

基于自适应带宽核密度估计的载荷外推方法研究

牛文铁 才福友 付景静

(天津大学机构理论与装备设计教育部重点实验室, 天津 300350)

摘要: 为了快速准确地得到玉米收获机车架的载荷谱,针对载荷谱编制过程中传统的载荷外推方法的局限性,提出一种基于二叉树算法的自适应带宽核密度估计(Kernel density estimation)算法,用来进行载荷外推。将经过预处理的实测原始载荷数据进行雨流计数统计,得到载荷循环均幅值矩阵,将小于载荷循环最大幅值10%的载荷滤除,其余的载荷循环均幅值数据根据二叉树分割算法进行不同区域的分割,选择高斯核函数,根据拇指法则计算各个区域的局部最优带宽,并根据数据区域内数据点的密集程度对核密度估计的输入进行优化,减少了核密度估计的计算量,最后结合蒙特卡洛模拟算法进行载荷外推。采用玉米收获机车架实测载荷数据进行实例验证,结果表明,与传统的固定带宽、自适应带宽核密度估计的载荷外推方法相比,本文提出的方法大大提高了计算效率,其概率密度函数图与实际载荷分布更为接近;载荷循环均幅值频次分布相关系数更接近于1,均方根误差更小;载荷循环幅值累积频次曲线的决定系数均大于0.99。

关键词: 玉米收获机; 车架; 载荷谱; 载荷外推; 自适应带宽核密度估计; 蒙特卡洛模拟

中图分类号: S225.5; TU413.4 **文献标识码:** A **文章编号:** 1000-1298(2021)01-0375-10 **OSID:**



Load Extrapolation Method Based on Adaptive Bandwidth Kernel Density Estimation

NIU Wentie CAI Fuyou FU Jingjing

(Key Laboratory of Mechanism Theory and Equipment Design, Ministry of Education, Tianjin University, Tianjin 300350, China)

Abstract: For the limitation of the traditional load extrapolation methods in the process of load spectrum compilation, an adaptive bandwidth kernel density estimation algorithm was proposed based on the quad-tree algorithm to obtain the load spectrum, which can obtain the corn harvester frame more accurately and quickly. Firstly, the rain flow counting method was applied to count the pretreated measured load data. The load cycles whose amplitudes were less than 10% of the maximum load cycle amplitude value were filtered. The remaining load data were segmented into different regions according to the quad-tree segmentation algorithm. The Gaussian kernel function was selected as the kernel function, and the local optimal bandwidth of the data in each region was calculated according to the rule of thumb. In addition, the input of the kernel density estimation was optimized according to the density of data points in the data area, which reduced the calculation consumption of kernel density estimation. The measured load data from the frame of the corn harvester were used for verifying the effectiveness of proposed method. Compared with the traditional load extrapolation methods of fixed and adaptive bandwidth kernel density estimation, the proposed method greatly improved the computational efficiency, the probability density calculated by the proposed method was closer to the actual load distribution. The correlation coefficient of frequency distribution of load cycle mean and amplitude was closer to 1, and the root mean square error was smaller. The determining coefficient of the amplitude cumulative frequency curve was greater than 0.99. The results showed that the research result can provide reference for load extrapolation and load spectrum compilation.

Key words: corn harvester; frame; load spectrum; load extrapolation; adaptive bandwidth kernel density estimation; Monte Carlo simulation

收稿日期: 2020-06-14 修回日期: 2020-09-27

基金项目: 国家重点研发计划项目(2017YFD0700300)

作者简介: 牛文铁(1971—),男,副教授,主要从事农业装备试验验证方法与技术 and 精密机床数字化设计研究, E-mail: niuwentie@tju.edu.cn

0 引言

载荷谱^[1-2]是进行零部件疲劳寿命预测和优化设计的重要依据。目前,载荷谱已经在航天、汽车、高铁等领域得到了广泛研究与应用^[3-5]。而我国大型玉米收获机关键零部件载荷谱的编制方法研究尚处于起步阶段,与国外发达国家相比还有较大的差距。国产玉米收获机故障频发,稳定性较差,其主要原因是设计水平不高,在研发设计初期大多以仿制和借鉴为主,很多零部件的结构设计并没有足够的载荷谱数据作为支撑。玉米收获机的车架是整车的核心承载部件,承受着割台、驾驶室、发动机、粮仓等传递的力和力矩,在实际工作过程中,经常会出现车架裂痕、甚至断裂的现象,严重影响了玉米收获机的整车性能。因此,采用合理的载荷外推方法进行多种工况下玉米收获机车架的载荷谱编制,对于玉米收获机车架的结构优化和整车性能提高具有重要意义。

在实际载荷测试试验中,由于时间、天气、试验场地和成本等多种原因,只能进行一定时间内载荷时间历程的测量,实测载荷仅反映试验期间的载荷时间历程,不能反映全寿命周期下的载荷分布,因此需要对有限的载荷时间历程进行合理有效的载荷外推^[6]。载荷外推是影响载荷谱编制准确性的关键环节。JOHANNESSON^[7]提出了基于 POT 参数模型的时域外推方法,通过设置峰值的阈值,估计峰值的分布函数,在最大限度保留原始载荷时间序列的基础上,重构得到外推后的载荷时间历程。杨子涵等^[8]对时域外推过程中的阈值选取方法进行了改进,提出了一种时域外推过程中阈值选取的量化方法,解决了时域外推过程中阈值选取主观性较强的问题。张英爽等^[9]将经过雨流计数法统计后的载荷循环均值和载荷循环幅值的分布分别用正态分布和威布尔分布进行参数拟合,并以此进行载荷外推。翟新婷等^[10]、GENG 等^[11]以混合分布函数作为函数拟合模型进行参数估计,该方法的拟合效果优于单分布的参数估计。但是,参数外推方法是用纯粹的分

布函数来描述载荷的分布规律,对于复杂、随机性较大的载荷,则会产生较大的误差,大大降低了外推载荷谱与实际载荷谱的等效性。随着研究的不断深入,基于核密度估计法的非参数估计方法被应用于载荷外推中,该方法既可很好地保留载荷数据本身的分布规律,又能实现对任意载荷分布的拟合。李凡松等^[12]采用基于自适应带宽的核密度估计方法进行载荷外推,每一个数据点都有其自身对应的带宽,该方法相较于固定带宽的核密度估计方法具有

更好的拟合效果。

本文针对核密度估计载荷非参数外推方法中带宽的选择问题,结合改进的四叉树分割算法,对核密度估计的带宽计算进行优化,提出一种改进的自适应带宽核密度估计的载荷外推方法。以玉米收获机车架为研究对象,通过与传统的基于核密度估计的固定带宽外推方法和自适应带宽外推方法进行比较,验证该自适应带宽核密度估计的载荷外推方法的准确性和合理性。

