

基于近红外光谱判断分析煤种的鉴别研究

洪子云^{1,2}, 严承琳², 闵红², 邢彦军^{1*}, 李晨², 刘曙^{2*}

1. 东华大学化学化工与生物工程学院生态纺织教育部重点实验室, 上海 201620
2. 上海海关工业品与原材料检测技术中心, 上海 200135

摘要 煤种信息为煤炭质量评价、进出口税收征管提供技术支持。传统煤种鉴别方法需测定煤炭样品干燥无灰基挥发分、低煤阶煤透光率、粘结指数、恒湿无灰基高位发热量等指标, 能耗大, 检测周期长, 不利于口岸快速通关。基于不消耗化学试剂、快速、低成本等优势, 采用近红外光谱鉴别煤种受到广泛关注, 但目前还未有针对全球不同产地来源煤炭的煤种鉴别应用, 煤炭近红外光谱特征与煤种的相关关系仍有待挖掘。采集了来自澳大利亚、俄罗斯、印度尼西亚等9个国家410批进口煤炭代表性样品, 涉及褐煤、烟煤和无烟煤3个煤种, 对比分析了不同煤种煤炭样品的漫反射近红外光谱特征, 发现不同煤种煤炭样品近红外光谱在吸光度、光谱斜率、特征峰上存在差异。结合样品成分信息、X射线衍射、近红外光谱进行数据挖掘, 发现近红外光谱吸光度与煤炭中固定碳含量呈正相关, 光谱斜率与煤炭芳构化呈负相关, 煤炭芳构化增加导致长波长方向的吸收系数增大, 光谱斜率变小, 光谱特征吸收峰主要为水分和有机物质含氢基团的特征信息, 特征峰强度取决于煤炭中水分和挥发分含量。采用主成分分析(PCA)进行数据降维, 光谱变量从1557个降到394个, 对前10个主成分进行逐步判别, 筛选出PC1, PC2, PC3, PC4, PC6, PC7, PC8, PC9和PC10代替原始数据作为模型输入变量, 建立煤种鉴别的Fisher判别分析模型, 建模样品验证准确率为98%, 交叉验证准确率为97.8%, 测试样品验证准确率为99.1%。PCA载荷图表明: PC1和PC2主要与煤炭挥发分含量相关, 其次是水分含量。判别函数1(57.7%)与PC1的相关性最强, 判别函数2(42.3%)与PC2的相关性最强, 说明不同煤种中挥发分含量和水分含量的差异是近红外光谱进行煤种鉴别的内在依据。

关键词 煤种鉴别; 近红外光谱; 判别分析

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)09-2800-07

引言

煤炭是古代植物埋藏在地下经历了复杂的生物化学和物理化学变化逐渐形成的固体可燃性矿物, 占我国一次能源消费总量的70%以上。中国是世界上最大的煤炭进口国, 进口煤炭种类包括无烟煤、烟煤、褐煤、其他煤, 在煤炭进口过程中, 为规避国家相关进口管理政策, 套用优惠贸易逃避税收征管, 存在不法企业伪报煤种的现象。传统煤种鉴别方法需要测定干燥无灰基挥发分、低煤阶煤透光率、粘结指数、恒湿无灰基高位发热量等多项指标^[1], 能耗大, 检测周期长, 不利于口岸快速通关。

煤炭的近红外光谱主要包括N—H, C—H和O—H等含氢基团的倍频与合频吸收, 反映的是煤炭中有机物质及水

分的信息, 基于不消耗化学试剂、快速、低成本等优势, 采用近红外光谱鉴别煤种受到广泛关注。有研究针对煤炭近红外谱图, 构建一种置信学习机—支持向量机(CM-SVM)分类器, 对来自山西的肥煤、焦煤、瘦煤、贫瘦煤进行分类, 留一交叉验证准确率为95.48%。有报道利用深度学习的卷积网络(CNN)提取煤炭样品的近红外光谱特征, 结合极限学习机算法, 测试集最高分类精度可达到96.51%。宋亮等^[2]采集了来自我国主要产煤地的烟煤、褐煤的可见-近红外光谱, 比较不同算法对于煤种的分类效果, 其中媒体遥感识别(MAO)模型分类时间最短, 随机森林(RF)模型交叉验证判别准确率最高。Mao等^[3]提出了一种改进的多层极限学习机算法, 结合可见-红外光谱, 建立了一种煤种快速鉴别模型, 测试集准确率达到92.25%。以上研究表明, 近红外光谱特征与煤种具有很好的相关关系, 结合模式识别可用于煤种鉴

收稿日期: 2021-07-26, 修订日期: 2021-11-09

基金项目: 国家重点研发计划项目(2018YFF0215400)资助

作者简介: 洪子云, 1997年生, 东华大学化学化工与生物工程学院硕士研究生 e-mail: 1772282817@qq.com

* 通讯作者 e-mail: liu_shu@customs.gov.cn; yjxing@dhu.edu.cn

别,但目前还未有针对全球不同产地来源煤炭的煤种鉴别应用,煤炭近红外光谱特征与煤种的相关关系仍有待挖掘。

研究中收集了来自澳大利亚、俄罗斯、印度尼西亚、蒙古、加拿大、哈萨克斯坦、菲律宾、越南、朝鲜9个国家410个进口煤炭代表性样品,涉及烟煤、褐煤、无烟煤3个煤种。对采集的煤炭样品进行漫反射近红外光谱分析,结合代表性样品成分信息、X射线衍射分析,解析了不同煤种煤炭样品的近红外光谱特征及差异。使用主成分分析(principal component analysis, PCA)对近红外光谱进行数据降维,结合逐步判别—Fisher判别分析建立煤种鉴别模型,采用交叉验证和测试样品验证进行模型评价。

