

油页岩含油率近红外光谱原位分析方法研究

刘杰, 张福东, 滕飞, 李军, 王智宏*

吉林大学仪器科学与电气工程学院, 吉林 长春 130026

摘要 为实现油页岩含油率的原位检测, 采用便携式近红外光谱分析技术, 针对吉林扶余油页岩基地2号钻井的66个岩芯样品开展了原位检测的分析建模方法研究。采用自制便携式近红外光谱仪器获得反射率、吸光度、K-M函数三种数据形式光谱数据, 结合主成分-马氏距离(PCA-MD)剔除异常样品、无信息变量消除法(UVE)波长筛选及二者组合的四种建模数据优化方法, 采用相同的数据预处理方法进行偏最小二乘(PLS)和反向传播神经网络(BPANN)两种方法的建模分析研究, 确定最佳分析模型及方法。结果表明(1)不论是否采用四种不同的数据优化方法, 两种建模方法所用建模数据库适合采用反射率或K-M函数的光谱数据形式;(2)两种建模方法, 采用四种不同的数据优化方法, 对相同数据库建模精度的影响不同: 采用PLS建模方法、以PCA-MD和UVE+PCA-MD两种方法进行数据优化、可以提高K-M函数光谱数据形式数据库的建模分析精度, 采用BPANN建模方法、以UVE、PCA-MD与UVE组合的三种方法进行数据优化、对三种数据形式数据库的建模精度均有所提高;(3)除以反射率光谱数据并进行PCA-MD数据优化外, 采用BPANN方法的建模精度好于PLS法。其中采用反射率光谱数据形式、只进行UVE数据优化外的BPANN建模精度最高, 预测相关系数为0.92、标准偏差为0.69%。

关键词 近红外光谱; 油页岩; 含油率; 原位分析; 数据形式; 建模方法; 数据优化

中图分类号: O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2014)10-2779-06

引言

油页岩作为一种石油的替代能源, 其勘查和开采对我国具有重要的战略意义^[1]。油页岩含油率是指其所含页岩油的质量分数, 是油页岩资源评价的关键指标之一。其常规的测定方法(低温干馏法^[2], SH/T 0508—92)需要在实验室进行, 过程繁琐、效率低。而现有的其他检测方法, 无法满足油页岩资源勘查和开采中评价指标快速检测的需要。如利用测井数据的含油率间接评价方法^[3,4], 对没有地质钻井数据的地区无法使用; 采用漫反射红外光谱实验室仪器的油页岩含油率分析^[5-8]方法需要样品粉碎、压片等制样前处理, 无法实现原位测量。

便携式近红外光谱分析技术可直接对目标进行现场和原位测定, 具有分析速度快、绿色环保等特点^[9], 而且油页岩的主要成分——页岩油、水分和矿物质^[1]对应的分子基团(C—H, O—H, C—O等)在近红外光谱区(760~2 526 nm)有各自不同的吸收特征^[10]。页岩油主要由液体烃类组成^[1],

其近红外区的特征分子键以C—H为主。因此要实现油页岩含油率的现场快速检测, 便携式近红外光谱分析是一种有效方法。

针对吉林扶余油页岩科研基地2号钻井(FK2)岩心的原位样品, 采用漫反射近红外光谱分析, 以反射率、吸光度、K-M三种不同的光谱数据表示形式, 结合异常样品剔除和波长优化筛选及二者组合的数据优化方法, 展开偏最小二乘(partial least squares, PLS)和反向传播神经网络(back propagating artificial neural network, BPANN)法建模分析实验, 研究不同光谱数据形式、数据优化和建模方法对油页岩含油率分析精度的影响, 以确定具有最高精度的分析方法。

1 实验部分

1.1 样品

扶科2号钻井(吉林扶余)359.7~390.8 m岩芯, 间隔0.1~0.5 m为一段, 每段内均匀取样约200 g, 装一袋, 共66袋固体样品, 恒温干燥24 h。

收稿日期: 2014-05-24, 修订日期: 2014-07-29

基金项目: 吉林省科技发展计划项目(20116014) 国家潜在油气资源(油页岩勘探开发利用)项目和产学研用合作创新项目(OSR-02-04)资助
作者简介: 刘杰, 女, 1969年生, 吉林大学仪器科学与电气工程学院副教授 e-mail: l_jie@jlu.edu.cn

