences, EDP Sciences, 2018, 23(2):00037.
26. Hu B, Li Y, Yang H, et al. Wind speed model based on kernel density estimation and its appli cation in reliability assessment of generating sy stems[J]. Journal of Modern Power Systems & Clean Energy, 2017, 5(2):220-227. https://doi.or g/10.1007/s40565-015-0172-5
27. Dias J A S, Borges C L T. A non-parametric stochastic model for river inflows based on kernel density estimation[C], 2014 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS). IEEE, 2014.
28. Diez-Olivan A, Pagan J A, Khoa N L D, et al. Kernel-based support vector machines for auto mated health status assessment in monitoring se nsor data[J]. International Journal of Advanced Manufacturing Technology, 2017, 95(1): 327-34 0. https://doi.org/10.1007/s00170-017-1204-2
29. Xu X, Yan Z, Xu S. Estimating wind speed pr obability distribution by diffusion-based kernel density method[J]. Electric Power Systems Rese arch, 2015, 121(apr.): 28-37. https://doi.org/10.1 016/j.epsr.2014.11.029
30. He, Jialong, et al. Time domain load extrapolati on method for CNC machine tools based on G RA-POT model[J]. The International Journal of Advanced Manufacturing Technology. 2019,103 (9): 3799-3812. https://doi.org/10.1007/s00170-0 19-03774-3
31. Macea L F, Márquez L, LLinás H. Improvemen t of axle load spectra Characterization by a mix ture of three distributions[J]. Journal of Transpo rtation Engineering, 2015, 141(12): 04015030. https://doi.org/10.1061/(ASCE)TE.1943-5436.00 00801