1 实验部分

1.1 样品与光谱采集

根据GB 474—2008《煤样的制备方法》,将煤炭样品制成粒度小于0.2 mm的化学分析样,装在密封袋中,置于阴凉处。使用Antaris II傅里叶变换近红外光谱仪收集光谱数据,扫描范围为 $10\ 000\sim 4\ 000\text{ cm}^{-1}$,波长点数为1 557。每个样品扫描32次,平均光谱用作原始光谱。实验样品来自澳大利亚、俄罗斯、印度尼西亚、蒙古、加拿大、哈萨克斯坦、菲律宾、越南、朝鲜这9个国家,包括烟煤、褐煤、无烟煤3个煤种,共计410个样品,具体情况如表1所示。

表1 实验样品数量汇总

Table 1 Summary of the number of experimental samples

煤种	国别	样品数	建模集	测试集	总计
烟煤	俄罗斯	72	173	74	247
	澳大利亚	95			
	印度尼西亚	29			
	蒙古	30			
	加拿大	15			
	哈萨克斯坦	6			
褐煤	印度尼西亚	126	94	40	134
	菲律宾	8			
无烟煤	俄罗斯	15	21	8	29
	越南	9			
	朝鲜	5			

煤炭样品的漫反射吸光度大小遵循Kubelka-Munk方程,如式(1)

$$A = \log(1/R_{c,\infty}) = -\log\left\{1 + \frac{K_c}{S_c} - \left[\left(\frac{K_c}{S_c}\right)^2 + 2\left(\frac{K_c}{S_c}\right)\right]^{1/2}\right\} \quad (1)$$

式(1)中, K_c 为吸光系数,其大小主要取决于煤样的化学组成与分子结构; S_c 为散射系数,其大小主要取决于煤样的物理状态,如颗粒的大小等。

1.2 煤质及XRD分析

为研究煤炭近红外光谱特征与煤种的相关关系,分别选取2个褐煤、烟煤、无烟煤代表性样品,进行煤质分析及X射线衍射分析。选取的代表性样品包括印度尼西亚褐煤

(YNH)、菲律宾褐煤(FH)、澳大利亚烟煤(AY)、蒙古烟煤(MY)、俄罗斯无烟煤(EWY)、朝鲜无烟煤(CWY)。煤质分析包括近似分析和最终分析,近似分析根据GB/T 212—2008测定煤炭中水分(M_{ad})、灰分(A_{ad})、挥发分(V_{ad})含量,再根据 M_{ad} , A_{ad} , V_{ad} , 计算固定碳(FC_{ad})含量,计算公式如式(2)

$$FC_{ad} = 100 - (M_{ad} + A_{ad} + V_{ad}) \quad (2)$$

式(2)中: FC_{ad} 为空气干燥基固定碳的质量分数,%; M_{ad} 为一般分析实验煤样水分的质量分数,%; A_{ad} 为空气干燥基灰分的质量分数,%; V_{ad} 为空气干燥基挥发分的质量分数,%。

最终分析根据GB/T 476—2008,采用电量重量法测定煤炭中C和H含量,再计算H/C。为防止煤炭中的无机矿物质对结构分析产生干扰,根据GB/T 7560—2001对6个煤炭代表性样品进行脱矿处理,X射线衍射分析实验仪器为德国布鲁克公司生产的D8 Focus X射线衍射仪,测量条件为:Cu靶辐射,电压40 kV,电流40 mA,扫描范围为 $5^\circ\sim 75^\circ$,扫描速度为 $0.5\text{ s}\cdot\text{步}^{-1}$ 。利用PeakFit软件对煤炭样品XRD谱图进行平滑处理,对位于 20° 和 25° 附近的 γ 峰、002峰进行拟合,从拟合曲线中获得峰位(θ)、面积(α)和半峰全宽(β)等参数。根据这些参数计算煤炭芳香度(f_a)的大小, f_a 为芳环中碳原子与脂肪族侧链中碳原子的比率,反映煤炭的芳构化趋势,计算如式(3),式(3)中 A_γ 和 A_{002} 分别是对应于 γ 峰和002峰的峰强度。

$$f_a = c_{ar}/(c_{ar} + c_{al}) = A_{002}/(A_{002} + A_\gamma) \quad (3)$$

1.3 建模方法

近红外光谱的变量数通常远远超过样本数,为解决样本容量小的问题,建立判别模型前采用PCA进行降维。PCA由皮尔逊于1901年提出,霍特林于1933年进一步发展。PCA主要用于将包含大量非显著成分的多维数据集简化为保留原始数据集中存在的最大变化的少数主成分,通常用于光谱变量的无监督降维。