*通讯联系人 e-mail: zhwang@jlu.edu.cn

1.2 仪器

自主研发的油页岩现场近红外光谱仪(PISA-OS), 光谱范围 1 300~2 500 nm, 采样间隔 2 nm, 取样方式: 积分球漫反射, 参比: 镀金白板; 光谱数据形式: 反射率、吸光度、K-M 函数可多选。

1.3 数据采集

1.3.1 样品光谱

每袋样品取 3 块, 选择一处较平整面置于仪器样品窗口, 测量其相对参比的反射率(同时保存样品的吸光度和 K-M 函数), 每袋样品得 3 个光谱数据, 取平均得该样品的反射率(以及吸光度和 K-M 函数)光谱数据。图 1—图 3 分别为 66 个样品三种数据形式的光谱图。

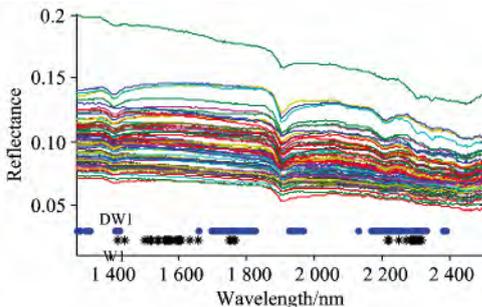


Fig 1 Reflectance spectra of samples

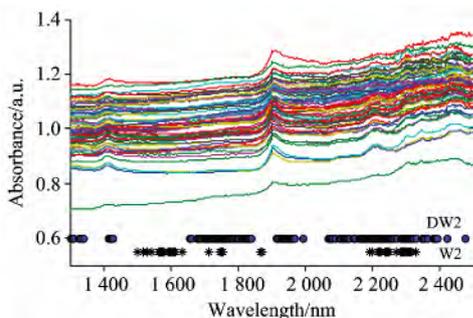


Fig 2 Absorbance spectra of samples

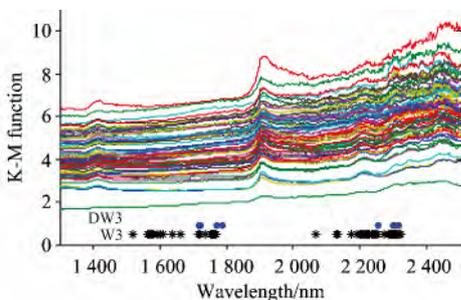


Fig 3 K-M function spectra of samples

1.3.2 含油率

每袋样品粉碎后量取 200 g, 采用低温干馏法测焦油产率 Tar. ad(%), 得 66 个样品的含油率(如图 4 所示), 范围为 0.07~10.87%, 平均 3.48%, 标准偏差 2.145%。

1.3.3 建模数据

建模数据包括 66 个样品的光谱和对应的含油率数据分别组成的 2 个矩阵: 光谱矩阵: 分别选用 66 个光谱数据的三个表示形式: 反射率、吸光度、K-M 函数, 构成 3 个光谱矩阵。

成分矩阵: 66 个样品的含油率数据。

分别选用 3 个光谱矩阵和成分矩阵组成 3 个建模数据库 FK2DB1, FK2DB2, FK2DB3。

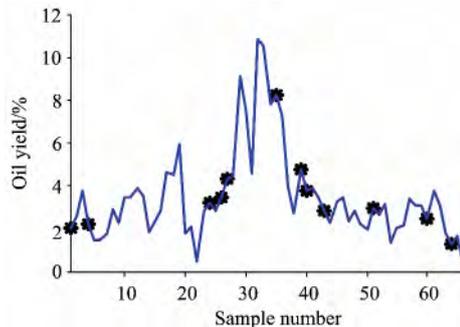


Fig 4 Oil yield of samples

1.4 方法

对建模数据库进行如下建模实验。

1.4.1 模型试建

(1)预处理: 对光谱矩阵中的每个光谱数据采用标准化(SNV)^[10]预处理后, 对处理后的光谱矩阵和成分矩阵分别按波长和成分进行标准化预处理。

(2)按主成分数 h ($=1, 2, \dots, 10$) 分别进行 10 次留一校验 PLS 建模。计算每次留一校验建模的预测结果的相关系数 R_h 和预测标准偏差 SEV_h ^[10]。

