

激光诱导击穿光谱结合随机森林的稀土矿石中钪元素定量分析

周嘉俊¹, 李茂刚¹, 张天龙^{1*}, 汤宏胜¹, 李华^{1,2}¹合成与天然功能分子教育部重点实验室, 西北大学化学与材料科学学院, 陕西 西安 710127;²西安石油大学化学化工学院, 陕西 西安 710065

摘要 钪(Sc)被广泛用于固体氧化燃料电池、陶瓷材料、催化剂与轻质高温合金等的制造,是不可或缺的重要战略资源,稀土矿石中Sc元素的定量分析对于稀土矿的勘探、开采具有重要意义。笔者提出了一种基于激光诱导击穿光谱(LIBS)结合随机森林(RF)算法的稀土矿石样品中Sc元素定量分析的方法。首先,考察了不同光谱预处理方法对RF校正模型预测性能的影响;然后,利用变量重要性测量(VIM)进行RF校正模型输入变量的选择与优化。为了进一步验证VIM-RF模型的预测性能,将其与标准曲线法、偏最小二乘(PLS)以及基于波段选择的RF模型进行了比较。最后,在最优化的光谱预处理(WT)和VIM阈值(阈值为0.016)等条件下,建立了基于小波变换结合VIM的RF校正模型。结果表明,VIM-RF校正模型表现出了良好的预测性能: R_{cv}^2 为0.9981, $RMSE_{cv}$ 为0.0430 mg/kg, MRE_{cv} 为0.0047, R_p^2 为0.9993, $RMSE_p$ 为0.4964 mg/kg, MRE_p 为0.0481。因此,LIBS技术结合RF算法可以有效实现稀土矿石中稀土元素Sc的定量分析,可为稀土矿石的品位分析与精准开采提供借鉴。

关键词 光谱学; 激光诱导击穿光谱; 稀土矿石; 定量分析; 随机森林; 变量选择

中图分类号 O657.31

文献标志码 A

DOI: 10.3788/CJL230671

1 引言

稀土元素具有独特的磁性、发光和催化等性能,在国防、冶金、农业、玻璃、石油化工、陶瓷等领域得到了广泛应用,特别是在电子产品、汽车工业、航空航天等高新产业有着不可替代的作用,被誉为“工业维生素”^[1]。在稀土元素中,钪(Sc)凭借其良好的物理化学属性常被用来制造半导体器件、计算机磁芯存储器、导弹、轻质高温合金和固体氧化燃料等,是至关重要的稀缺战略资源^[2]。自然界中的Sc主要存在于稀土矿、铝土矿、铀矿和锰铁矿等矿物中^[3]。然而,其稀缺性(Sc在地壳中的质量分数约占0.0005%)以及提取Sc的高昂成本导致目前其在稀土市场上的供应比较紧张^[4]。因此,稀土矿中Sc元素的定量分析对于稀土产业勘探、采矿、工业原料提取和产品质量检查等具有重要意义。

激光诱导击穿光谱(LIBS)是一种通过激光烧蚀样品表面产生等离子体进而对等离子体的原子发射光谱进行检测的技术,通过分析激光烧蚀样品表面产生的特征发射光谱可以确定样品的元素组成以及待分析元素的含量。与其他技术相比,LIBS具有现场快速检测、远程检测、实时在线检测和微损检测等特点,而且

不需要制备样品,可以同时进行多元素分析,已被广泛应用于冶金分析、地质勘探、环境监测、太空探索^[5-8]等领域。特别是在地质勘探领域,LIBS可以获得矿物中元素的成分信息并支持矿物的分类与鉴别^[9],已被成功应用于铝土矿、铁矿石、锰矿石、铬矿石等矿石的定性及定量分析^[9-10]。近年来,LIBS在稀土元素分析中的应用越来越多,如:刘小亮等^[11]将LIBS应用于稀土矿区土壤中Sm含量的测定,该技术对Sm元素两条分析线的预测相对误差分别为0.97%和0.26%。Devangad等^[12]将LIBS技术应用于磷酸盐玻璃中稀土元素Sm、Tm和Yb的定量分析,三种元素的相关系数 R^2 分别为0.9980、0.9930、0.9780。上述研究表明,LIBS是一种可行的稀土元素分析方法。

在LIBS检测过程中,光谱的稳定性会受到样品基体效应、样品表面不均匀性及检测环境变化等的影响。因此,传统的单变量分析方法在LIBS数据处理时无法获得理想的分析结果。近年来,多元校正方法,如随机森林(RF)、偏最小二乘(PLS)、人工神经网络(ANN)、支持向量机(SVM)等,已被应用于提高LIBS光谱分析的精确度及准确度,并取得了较好的分析结果^[13-17]。RF是一种基于回归树的集成学习方法,具有不易过拟合、准确率高、模

收稿日期: 2023-03-30; 修回日期: 2023-05-18; 录用日期: 2023-05-24; 网络首发日期: 2023-07-04

通信作者: *tlzhang@nwu.edu.cn

型参数优化简单等优势,已被广泛应用于 LIBS 光谱分析。Li 等^[18]将 LIBS 与 RF 算法结合并应用于含油污泥样本污染风险因子的预测,预测集的 R^2 和均方根误差 (RMSE_p) 分别为 0.9681 和 0.6009。李铸等^[19]将 LIBS 与 RF 算法结合并应用于高碳铬轴承钢 (GCr15 钢) 硬度的定性分析与识别,对于未知硬度的样品,RF 模型的平均预测准确率可达 96%。LIBS 光谱数据通常包含着大量的变量数据,其中包括待分析元素的特征信息以及大量的冗余变量。因此,在构建基于 LIBS 数据的分析模型前需要进行变量筛选。变量重要性测量 (VIM) 是一种基于 RF 算法的变量选择方法,已被成功应用于 LIBS 光谱数据的变量筛选上。它计算所有变量对模型预测性能的贡献度,即变量的重要性,然后通过不断优化变量的重要性阈值来筛选、保留对模型预测性能贡献度高的变量,从而提高模型的预测性能并缩短建模时间。丁宇等^[20]将 VIM 应用于铝合金中 Mg 元素的定量分析,结果表明,基于 VIM 筛选所构建的 RF 校正模型的预测性能优于基于全谱构建的 RF 校正模型的预测性能,其预测集 R_p^2 为 0.9461, RMSE_p 为 0.9534。