逐步判别-Fisher判别分析是一种常用的有监督的分类方法,先建立已知样本的分类模型,再对未知样本进行预测。通过协方差分析的F检验的显著性水平选择最佳的预测变量,当加入或删除一个变量后,对判别函数中的变量进行方差分析,若计算的F值大于指定的进入值(默认值是3.84),则保留该变量;若计算的F值小于指定的删除值时(默认值为2.71),则剔除该变量。判别的基本原理是将高维空间的样本投影到低维空间上,使得投影后的样本数据在新的子空间上有最小的类内距离以及最大的类间距离,保证最大的可分性。

建模过程中随机选取不同煤种光谱的70%作为训练样本,剩余30%光谱数据作为测试样本,如表1所示。使用判别准确率来评估模型性能,判别准确率即不同煤种中正确分类的煤炭样本相对于该类样本总数的百分比。

2 结果和讨论

2.1 近红外光谱特征分析

使用Antaris II傅里叶变换近红外光谱仪,采用漫反射

积分球模式收集了来自澳大利亚、俄罗斯、印度尼西亚、蒙古、加拿大、哈萨克斯坦、菲律宾、越南、朝鲜 9 个国家, 烟煤、褐煤、无烟煤 3 个煤种, 410 个进口煤炭样品的光谱数据, 样品粒度均小于 0.2 mm。图 1 显示了所有样本在 10 000 ~ 4 000 cm^{-1} 范围内的平均光谱。总的来说, 不同煤种的近红外谱图差异十分明显, 主要体现在吸光度大小、光谱斜率、主要吸收峰三方面。

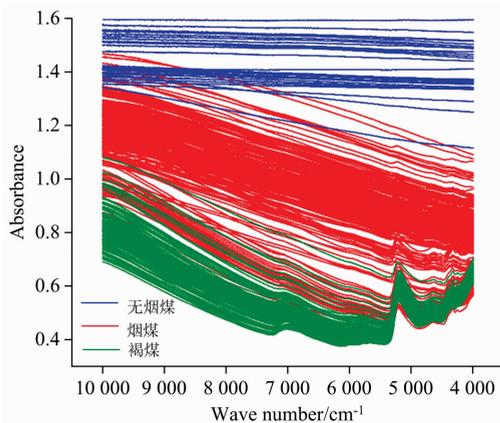


图 1 不同煤种近红外谱图

Fig. 1 Near-infrared spectra of different coals

褐煤的吸光度最低, 大部分褐煤样品的吸光度在 0.35 ~ 1.1 Abs 范围内, 烟煤的吸光度在 0.43 ~ 1.5 Abs 范围内波动, 整体高于褐煤, 无烟煤的吸光度整体最高, 在 1.1 ~ 1.6 Abs 范围内波动; 褐煤的吸光度随波数增加明显降低, 光谱斜率最大, 烟煤次之, 无烟煤的吸光度随波数增加变化缓慢, 光谱斜率几乎为零; 褐煤在 5 199 cm^{-1} (O—H 伸缩和 O—H 变形组合振动) 处有较强的吸收峰, 在 7 021 cm^{-1} (H_2O 的合频) 处有小而明显的吸收峰, 在 4 655 ~ 4 000 cm^{-1} 波段有明显的吸收带(各种有机组合和泛音带)产生^[4-5]。少数烟煤在 5 199 cm^{-1} 处有较强的水峰, 无烟煤整体谱线趋于平坦, 无明显吸收峰出现。

为了找出造成不同煤种的近红外吸光度大小、光谱斜率和吸收峰显著差异的原因, 选取褐煤(YNH、FH)、烟煤(AY、MY)、无烟煤(EWY、CWY), 结合煤炭中水分、挥发分、固定碳含量、H/C 信息及 XRD 提供的煤炭芳香度信息进行分析。

2.1.1 吸光度大小

表 2 是煤质标准分析结果, 展现了不同煤种中水分、挥发分、固定碳含量的差异。图 2 (a) 是六个代表性样品的近红外谱图, 从图中可以看出不同煤种吸光度大小有显著差异, 无烟煤样品的吸光度最高, 烟煤次之, 褐煤最低。漫反射吸光度大小主要和 K_c 和 S_c 有关^[6], S_c 的大小主要取决于样品的物理状态, K_c 的大小主要取决于样品的化学组成。实验中所用样品均为粒度小于 0.2 mm 的粉末, 样本之间物理性质相差不大, 实验中煤炭漫反射吸光度大小主要与 K_c 有关。而 K_c 大小主要取决于样品种类, 煤化程度越高的煤种, 其芳香环的缩合度越高, 芳香环之间的脂肪侧链不断脱落, 煤的有序结构不断增强^[7], 最终形成完整并具有稳定的共轭

环结构和共轭结构的高碳化合物, 导致对近红外的吸收随之增强, K_c 增大。图 2 (b) 是 10 000 cm^{-1} 处吸光度大小和固定碳含量的关系, 从图 2 (b) 可以看出, 无烟煤固定碳含量最高, 吸光度最大, 褐煤固定碳含量最低, 吸光度最小。说明煤样近红外漫反射吸光度大小与固定碳含量有关, 固定碳含

表 2 褐煤(YNH、FH)、烟煤(AY、MY)、无烟煤(EWY、CWY)样品近似分析和最终分析结果

Table 2 Proximate analyse and ultimate analyses of lignite (YNH, FH), bituminous coal (AY, MY) and anthracite (EWY, CWY)