(3)确定建模推荐参数

求 1~10 不同的主成分数留一校验 PLS 模型的 R_h 最大值 R_{max} , 以其对应的模型为最佳模型, 其预测标准偏差 SEV 为 SEV_h 中的最小值, 其主成分数为 PLS 主成分数推荐值 H_{opt} 。

1.4.2 数据优化

(1)异常样品剔除: 采用主成分—马氏距离(PCA-MD)分析方法^[10, 11], 分别对数据库的光谱数据标准化后计算 H_{opt} 主分的马氏距离。根据马氏距离, 确定异常样品, 并将其从数据库中剔除, 得剔除异常样品后的数据库。

(2)波长优化: 采用无信息变量消除法(UVE)^[10, 12]对数据库数据进行波长筛选。计算数据库中附加噪声的光谱(标准化后附加等波长长度、 10^{-10} 幅值的噪音)和含油率数据的 PLS 回归系数, 以回归系数的均值与标准偏差比值的绝对值 B 的稳定性选择建模波长。选择波长对应 B 大于噪声对应 B 的最大值 A_{max} 为有效波长, 以有效波长对应的光谱数据值重新组成光谱矩阵, 得波长优化后的数据库。

(3)组合优化: 对数据库数据先进行异常样品剔除再进行波长优化(PCA-MD+UVE)处理, 或先进行波长优化再进行异常样品剔除(UVE+PCA-MD)处理。

1.4.3 建模及对比

将数据库样品按含油率排序, 等间隔取中间的 12 个样

品(图 4 中星点对应的样品)的数据为验证集,其余为校正集。根据 1.4.1 得到的推荐参数,采用与 1.4.1 模型试建时相同的预处理,对建模数据库校正集进行 PLS 建模,以此模型分别计算校正集和验证集的预测标准偏差 SEC 和 SEP、相关系数 R_C 和 R_P ^[10]。

对建模数据库,采用 1.4.1 相同的预处理方法、相同的校正集和验证集进行 30 次 BPANN 建模(构建 3 层网络,隐含层节点数为 20,传递函数分别为: 'tansig' 和 'purelin',训练函数为 'trainsecg',训练迭代次数为 300,目标为验证集相关系数大于 99% 或训练次数超过 300),得对应的校正集和验证集的分析误差 SEC_i 和 SEP_i 、相关系数 R_{Ci} 和 R_{Pi} ($i=1, 2, \dots, 30$)。计算 30 个模型对应精度参数平均值、预测最优模型(SEP_i 最小值对应的模型)对应的精度参数,以此对比

不同数据库所建模型的精度。

上述过程采用 Matlab 编制程序实现。

2 结果与讨论

2.1 异常样品剔除

采用主成分—马氏距离(PCA-MD)分析方法,分别对三个数据库 FK2DB1~3 的光谱数据计算 H_{opt} (数据库 FK2DB1~3 试建 PLS 模型推荐主成分数,见表 1 第 1~3 行第 8 列,均为 4)主成分的马氏距离,如图 5(a)~(c)所示。图 5 中标注值为距离值大的样品号,视为异常样品,分别将其对应的数据从数据库中剔除,得异常样品剔除后的数据库 DFK2DB1~3。

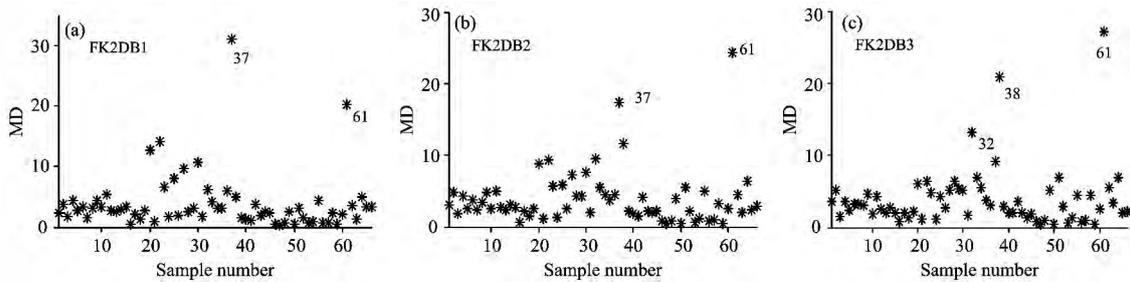


Fig 5 PCA-MD of different data format spectra (MD: Mahalanobis distance)