Figures

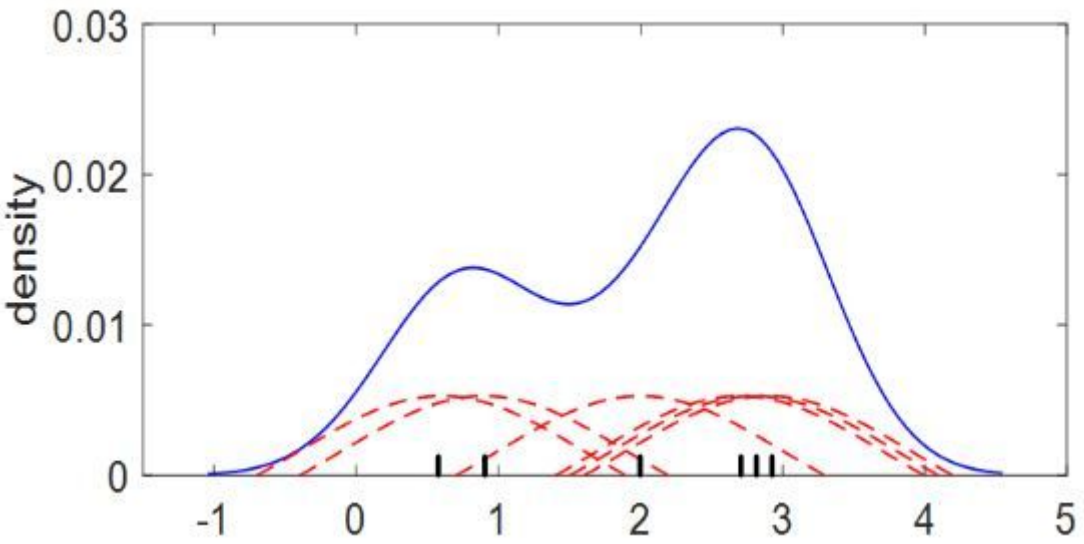

Figure 1

Gaussian function to estimate kernel density

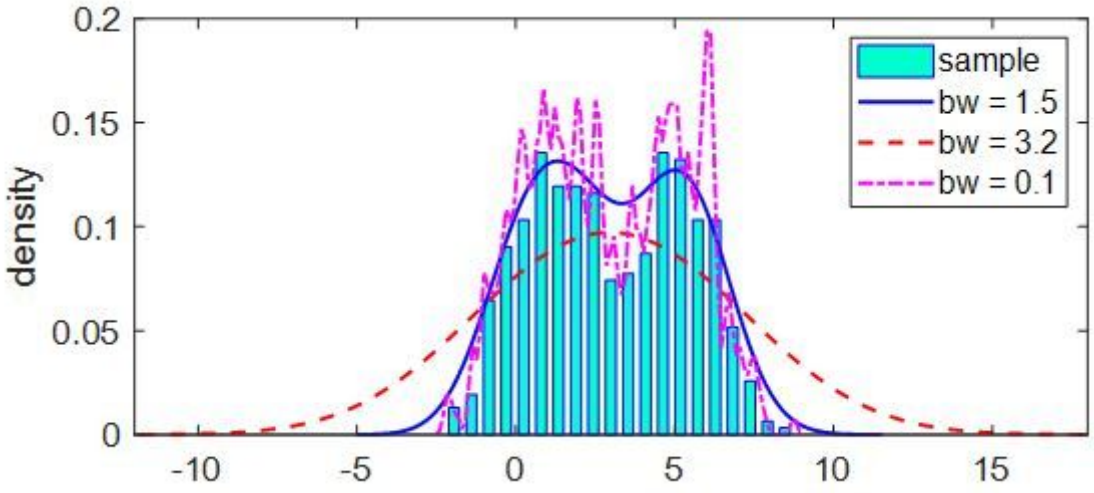

Figure 2

