

DOI: 10.13718/j.cnki.xdzk.2023.03.020

基于便携式近红外光谱仪的柴油硫质量分数检测

郑权¹, 祝诗平¹, 齐宝华², 唐鑫¹, 黄华¹

1. 西南大学 工程技术学院, 重庆 400715; 2. 重庆航天楷世科技有限公司, 重庆 400718

摘要: 柴油中的硫是机动车排放造成大气污染的主要成分之一, 目前各国政府制定了限定柴油硫质量分数的措施, 因此研究对柴油中硫质量分数的快速检测方法具有重要意义。应用便携式近红外光谱仪采集不同硫质量分数柴油的光谱, 共采集 261 份柴油的近红外光谱数据, 利用 Kennard-Stone(K-S) 算法以 3:1 比例将样本划分为校正集和预测集, 对原始光谱在全谱区间采用去中心化、归一化、多元散射校正和 15 点 2 次平滑等多种预处理, 实验结果表明, 去中心化预处理方法对建立柴油硫质量分数的偏最小二乘回归(PLSR)模型效果最优, 其决定系数(R^2)为 0.894 和预测均方根误差(RMSEP)为 89.17, 相对分析误差(RPD)为 3.089。比较了蒙特卡罗无信息变量消除(MCUVE)和竞争性自适应加权抽样(CARS)两种波长选择算法, 最终使用 CARS 算法提取得到 35 个特征波长点进行高斯过程回归(GPR)建模的结果最佳, 其 R^2 为 0.967, 预测均方根误差为 45.378, 相对分析误差为 5.616。结果表明, 利用便携式近红外光谱技术建立柴油硫质量分数定量预测模型, 实现对柴油中硫质量分数快速和无损的近红外定量检测具有可行性。

关键词: 近红外光谱; 柴油硫质量分数; 便携式; 波长选择;

高斯过程回归

中图分类号: O433.1; U473.1⁺2

文献标志码: A

开放科学(资源服务)标识码(OSID):



文章编号: 1673-9868(2023)03-0232-07

Detection of Sulfur Content in Diesel Based on Portable Near Infrared Spectroscopy

ZHENG Quan¹, ZHU Shiping¹, QI Baohua²,
TANG Xin¹, HUANG Hua¹

1. College of Engineering and Technology, Southwest University, Chongqing 400715, China;

2. Chongqing Hangtian Kaishi Co. Ltd, Chongqing 400718, China

Abstract: Air pollution is a serious problem, of which the emissions of motor vehicles are their main pollution sources. Sulfur in diesel is one of the main components of atmospheric pollutants. At present, the

收稿日期: 2022-08-01

基金项目: 重庆市自然科学基金项目(CSTC, 2008BB1091).

作者简介: 郑权, 硕士研究生, 主要从事近红外光谱研究.

通信作者: 黄华, 讲师.

measures to limit the sulfur content in diesel are formulated by governments all over the world. Therefore, the research on the rapid detection method of sulfur content in diesel is of great significance. A portable near infrared spectrometer (NIR spectrometer) was used to collect the spectra of diesel with different sulfur contents. A total of 261 diesel NIR data were obtained. Kennard stone (K-S) algorithm was used to divide the scale sample set to correction set and prediction set at a 3 : 1 ratio. The original spectra were processed by various pretreatment methods such as decentralization, normalization, multivariate scattering correction, and 15 point twice smoothing. The experimental results showed that decentralized pretreatment method was proved to be the best for establishing the partial least squares regression (PLSR) model of diesel sulfur content, with the determination coefficient (R^2) of 0.894, the root mean square error of prediction (RMSEP) of 89.17, and the relative analysis error (RPD) of 3.089. Monte Carlo uninformative variable elimination (MCUVE) and competitive adaptive weighted sampling (CARS) were compared. Finally, Gaussian process regression (GPR) modeling with 36 characteristic wavelength points extracted by cars algorithm were the best, with R^2 of 0.967, RMSEP of 45.378 and RPD of 5.616. The results show that it is feasible to establish the quantitative prediction model of sulfur content of diesel by using portable near infrared spectroscopy technology, and realize the rapid and nondestructive near infrared quantitative detection of sulfur content in diesel.