本文拟探究 LIBS 技术结合 RF 算法用于稀土矿石中 Sc 元素定量分析的可行性。首先,采集稀土矿石标准样品的 LIBS 光谱,并基于美国国家标准与技术研究院 (NIST) 的原子光谱数据库对稀土矿石样本中 Sc 元素的 LIBS 发射谱线进行识别;其次,考察不同光谱预处理方法 (标准正态变换、归一化、一阶导数、二阶导数和小波变换) 对 RF 校正模型预测性能的影响;然后,利用 VIM 进行 RF 校正模型输入变量的选择与优化,并基于留一法交叉验证对 VIM 的阈值进行优化。为了进一步验证该模型的预测性能,将其与标准曲线法、偏最小二乘 (PLS) 以及基于波段选择的 RF 模型进行了比较。最后,在最优化的光谱预处理和 VIM 阈值等条件下,建立针对稀土矿石中 Sc 定量分析的 RF 校正模型。

2 实验过程

2.1 样品制备

6 个稀土矿石标准样品购于武汉睿辰标物科技有限公司,稀土矿石样品中 Sc 含量的参考值见表 1。由于稀土矿石标准样品为粉末状,为了防止激光冲散粉末从而导致光谱信号不稳定,在采集 LIBS 光谱前,使用 PC-24 压片机将粉末样品在 20 MPa 压力下持续压制 5 min,制成薄片 (6 个稀土矿石样品各制备 1 个薄片)。

2.2 LIBS 光谱采集

实验中稀土矿石样品 LIBS 光谱的采集装置主要由 Nd:YAG 激光器 (Nano, Litron, UK, 波长为 1064 nm, 脉冲能量为 80 mJ, 激光频率为 5 Hz)、样品台、光

表 1 稀土矿石样品中 Sc 元素含量的参考值

Table 1 Reference value of Sc content in rare earth ore samples

Sample No.	Sample name	Mass fraction reference value / 10^{-6}
#1	GBW07158	11.30
#2	GBW07159	9.50
#3	GBW07160	8.90
#4	GBW07161	11.80
#5	GBW07187	10.10
#6	GBW07188	7.20

路系统、光纤、光谱仪 (Aryelle 200, LTB, Germany, 光谱范围为 190~840 nm, 分辨率为 0.015 nm, 门宽为 10 μ s)、增强型电荷耦合检测器 (iStar 340t, Andor Technology, UK) 及计算机系统等构成。在检测过程中,将稀土样品放置于 X-Y-Z 手动可调样品台上,在实验室环境下进行 LIBS 光谱的采集,延迟时间为 1 μ s。

为了提高 LIBS 光谱的稳定性,降低样品不均匀性对稀土矿石样品 LIBS 光谱准确度和精度的影响,每个压片随机选择 16 个采样点进行 LIBS 光谱的采集 (点位尽可能在样品表面上均匀分布),每个采样点通过累计进行 5 次激光脉冲作用后求 5 张光谱平均值的方式获得 LIBS 光谱,以降低激光脉冲波动的影响。6 个稀土矿石样品共采集 96 条光谱数据 (每个样品采集 16 条光谱)。构建 RF 模型时,选取 #3、#4、#5 及 #6 样品作为校正集,用于构建 RF 校正模型;选取 #1 及 #2 样品作为预测集,用于验证 RF 模型的预测性能。

2.3 随机森林

RF 是 Breiman^[21]在 2001 年提出的一种将多个决策树集成来构建预测模型的统计学习理论,具有较好的模型稳定性,常被用于处理分类和回归问题。RF 在模型构建过程中有两个重要的参数,即决策树节点数 (m_{try}) 和决策树数 (n_{tree})。本文的研究思路及 RF 建模流程如图 1 所示。RF 建模过程可以简要概括为: 1) 在原始数据集 M 中通过 Bootstrap 重采样策略有放回地取出 n 个数据作为校正集; 2) 从每个训练数据集中随机选择 m 个特征,并不断地根据这 m 个特征建立 n 棵决策树; 3) 在回归问题中,RF 模型的最终预测结果是每棵决策树结果的平均值。在本研究中,将稀土矿石样品的 LIBS 光谱强度与相应的 Sc 元素含量相对应,用于构建 RF 校正模型;同时,将所有决策树预测结果的平均值作为 Sc 含量的最终预测结果。建模时决策树节点数和决策树数使用默认值,即 $n_{\text{tree}}=500$ 和 $m_{\text{try}}=M/3$,其中 M 为稀土矿石样品 LIBS 光谱的变量数, $M=43399$ ^[18]。在模型优化过程中,采用 R^2 、RMSE 和平均相对误差 (MRE) 作为模型性能的评价指标。

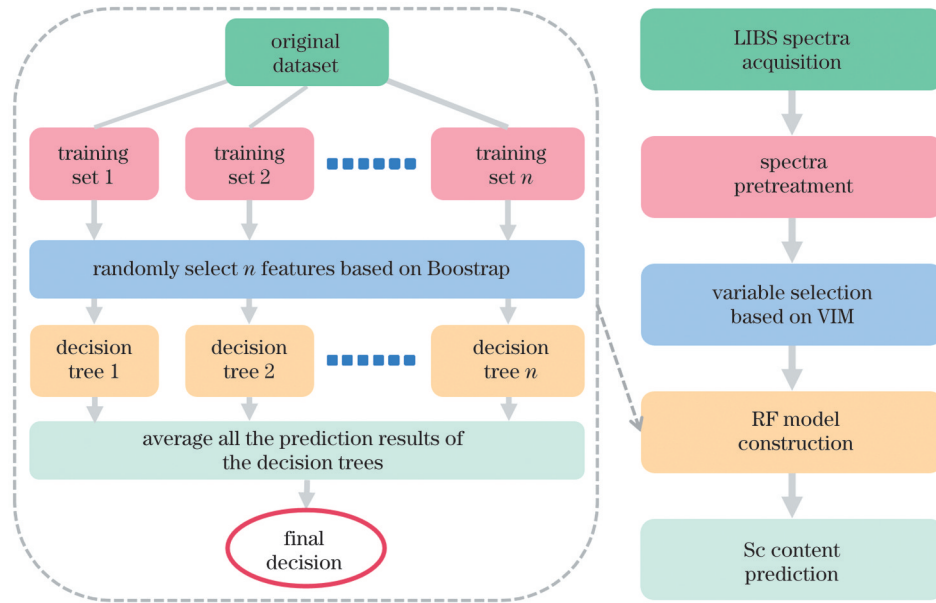


图 1 RF 建模流程图及研究简图

Fig. 1 RF modeling flow chart and research schematic

3 分析与讨论

3.1 稀土矿石样品的 LIBS 光谱分析

在 RF 校正模型构建之前,需要对 Sc 的 LIBS 光谱特征谱线进行识别。本研究根据 NIST 的原子光谱数据库对稀土矿石样品 LIBS 光谱中 Sc 元素的特征谱线

进行识别。图 2 为 #4 稀土矿石样品的 LIBS 光谱图。从图 2 可以看出,LIBS 光谱可以识别 Si、Al、Fe、Mg、K、Ti、Y 与 Sc 等的特征谱线,其中 Sc 在稀土矿石样品中的含量相对较低,所识别的 Sc 元素的特征谱线主要集中在 300~400 nm 波段内。为了保证校正模型的稳定性与普适性,在后续建模分析过程中采用 190~