KDE results under different bandwidths

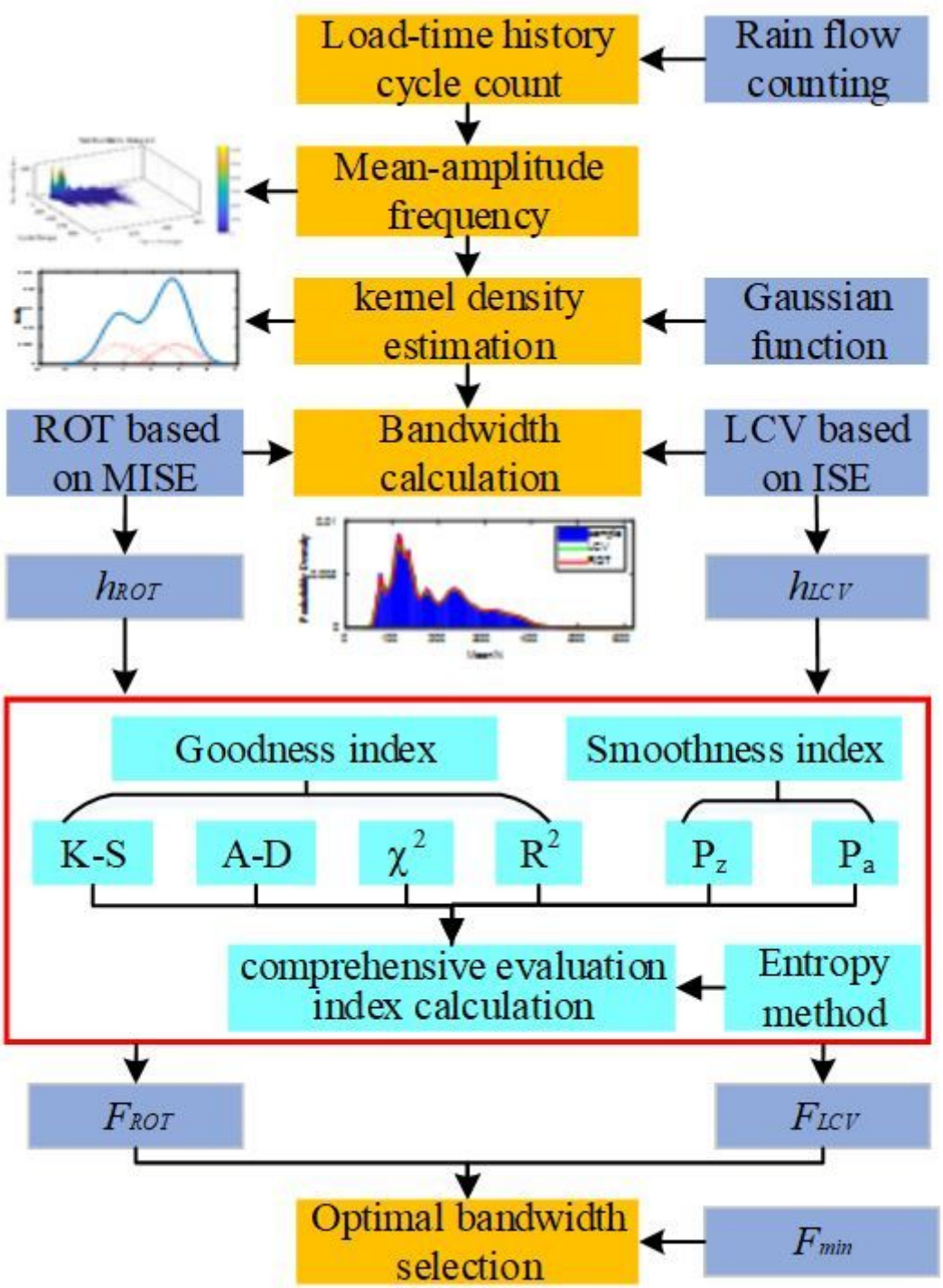

Figure 3

flow chart of G-SCE

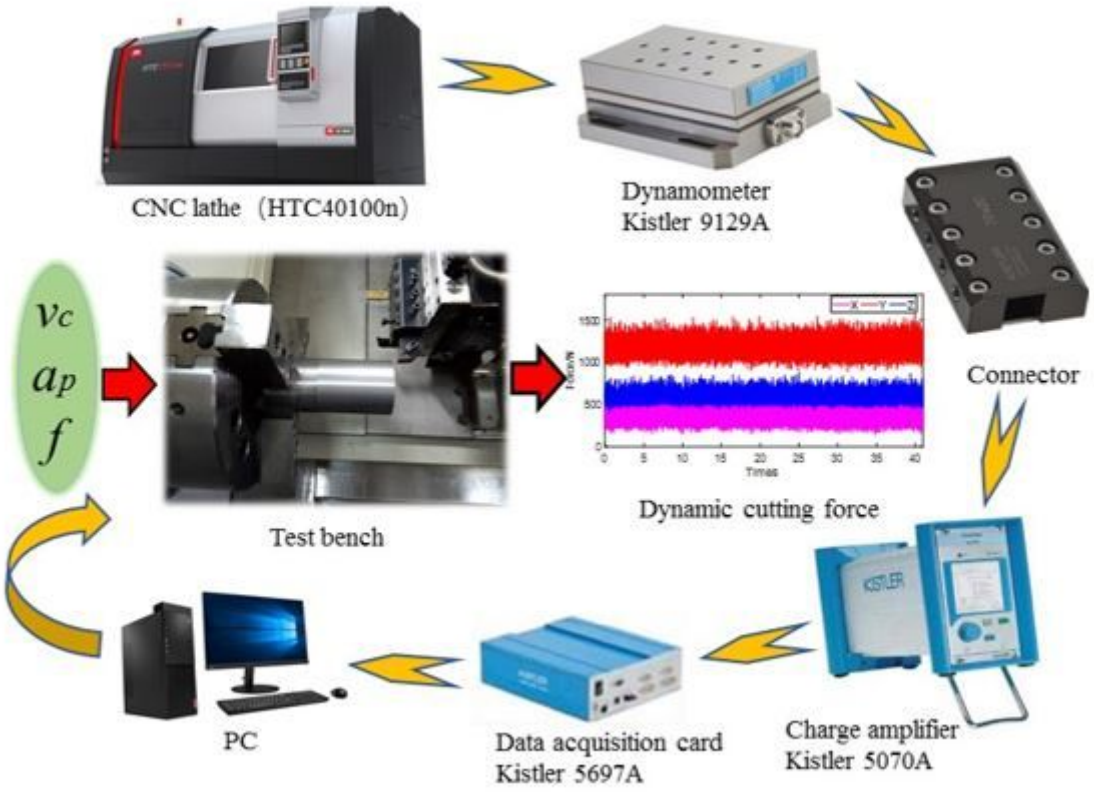

Figure 4