Key words: near infrared spectroscopy (NIRS); sulfur content in diesel fuel; portable; wavelength selection; gaussian process regression

柴油属于原油加工获得的一类产品,一般在(175~375) °C范围内沸腾,主要表现为含有8~24个碳原子的化合物^[1-2].柴油中的硫及硫化物在使用过程中会产生大量二氧化硫、三氧化硫等有害气体.在温度较高的情况下,这些有害气体可能会和发动机表面的某些金属物质发生一系列化学反应,不仅会使润滑油老化变质,还会造成机械表面腐蚀,严重影响发动机的寿命.同时,大量的硫燃烧会对大气造成严重的污染,也会对人体健康产生危害^[3-5].为了改善空气质量,世界许多国家都制定了环境法规,通过限制排放水平来减少SO_x的排放.因此,快速测定柴油的硫质量分数,对于柴油品质的正确评价具有重要意义.目前国内外对柴油硫质量分数检测采用的主要方法为荧光光谱,Ranzan等^[6]使用激励-发射矩阵(EEM)荧光光谱法来测量柴油硫质量分数,对硫质量分数小于10 mg/kg组的柴油进行测量,其平均绝对百分比误差(MAPE)小于0.24%,预测均方根误差为0.015.对平均硫质量分数为100 mg/kg组的柴油进行测量,MAPE小于4%,预测均方根误差为4.68.Ranzan等^[7]直接利用二维荧光光谱技术能够将柴油总硫质量分数小于10 mg/kg和柴油硫平均质量分数在100.3 mg/kg的样品组进行分类,并将单个荧光光谱成分结合多元线性模型来定量预测柴油中的硫质量分数.Aburto等^[8]提出了在测量荧光光谱之前用酶氧化对样品进行预处理,能够准确和可重复地测量超低硫柴油,其预测硫质量分数与实际硫质量分数的 R^2 为0.95.Cruz等^[9]应用电感耦合等离子体发射光谱法来测定柴油硫质量分数,其测量结果与标定测量符合度为95%.便携式近红外光谱仪具有分析速度快、对样品无损、操作简单等优点.2012年,Zamora-Rojas等^[10]应用便携式近红外光谱技术对肉类中的脂肪、水分和蛋白质质量分数进行了量化,精确值分别为0.72%,0.73%和0.66%.Paiva等^[11]通过PLS模型评估了使用超紧凑型分光光度计(MicroNIR)和傅里叶变换台式仪器(FT-NIR)对生物柴油/柴油混合物中的植物油进行量化的方法,建立了两种仪器之间的预测均方根误差相似值,证明了这种便携式设备的可行性.Correia等^[12]利用microNIR光谱数据建立了化学计量学模型并对柴油中生物柴油、硫的质量分数进行了预测.综上所述,基于便携式近红外光谱技术在燃油质量控制方面具有优异的性能.本文应用便携式近红外光谱仪采集不同硫质量分数的柴油的近红外光谱数据,采

用不同的预处理方法和谱区选择方法,建立了柴油硫质量分数定量预测模型,实现了对柴油硫质量分数快速和无损的近红外定量检测。

1 材料和方法

1.1 样品制备

实验使用了 5 种不同硫质量分数的柴油样品,其编号分别为:s10,s100,s200,s500,s1000,根据《轻质烃及发动机燃料和其他油品的总硫含量测定法》(SH/T 0689-2000)^[13]对样品中的硫质量分数进行测定,其硫质量分数分别为:10.3,104.0,209.0,524.0,1 038.0 mg/kg.为了扩大柴油硫质量分数的浓度范围,将上述 5 个样品按已制定好的方案进行配制.取样采用移液枪,利用分析天平(岛津 AUW220)进行配制;然后用磁力搅拌器(LAB FISH 的 MS5)在 1 000 r/min 下将样品搅拌 10 min,以便混合均匀,之后装瓶贴上样本标签,一共获得了 261 份不同硫质量分数的柴油样品,其质量分数范围为 10.3~1 038.0 mg/kg.样品统计结果如表 1 所示.