	近似分析 (db, Wt%)			最终分析 (daf, Wt%)		H/C
	水分	挥发分	固定碳	C	H	
YNH	19.35	38.28	24.34	67.12	3.19	0.570
FH	10.12	35.27	24.14	61.37	3.94	0.770
AY	6.94	23.69	39.07	62.45	2.89	0.555
MY	1.15	14.56	65.36	75.45	3.56	0.566
EWY	6.07	2.71	84.99	76.18	3.33	0.525
CWY	2.77	5.38	61.60	82.45	3.19	0.464

注: db; 干燥基; daf: 干燥无灰基

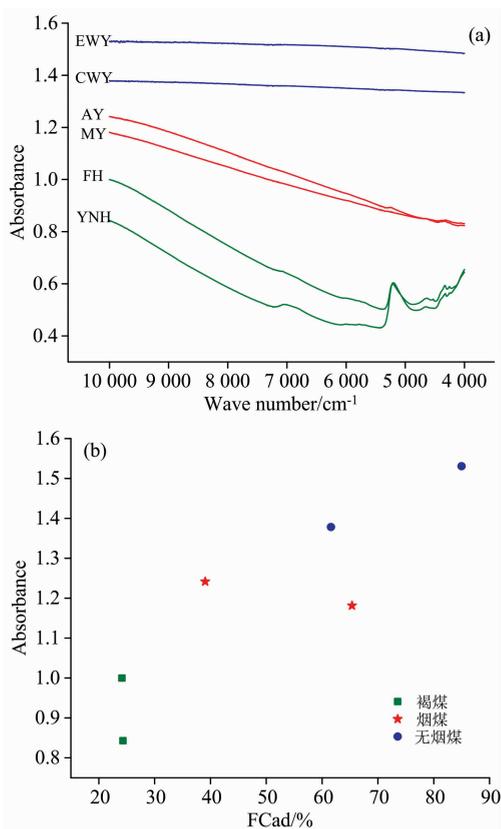


图 2 (a) 褐煤(YNH、FH)、烟煤(AY、MY)、无烟煤(EWY、CWY)样品的近红外谱图; (b) 10 000 cm^{-1} 处吸光度大小固定碳含量的关系

Fig. 2 (a) The near-infrared spectroscopy of lignite (YNH, FH), bituminous coal (AY, MY) and anthracite (EWY, CWY); (b) The relationship between the absorbance at 10 001 cm^{-1} and the fixed carbon content

量越高的煤种，对近红外光的吸收越强，整体吸光度越高^[8]。

2.1.2 光谱斜率

六个代表性样品的 XRD 谱图如图 3(a)所示，图 3(a)中每个煤样都具有相同的石墨特征，并显示出高的背景强度。表明煤样含有一定比例的无定形碳形式的高度无序的物质^[9]。以 EWY 样品为例，图 3 (b)显示了 EWY 在 10°~35° 范围内位于 20°和 25°附近的 002 峰和 γ 峰的拟合谱图，通过

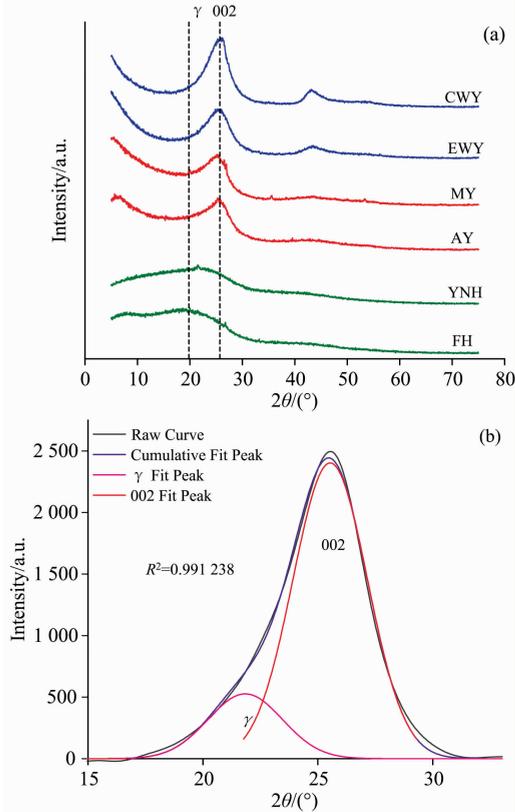


图 3 (a) 褐煤(YNH、FH)，烟煤(AY、MY)，无烟煤(EWY、CWY)的 XRD 谱图；(b) EWY 样品在 10°~35°范围内的 γ 峰和 002 峰的高斯拟合曲线

Fig. 3 (a) The XRD spectra of lignite (YNH, FH), bituminous coal (AY, MY) and anthracite (EWY, CWY); (b) Gaussian fitting curve of γ peak and 002 peak of EWY coal sample in 2θ range of 10°~35°

表 3 XRD 结构参数

Table 3 Structure parameters of XRD

样品	γ 峰			002 峰			f_a
	峰位	峰面积	半峰全宽	峰位	峰面积	半峰全宽	
YNH	18.33	5 150.66	9.52	24.04	5 116.69	7.58	0.498
FH	19.16	7 480.88	8.66	24.71	2 672.72	5.9	0.263
AY	21.58	3 524.84	5.92	25.73	6 763.85	4.24	0.657
MY	20.28	1 056.22	3.93	25.01	8 302.36	4.68	0.887
EWY	19.38	1 102.79	4.73	25.56	10 095.72	4.73	0.902
CWY	19.57	1 923.67	5.17	25.86	18 423.64	5.17	0.905