1 核密度估计

核密度估计是由 ROSENBLATT^[13] 和 PARZEN^[14]提出的一种由实测分布函数未知的随机变量来估计其概率密度函数的非参数估计方法。该方法不需要知道数据的先验分布,也不必对数据进行任何的假设,只需要确定输入的数据变量、核函数以及带宽就可以估计输入数据的概率密度函数。目前,核密度估计已经在电力、医疗、地理等领域得到了非常广泛的应用^[15-18]。玉米收获机工作时由于其结构复杂,工况多样,导致负载波动大,实测载荷信号具有很强的分散性和随机性,相较于一般的参数估计法来说,核密度估计能更好地描述载荷数据的分布规律,从而使得载荷外推更加准确。

基于雨流计数矩阵的载荷外推是关于载荷循环均值和幅值的二维问题,二维核密度估计表达式为

$$\hat{f}(x,y)=\frac{1}{nh_xh_y}\sum_{i=1}^nK\left(\left|\frac{x-x_i}{h_x}\right|,\left|\frac{y-y_i}{h_y}\right|\right)\tag{1}$$

其中

$$\begin{cases} K(\cdot)\geq 0 \\ \iint K(\cdot)dx=1 \end{cases}\tag{2}$$

式中 n ——输入数据点的个数
 $h_x、h_y$ ——核密度估计载荷循环幅值、均值的带宽
 $x_i、y_i$ ——输入的第 i 个载荷循环幅值、均值
 $K(\cdot)$ ——核函数

只要选取的核函数和带宽合适,就可以无限制地去逼近任何随机变量真实的概率密度函数。大量研究表明,当输入的样本数据量足够大时,核函数的具体形式对概率密度估计的准确性产生的影响相对较小^[19],本文选择光滑且连续、明显单峰分布的高斯核函数。二维高斯核函数表达式为

$$K\left(\left|\frac{x-x_i}{h_x}\right|,\left|\frac{y-y_i}{h_y}\right|\right)=\frac{1}{2\pi}\exp\left(-\frac{(x-x_i)^2}{2h_x^2}-\frac{(y-y_i)^2}{2h_y^2}\right)\tag{3}$$

与核函数相比,带宽 h 的选择对核密度估计的准确性影响更大^[20]。带宽 h 决定了 $\hat{f}(x, y)$ 的光滑程度,若 h 较大,则有较多的数据点影响此处的概率密度计算, $\hat{f}(x, y)$ 曲线在此处较光滑,但是其与实际概率密度曲线的偏差较大;若 h 较小,则有较少的数据点影响此处的概率密度计算, $\hat{f}(x, y)$ 曲线在此处较陡峭,但是其与实际概率密度曲线的偏差较小。因此,为了更加准确地进行核密度估计,带宽 h 的选择尤为重要。为了量化实现最优带宽计算,提出了平均积分平方误差 (MISE), MISE 表达式为

$$M_{ISE}(h) = \int E(\hat{f}(x) - f(x))^2 dx \tag{4}$$

式中 $\hat{f}(x)$ ——估计的概率密度
 $f(x)$ ——实测数据真实的概率密度
 $E(\cdot)$ ——求均值函数

拇指法则是目前应用最多的固定最优带宽的计算法则,拇指法则的最优带宽计算公式为

$$h = \left(\frac{4}{d+2} \right)^{\frac{1}{d+4}} \sigma n^{-\frac{1}{d+4}} \tag{5}$$

式中 d ——核密度估计的维数,取 2
 σ ——输入二维数据样本的标准差

此外,还可用无偏交叉验证 (Unbiased cross validation)^[21]、插入法 (Plug-in)^[22] 等进行最优带宽的计算。上述方法计算的最优带宽是固定的,即每一个数据都有着同样的带宽,不能自动调节和变化,然而,在核密度估计的实际应用中,由于数据的随机性较强,分布不均匀,通过固定带宽核密度估计计算得到的概率密度可能与实际的分布相差较大,所以,希望带宽能够随数据的变化而变化,在数据密集的地方取得小一些,在数据稀疏的地方取得大一些。因此,自适应带宽核密度估计^[23-24] 得到了广泛的应用,该方法的具体计算公式为

$$\hat{f}(x, y) =$$

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{(\lambda_i h_x)(\lambda_i h_y)} K \left(\left| \frac{x - x_i}{\lambda_i h_x} \right|, \left| \frac{y - y_i}{\lambda_i h_y} \right| \right) \tag{6}$$

其中 $\lambda_i = (g^{-1} \hat{f}(x_i, y_i))^{-\alpha} \tag{7}$

$$g = \sqrt{\prod_{i=1}^n \hat{f}(x_i, y_i)} \tag{8}$$

式中 λ_i ——带宽的自适应修订系数
 α ——敏感性参数,取 0 ~ 1
 $\hat{f}(x_i, y_i)$ ——固定带宽初步估计的概率密度
一般情况下当 α 取 0.5 时,核密度估计的拟合效果较好^[25],相对于固定带宽的核密度估计来说,自适应带宽在一定程度上提高了核密度估计的准确

性,但由于数据集中的每一个数据都会影响彼此的带宽,当输入数据集较大时,会大大降低核密度估计的计算效率。当一定数量的数据聚集在一个很小的区域时,此区域内带宽的差异对核密度估计的影响十分有限,因此可以根据数据的聚集程度将同一区域内所有数据的带宽设置成相同的,就可以避免对这些点进行不必要的详细的带宽计算,在保证一定核密度估计准确性的前提下,提高了计算效率,所以针对核密度估计最优带宽的选择问题,为了更好兼顾核密度估计的计算效率与准确性,本文在上述方法的基础上进行改进,提出了一种新的自适应带宽计算方法。

2 基于四叉树算法的自适应带宽核密度估计载荷外推

基于四叉树算法的自适应带宽核密度估计载荷外推主要由 3 个步骤组成,包括:核密度估计的数据输入、基于四叉树算法的数据区域分割以及基于四叉树算法的自适应带宽核密度估计。具体的载荷外推流程图如图 1 所示。

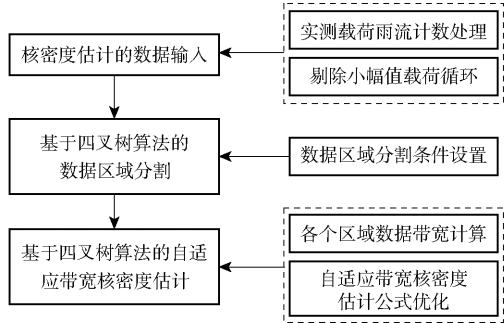


图 1 基于四叉树算法的自适应带宽核密度估计载荷外推流程图

Fig. 1 Adaptive bandwidth kernel density estimation load extrapolation flow chart based on quad-tree algorithm