Test bench for cutting force measurement

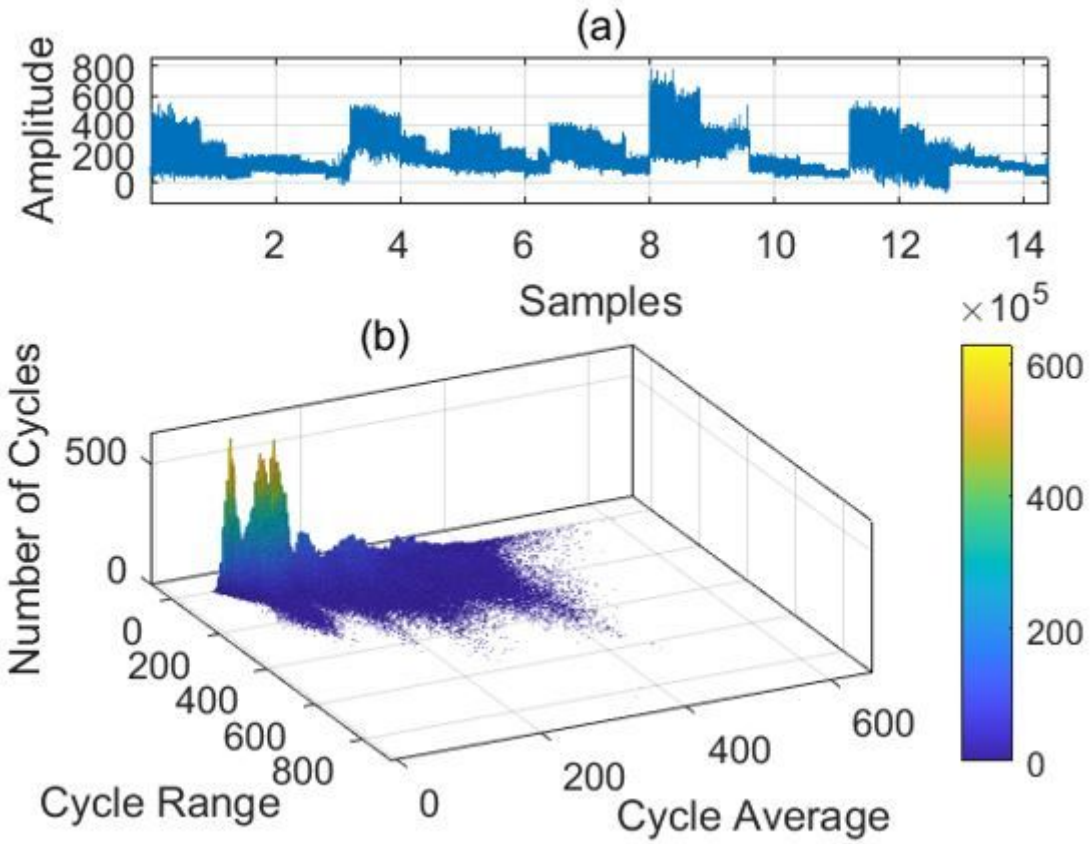

Figure 5

Dynamic cutting force in X direction(a) and its rainflow counting result(b).