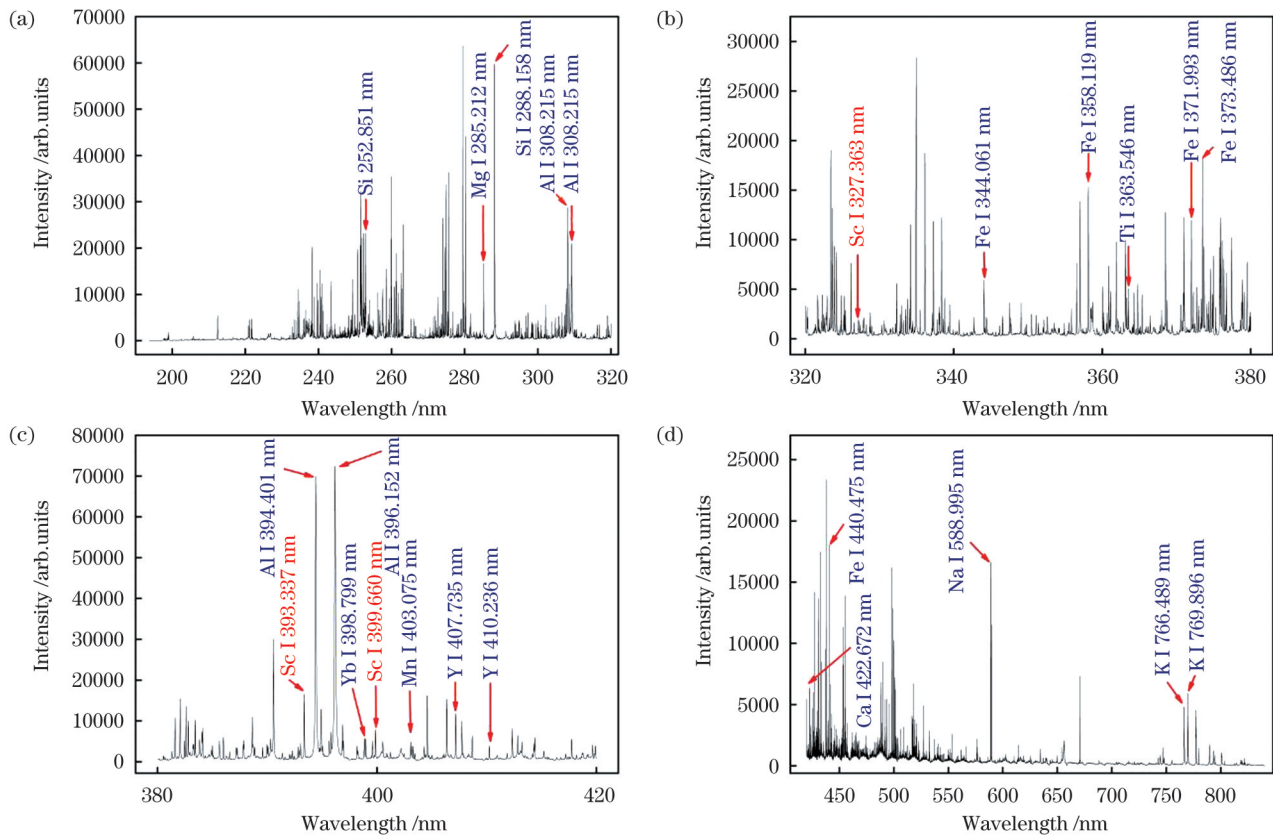


图 2 #4 稀土矿石样品的 LIBS 光谱图。(a) 200~320 nm; (b) 320~380 nm; (c) 380~420 nm; (d) 420~840 nm

Fig. 2 LIBS spectrum of #4 rare-earth ore sample. (a) 200~320 nm; (b) 320~380 nm; (c) 380~420 nm; (d) 420~840 nm

840 nm 全波段进行建模。在稀土矿石样品 LIBS 光谱中,识别出的 Sc 元素的特征线普遍较弱,这是由基体效应、噪声干扰、实验过程中的测量误差以及 Sc 元素在稀土矿石样品中的含量相对较低造成的。如果不对 LIBS 光谱进行处理,不仅会影响元素谱峰的认识,还会影响定量分析模型的准确性。因此,需要对样品的 LIBS 原始光谱进行预处理,尽可能地消除干扰因素,从而最大限度地提取 LIBS 光谱中的有效信息。

3.2 预处理方法的选择与优化

LIBS 光谱常用的预处理方法主要包括标准正态变换(SNV)、归一化(Nor)、一阶导数(D1st)、二阶导数(D2nd)和小波变换(WT)等。其中:SNV 常用于校正由样品颗粒大小不同及颗粒分布不均匀等导致的散射所引起的光谱误差;Nor 常用于校正由光谱数据尺度差异过大带来的误差;D1st 与 D2nd 常用于降低仪器背景或基线漂移对信号的影响;WT 可同时对信号的时间和频率进行分析,常用于光谱数据的平滑滤波、去除背景等。从图 2 中可以看出稀土矿石样品 LIBS 光谱中仍然存在特征谱线强度较弱及噪声等问题,因此,对光谱进行预处理是非常有必要的。

本研究将探讨上述 5 种预处理方法对稀土矿石样品 LIBS 光谱的影响。首先分别采用上述 5 种方法对稀土矿石的 LIBS 光谱进行预处理;然后,将经过预处理优化的 LIBS 光谱作为输入变量,构建 RF 校正模型,并通过对比不同 RF 校正模型的性能来寻找最优的预处理方法。在光谱预处理中,Nor 将 LIBS 光谱的强度最大值作为参考进行归一化;D1st 与 D2nd 中可选择的平滑点数为 3、5、7、9、11;WT 中可选择的小波基函数为 Coif1、Coif2、Coif3、Coif4、Coif5、Db2、Db3、Db4、Db5,分解层数为 1、2、3、4、5。本研究采用留一法交叉验证对上述预处理方法的参数进行优化。图 3 为 D1st 与 D2nd 在不同平滑点数优化下对 RF 校正模型留一法交叉验证结果与校正集内部验证结果的影响。由图 3 可以看出,在平滑点数优化过程中,D1st-RF 和 D2nd-RF 校正模型在平滑点数分别为 7 和 5 时获得了最好的性能。图 4 为 WT 在不同小波基函数与分解层数下对 RF 校正模型留一法交叉验证结果与校正集内部验证结果的影响。由图 4 可以看出,在使用 WT 法时,WT-RF 校正模型在小波基函数选择为 Coif5、分解层数选择为 1 时获得了最好的性能。