表 1 柴油样品数据统计结果

样品名称	最小值	最大值	平均值	标准差
硫质量分数/(mg·kg ⁻¹)	10.3	1 038.0	387.2	273.8

1.2 光谱采集

采用天津九光科技发展责任有限公司的便携式近红外光谱仪(NIR spectrometer),如图 1 所示,对每份样品进行光谱采集,采集条件为:仪器数据采集软件为 Milk-new,近红外光谱波长范围为 900~1 700 nm,积分时间为 2 ms,波长间隔为 8 nm,光谱点数为 101,室内环境温度为 25 °C.261 份柴油硫质量分数原始近红外光谱如图 2 所示.



图 1 便携式近红外光谱仪

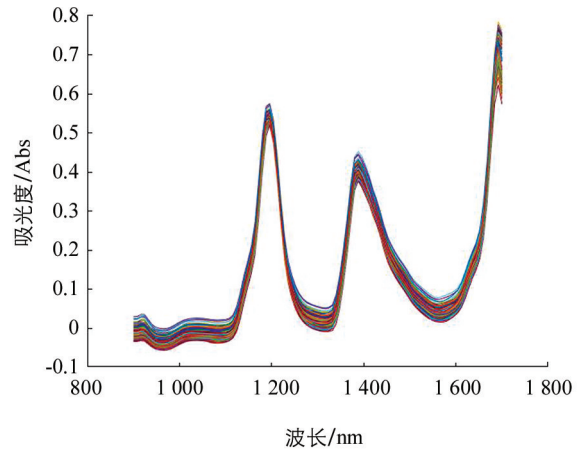


图 2 261 份柴油原始光谱

1.3 实验算法原理

1.3.1 基于偏最小二乘法(PLS)的近红外光谱定量分析方法

PLS 分析是一种多元统计数据分析法^[14].利用 PLS 建立光谱矩阵(\mathbf{X})与浓度矩阵(\mathbf{Y})之间的关系,来预测具有最小潜在变量的一组样品的浓度.首先将包括光谱矩阵和浓度矩阵的数据集分成校正集和预测集,PLS 对光谱矩阵和浓度矩阵进行分解,其模型为

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}_X = \sum_{k=1}^f t_k p_k^T + \mathbf{E}_X \quad (1)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{E}_Y = \sum_{k=1}^f u_k q_k^T + \mathbf{E}_Y \quad (2)$$

式中, t_k 为光谱矩阵的第 k 个主因子的得分; p_k 为吸光度矩阵的第 k 个主因子的载荷; u_k 为浓度矩阵的第 k 个主因子的得分; q_k 为浓度矩阵的第 k 个主因子的载荷; f 为主因子数. 即: \mathbf{T} 和 \mathbf{U} 分别为 \mathbf{X} 矩阵和 \mathbf{Y} 矩阵的得分矩阵, \mathbf{P} 和 \mathbf{Q} 分别为 \mathbf{X} 矩阵和 \mathbf{Y} 矩阵的载荷矩阵, \mathbf{E}_X 和 \mathbf{E}_Y 分别为 \mathbf{X} 矩阵和 \mathbf{Y} 矩阵的 PLS 拟合残差矩阵.

然后将 \mathbf{T} 和 \mathbf{U} 做线性回归:

$$\mathbf{U} = \mathbf{T}\mathbf{B} \quad (3)$$

$$\mathbf{B} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{Y} \quad (4)$$

在预测时, 首先根据 \mathbf{P} 求出未知样品的光谱矩阵 $\mathbf{X}_{\text{未知}}$ 的得分 $\mathbf{T}_{\text{未知}}$, 然后由下式得到浓度矩阵的预测值:

$$\mathbf{Y}_{\text{未知}} = \mathbf{T}_{\text{未知}}\mathbf{B}\mathbf{Q} \quad (5)$$