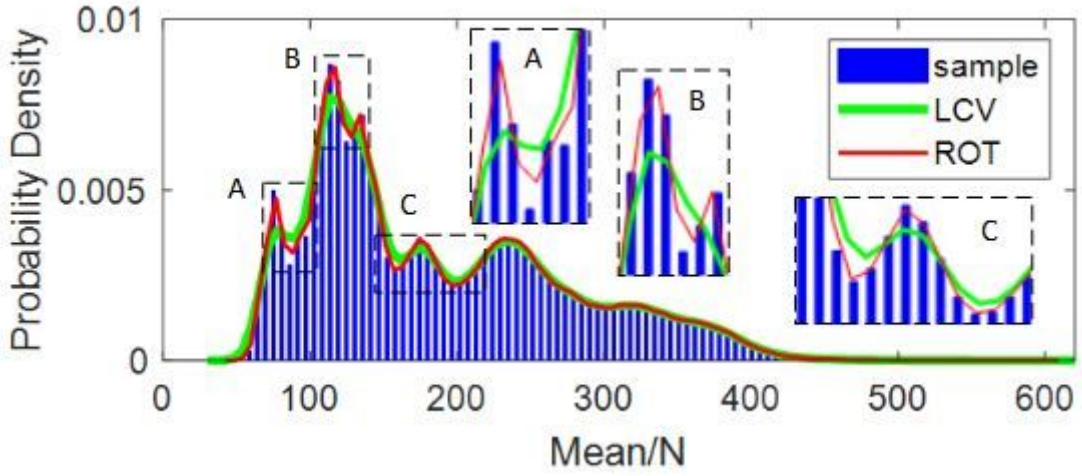

Figure 6

Probability density Estimation of the mean of dynamic cutting force in X direction

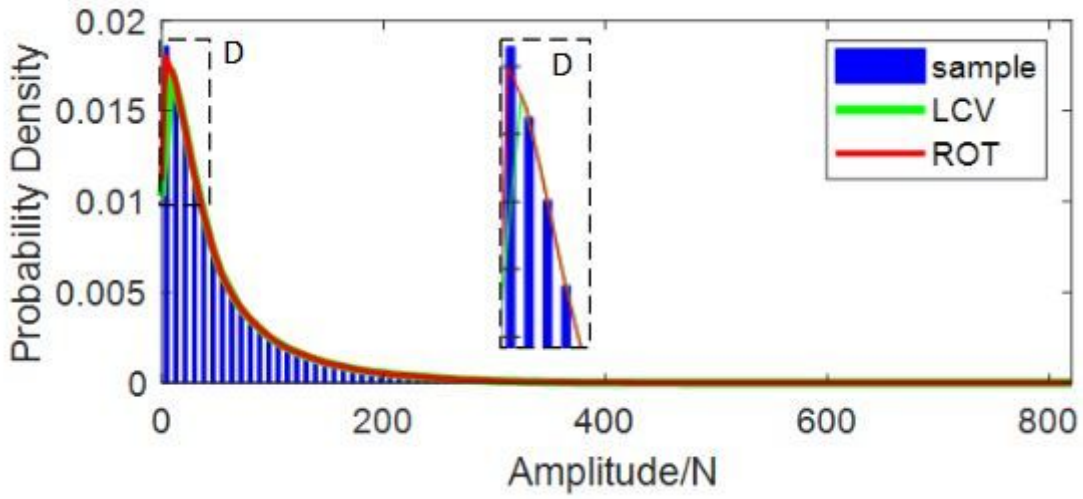

Figure 7

Probability density estimation of amplitude of dynamic cutting force in X direction