分峰拟合可以得到峰中心(θ)和半峰全宽(FWHM)等参数，如表 3 所示，根据式(3)计算出芳香度。采用表 2 中的 H/C 反映不同种类样品煤化程度的差异，无烟煤煤化程度最高，H/C 最低。

不同煤种 f_a 和 H/C 的关系如图 4(a)所示，随着 H/C 的增加， f_a 在不断降低，说明煤化程度越高的煤种，芳构化趋势越明显。芳构化程度的增加会导致芳香簇和石墨微晶的尺寸不断增大，煤中的微晶结构逐渐完善，更接近石墨结构^[10]，芳香分子中的电子跃迁移动到更长的波长，吸收边缘的产生使得长波方向的吸收系数不断增大^[5]，导致芳构化最强的无烟煤其近红外谱图几乎趋近于水平，近红外谱图的斜率大小和 f_a 的相关性如图 4(b)所示，可以看出，随着芳构化程度的增加，曲线光谱斜率不断减小。

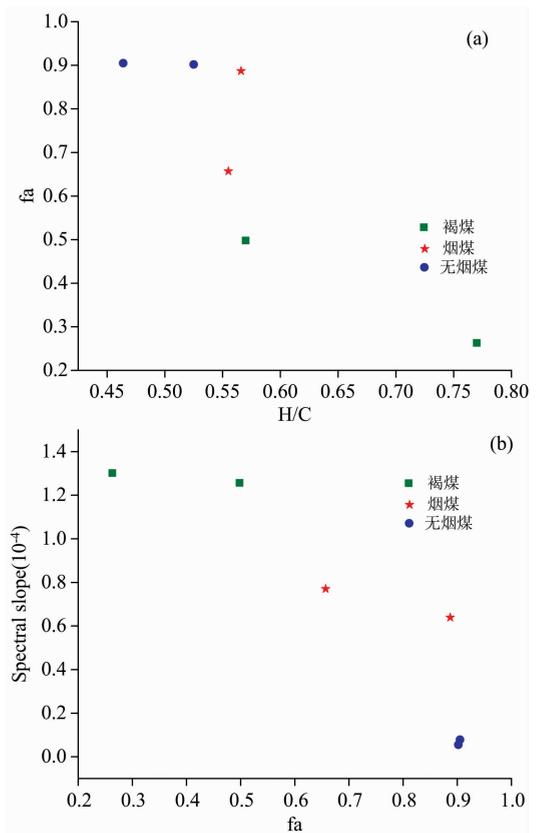


图 4 (a) f_a 和 H/C 的关系；(b) 光谱斜率和 f_a 的相关性
Fig. 4 (a) The relationship between the f_a and the H/C; (b) The relationship between the spectral slope and the f_a

2.1.3 特征吸收峰

煤炭近红外光谱的主要特征，是由以脂肪侧链为主的各有机基团对近红外光产生的倍频与合频吸收所导致。无烟煤由于碳化程度高，煤样中有机质含量少，对近红外光吸收少，无明显吸收特征峰出现，谱图趋于平坦。褐煤样品大多为印度尼西亚褐煤，由于印度尼西亚煤炭大部分来自新生代，煤层埋藏较浅，煤炭变质程度普遍较低，孔裂隙较大，毛细孔对水的吸附力强，使得内在水分经空气干燥难以去除^[11]，导致褐煤在 7 021 cm^{-1} (H_2O 的合频)和 5 199 cm^{-1}

(O—H 伸缩和 O—H 变形组合振动)处有明显的吸收峰,且 $5\ 199\ \text{cm}^{-1}$ 比 $7\ 021\ \text{cm}^{-1}$ 处的吸收特征更明显。在 $5\ 199\ \text{cm}^{-1}$ 处产生特征吸收的烟煤样品均为印度尼西亚烟煤,显示了印啡尼西亚煤炭这个特点。表 2 六个样品的水分数据显示,烟煤和无烟煤的水分含量相差不大,但褐煤的水分含量明显高于烟煤和无烟煤。 $4\ 655\sim 4\ 000\ \text{cm}^{-1}$ 处的吸收峰是有机质光吸收所产生的,其吸收强弱取决于挥发分大小,从表 2 可以看出,挥发分含量明显褐煤>烟煤>无烟煤,反映在 NIR 谱图上吸收峰强弱也是褐煤>烟煤>无烟煤。

2.2 煤种鉴别

上文结合煤炭中水分、挥发分、固定碳含量信息及 XRD 提供的煤炭芳香度信息对煤炭近红外光谱特征与煤种之间的相关性进行了解释,不同煤种漫反射近红外吸光大小与煤炭中固定碳的含量有关,近红外光谱斜率与样品芳构化趋势有关,近红外光谱主要特征吸收峰的强弱与煤炭中水分和挥发分含量有关。基于这种相关关系,结合 Fisher—逐步判别分析对全球不同产地来源煤炭的煤种鉴别应用进行研究。