1.3.2 基于高斯过程回归(GPR)的近红外光谱定量分析方法

高斯过程回归^[15]是基于贝叶斯方法非参数概念模型, 任意有限数量随机变量的组合都具有联合高斯分布, GPR 的表达式为

$$y = f(x) + \varepsilon \quad (6)$$

其中: x 为满足多元高斯分布的 N 组独立的学习样本, ε 为独立同分布的噪声变量, 其服从:

$$\varepsilon \approx N(0, \delta_n^2) \quad (7)$$

在高斯过程回归模型中, 回归函数 $f(X)$ 为一个高斯过程, 由其均值函数 $m(X)$ 和协方差函数 $k(X, X')$ 来描述. 如公式(8)所示:

$$f(X) \sim GP(m(X), k(X, X')) \quad (8)$$

其中:

$$m(X) = E[f(x)] \quad (9)$$

$$k(X, X') = E[(f(X) - m(X))(f(X') - m(X'))^T] \quad (10)$$

在 GPR 中, 我们假设 f 服从均值为 0 的高斯分布.

$$y \sim N(0, k(X, X') + \delta_n^2 I) \quad (11)$$

当给定测试值, 预测输出 y^* 与输出 y 的联合分布也服从高斯分布, 即:

$$\begin{pmatrix} y \\ y^* \end{pmatrix} \sim N\left(0, \begin{bmatrix} k(X, X) + \delta_n^2 I, & k(X, X^*) \\ k(X^*, X), & k(X^*, X^*) \end{bmatrix}\right) \quad (12)$$

在已知取得训练集 $D = \{X, y\}$ 的条件下, y^* 的后验条件分布为

$$p(y^* | X^*, X, y) = N(m(y^*), \text{cov}(y^*)) \quad (13)$$

$$m(y^*) = (k(X^*, X)[k(X, X) + \delta_n^2 I]^{-1} y \quad (14)$$

$$\text{cov}(y^*) = k(X^*, X) - k(X^*, X)[k(X, X) + \delta_n^2 I]^{-1} k(X^*, X)^T \quad (15)$$

2 结果与分析

2.1 光谱数据预处理

对于 261 份不同柴油硫质量分数的样品, 利用 Kennard-Stone(K-S)算法^[16]对不同硫质量分数柴油的数据进行划分, 按接近 3:1 分配, 即校正集和预测集分别包含 174 份和 87 份柴油硫质量分数样品.

在原始光谱中, 除了化学信息外, 光谱还包含其他不相关或不需要的信息, 如电噪声、基质背景、杂散光、散射效应等. 在用化学计量学方法建立校准模型时, 利用预处理方法消除光谱数据中的无关信息和噪声是非常必要的. 分别对原始光谱采用去中心化、归一化、多元散射校正、15 点 2 次平滑、一阶微分、二阶微分等预处理方法, 然后运用 PLS 方法进行建模, 其结果如表 2 所示.

表 2 不同预处理方法柴油硫质量分数预测结果

预处理方法	校正集		预测集		相对分析误差 (RPD)
	R^2	RMSEC 值	R^2	RMSEP 值	
无预处理	0.878	95.51	0.881	93.69	2.903
去中心化	0.894	88.43	0.894	89.17	3.089
归一化	0.901	86.26	0.893	88.84	3.063
多元散射校正	0.901	86.30	0.893	88.95	3.059
15 点 2 次平滑	0.776	128.79	0.760	134.19	2.116
一阶微分+15 点 2 次平滑	0.856	103.11	0.827	113.95	2.641
二阶微分+15 点 2 次平滑	0.855	103.43	0.846	107.59	2.630

由表 2 知: 经过 15 点 2 次平滑, 一阶微分+15 点 2 次平滑, 二阶微分+15 点 2 次平滑的预处理后, 建的模型质量不佳. 而经过去中心化、归一化、多元散射校正等预处理方法后, 模型的预测能力有明显的提升, 其中去中心化结果最优, 其 R^2 由 0.881 提升到 0.893, RMSEP 值由 93.69 降到 89.17, RPD 值由 2.903 提升到 3.089, 故选择去中心化的光谱数据预处理方法, 其预处理后的光谱数据如图 3 所示.