2.2 波长优化

分别对三个数据库 FK2DB1~3 进行 UVE 波长筛选,得光谱数据与含油率的 H_{opt} (均为 4)主成分 PLS 回归系数的均值与标准偏差比值的绝对值 B 和噪声对应 B 值的最大值 A_{max} (分别为 23.91, 23.00, 22.996),分别如图 6(a)~(c)所示。取 B 值大于 A_{max} 对应的波长为选用波长,分别有 52, 57, 69 个波长点(如图 1—图 3 中 W1~3 所示),得波长优化(UVE)后的数据库 WFK2DB1~3。

2.3 异常样品剔除后波长优化

分别对三个数据库 DFK2DB1~3 进行 UVE 波长筛选,得光谱数据与含油率的 H_{opt} (数据库 DFK2DB1~3 试建 PLS 模型推荐主成分数,见表 1 第 4~6 行第 8 列,分别为 3, 3, 5)主成分 PLS 回归系数的均值与标准偏差比值的绝对值 B 和噪声对应 B 值的最大值 A_{max} (分别为 26.86, 24.36, 21.51),分别如图 7(a)~(c)所示。取 B 值大于 A_{max} 对应的波长为有效波长,分别有 143, 213, 11 个波长点(如图 1~图 3 中 DW1~3 所示),得 PCA-MD+UVE 后的数据库 WDFK2DB1~3。

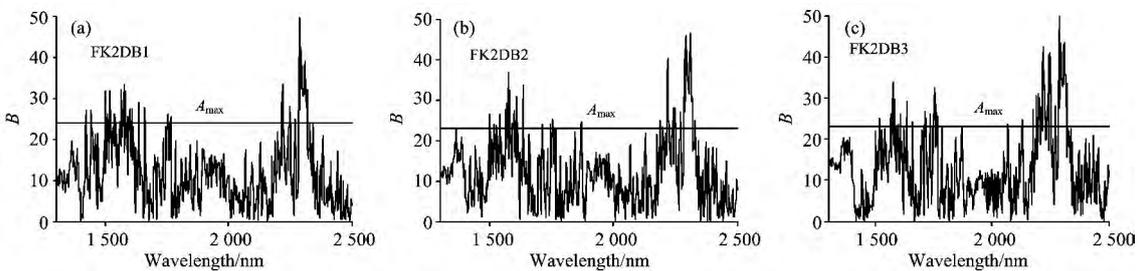


Fig 6 Results of UVE wavelength optimization for different data format bases

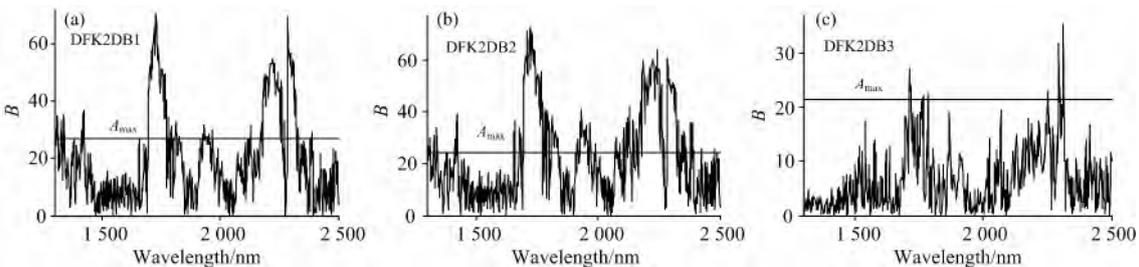


Fig 7 Results of PCA-MD+UVE wavelength optimization for different data format bases

2.4 波长优化后异常样品剔除

采用主成分—马氏距离(PCA-MD)分析方法,分别对数据库 WFK2DB1~3 的光谱数据计算 H_{opt} (数据库 WFK2DB1~3 试建 PLS 模型推荐主成分数,见表 1 第 7~9 行第 8 列,分别为 2, 2, 4)主成分的马氏距离,确定异常样品同 2.1,分别将其对应数据从数据库中剔除,得数据库 WDFK2DB1~3。