首先使用 PCA 对光谱进行降维,变量的数量从 1 557 个减少到 394 个,前十个主成分的累计贡献率如图 5(a)所示,其中,PC1 (98.11%)和 PC2 (1.71%)这两个主成分解释了 99.82%的数据差异,表明这两个主成分能够最大限度地分离几乎所有样本。获得的 PC1 和 PC2 的载荷图如图 5(b)和 (c)所示,载荷图本质上描述的是主成分与全波数变量的相关程度,载荷系数的绝对值越大,对主成分的贡献越大。图 5(b)中 PC1 的载荷图显示,对 PC1 影响最大的是位于 $6\ 205\ \text{cm}^{-1}$ 处的特征吸收峰,其次是位于 $7\ 250$ 和 $5\ 458\ \text{cm}^{-1}$ 处的吸收峰。 $6\ 205\ \text{cm}^{-1}$ 处的特征吸收属于 C—H 伸缩振动的二倍频吸收,与煤炭中挥发分对近红外的吸收有关, $7\ 250\ \text{cm}^{-1}$ 处的特征吸收属于 C—H 伸缩振动的二倍频和 C—H 变形振动的组合吸收峰,也与挥发分的吸收有关, $5\ 458\ \text{cm}^{-1}$ 处的特征吸收属于 O—H 伸缩振动,与水分的吸收有关,说明 PC1 主要取决于煤炭中挥发分的含量,其次是水分的含量。图 5(c)显示 PC2 与 $4\ 345\sim 4\ 000\ \text{cm}^{-1}$ 波段的吸光度相关性最强,其次是位于 $5\ 174\ \text{cm}^{-1}$ 处吸收峰, $4\ 655\sim 4\ 000\ \text{cm}^{-1}$ 波段是各种有机组合和泛音带的吸收带,与挥发分对近红外光的吸收有关, $5\ 174\ \text{cm}^{-1}$ 处的吸收峰表征 O—H 伸缩和 O—H 变形组合振动,与水分对近红外的吸收有关,说明 PC2 主要取决于煤炭中挥发分和水分的含量。

将前十个主成分作为输入变量进行逐步判别,通过剔除不合适的变量来提高判别准确率。PC1, PC2, PC3, PC4, PC6, PC7, PC8, PC9 和 PC10 作为有效变量,进入最终模型。使用 PC1, PC2, PC3, PC4, PC6, PC7, PC8, PC9 和 PC10 建立 Fisher 判别模型,共得到 2 组判别函数和相应的组质心处的函数。判别函数方程是用非标准系数发展起来的,要建立一个好的模型,必须具备高特征值^[12]。判别函数 1(57.7%)的特征值为 10.63,判别函数 2(42.3%)的特征值为 7.802,均满足条件。因此,建立的方程将阐明煤炭样品的美好判别模型。具体判别函数如式(4)和式(5)所示。

$$F_1 = 0.301X_1 + 0.212X_2 - 0.211X_3 - 8.169X_4 + 8.917X_5 - 19.393X_6 - 25.721X_7 - 14.089X_8 +$$

$$24.106X_9 - 0.079(57.7\%) \quad (4)$$

$$F_2 = -0.031X_1 + 2.001X_2 + 2.083X_3 - 1.830X_4 - 1.799X_5 - 3.920X_6 + 3.020X_7 + 9.314X_8 - 28.885X_9 - 0.003(42.3\%) \quad (5)$$

式中 $X_1\sim X_9$ 分别代表 PC1, PC2, PC3, PC4, PC6, PC7, PC8, PC9 和 PC10 的得分。

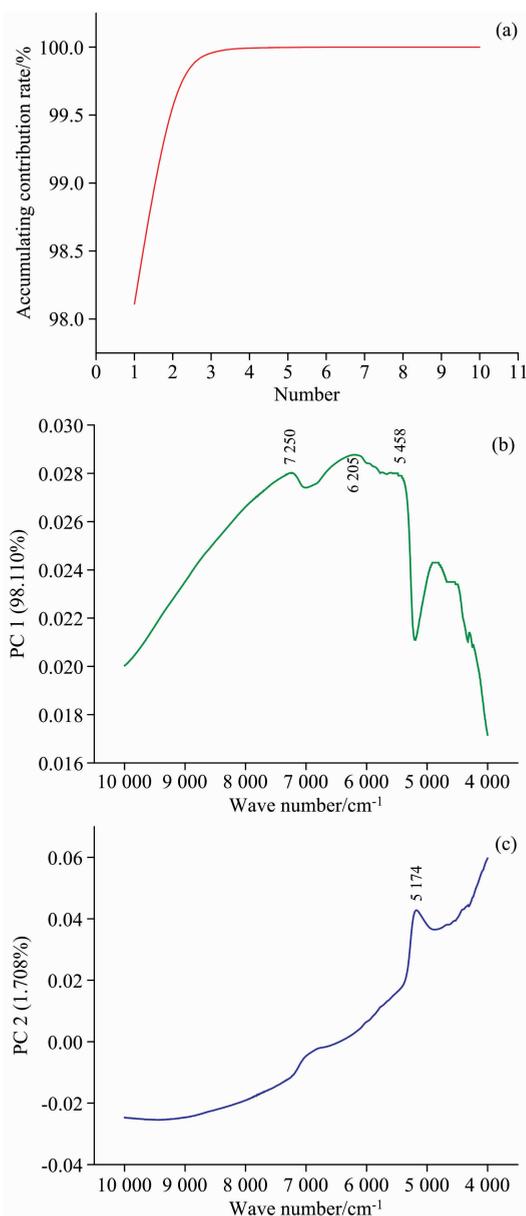


图 5 (a) 累计贡献率; (b) PC1 (98.11%) 的载荷图; (c) PC2 (1.708%) 的载荷图

Fig. 5 (a) Cumulative contribution rate of PCs; Loading plot of (b) PC1 (98.11%), (c) PC2 (1.708%)