2.1 核密度估计的数据输入

雨流计数法得到的应变与材料的应力-应变迟滞回线具有很好的一致性,并且考虑了载荷循环幅值和均值 2 个变量,符合疲劳载荷本身固有的特性,能够满足疲劳寿命预估以及载荷谱编制的条件,是目前应用最为广泛的实测载荷时间历程的计数统计方法。首先将已经预处理的实测载荷利用该方法进行计数统计处理,得到载荷循环均幅值矩阵 $M(M_m, M_a)$,其中 M_m 表示载荷循环均值, M_a 表示载荷循环幅值;在实际工作中,小幅度载荷循环的数量比大幅度载荷循环大,但是其造成的疲劳损伤非常小^[26],所以可以将小于最大载荷循环幅值 10% 的载荷循环只进行简单的线性比例外推,这很大程度上减少了核密度估计的数据输入量,从而降低核密度估计

的计算量,提高了计算效率。已经过滤掉小幅值载荷循环的载荷循环均幅值矩阵记为 $M_n(M_{mn}, M_{an})$, 并以此作为核密度估计的数据输入。

2.2 基于四叉树算法的数据区域分割

四叉树算法^[27]是一种经典的空间分割与索引技术,目前在图像分割与空间索引领域已经得到了广泛的应用^[28-29]。该算法可以通过对数据空间进行递归四等分割将数据区域分割成密度不等的块,数据越密集的地方,块的数量越多,块相对较小;而数据越稀疏的地方,块的数量越少,块相对较大,分割出的块的数量与大小从一定程度上反映出数据的密集程度。数据密集的地方采用较小的带宽,能够更好地反映出该区域分布的细节,与实际分布更加接近,拟合效果较好。根据四叉树分割算法分割出不同的数据块,计算各自的局部最优带宽,达到自适应带宽核密度估计的目的。将一个区域进行四叉树分割的流程如图 2 所示。

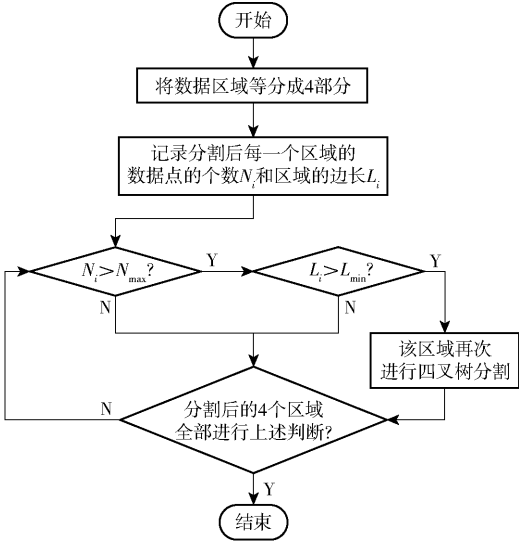


图 2 四叉树分割流程图

Fig. 2 Flow chart of quad-tree segmentation

通过对区域不断执行上述四叉树分割流程,直到所有区域都不满足区域分割的条件,最后可以将原始数据区域分割成大小不尽相同的数据块。判断数据区域是否需要分割的阈值,数据区域内数据点个数的最大值 $N_{\max}$ 和数据区域最小宽度 $L_{\min}$ 对分割完成后数据块的大小和数量有着重要的影响,不同 $N_{\max}$ 对概率密度函数的影响如图 3 所示。

由图 3 可知, $N_{\max}$ 较大时,图 3b 不能很好地反映出该区域内数据点分布的细节信息; $N_{\max}$ 较小时,图 3c 反映的细节信息和图 3a 没有明显的差别,但是 $N_{\max}$ 越小,四叉树分割的层次就越深,计算量越大。

四叉树分割相当于根据数据的密集程度进行数据的分类,数据块内的数据点影响着该区域的带宽,

而数据块之间不互相影响。这与 KNN-KDE 理论^[30]中根据每个数据点由相同个数的相邻点来确定区域有些相似,在每个区域内包含的点数相同的情况下,自适应带宽的大小与数据区域的大小呈正比。为了更好地进行数据区域的划分,根据 KNN-KDE 理论中二维数据的最优相邻个数将数据区域分割的阈值 $N_{\max}$ 设置成 $\sqrt{n}$, n 为进行四叉树分割数据点的总个数。 $L_{\min}$ 的设置是为了避免四叉树分割进入太深层次。在数据特别密集的地方,带宽的变化很小,对最后分布拟合的影响有限,为了减少计算量,不需要进行太深层次的四叉树分割。通过雨流计数法得到的载荷均幅值矩阵通常按照 64 级来统计计数,每一级数据区域的边长定义为

$$\Delta_{an} = (\max M_{an} - \min M_{an}) / 64 \tag{9}$$

$$\Delta_{mn} = (\max M_{mn} - \min M_{mn}) / 64 \tag{10}$$

式中 Δ_{an} ——载荷循环幅值每一级的长度
 Δ_{mn} ——载荷循环均值每一级的长度

因此定义四叉树分割数据块的最小宽度 $L_{\min} = \min(\Delta_{an}, \Delta_{mn})$, 这既能提高四叉树分割的计算效率,也能最小限度影响核密度估计的准确性。

2.3 基于四叉树算法的自适应带宽核密度估计

利用四叉树分割算法将数据区域分割成不同的数据块后,每个数据块内数据点的个数不相同,数据点带宽的计算流程如图 4 所示。其中, $N_{\min}$ 是判断该数据块内数据点带宽计算的阈值条件, $N_{\min} = \sqrt{N_{\max}}$, $N_{\min}$ 向下取整; h_{xi} 、 h_{yi} 分别为数据块内载荷循环幅值和载荷循环均值的带宽, L_{xi} 、 L_{yi} 分别为数据块载荷循环幅值和载荷循环均值的边长。