2.2 基于 CARS-PLS 的柴油硫质量分数预测模型

波长(变量)选择可以提高模型的预测能力, 简化模型, 将去中心化预处理后的柴油硫质量分数光谱数据用竞争性自适应加权抽样(CARS)^[17-18]、蒙特卡罗无信息变量消除(MCUVE)^[19]特征波段筛选方法对 101 个波长点进行提取, 将与预测柴油硫质量分数无关的特征变量进行去除, CARS 算法共去除了 66 个无关信息波长点, 保留信息变量 35 个, 最终所选取的波段如图 4 所示. 用全部波段的 34.65% 建立柴油硫质量分数的定量分析模型, 预测结果如图 5 所示, 其预测模型 R^2 的值为 0.939, RMSEP 的值为 61.67, RPD 的值为 4.084. MCUVE 算法去除了 61 个无关信息波长点, 保留信息波长 40 个, 其选取的波段如图 4 所示, 用全部波段的 39.60% 建立柴油硫质量分数的定量分析模型, 预测结果如图 5 所示, 其预测模型 R^2 的值为 0.932, RMSEP 的值为 65.23, RPD 的值为 3.895.

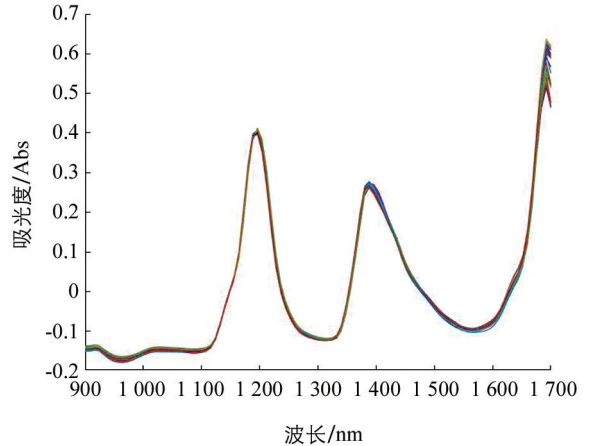


图 3 柴油硫质量分数预处理光谱图

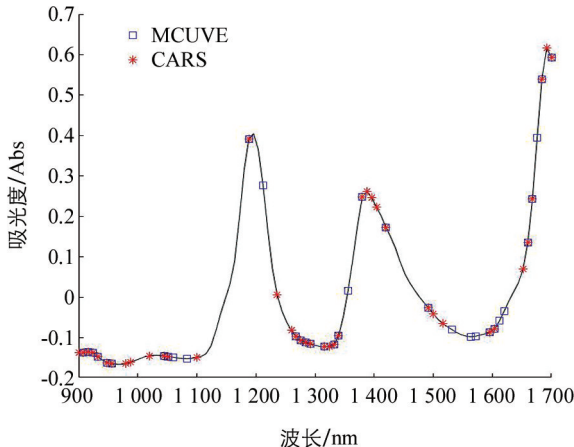


图 4 柴油硫质量分数变量筛选图

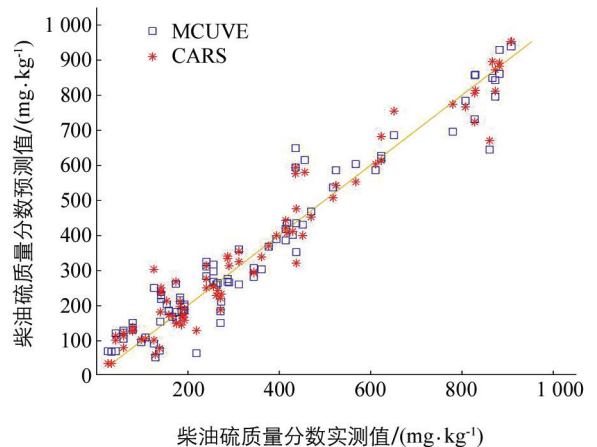


图 5 基于不同变量选择的柴油硫质量分数预测图

将建立的 MCUVE 与 CARS 模型与整个波段回归模型进行比较,由表 3 可知,经波长筛选模型质量明显提高,其中 CARS 模型质量最优且选择的变量数更少,故采用 CARS 算法选取的变量进行建模分析。