将函数 1 得分作为横坐标,函数 2 得分作为纵坐标,用函数 1 和函数 2 的判别得分作散点图,如图 6(a),可以看出不同煤种组质心分离明显,褐煤样品和烟煤样品有部分重叠,判别函数 1 对无烟煤和烟煤、褐煤之间的分离起主要作用,判别函数 2 对烟煤和褐煤、无烟煤之间的分离起主要作用。

用。判别变量和判变函数之间的相关性可以用绝对相关系数表示^[13]，绝对相关系数越高，说明变量对判别函数的贡献越大。图 6(b)表明，PC1 和第一个判别函数相关性最强，PC2 和第二个判别函数相关性最强，PC1 和 PC2 主要与煤炭中挥发分含量有关，其次是水分，说明 Fisher 判别模型对不同煤种进行判别的主要依据是不同煤种中挥发分和水分含量的差异。

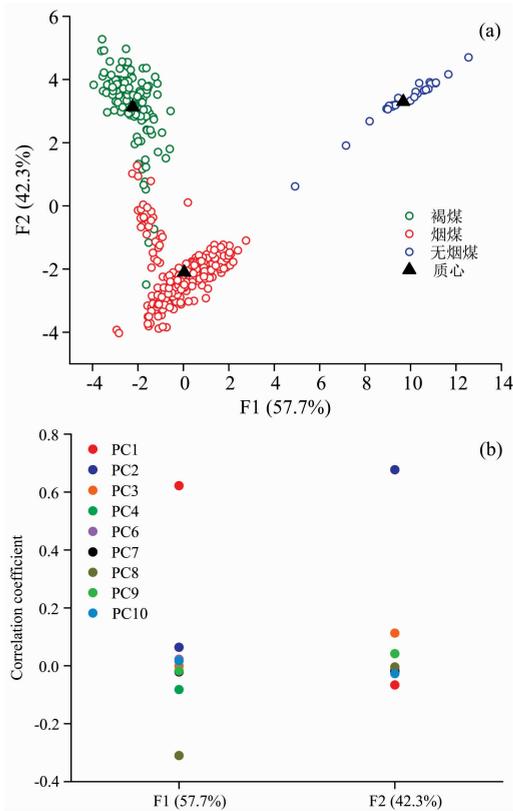


图 6 (a) 判别函数 (F_1, F_2) 得分散点图; (b) 判别函数 (F_1, F_2) 中每个变量的相关系数

Fig. 6 (a) Discriminant function (F_1, F_2) decentralized point map; (b) Correlation coefficient of each variable in the discriminant function (F_1, F_2)

具体判别结果如表 4 所示，建模样品集显示了 98% 的判别结果，采用留一交叉验证来验证这些输出，交叉验证的结果显示 97.8% 的煤炭样品判别准确，通过将 123 个未知样本作为该模型的测试集，进一步验证了所开发的模型方程，测

试样集判别准确率为 99.1%。说明近红外光谱结合逐步判别—Fisher 判别分析，能实现对全球范围内煤炭的煤种识别。

表 4 判别准确率 (%)

Table 4 Discrimination accuracy (%)			
	建模样品集	交叉验证	测试样品集
烟煤	98.3	97.7	97.3
褐煤	95.7	95.7	100
无烟煤	100	100	100
总	98	97.8	99.1

3 结 论

利用傅里叶变换近红外光谱收集了来自澳大利亚、俄罗斯、印度尼西亚、蒙古、加拿大、哈萨克斯坦、菲律宾、越南、朝鲜 9 个国家，涉及烟煤、褐煤、无烟煤 3 个煤种，共计 410 个煤种样品的近红外漫反射光谱数据。近红外谱图表明，不同煤种的吸光度、光谱斜率、特征吸收峰存在差异。无烟煤的吸光度整体最高，烟煤次之，褐煤最低；无烟煤光谱斜率几乎趋于零，而褐煤的光谱斜率最大；三个煤种中，仅褐煤的特征吸收峰最明显，无烟煤几乎无特征吸收峰产生。为对煤种和近红外光谱特征之间的相关性进行解释，结合煤炭中水分、挥发分、固定碳含量信息及 XRD 提供的煤炭芳香度信息进行分析，发现煤炭近红外漫反射吸光度大小与煤炭中固定碳含量呈正相关，光谱斜率与煤炭芳构化呈负相关，主要特征峰的强弱与煤炭中水分和挥发分含量呈正相关。基于近红外光谱特征与煤种之间的相关关系，结合 Fisher—逐步判别分析对全球不同产地来源煤炭的煤种进行鉴别。首先使用 PCA 对光谱降维，然后采用逐步判别筛选出 PC1, PC2, PC3, PC4, PC6, PC7, PC8, PC9 和 PC10 作为有效变量建立煤种鉴别的 Fisher 判别分析模型，共得到两组判别函数，判别函数 1 和判别函数 2 分别与 PC1 和 PC2 相关性最强，PC1 和 PC2 的载荷图又表明 PC1 和 PC2 主要与煤炭中挥发分和水分含量有关，说明 Fisher 判别分析对不同煤种进行鉴别的内在依据是煤炭中挥发分和水分含量的差异。最终测试样品集判别准确率可达到 99.5%，说明近红外结合 Fisher—逐步判别分析是一种有效的针对全球范围内煤炭煤种鉴别的方法。