不同数据块内数据点的带宽因为数据的差别而有所不同,但同一数据块内数据点的带宽是相同的。当一个数据块内的数据点密集程度比较高,彼此相差不大时,每一个数据点贡献的核密度估计分量的差别很微小,数据落入此数据块不同位置的概率近似相同。因此,为了进一步提高自适应核密度的计算效率,在数据密集程度比较高的块,不需要每一个数据点都进行核密度估计的计算,只需要将该数据区域数据点的平均值作为核密度估计的数据输入,再乘以该区域数据点的个数即可。在此将表示数据块内数据点密集程度的参数定义为

$$\gamma = \frac{L_{side}}{H} \tag{11}$$

其中 $L_{side} = \min(\Delta_{an}, \Delta_{mn}) / 2 \tag{12}$

$$H = \max(h_x, h_y) \tag{13}$$

当 $\gamma > 1$ 时,表示该数据块内的数据点密集程度高,可以对此数据块内的核密度估计计算进行优化。依

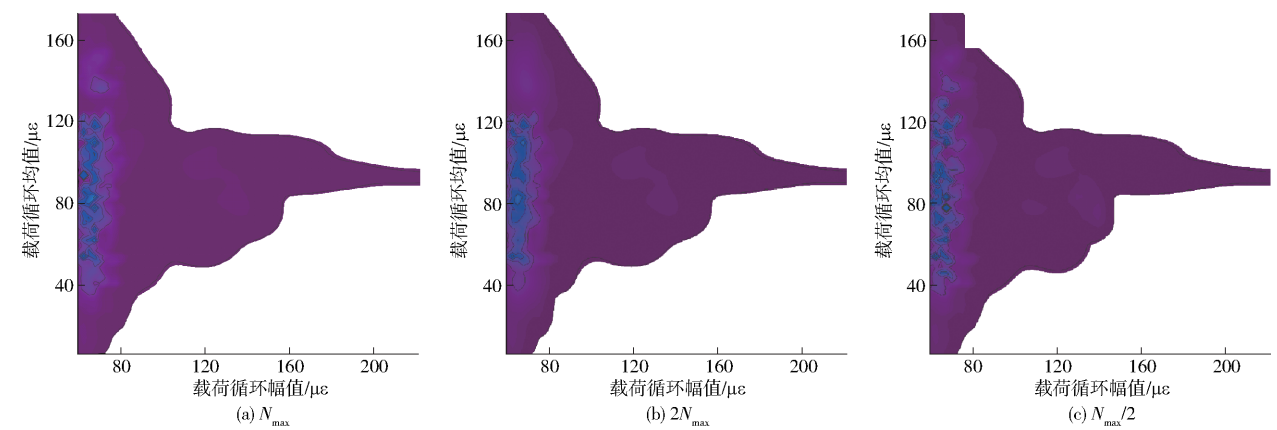


图 3 不同 $N_{\max}$ 对概率密度函数的影响

Fig. 3 Effects of different $N_{\max}$ on probability density function

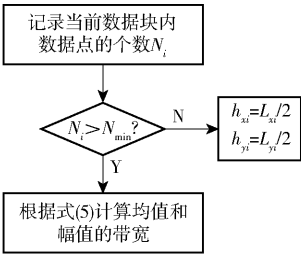


图 4 数据点带宽计算流程图

Fig. 4 Flow chart of data point bandwidth calculation

此改进的自适应核密度估计公式为

$$\hat{f}(x,y)=\frac{1}{N}\sum_{i=1}^{N_0}\frac{1}{h_{x_i}h_{y_i}}n(x_i,y_i)K\left(\left|\frac{x_i-x}{h_{x_i}}\right|,\left|\frac{y_i-y}{h_{y_i}}\right|\right)$$

(14)

式中 N ——优化前载荷数据点总个数

N_0 ——优化后载荷数据点总个数

当 $\gamma > 1$ 时, $n(x_i, y_i)$ 表示此数据块内数据点的个数, x_i, y_i 分别表示此数据块内数据点的载荷循环幅值以及载荷循环均值的平均值; 当 $\gamma < 1$ 时, $n(x_i, y_i)$ 等于 1, x_i, y_i 分别表示此数据块内每一个数据点的载荷循环幅值以及载荷循环均值。

通过上述方法进行自适应带宽核密度估计后, 再结合 Monte Carlo 模拟算法进行载荷外推。

3 实例验证

3.1 试验方案

玉米收获机在实际工作中, 工况较多且复杂多变, 由于地形、车辆载重以及操作行为不规范等原因, 车架时常会出现裂纹, 大大降低了车架的疲劳强度, 影响整车性能, 受到很大冲击时, 甚至会出现车架整根断裂的问题, 严重影响了粮食收获。由于试验环境较恶劣, 并且测试部位空间有限, 需要选用安装方便并且固定牢靠的传感器。因此本试验采用中航工业公司 BE350-4AA 型应变片进行数据测取,

并通过有线连接的方式与 HBM 公司的 SoMat eDAQ 型数据采集仪连接。该数据采集仪具有极佳的密封性能和抗震性能, 能有效应对水溅、扬尘、颠簸等测试环境, 有线连接的方式能够确保信号传输的稳定性, 通过数采的计算机终端软件将采样频率设置为 500 Hz。应变片的安装如图 5 所示, 现场试验如图 6 所示, 数据采集系统组成如图 7 所示。

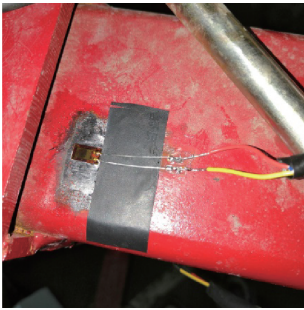


图 5 应变片安装图

Fig. 5 Installation drawing of strain gauge sensor



图 6 现场试验图

Fig. 6 Photo of field test

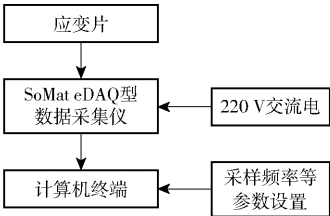


图 7 数据采集系统组成

Fig. 7 Composition of data acquisition system

砂石路面工况下行驶较为平稳, 载荷变化相对较小, 大幅值载荷循环较少, 载荷循环的分布相对集

中;田间收获工况下载荷容易受田间地面的软硬程度以及粮仓的使用率等因素的影响,载荷的随机性较高,平稳性较差,大幅值载荷循环相对多一些,载荷循环的分布也更为分散。根据通过四叉树分割算法得到的数据块的大小以及密集程度,能够很好地

反映出这两种工况下数据的密集程度以及分布情况,并且这两种工况占玉米收获机实际工作工况中很大的比重,具有一定的代表性,为了降低结论的偶然性,本文采用 2 种工况下各 3 组样本进行对比验证,经过预处理的试验数据如图 8 所示。