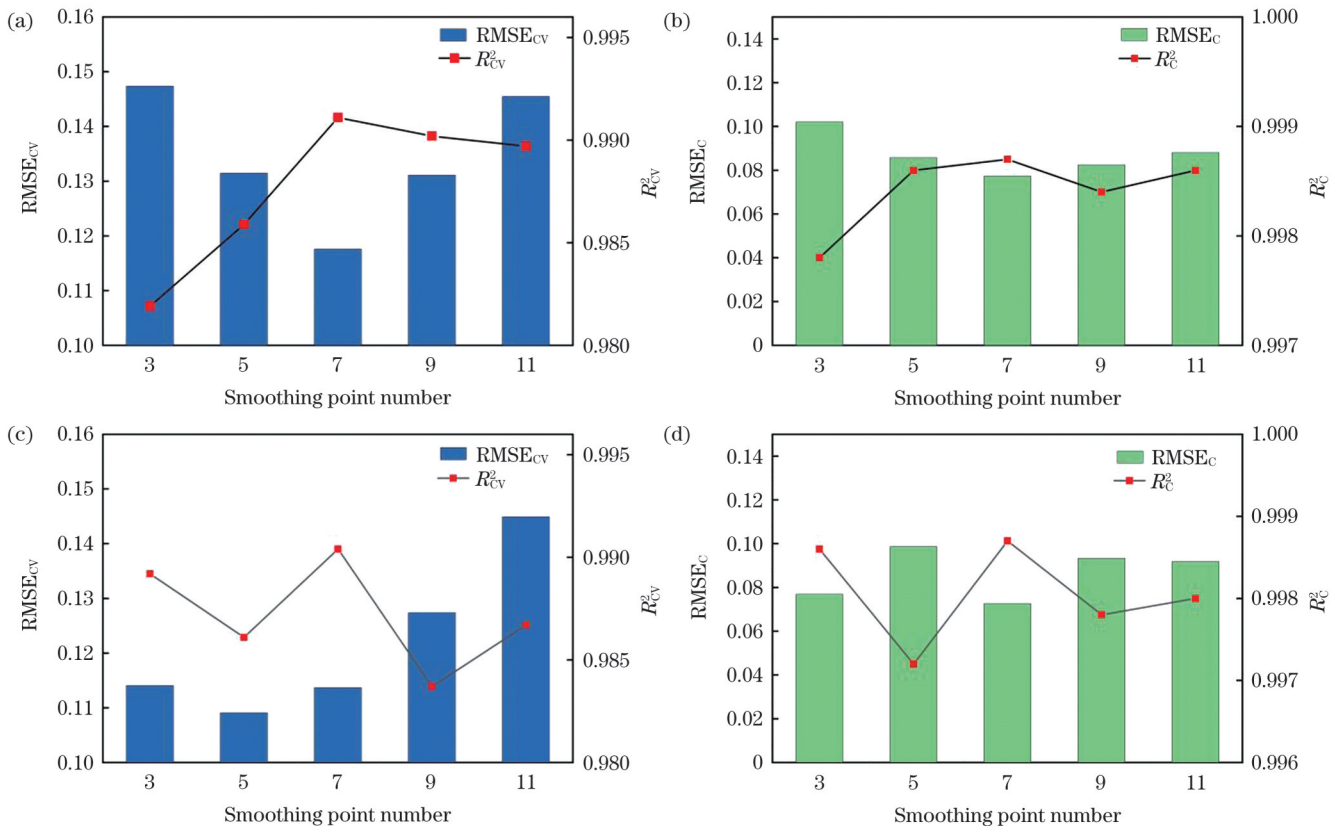


图 3 不同参数条件下 D1st-RF 与 D2nd-RF 校正模型的预测性能。(a)D1st-RF 交叉验证结果;(b)D1st-RF 内部验证结果;(c)D2nd-RF 交叉验证结果;(d)D2nd-RF 内部验证结果

Fig. 3 Predictive performance of first derivative-random forest (D1st-RF) and second derivative-random forest (D2nd-RF) calibration models under different parameters. (a) D1st-RF cross validation; (b) D1st-RF internal validation; (c) D2nd-RF cross validation; (d) D2nd-RF internal validation