References

[1] Russell M C, Belle J H, Liu Y. Journal of the Air and Waste Management Association, 2017, 67(1): 3.
 [2] SONG Liang, LIU Shan-jun, MAO Ya-chun, et al(宋 亮, 刘善军, 毛亚纯, 等). Journal of Northeastern University • Natural Science (东北大学学报 • 自然科学版), 2017, 38(10): 1473.
 [3] Mao Y, Le B T, Xiao D, et al. Optics & Laser Technology, 2019, 114: 10.
 [4] Kim D W, Lee J M, Kim J S. Korean Journal of Chemical Engineering, 2009, 26(2): 489.
 [5] Cloutis E A. Fuel, 2003, 82(18): 2239.
 [6] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al(严衍禄, 赵龙莲, 韩东海, 等). Fundamentals and Applications of Near Infrared Spectroscopy Analysis(近红外光谱分析基础及应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005. 1.

- [7] Yan J, Lei Z, Li Z, et al. *Fuel*, 2020, 268(117038): 1.
- [8] Song Z, Kuenzer C. *International Journal of Coal Geology*, 2017, 171: 142.
- [9] Jiang J, Yang W, Cheng Y, et al. *Fuel*, 2019, 239: 559.
- [10] He X, Liu X, Nie B, et al. *Fuel*, 2017, 206: 555.
- [11] Feng L, Yuan C, Mao L, et al. *Fuel*, 2018, 219: 288.
- [12] Chauhan R, Kumar R, Kumar V, et al. *Forensic. Sci. Int.*, 2021, 319: 110655.
- [13] Yan C, Zhu Z, Wu X, et al. *Analytical and Bioanalytical Chemistry*, 2021, 413(17): 4521.

Research on Coal Species Identification Based on Near-Infrared Spectroscopy and Discriminant Analysis

HONG Zi-yun^{1,2}, YAN Cheng-lin², MIN Hong², XING Yan-jun^{1*}, LI Chen², LIU Shu^{2*}

1. Chemical Engineering and Biotechnology, Key Laboratory of Science and Technology of Eco-Textile, Ministry of Education, Donghua University, Shanghai 201620, China
2. Technical Center for Industrial Product and Raw Material Inspection and Testing, Shanghai Customs, Shanghai 200135, China

Abstract The information on coal species provides technical support for evaluation of coal quality and import and export tax. Traditional coal identification methods require the determination of indicators such as dry ash-free volatile matter, the light transmittance of low-rank coal, bonding index, the gross calorific value on the moist ash-free basis of coal samples and other indicators, with large energy consumption and long detection cycle, which is not conducive to the rapid customs clearance at ports. Due to the advantages of no chemical reagents consumption, and fast and low cost, the research on coal identification by near-infrared spectra has attracted extensive attention. However, there has not been any application for the identification of coal from different sources in the world so far, and the correlation between NIR spectral characteristics of coal and coal species remains to be explored. This research collected 410 representative samples of imported coal from 9 countries, including Australia, Russia and Indonesia, etc. involving lignite, bituminous coal and anthracite. By analyzing the near-infrared spectrum, it is found that the differences in NIR spectra of different coal species mainly focus on absorbance, spectral slope and characteristic peak. Combining sample composition information, X-ray diffraction analysis and near-infrared spectra to analyze the reasons for these differences shows that the NIR absorbance is positively correlated with the fixed carbon content in coal, and the spectral slope is negatively correlated with this the aromatization of coal. The increase of coal aromatization leads to the increase of the absorption coefficient in the long-wavelength direction and the decrease of the spectral slope. Spectral characteristic absorption peaks are mainly the characteristic information of water and hydrogen-containing groups of organic substances, and the intensity of characteristic peaks depends on the content of water and volatile matter in coal. Principal component analysis (PCA) was used for data dimension-reduction, and the spectral variables were reduced from 1 557 to 394. The first 10 principal components were discriminated step by step, and PC1, PC2, PC3, PC4, PC6, PC7, PC8, PC9 and PC10 were selected as input variables to establish the Fisher discriminant analysis model for coal species identification. The verification accuracy of modeling sample was 98%, the cross-validation accuracy was 97.8%, and the verification accuracy of the test sample was 99.1%. PCA load diagram shows that PC1 and PC2 are mainly related to the volatile content of coal, followed by moisture content. The correlation between the discriminant function 1 (57.7%) and PC1 was the strongest, and the correlation between the discriminant function 2 (42.3%) and PC2 was the strongest, which indicated that the difference between volatile content and moisture content in different coal species was the internal basis for the identification of coal species by NIR.

Keywords Coal species identification; Near-infrared spectroscopy; Discriminant analysis

(Received Jul. 26, 2021; accepted Nov. 9, 2021)

* Corresponding authors