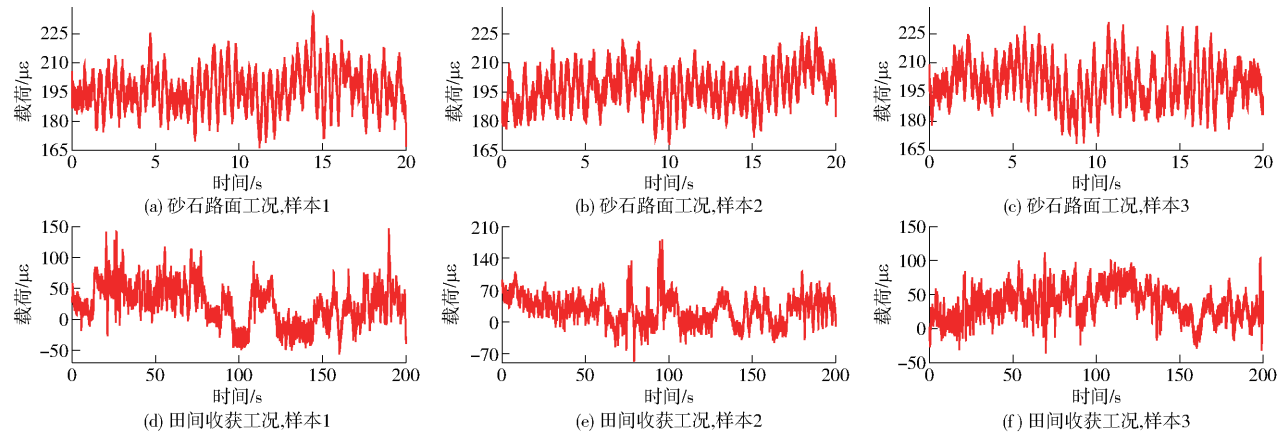


图 8 预处理后的试验数据

Fig. 8 Preprocessed experimental data

3.2 结果验证

通过雨流计数法对上述已经完成预处理的车架实测原始载荷进行统计计数处理,得到的部分载荷循环均幅值频次分布图如图 9 所示。

输入,本文方法的核密度估计的数据输入在过滤的基础上还需要通过改进的四叉树算法进行优化。首先将已经筛选好的载荷循环均幅值矩阵中的数据点绘制成如图 10 所示的散点图,并在此散点图上利用改进的四叉树算法进行数据区域的分割操作,数据区域信息如表 1 所示。

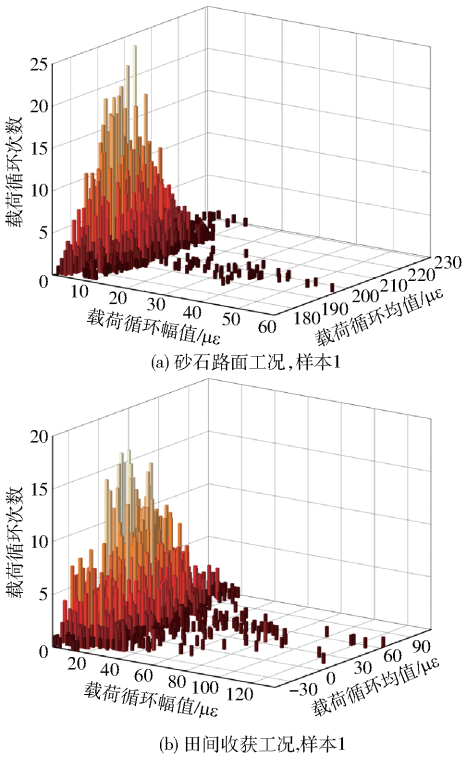


图 9 载荷循环均幅值频次分布图

Fig. 9 Frequency distribution histograms of load cycle mean and range

由图 9 可知,幅值较小的载荷循环占比较大,将小于最大幅值 10% 的载荷循环过滤掉,过滤完成的数据作为固定带宽和自适应带宽核密度估计的数据

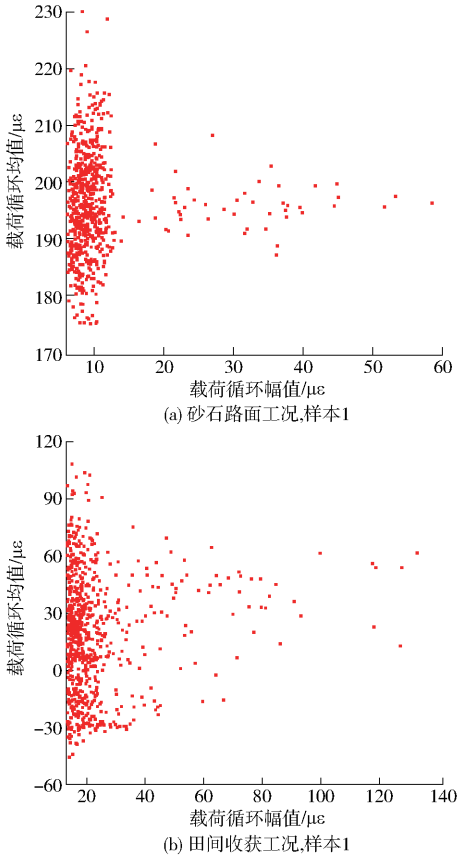


图 10 载荷循环均幅值散点图

Fig. 10 Scatter diagrams of load cycle mean and range

表 1 数据区域信息
Tab.1 Information of data area

工况	样本	分割后的数据块数	分割前核密度估计输入数	分割后核密度估计输入数	计算减少率/%
砂石路面	1	64	648	288	55.6
	2	70	662	345	47.9
	3	61	648	412	36.5
田间收获	1	68	797	378	52.6
	2	61	799	320	59.9
	3	64	753	414	45.0

经过基于四叉树算法的数据区域分割的计算,数据区域被分割成大小不相同的数据块,每个数据块都有各自的带宽,利用式(14)对数据块的自适应带宽核密度估计进行优化,在之前滤除小幅值载荷循环的基础上又大幅减少了核密度估计的输入,进一步提高了核密度估计的计算效率。

为了验证本文提出的基于四叉树算法的自适应带宽核密度估计的载荷外推方法的准确性与合理

性,与传统的基于核密度估计的固定带宽载荷外推和自适应带宽载荷外推以及实测原始载荷进行对比论证。

由 3 种带宽选择方法计算得到的载荷循环均幅值概率密度分布如图 11 所示。将图 11 与图 10 进行对比,可以看出传统的固定带宽与自适应带宽核密度估计的曲线较为平滑,而本文提出的自适应带宽核密度估计的曲线较为曲折,能够很好地反映图 10 数据的密集程度,并且数据密集点的分布情况划分更加具体,细节信息表达更为清楚,能很好地拟合真实概率密度。为了进一步评价概率密度函数的拟合程度,将固定带宽、自适应带宽以及本文提出的自适应带宽核密度估计的概率密度函数的结果分别结合 Monte Carlo 模拟算法进行载荷外推。将实测原始载荷数据和上述 3 种方法进行载荷外推后得到的载荷循环均幅值分别进行分级统计处理,并计算实测原始数据与 3 种带宽选择外推方法的均方根误差(RMSE),计算结果如表 2 所示。