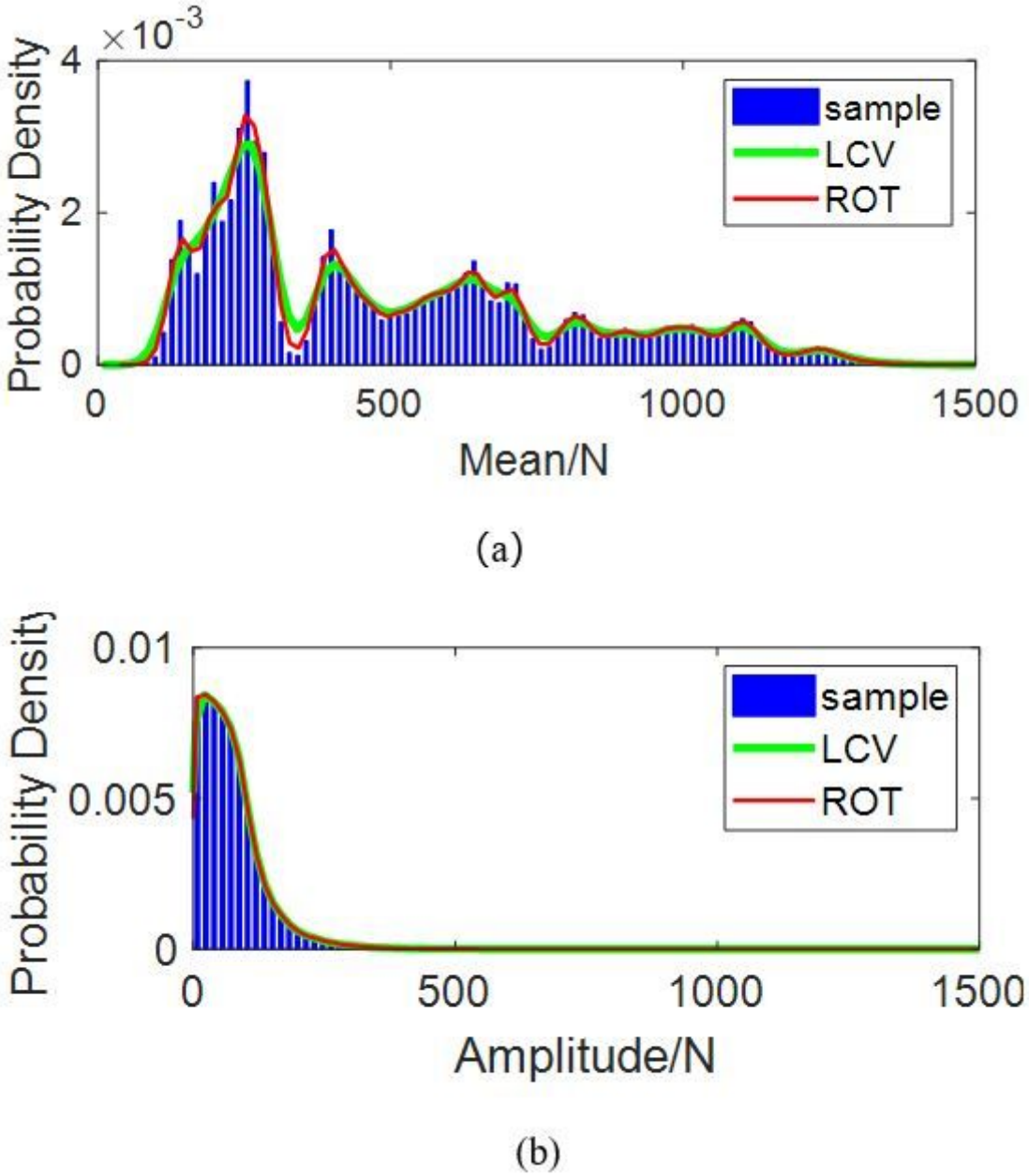

Figure 8

KDE of mean(a) and amplitude(b) of dynamic Cutting Force in Y Direction.