表 2 为不同光谱预处理方法下 RF 校正模型的留一法交叉验证与校正集内部验证结果。可以看出,当

使用原始光谱直接进行 RF 校正模型的构建时,RF 校正模型的预测性能为: R_{cv}^2 为 0.9940, RMSE_{cv} 为

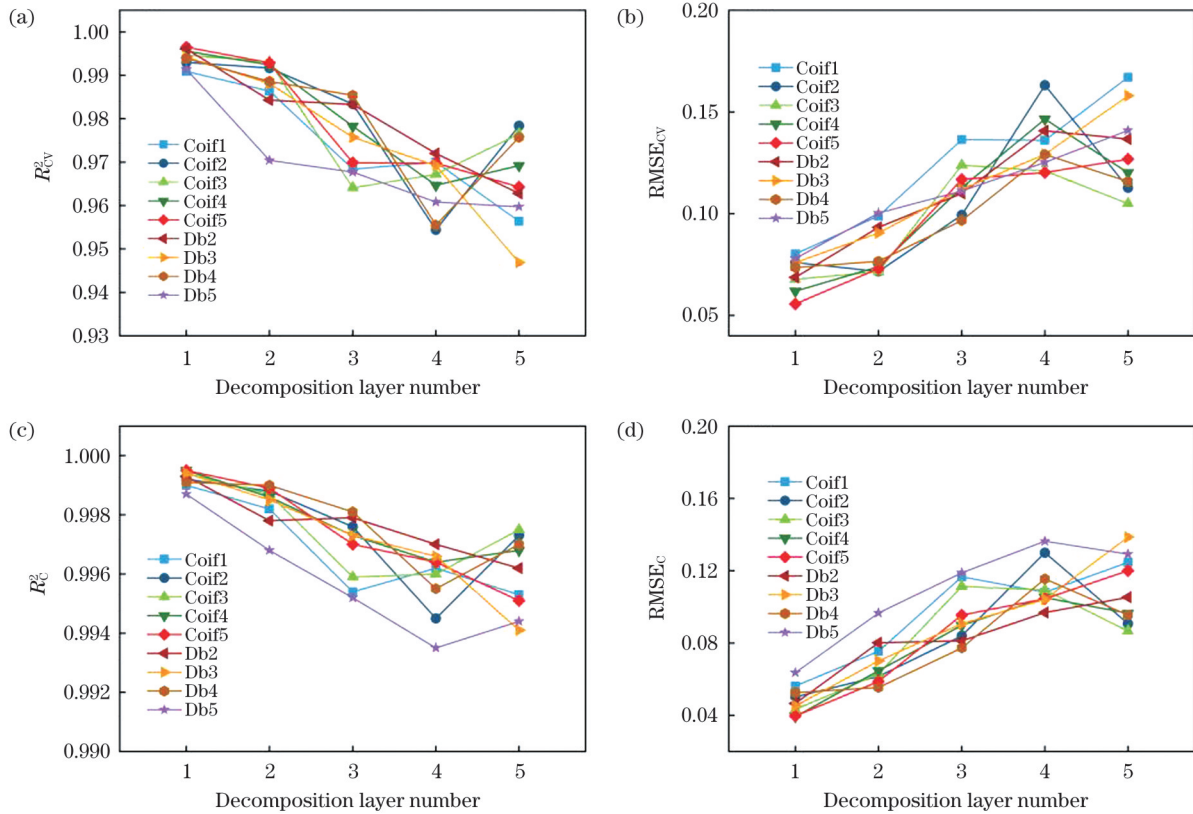


图 4 不同参数条件下 WT-RF 校正模型的预测性能。(a)(b)WT-RF 交叉验证结果;(c)(d)WT-RF 内部验证结果

Fig. 4 Predictive performance of wavelet transform-random forest (WT-RF) calibration model under different parameters. (a)(b) WT-RF cross validation; (c)(d) WT-RF internal validation

0.0893 mg/kg, MRE_{cv} 为 0.0104。使用 SNV、Nor、D1st、D2nd 和 WT 对光谱进行预处理后,校正模型的留一法交叉验证结果和校正集内部验证结果均获得了不同程度上的提升;除了 WT 以外,经 SNV、Nor、D1st 和 D2nd 预处理后,RF 校正模型的留一法交叉验证结果和校正集内部验证结果均有不同程度的下降;经 WT(小波基函数为 Coif5,分解层数为 1)处理后,RF 校正模型获得了更好的留一法交叉验证结果与校正集内部验证结果。相较于原始光谱建模,WT-RF 校正模型的交叉验证结果得到了一定改善(R_{cv}^2 从

0.9940 提升至 0.9965, $RMSE_{cv}$ 从 0.0893 mg/kg 减小为 0.0556 mg/kg, MRE_{cv} 从 0.0104 减小为 0.0060),校正集性能也得到了一定改善(R_c^2 从 0.9993 提升至 0.9995, $RMSE_c$ 从 0.0538 mg/kg 减小为 0.0397 mg/kg, MRE_c 从 0.0038 减小为 0.0024)。从图 2 中可以看出稀土矿石样品的 LIBS 光谱存在噪声和基线漂移现象,需要对 LIBS 光谱进行滤噪和背景扣除处理。综上所述可知,选择 WT(Coif5, 1)作为稀土矿石样品 LIBS 光谱预处理方法时,RF 校正模型对 Sc 元素的预测性能较好。

表 2 基于不同预处理方法的 RF 校正模型的预测性能

Table 2 Predictive performance of RF calibration model based on different pretreatment methods

Pretreatment method	Leave-one-out cross validation			Calibration set		
	R_{cv}^2	$RMSE_{cv} / (mg \cdot kg^{-1})$	MRE_{cv}	R_c^2	$RMSE_c / (mg \cdot kg^{-1})$	MRE_c
Raw	0.9940	0.0893	0.0104	0.9993	0.0538	0.0038
SNV	0.9962	0.0955	0.0103	0.9994	0.0593	0.0040
Nor	0.9875	0.1299	0.0149	0.9988	0.0684	0.0048
D1st(7)	0.9911	0.1176	0.0139	0.9987	0.0774	0.0051
D2nd(5)	0.9861	0.1091	0.0132	0.9972	0.0987	0.0051
WT(Coif5, 1)	0.9965	0.0556	0.0060	0.9995	0.0397	0.0024

3.3 输入变量的选择与优化

对稀土矿石样品的 LIBS 原始光谱进行预处理并不会减少数据的变量数,变量数依然为 43399,这些变

量不仅包括 Sc 含量定量分析所需的有效信息,还包括大量的无效或噪声信息。这些无效信息不仅会影响模型的预测精度,还会增加建模时间。因此,在模型构建

前,需要对输入变量进行筛选,以确保模型的预测性能。本研究考察了 VIM 对基于稀土矿石样品 LIBS 光谱的 RF 校正模型预测性能的影响。图 5 展示了 VIM 筛选的输入变量的重要性。

从图 5 可以看出,LIBS 光谱变量对 RF 模型的贡献是不同的,变量重要性越大,表示该变量对模型的贡献越大。当设置不同的变量重要性阈值时,可以保留高于此阈值的特征变量,同时删除低于该阈值的变量。通过优化变量重要性阈值可以获得一些具有代表性的特征变量,从而提高 RF 模型的预测性能。表 3 为基于 VIM 的 WT-RF 校正模型的预测结果对比。从表 3 可以看出:保留的变量数随着 VIM 阈值的增大而显著减少;当 VIM 阈值选择为 0.016 时,RF 校正模型的输入变量数为 795,获得了最佳的交叉验证结果与校正集内部验证结果。与未经变量选择的 RF 模型预测性能相比,VIM-RF 校正模型的交叉验证结果得到改善, R_{cv}^2 从 0.9965 提升至 0.9981, $RMSE_{cv}$ 从 0.0556 mg/kg 减小为 0.0430 mg/kg, MRE_{cv} 从 0.0060 减小为 0.0047。

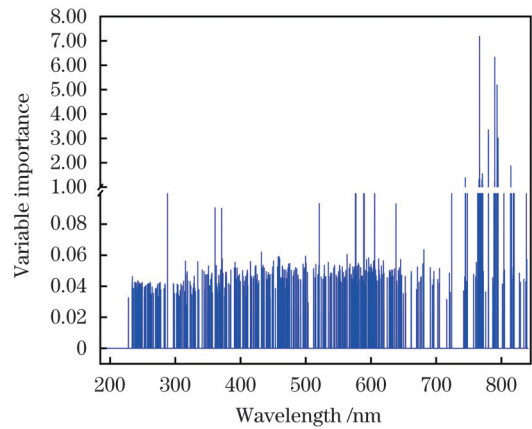


图 5 RF 校正模型输入变量的重要性

Fig. 5 Importance of the input variables of RF calibration model

变量个数的减少会使建模时间缩短。因此,选择 0.016 作为 VIM 阈值对稀土样本的 LIBS 光谱进行特征变量选择,最终保留的特征变量数目为 795。

表 3 基于 VIM 的 WT-RF 校正模型的预测性能对比

Table 3 Comparison of the predictive performance of WT-RF calibration models based on VIM

Threshold	Reserved variable quantity	Leave-one-out cross validation		
		R_{cv}^2	$RMSE_{cv} / (mg \cdot kg^{-1})$	MRE_{cv}
-	43399	0.9965	0.0556	0.0060
0	2072	0.9976	0.0467	0.0052
0.016	795	0.9981	0.0430	0.0047
0.020	794	0.9974	0.0460	0.0051
0.032	789	0.9977	0.0455	0.0050
0.040	681	0.9977	0.0453	0.0050
0.050	396	0.9974	0.0483	0.0053
0.100	285	0.9968	0.0445	0.0049
0.400	110	0.9975	0.0406	0.0044
0.800	34	0.9975	0.0378	0.0041
1.000	21	0.9969	0.0396	0.0042
4.000	3	0.8201	0.5679	0.0645

Notes: symbol “-” denotes that the variable selection is not performed.