表 3 不同波长筛选方法的结果比较

波长筛选方法	变量数	R^2	RMSEP 值	RPD 值
None	101	0.881	93.69	2.903
CARS	35	0.939	61.67	4.084
MCUVE	40	0.932	65.23	3.895

2.3 基于高斯过程回归的柴油硫质量分数预测模型

将 CARS 算法选取的 35 个特征波长点,利用高斯过程回归作为建模算法,按照上述相同的方法对全部样品进行校正集和验证集划分,用校正集中的 174 个样品建立 GPR 模型,对预测集中的 87 个样品进行验证.应用 GPR 建立柴油硫质量分数定量模型,本实验中核函数采用 Matern5/2 核,基函数为常量,其预测模型的 R^2 为 0.967、RMSEP 值为 45.378、RPD 值为 5.616.其预测结果如图 6 所示。

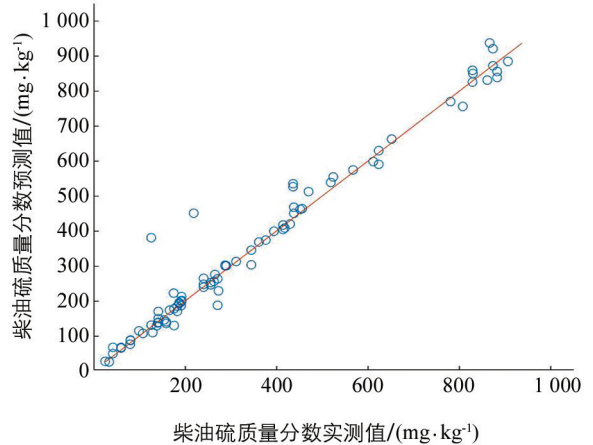


图 6 基于高斯过程回归的柴油硫质量分数预测图

3 结论

本文利用不同柴油硫质量分数在 900~1 700 nm

波段范围内的便携式近红外光谱信息,在全谱区选择不同的预处理方法,建立了柴油硫质量分数的近红外光谱定量分析建模.实验结果表明,去中心化光谱预处理方法表现最优,其 R^2 为 0.894, RMSEP 值为 89.17, RPD 值为 3.089.采用 MCUVE 与 CARS 方法选取特征波长来提升模型预测能力,所得结果明显优于全谱建模结果.经过 CARS 特征波长选择后,从 101 个波长点中筛选出 35 个波长点,用于高斯过程回归建立柴油硫质量分数的预测模型,其预测模型的 R^2 为 0.967, RMSEP 值为 45.378, RPD 值为 5.616.实验结果表明,基于便携式近红外光谱仪的柴油硫质量分数检测具有可行性。

参考文献:

- [1] MA Y Y, LIU P P, HU X H, et al. Fast Determination of Trace Elemental Sulfur Residue in Rare Earth Sulfide Samples by Inductively Coupled Plasma Optical Emission Spectrometry [J]. Atomic Spectroscopy, 2020, 41(2): 81-86.
- [2] HRADECKÁ I, VELVARSKÁ R, JAKLOVÁ K D, et al. Rapid Determination of Diesel Fuel Properties by Near-Infrared Spectroscopy [J]. Infrared Physics and Technology, 2021, 119: 103933.
- [3] STANISLAUS A, MARAFI A, RANA M S. Recent Advances in the Science and Technology of Ultra Low Sulfur Diesel (ULSD) Production [J]. Catalysis Today, 2010, 153(1-2): 1-68.
- [4] CAMPOS-MARTIN J M, CAPEL-SANCHEZ M C, PEREZ-PRESAS P, et al. Oxidative Processes of Desulfurization of Liquid Fuels [J]. Journal of Chemical Technology & Biotechnology, 2010, 85(7): 879-890.
- [5] SHEN F X, LI X B. Effects of Fuel Types and Fuel Sulfur Content on the Characteristics of Particulate Emissions in Marine Low-Speed Diesel Engine [J]. Ecology Environment & Conservation, 2020, 27(30): 37229-37236.
- [6] RANZAN L, TRIERWEILER L F, TRIERWEILER J O. Prediction of Sulfur Content in Diesel Fuel Using Fluorescence Spectroscopy and a Hybrid ant Colony-Tabu Search Algorithm with Polynomial Bases Expansion [J]. Chemometrics and Intelligent Laboratory Systems, 2020, 206: 104161.
- [7] RANZAN L, RANZAN C, TRIERWEILER L F, et al. Classification of Diesel Fuel Using Two-Dimensional Fluores-