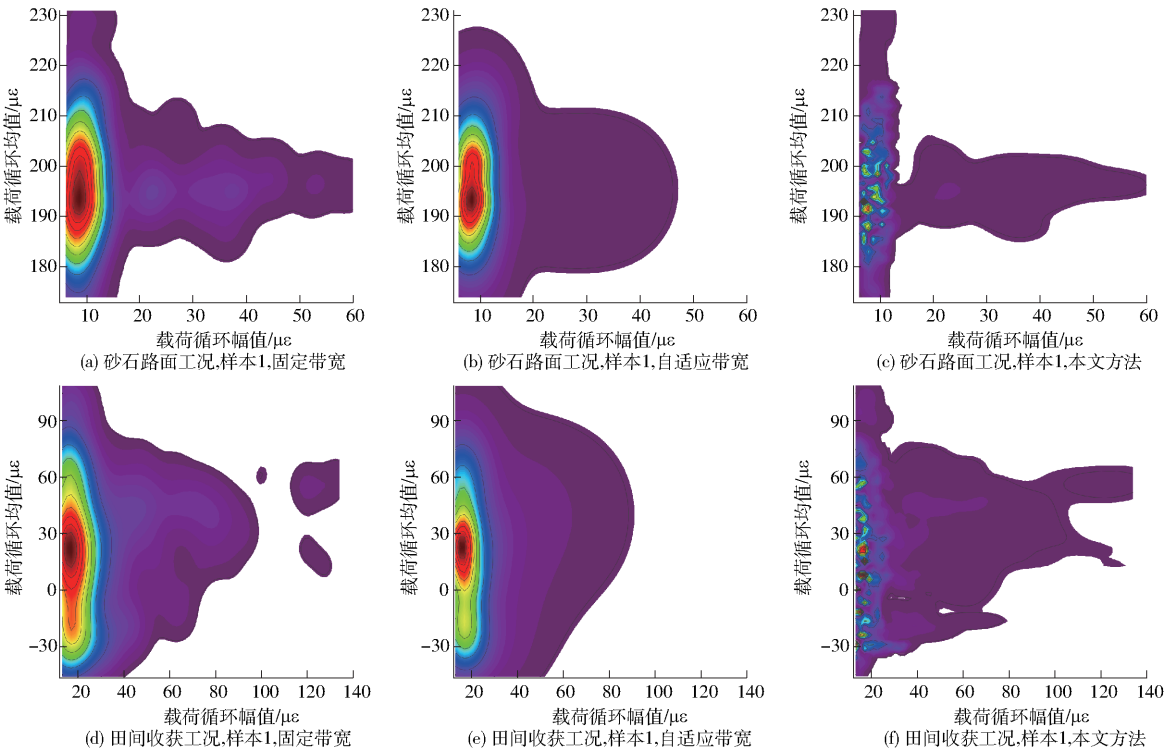


图 11 3 种不同方法的载荷循环均幅值概率密度分布
Fig. 11 Mean and range probability density distributions of three different methods of load cycle

表 2 载荷循环均幅值的均方根误差

Tab.2 RMSE of load cycle mean and range με				
工况	样本序号	固定带宽	自适应带宽	本文方法
砂石路面	1	3.613 1	3.468 2	2.644 1
	2	3.232 8	3.052 2	2.409 0
	3	3.399 7	3.222 8	2.465 6
田间收获	1	4.243 3	3.960 4	2.943 7
	2	4.516 0	4.207 2	2.656 9
	3	3.982 7	3.709 2	2.687 3

由表 2 可知,相较于传统的固定带宽和自适应带宽核密度估计,本文提出的自适应带宽核密度估计载荷外推方法的均方根误差 RMSE 最小,表明通过本文方法计算得到的概率密度更加接近于真实的概率密度。

将实测原始载荷数据和上述 3 种方法进行载荷外推后得到的载荷循环均值和幅值数据分别进行分

级统计处理,结果如图 12 所示。为进一步分析频次分布的相关性和拟合程度,计算实测原始数据与 3 种带宽选择外推方法的相关系数(R)以及均方根误差(RMSE),结果如表 3、4 所示。

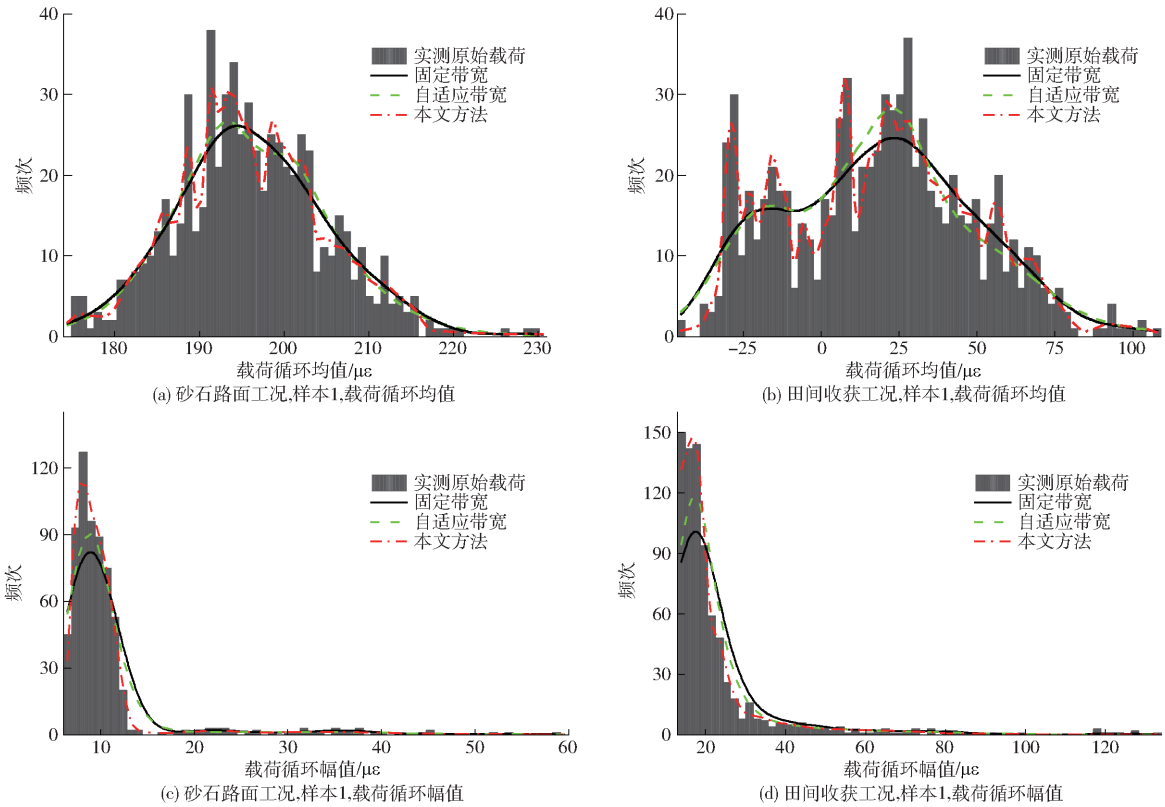


图 12 实测原始数据与 3 种不同方法的载荷循环均值、幅值频次分布比较

Fig. 12 Measured original data compared with mean and range frequency distribution of three different methods