3.4 不同校正模型预测性能的比较

为了进一步验证 VIM-RF 校正模型对稀土矿中 Sc 元素含量的预测性能,将 VIM-RF 校正模型与标准曲线法、PLS、RF 校正模型的预测性能进行比较。同时,为了验证变量选择方法对 Sc 元素含量的预测性能,考察了基于波段的变量选择方法对 RF 校正模型预测性能的影响。在标准曲线法构建过程中,分析线通常优先选择具有相对强度较大、跃迁概率大的特征谱线,而且应避免选择距离较近、易出现叠加的特征谱线。笔者根据 NIST 原子光谱数据库选择 Sc I 327.362 nm 作为分析线。在 PLS 校正模型构建过程中,将 WT(小波基函数为 Coif5,分解层数为 1)预处理后的光谱作为 PLS 校正模型的输入变量,并采用留一

法交叉验证对模型参数潜变量进行优化,最终选择最优潜变量为 6。Sc 元素的特征峰主要处于 200~650 nm 范围内,因此在基于波段选择的 RF 校正模型的构建过程中,在 200~650 nm 范围内对输入波段进行筛选,将 WT(小波基函数为 Coif5,分解层数为 1)预处理后的光谱作为输入变量并采用留一法交叉验证对其波段进行优化。对于 Sc 元素来说,其对应的最优波段为 300~400 nm。

图 6 与图 7 分别展示了上述 5 种校正模型的预测性能以及 VIM-RF 校正模型对稀土矿中 Sc 含量的预测结果。通过对比图 6 中 5 种校正模型对 Sc 含量的预测性能可知:在建模方法上,相较于标准曲线法与 PLS 校正模型,RF 校正模型的留一法交叉验证结果及预测

结果均更优;在变量选择方法上,相较于波段选择的 RF 模型, VIM-RF 模型获得了更好的预测性能。由图 7 可知, VIM-RF 校正模型的预测性能为: R_{CV}^2 为 0.9981, $RMSE_{CV}$ 为 0.0430 mg/kg, MRE_{CV} 为 0.0047,

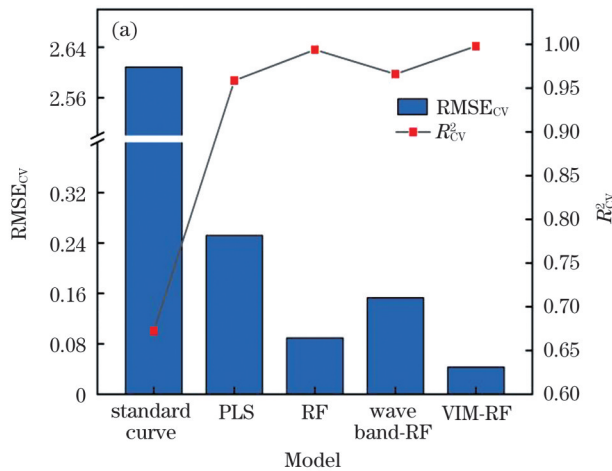


图 6 不同校正模型预测性能的比较。(a)留一法交叉验证;(b)预测集

Fig. 6 Comparison of predictive performance of the different calibration models. (a) Leave-one-out cross validation; (b) prediction set

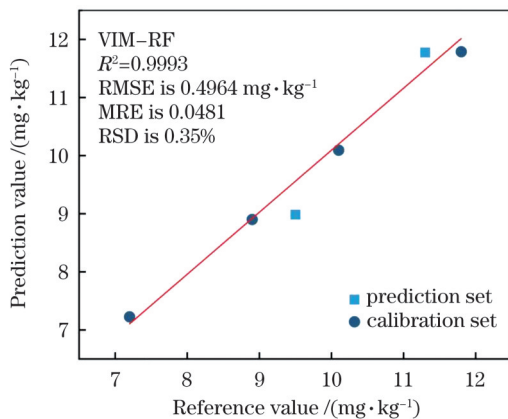


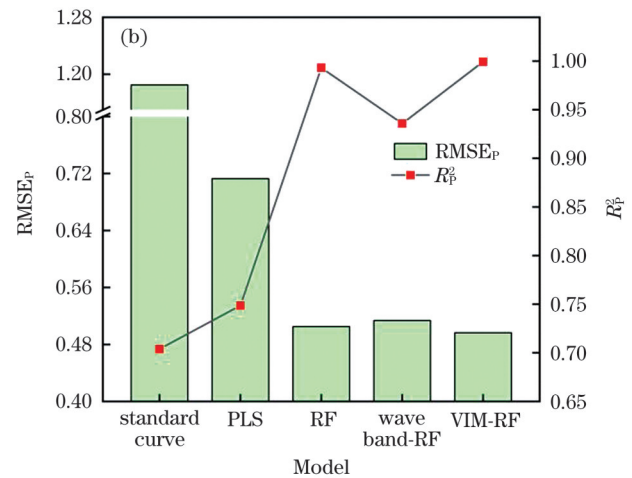
图 7 VIM-RF 校正模型对稀土矿中 Sc 含量的预测结果

Fig. 7 Predictive content of Sc in the rare-earth ores with VIM-RF calibration model

4 结 论

笔者建立了一种基于 LIBS 结合 VIM-RF 的稀土矿石中 Sc 元素含量的检测方法。首先,探究不同光谱预处理方法对 RF 模型性能的影响,以获得一种最优的预处理方法。当选择 WT(小波基函数为 Coif5,分解层数为 1)作为预处理方法时,RF 模型的交叉验证结果相较于原始光谱构建的 RF 模型有较好的提升, R_{CV}^2 从 0.9940 提升至 0.9965, $RMSE_{CV}$ 从 0.0893 mg/kg 减小为 0.0556 mg/kg, MRE_{CV} 从 0.0104 减小为 0.0060。然后,为了进一步提高建模效率与模型的预测性能,将 VIM 应用于 LIBS 光谱的特征筛选,基于留一法交叉验证对 VIM 的重要性阈值进行优化。为了进一步验证 VIM-RF 模型的预测性能,将其与标准曲线法、偏最小二乘(PLS)以及基于波段选择的 RF 模型进行比