- cence Spectroscopy [J]. *Energy & Fuels*, 2017, 31(9): 8942-8950.
- [8] ABURTO P, ZUNIGA K, CAMPOS-TERAN J, et al. Quantitative Analysis of Sulfur in Diesel by Enzymatic Oxidation, Steady-State Fluorescence, and Linear Regression Analysis [J]. *Energy & Fuels*, 2014, 28(1): 403-408.
- [9] CRUZ S M, TIRK P, DALLA NORA F M, et al. Feasibility of Sulfur Determination in Diesel Oil by Inductively Coupled Plasma Optical Emission Spectrometry After Microwave-Induced Combustion Using Flame Retardant [J]. *Fuel*, 2015, 160: 108-113.
- [10] ZAMORA-ROJAS E, PÉREZ-MARÍN D, PEDRO-SANZ E D, et al. Handheld NIRS Analysis for Routine Meat Quality Control: Database Transfer from At-Line Instruments [J]. *Chemometrics and Intelligent Laboratory Systems*, 2012, 114: 30-35.
- [11] PAIVA E M, ROHWEDDER J J R, PASQUINI C, et al. Quantification of Biodiesel and Adulteration with Vegetable Oils in Diesel/Biodiesel Blends Using Portable Near-Infrared Spectrometer [J]. *Fuel*, 2015, 160: 57-63.
- [12] CORREIA R M, TOSATO F, DOMINGOS E, et al. Portable Near Infrared Spectroscopy Applied to Quality Control of Brazilian Coffee [J]. *Talanta*, 2018, 176: 59-68.
- [13] 国家石油和化学工业局. 轻质烃及发动机燃料和其他油品的总硫含量测定法(紫外荧光法): SH/T 0689-2000 [S]. 北京: 中国标准出版社, 2000.
- [14] 祝诗平, 梁晶, 屠大伟, 等. 基于近红外光谱与 DPLS 的泔水油快速鉴别方法 [J]. *西南大学学报(自然科学版)*, 2012, 34(5): 1-6.
- [15] SEEGER M. Gaussian Processes for Machine Learning [J]. *International Journal of Neural Systems*, 2008, 14(2): 69-106.
- [16] ZHANG L, LI G, SUN M X, et al. Kennard-Stone Combined with Least Square Support Vector Machine Method for Noncontact Discriminating Human Blood Species [J]. *Infrared Physics & Technology*, 2017, 86: 116-119.
- [17] HONG Y S, CHEN S C, CHEN Y Y, et al. Comparing Laboratory and Airborne Hyperspectral Data for the Estimation and Mapping of Topsoil Organic Carbon: Feature Selection Coupled with Random Forest [J]. *Soil and Tillage Research*, 2020, 199: 104589.
- [18] 吴习宇, 曾凯芳, 郭启高, 等. 基于高光谱成像的枇杷果实品质检测 [J]. *西南大学学报(自然科学版)*, 2022, 44(9): 52-60.
- [19] 郝勇, 孙旭东, 潘圆媛, 等. 蒙特卡罗无信息变量消除方法用于近红外光谱预测果品硬度和表面色泽的研究 [J]. *光谱学与光谱分析*, 2011, 31(5): 1225-1229.

责任编辑 汤振金