表 3 载荷循环均值的相关系数及均方根误差

工况	样本序号	固定带宽		自适应带宽		本文方法	
		R	RMSE/ $\mu\epsilon$	R	RMSE/ $\mu\epsilon$	R	RMSE/ $\mu\epsilon$
砂石路面	1	0.928 5	3.755 5	0.930 2	3.694 7	0.983 6	1.815 6
	2	0.922 1	3.435 8	0.927 5	3.304 1	0.965 4	2.294 2
	3	0.940 2	2.613 7	0.942 7	2.532 8	0.969 1	1.871 1
田间收获	1	0.859 4	5.025 7	0.862 3	4.933 5	0.973 0	2.249 3
	2	0.914 8	4.368 8	0.923 7	4.139 6	0.980 7	2.089 3
	3	0.929 4	3.211 9	0.935 3	3.033 3	0.963 2	2.303 2

表 4 载荷循环幅值的相关系数及均方根误差

工况	样本序号	固定带宽		自适应带宽		本文方法	
		R	RMSE/ $\mu\epsilon$	R	RMSE/ $\mu\epsilon$	R	RMSE/ $\mu\epsilon$
砂石路面	1	0.958 2	8.443 3	0.973 3	6.817 2	0.995 5	2.555 0
	2	0.963 3	8.142 3	0.983 9	5.575 3	0.997 9	1.825 8
	3	0.947 4	9.419 2	0.972 1	7.022 3	0.994 0	2.964 4
田间收获	1	0.933 6	12.646 3	0.959 8	9.682 0	0.995 0	3.293 9
	2	0.900 5	16.770 9	0.925 6	14.372 7	0.994 6	3.830 4
	3	0.923 2	14.017 9	0.949 1	11.277 5	0.993 5	4.016 1

由表 3、4 可知,相较于传统的固定带宽和自适应带宽核密度估计,本文提出的自适应带宽核密度估计载荷外推与实测原始载荷均值和幅值的

相关系数 R 最大,均方根误差最小,表明通过本文提出的自适应带宽核密度估计进行载荷外推得到的载荷循环均值和幅值分布与实测原始载荷的载

荷循环均值和幅值分布具有很强的相似性,分布拟合效果好,更加接近实测原始载荷的真实分布规律。

载荷循环幅值累积频次曲线为后续程序载荷谱的编制提供数据支撑,影响程序载荷谱编制的准确性。将实测原始数据与传统的固定带宽载荷外推、自适应带宽载荷外推以及本文提出的自适应带宽载荷外推的载荷循环幅值累积频次曲线进行对比验证,如图 13 所示。由图 13 可以看出,通过本文提出的带宽选择方法进行载荷外推得到的曲线与实测原始载荷曲线更为接近,拟合效果更好。为了更加直观进行对比,更好评价载荷循环幅值累积频次曲线的拟合效果,采用决定系数 R^2 作为曲线拟合的检验指标,计算结果如表 5 所示。

表 5 载荷循环幅值累积频次曲线的决定系数				
Tab. 5 Coefficient of determination of load cycle range cumulative frequency curve				
工况	样本序号	固定带宽	自适应带宽	本文方法
砂石路面	1	0.966 9	0.982 0	0.999 8
	2	0.983 0	0.992 8	0.999 8
	3	0.966 5	0.985 2	0.999 7
田间收获	1	0.926 7	0.971 9	0.999 2
	2	0.896 8	0.950 2	0.999 6
	3	0.918 0	0.964 5	0.998 3

由表 5 可知,本文提出的自适应带宽核密度估计外推方法的 R^2 更加接近于 1,这表明,与另外两种方法相比,本文提出的方法与实测原始载荷的载荷循环幅值累积频次曲线的拟合度更高,更为相似,能够为程序载荷谱编制提供更加接近实际的数据支撑。

4 结论

(1)提出了一种基于四叉树算法的自适应带宽核密度估计载荷外推方法,该方法利用四叉树分割算法将载荷循环数据区域进行分割,以分割完成后的数据块为依据计算带宽,并根据数据块的密集程度对核密度估计的数据输入进行优化。

(2)以玉米收获机车架为研究对象,在砂石路面工况和田间收获工况下,采用本文提出的方法大幅降低了核密度估计的数据输入,极大提高了核密度估计的计算效率。

(3)将固定带宽、自适应带宽与本文提出的基于四叉树算法的自适应带宽核密度估计载荷外推方法进行了对比验证。结果表明,通过本文提出的载荷外推方法计算得到的载荷循环均值和幅值分布的相关系数均更加接近于 1,均方根误差更小;载荷循环幅值累积频次曲线的决定系数均大于 0.99。与传统的固定带宽以及自适应带宽的核密度估计方法相比,本文提出的方法具有更高的准确性,载荷外推结果更接近实际载荷的分布情况,能够为载荷谱的编制、零部件结构优化提供参考。

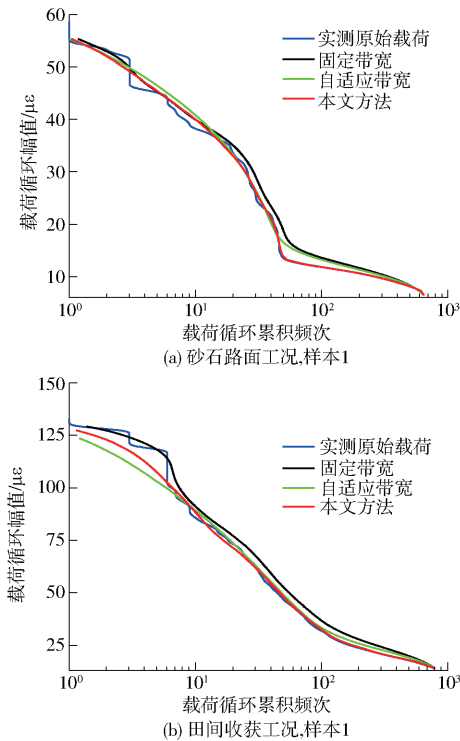


图 13 实测原始数据与 3 种不同方法的载荷循环幅值累积频次曲线

Fig. 13 Measured original data and three different methods of load cycle range cumulative frequency curves

参 考 文 献

[1] 闫祥海,周志立,贾方. 拖拉机动力输出轴动态转矩载荷谱编制与验证[J]. 农业工程学报,2019,35(19): 74-81.
YAN Xianghai, ZHOU Zhili, JIA Fang. Compilation and verification of dynamic torque load spectrum of tractor power take-off [J]. Transactions of the CSAE,2019,35(19): 74-81. (in Chinese)

[2] WANG M L, LIU X T, WANG X L, et al. Research on load-spectrum construction of automobile key parts based on Monte Carlo sampling[J]. Journal of Testing and Evaluation,2018,46(3): 1099-1110.