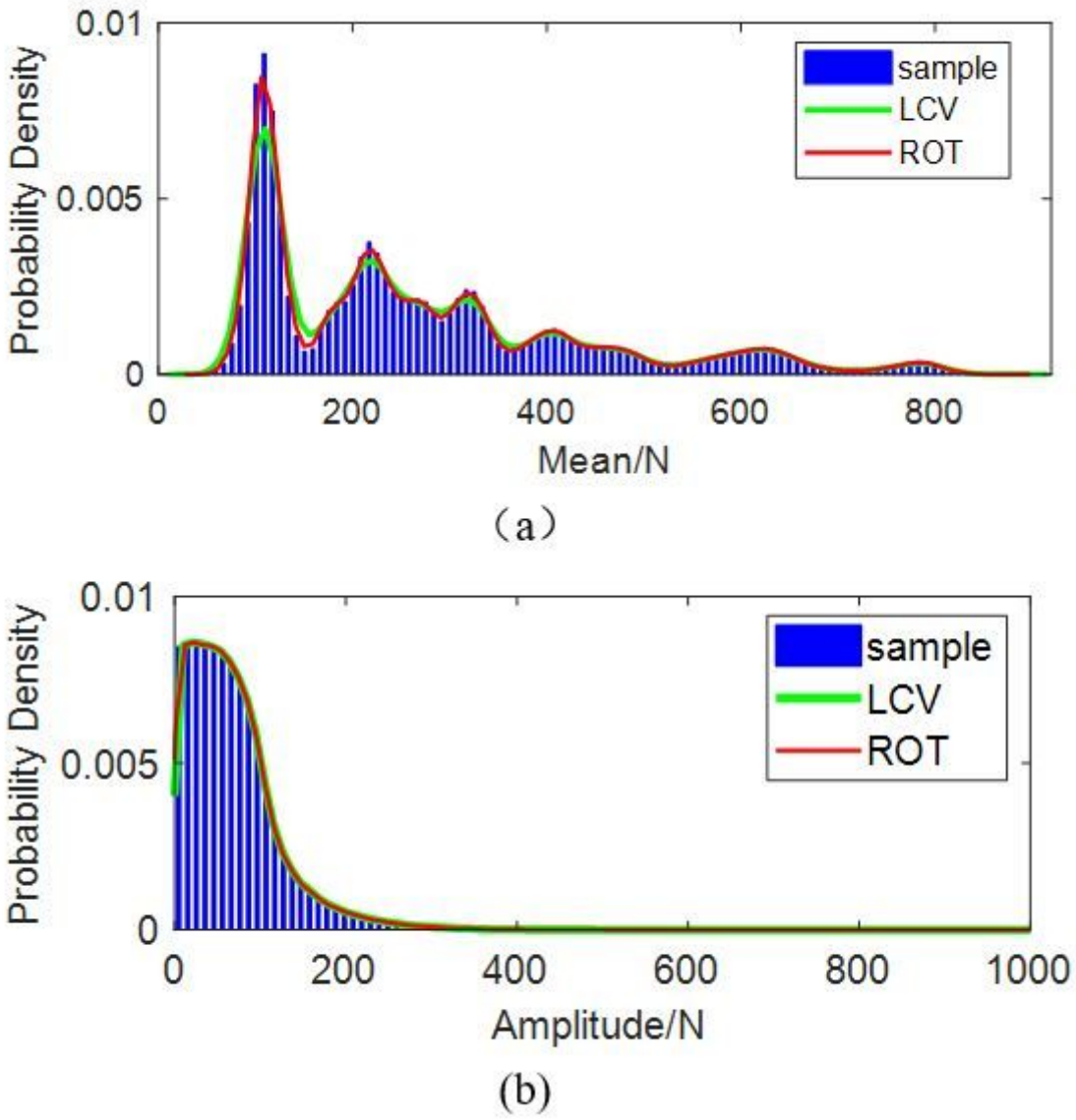

Figure 9

KDE of mean(a) and amplitude(b) of dynamic cutting force in Z Direction