2.5 PLS 建模结果

对原建模数据库 FK2DB1~3、异常剔除后数据库 DFK2DB1~3、波长优化后数据库 WFK2DB1~3、组合优化数据库 DWFK2DB1~3、WDFK2DB1~3 共 15 个,分别进行预处理、PLS 试建模型,得 1~10 主成分数留一校验预测结果的相关系数 R_h 与预测标准偏差 SEV_h 。 R_h 的最大值 R_{max} 和 SEV_h 最小值 SEV ,以及根据 R_{max} 确定的推荐主成分数 H_{opt} 分别列于表 1 第 6, 7 和 8 列。

Table 1 PLS modeling results by different bases

组号	序号	数据库	数据形式	优化参数	试建建模			PLS 建模/%	
					$R_{max}/\%$	$SEV/\%$	H_{opt}	SEC	SEP
1	1	FK2DB1	反射率	—	85.71	1.10	4	0.91	1.22
	2	FK2DB2	吸光度	—	85.76	1.10	4	0.90	1.21
	3	FK2DB3	K-M 函数	—	85.63	1.11	4	0.88	1.18
2	4	DFK2DB1	反射率	剔除 37, 61	86.68	1.09	3	0.92	1.27
	5	DFK2DB2	吸光度	剔除 37, 61	86.52	1.09	3	0.92	1.29
	6	DFK2DB3	K-M 函数	剔除 32, 38, 61	82.63	1.12	5	0.77	1.13
3	7	WFK2DB1	反射率 4	52 个波长点	83.32	1.19	2	1.04	1.35
	8	WFK2DB2	吸光度 4	52 个波长点	84.37	1.15	2	1.02	1.26
	9	WFK2DB3	K-M 函数 4	69 个波长点	86.84	1.06	4	0.80	1.19
4	10	DWFK2DB1	反射率 3	143 个波长点	88.09	1.03	3	0.87	1.22
	11	DWFK2DB2	吸光度 3	213 个波长点	88.11	1.03	2	0.95	1.28
	12	DWFK2DB3	K-M 函数 3	11 个波长点	80.65	1.18	3	0.99	1.38
5	13	WDFK2DB1	反射率	剔除 37, 61	84.31	1.17	3	0.88	1.41
	14	WDFK2DB2	吸光度	剔除 37, 61	85.57	1.13	2	0.99	1.28
	15	WDFK2DB3	K-M 函数	剔除 32, 38, 61	83.35	1.09	4	1.024	1.09

对上述 15 个数据库分别按表 1 第 8 列的 H_{opt} , 进行交叉校验 PLS 建模,得对应的校正集和验证集的分析误差 SEC 和 SEP 列于表 1 第 9 和 10 列。表 1 中每 3 行数据为一组,以 PLS 模型对应的 SEP 值分析数据。

(1)组内数据对比:由表 1 中第 1 组数据可知,三种数据格式的 PLS 模型的 SEP 值为 1.18%~1.22%,相差 0.04%,反射率的最大、K-M 函数的最小。表明三种格式光谱数据与其含油率 PLS 模型的分析精度,以 K-M 函数最好、其次是吸光度,反射率最差。

5 组数据中第 3 组(序号第 7~9 行)和第 5 组(第 13~15 行)与第 1 组有相同的结论,而第 2 组(第 4~5 行)和第 4 组(第 10~12 行)的三种光谱数据形式 PLS 模型分析精度与第一组不同,最小值分别出现在第 6 和 10 行,对应 K-M 函数和反射率。表明对于不采用数据优化、或采用异常样品剔除(PCA-MD)、波长优化(UVE)和先波长优化再异常样品剔除(UVE+PCA-MD)三种数据优化处理后,以 K-M 函数光谱数据形式的 PLS 模型分析精度最高,而采用先异常样品剔除再波长优化(PCA-MD+UVE)数据优化处理后,以反射率光谱数据形式的 PLS 模型分析精度最高。

(2)各组间数据对比:与第一组数据对比,只有第 2 和 5 组的 K-M 函数(第 6、15 行)SEP 值由 1.18% 分别降到 1.13% 和 1.09%,其他各行均有所增大。表明采用 PCA-MD 和 UVE+PCA-MD 数据优化后,可以提高 K-M 函数光谱数据形式数据库 PLS 模型分析精度,而采用 UVE 和 PCA-MD