[3] 刘克格,闫楚良. 飞机起落架载荷谱实测与编制方法[J]. 航空学报,2011,32(5): 841-848.
LIU Kege, YAN Chuliang. Load measurement and compilation of landing gear of airplane[J]. Acta Aeronautica et Astronautica Sinica,2011,32(5): 841-848. (in Chinese)

[4] 李雯,张承宁,宋强,等. 混合动力汽车电机驱动系统二维载荷谱研究[J]. 农业机械学报,2010,41(7): 21-24.
LI Wen, ZHANG Chengning, SONG Qiang, et al. 2-D for motor propulsion system on hybrid electric vehicle[J]. Transactions of the Chinese Society for Agricultural Machinery,2010,41(7): 21-24. (in Chinese)

- [5] MA S, SUN S G, WANG B J, et al. Estimating load spectra probability distributions of train bogie frames by the diffusion-based kernel density method[J]. *International Journal of Fatigue*, 2020, 132: 105352.
- [6] WANG J X, CHEN H B, LI Y, et al. A review of the extrapolation method in load spectrum compiling[J]. *Strojniski Vestnik- Journal of Mechanical Engineering*, 2016, 62(1): 60–75.
- [7] JOHANNESSON P. Extrapolation of load histories and spectra[J]. *Fatigue & Fracture of Engineering Materials & Structures*, 2006, 29(3): 201–207.
- [8] 杨子涵, 宋正河, 尹宜勇, 等. 基于 POT 模型的大功率拖拉机传动轴载荷时域外推方法[J]. *农业工程学报*, 2019, 35(15): 40–47.
YANG Zihan, SONG Zhenghe, YIN Yiyong, et al. Time domain extrapolation method for load of drive shaft of high-power tractor based on POT model[J]. *Transactions of the CSAE*, 2019, 35(15): 40–47. (in Chinese)
- [9] 张英爽, 王国强, 王继新, 等. 工程车辆传动系载荷谱编制方法[J]. *农业工程学报*, 2011, 27(4): 179–183.
ZHANG Yingshuang, WANG Guoqiang, WANG Jixin, et al. Compilation method of power train load spectrum of engineering vehicle[J]. *Transactions of the CSAE*, 2011, 27(4): 179–183. (in Chinese)
- [10] 翟新婷, 张晓晨, 江柱锦, 等. 基于混合分布的轮式装载机半轴载荷谱编制[J]. *农业工程学报*, 2018, 34(8): 78–84.
ZHAI Xinting, ZHANG Xiaochen, JIANG Zhujin, et al. Load spectrum compiling for wheel loader semi-axle based on mixed distribution[J]. *Transactions of the CSAE*, 2018, 34(8): 78–84. (in Chinese)
- [11] GENG S L, LIU X T, YANG X B, 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, 223(14): 3707–3720.
- [12] 李凡松, 邬平波, 曾京. 基于核密度估计法的载荷历程外推研究[J]. *铁道学报*, 2017, 39(7): 25–31.
LI Fansong, WU Pingbo, ZENG Jing. Extrapolation of load histories based on kernel density method[J]. *Journal of the China Railway Society*, 2017, 39(7): 25–31. (in Chinese)
- [13] ROSENBLATT M. Remarks on some nonparametric estimates of a density function[J]. *Annals of Mathematical Statistics*, 1956, 27: 832–837.
- [14] PARZEN E. On estimation of probability density function and mode[J]. *The Annals of Mathematical Statistics*, 1962, 33(3): 1065–1076.
- [15] HE Y Y, LI H Y. Probability density forecasting of wind power using quantile regression neural network and kernel density estimation[J]. *Energy Conversion and Management*, 2018, 164: 374–384.
- [16] LU X G, QIAN X, LI X, et al. DMCM: a data-adaptive mutation clustering method to identify cancer-related mutation clusters [J]. *Bioinformatics*, 2019, 35(3): 389–397.
- [17] ZHANG J D, CHOW C Y. GeoSoCa: exploiting geographical social and categorical correlations for point-of-interest recommendations[C] // 38th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), Santiago: Assoc Computing Machinery, 2015.
- [18] 宋怀波, 阴旭强, 吴岷华, 等. 基于自适应无参核密度估计算法的运动奶牛目标检测[J/OL]. *农业机械学报*, 2019, 50(5): 196–204.
SONG Huaibo, YIN Xuqiang, WU Dihua, et al. Detection of moving cows based on adaptive kernel density estimation algorithm[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(5): 196–204. http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190523&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2019.05.023. (in Chinese)
- [19] WALTERS R M. Density estimation for statistics and data analysis[M]. London: Chapman and Hall, 1986.
- [20] SCOTT D W. Multivariate density estimation: theory, practice, and visualization[M]. New Jersey: John Wiley & Sons, 2015.
- [21] RUDEMO M. Empirical choice of histograms and kernel density estimators[J]. *Scandinavian Journal of Statistics*, 1982, 9(2): 65–78.
- [22] DELAIGLE A, GIJBELS I. Estimation of integrated squared density derivatives from a contaminated sample[J]. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 2002, 64(4): 869–886.
- [23] ZOUGAP N, ADJABI S, KOKONENDJI C C. Bayesian estimation of adaptive bandwidth matrices in multivariate kernel density estimation[J]. *Computational Statistics & Data Analysis*, 2014, 75(1): 28–38.
- [24] WAND M P, JONES M C. Kernel smoothing[M]. Boca Rtona: Chemical Rubber Company Press, 1994.
- [25] ABRAMSON, IAN S. On bandwidth variation in kernel estimates-a square root law[J]. *Annals of Statistics*, 1982, 10(4): 1217–1223.
- [26] WANG J X, WANG N X, WANG Z Y, et al. Determination of the minimum sample size for the transmission load of a wheel loader based on multi-criteria decision-making technology[J]. *Journal of Terramechanics*, 2012, 49(3–4): 147–160.
- [27] MA T. Low-complexity and efficient image coder/decoder with quad-tree search model for embedded computing platforms[J]. *IET Image Processing*, 2018, 12(2): 235–242.
- [28] FAN W Q, XIAO W S. Image segmentation based on blob analysis and quad-tree algorithm[C] // 3rd IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing: IEEE, 2018.
- [29] LAUE S, BOMAS H. Spectrum fatigue life assessment of notched specimens based on the initiation and propagation of short cracks[J]. *International Journal of Fatigue*, 2006, 28(9): 1011–1021.
- [30] BREIMAN L, MEISEL W, PURCELL E. Variable kernel estimates of multivariate densities[J]. *Technometrics*, 1977, 19(2): 135–144.