较。结果表明, VIM-RF 校正模型具有更好的预测性能,当 VIM 阈值选择为 0.016 时,预测性能为: R_{CV}^2 为 0.9981, $RMSE_{CV}$ 为 0.0430 mg/kg, MRE_{CV} 为 0.0047, R_p^2 为 0.9993, $RMSE_p$ 为 0.4964 mg/kg, MRE_p 为 0.0481, RSD_p 为 0.35%。研究结果表明,基于 LIBS 光谱的 VIM-RF 校正模型是稀土矿石中 Sc 元素定量分析的一种可行方法。



较。结果表明, VIM-RF 校正模型具有更好的预测性能,当 VIM 阈值选择为 0.016 时,预测性能为: R_{CV}^2 为 0.9981, $RMSE_{CV}$ 为 0.0430 mg/kg, MRE_{CV} 为 0.0047, R_p^2 为 0.9993, $RMSE_p$ 为 0.4964 mg/kg, MRE_p 为 0.0481。LIBS 结合 VIM-RF 算法是一种可行的稀土矿石中 Sc 元素定量分析的方法,该方法可为稀土矿石中 Sc 元素的快速检测提供一些思路与借鉴。

参考文献

- [1] Luo X P, Zhang Y B, Zhou H P, et al. Review on the development and utilization of ionic rare earth ore[J]. Minerals, 2022, 12(5): 554.
- [2] 刘峰, 胡文鑫. 钪在铝、镁合金中的应用研究进展[J]. 稀土, 2019, 40(2): 130-141.
Liu F, Hu W X. Research progress in application of Sc-containing aluminum and magnesium alloys[J]. Chinese Rare Earths, 2019, 40(2): 130-141.
- [3] 范亚洲, 周伟, 王子玺, 等. 稀散元素 Sc 的矿床类型及找矿前景[J]. 西北地质, 2014, 47(1): 234-243.
Fan Y Z, Zhou W, Wang Z X, et al. The types of rare and disperse scandium deposits and prospecting potential[J]. Northwestern Geology, 2014, 47(1): 234-243.
- [4] Dutta T, Kim K H, Uchimiya M, et al. Global demand for rare earth resources and strategies for green mining[J]. Environmental Research, 2016, 150: 182-190.
- [5] Dong F Z, Chen X L, Wang Q, et al. Recent progress on the application of LIBS for metallurgical online analysis in China[J]. Frontiers of Physics, 2012, 7(6): 679-689.
- [6] Harmon R S, Lawley C J M, Watts J, et al. Laser-induced breakdown spectroscopy—an emerging analytical tool for mineral exploration[J]. Minerals, 2019, 9(12): 718.
- [7] Zhang Y, Zhang T L, Li H. Application of laser-induced breakdown spectroscopy (LIBS) in environmental monitoring[J]. Spectrochimica Acta Part B: Atomic Spectroscopy, 2021, 181: 106218.
- [8] Gasda P J, Acosta-Maeda T E, Lucey P G, et al. Next generation laser-based standoff spectroscopy techniques for Mars exploration

- [J]. Applied Spectroscopy, 2015, 69(2): 173-192.
- [9] 邱苏玲, 李安, 王宪双, 等. 基于激光诱导击穿光谱的矿石中铁含量的高准确度定量分析[J]. 中国激光, 2021, 48(16): 1611002. Qiu S L, Li A, Wang X S, et al. High-accuracy quantitatively analysis of iron content in mineral based on laser-induced breakdown spectroscopy[J]. Chinese Journal of Lasers, 2021, 48(16): 1611002.
- [10] Carvalho A A C, Alves V C, Silvestre D M, et al. Comparison of fused glass beads and pressed powder pellets for the quantitative measurement of Al, Fe, Si and Ti in bauxite by laser-induced breakdown spectroscopy[J]. Geostandards and Geoanalytical Research, 2017, 41(4): 585-592.
- [11] 刘小亮, 孙少华, 孟祥厅, 等. 激光诱导击穿光谱法测定稀土矿区土壤中钐含量[J]. 中国光学, 2022, 15(4): 712-721. Liu X L, Sun S H, Meng X T, et al. Measurement of Sm in rare earth mineral soil using laser-induced breakdown spectroscopy[J]. Chinese Optics, 2022, 15(4): 712-721.
- [12] Devangad P, Unnikrishnan V K, Nayak R, et al. Performance evaluation of laser induced breakdown spectroscopy (LIBS) for quantitative analysis of rare earth elements in phosphate glasses[J]. Optical Materials, 2016, 52: 32-37.
- [13] 李正风, 徐广晋, 王家俊, 等. 模型诊断用于近红外光谱建模校正集中奇异样本的识别[J]. 分析化学, 2016, 44(2): 305-309. Li Z F, Xu G J, Wang J J, et al. Outlier detection for multivariate calibration in near infrared spectroscopic analysis by model diagnostics[J]. Chinese Journal of Analytical Chemistry, 2016, 44(2): 305-309.
- [14] 袁自然, 魏立飞, 张杨熙, 等. 优化 CARS 结合 PSO-SVM 算法农田土壤重金属砷含量高光谱反演分析[J]. 光谱学与光谱分析, 2020, 40(2): 567-573. Yuan Z R, Wei L F, Zhang Y X, et al. Hyperspectral inversion and analysis of heavy metal arsenic content in farmland soil based on optimizing CARS combined with PSO-SVM algorithm[J]. Spectroscopy and Spectral Analysis, 2020, 40(2): 567-573.
- [15] Borin A, Ferrão M F, Mello C, et al. Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk[J]. Analytica Chimica Acta, 2006, 579(1): 25-32.
- [16] Liu K, Tian D, Xu H Y, et al. Quantitative analysis of toxic elements in polypropylene (PP) via laser-induced breakdown spectroscopy (LIBS) coupled with random forest regression based on variable importance (VI-RFR)[J]. Analytical Methods, 2019, 11(37): 4769-4774.
- [17] 孟德硕, 赵南京, 马明俊, 等. LIBS 结合 ANN 对不同类型土壤中的 Cu 的定量检测[J]. 光电子·激光, 2015, 26(10): 1984-1989. Meng D S, Zhao N J, Ma M J, et al. Quantitative detection of Cu in different types of soils using laser induced breakdown spectroscopy combined with artificial neural network[J]. Journal of Optoelectronics·Laser, 2015, 26(10): 1984-1989.
- [18] Li M G, Fu H, Du Y, et al. Laser induced breakdown spectroscopy combined with hybrid variable selection for the prediction of the environmental risk Nemerow index of heavy metals in oily sludge[J]. Journal of Analytical Atomic Spectrometry, 2022, 37(5): 1099-1108.
- [19] 李铸, 张庆永, 孔令华, 等. 基于激光诱导击穿光谱与随机森林识别 GCr15 钢的硬度[J]. 中国激光, 2022, 49(9): 0911002. Li Z, Zhang Q Y, Kong L H, et al. Identification of hardness of GCr15 steel based on laser-induced breakdown spectrum and random forest[J]. Chinese Journal of Lasers, 2022, 49(9): 0911002.
- [20] 丁宇, 杨淋玉, 陈靖, 等. 基于激光诱导击穿光谱法的铝合金中 Mg 元素定量分析[J]. 激光与光电子学进展, 2022, 59(13): 1314006. Ding Y, Yang L Y, Chen J, et al. Quantitative analysis of Mg in aluminum alloy based on laser-induced breakdown spectroscopy[J]. Laser & Optoelectronics Progress, 2022, 59(13): 1314006.
- [21] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.

Quantitative Analysis of Sc in Rare-Earth Ores via Laser-Induced Breakdown Spectroscopy Combined with Random Forest

Zhou Jiajun¹, Li Maogang¹, Zhang Tianlong^{1*}, Tang Hongsheng¹, Li Hua^{1,2}

¹Key Laboratory of Synthetic and Natural Functional Molecule of the Ministry of Education, College of Chemistry and Materials Science, Northwest University, Xi'an 710127, Shaanxi, China;

²College of Chemistry and Chemical Engineering, Xi'an Shiyong University, Xi'an 710065, Shaanxi, China

Abstract

Objective Sc element is an indispensable strategic resource that is widely used in the manufacture of solid oxide fuel cells, ceramic materials, catalysts, and lightweight high-temperature alloys. Sc mainly exists in minerals, such as rare-earth ores, bauxites, uranium ores, and manganese iron ores. However, its scarcity (the Sc content in the crust accounting for approximately 0.0005%) and the high cost of extraction lead to the current shortage of Sc in the rare-earth market. Therefore, quantitatively analyzing the Sc content in rare-earth ores holds great significance for exploring and mining these ores. Laser-induced breakdown spectroscopy (LIBS) is an analytical detection technology based on the atomic emission spectrum of plasma generated by laser ablation of samples. The characteristic emission spectrum produced by the laser ablation of the sample surface determines the elemental composition and content of the samples. Compared to other technologies, LIBS offers numerous advantages such as no sample preparation, simultaneous multi-element analysis, on-site rapid detection, remote detection, real-time online detection, and microdamage detection. It is extensively used in various fields such as metallurgical analysis, geological exploration, environmental monitoring, and space exploration. However, the spectrum's stability is affected by the sample's matrix effect, the unevenness of the sample surface, and changes in the detection environment during the LIBS detection process. In recent years, multivariate correction methods, such as random forest (RF), partial least squares (PLS), artificial neural network (ANN), and support vector machine (SVM) successfully address and resolve these issues. These methods enhance the accuracy and precision of LIBS spectral analysis, yielding impressive results. RF is

an integrated learning method that is based on a regression tree, stands out for its resistance to overfitting, high accuracy, and straightforward optimization of model parameters. It is in widespread use in LIBS spectral analysis. This study aims to explore the feasibility of combining LIBS with RF for the quantitative analysis of Sc in rare-earth ores.

Methods In this study, a method for quantitative analysis of Sc in rare-earth ores based on laser-induced breakdown spectroscopy (LIBS) combined with a random forest (RF) algorithm had been proposed. Six rare-earth ore standard samples were prepared, and the reference values for the Sc content in rare-earth ore samples are shown in Table 1. Each powder sample was compressed into thin slices using a tablet press at 20 MPa for 5 min before the LIBS spectra were collected. Next, an LIBS device was set up to collect the spectra. To enhance the stability of the LIBS spectra of rare-earth ore samples, 16 sampling points were randomly chosen for LIBS spectra collection for each compressed piece. The LIBS spectra were obtained by accumulating five laser pulses and averaging them to minimize the effect of laser pulse fluctuations. A total of 96 spectral data points were collected from the six rare-earth ore samples (16 spectra for each sample). The influence of different spectral preprocessing methods on the performance of the RF model was then investigated, and variable importance measurement (VIM) was applied for feature screening and parameter optimization for the RF calibration model. To further validate the prediction performance of the RF model, it was compared with other models. Finally, a VIM-RF calibration model was established based on the optimized spectral preprocessing and VIM threshold conditions.

Results and Discussions First, the influence of different spectral preprocessing methods on the performance of the RF model is investigated. As Table 2 demonstrates, the RF correction model processed by the WT (where the wavelet basis function is Coif5, and the decomposition level is 1) realizes better cross-validation and internal validation results within the correction set. Compared to the original spectral model, the cross-validation results of the WT-RF correction model show an increase in R_{CV}^2 from 0.9940 to 0.9965, a decrease in $RMSE_{CV}$ from 0.0893 to 0.0556 mg/kg, and a decrease in MRE_{CV} from 0.0104 to 0.0060. Subsequently, VIM is applied for feature screening and parameter optimization for the RF calibration model. As Table 3 demonstrates, when the VIM threshold is chosen as 0.016, the RF correction model (with 795 input variables) realizes the best cross-validation and internal validation results within the correction set. Finally, to further verify the prediction performance of the RF model, it is compared with other models. As Fig.6 and Fig.7 demonstrate, the results show that the VIM-RF model exhibited excellent prediction performance (R_{CV}^2 is 0.9981, $RMSE_{CV}$ is 0.0430 mg/kg, MRE_{CV} is 0.0047, R_p^2 is 0.9993, $RMSE_p$ is 0.4964 mg/kg, and MRE_p is 0.0481), indicating that the VIM-RF calibration model based on LIBS spectroscopy is a feasible method for the quantitative analysis of Sc in rare-earth ores.

Conclusions This study establishes a method for detecting Sc in rare-earth ores based on LIBS combined with VIM-RF. First, the influence of different spectral preprocessing methods on the performance of the RF model is investigated. Then, variable importance measurement (VIM) is applied for feature screening and parameter optimization for the RF calibration model. To further verify the prediction performance of the RF model, it is compared with other models. Finally, a VIM-RF calibration model is established based on the optimized spectral preprocessing (WT) and VIM threshold (0.016) conditions. The results show that the VIM-RF model exhibits excellent prediction performance (R_{CV}^2 is 0.9981, $RMSE_{CV}$ is 0.0430 mg/kg, MRE_{CV} is 0.0047, R_p^2 is 0.9993, $RMSE_p$ is 0.4964 mg/kg, MRE_p is 0.0481). Therefore, combining LIBS with RF is a feasible method for the quantitative analysis of Sc in rare-earth ores and provides insights and methods for grade analysis and accurate mining of rare-earth ores.

Key words spectroscopy; laser-induced breakdown spectroscopy; rare-earth ore; quantitative analysis; random forest; variable selection