+UVE 后,却降低 K-M 函数光谱数据形式数据库 PLS 模型分析精度,对于反射率和吸光度光谱数据形式数据库无论采用何种数据优化方法均不能提高 PLS 模型分析精度。

(3)15 行数据中,SEP 值最小的在第 15 行(1.09%)。表明针对实验油页岩样品,采用 K-M 函数光谱数据形式数据库,并进行 UVE+PCA-MD 数据优化后的 PLS 模型分析精度最高。

2.6 BPANN 建模结果

分别对上述 15 个数据库进行 30 次交叉校验 BPANN 建模,得对应 30 个模型的校正集和验证集的分析误差 SEC_i 和 SEP_i 、相关系数 R_{Ci} 和 R_{Pi} 平均值和预测最优模型对应值,其中 $SEC_i=0$, $R_{Ci}=100\%$,其余参数列于表 2。表 2 中每 3 行数据为一组,以 30 个模型平均和预测最优模型对应的 SEP 值分析数据。

(1)组内数据对比:对比表 2 中 5 组数据可知,SEP 值最小对应的模型采用光谱数据形式分别是反射率(平均为 1.36%,最优为 0.98%)或 K-M 函数(平均为 1.39%,最优为 0.97%)、K-M 函数(平均为 1.42%,最优为 1.0%)、反射率(平均为 1.12%,最优为 0.69%)、K-M 函数(平均为 1.29%,最优为 0.90%)、K-M 函数(平均为 1.17%,最优为 0.89%)。表明不采用数据优化和采用 UVE,建模数据库的光谱数据适合采用反射率;不采用数据优化或采用 PCA-MD 和组合数据优化,建模数据库的光谱数据适合采用 K-M 函数。

Table 2 BPANN modeling Results by different bases

组号	序号	数据库	所有模型平均/%		预测最优模型/%	
			R_p	SEP	R_p	SEP
1	1	FK2DB1	63.77	1.36	83.81	0.98
	2	FK2DB2	56.51	1.47	79.57	1.09
	3	FK2DB3	62.38	1.39	84.06	0.97
2	4*	DFK2DB1	75i	2.23	59.45	1.44
	5	DFK2DB2	52.82	1.52	81.58	1.04
	6*	DFK2DB3	60.29	1.42	83.14	1.00
3	7	WFK2DB1	76.97	1.12	92.41	0.69
	8	WFK2DB2	68.34	1.28	90.01	0.78
	9	WFK2DB3	67.53	1.30	87.92	0.86
4	10	DWFK2DB1	68.83	1.28	86.09	0.91
	11	DWFK2DB2	63.05	1.37	85.69	0.93
	12	DWFK2DB3	68.56	1.29	86.47	0.90
5	13	WDFK2DB1	51.24	1.52	84.30	0.97
	14	WDFK2DB2	61.68	1.40	86.21	0.91
	15	WDFK2DB3	75.11	1.17	86.98	0.89

(2)组间数据对比:5组数据对应数据对比(组内序号相同,即采用相同数据形式的数据库),第3组的2列SEP数值均为最小,表明在四种数据优化方法中最有效的是 UVE 波长优化。

第2~5组与第1组对应数据对比,只有第2组反射率和 K-M 函数(第4和6行)的 SEP 值增大、其他3组 SEP 值均减小。表明采用 PCA-MD 数据优化方法,对于采用反射率和 K-M 函数光谱数据形式建模数据库进行 BPANN 建模,会降低分析精度,而其他三种数据优化方法对三种数据形式的数据库的 BPANN 模型的分析精度均有所提高。

(3)综合表2中2列 SEP 数据,SEP 值最小均在第7行(分别为 1.12%和 0.69%),表明采用反射率光谱数据形式、UVE 波长优化处理后,BPANN 模型的分析精度最高。

2.7 两种建模方法对比

对比表1(第10列)和表2中最优模型(第7列)对应的 SEP 值,表2中除了第4行的数据大于表1中的值,其余均小于表1中的对应值,表明针对实验油页岩样品,除采用反射率光谱数据形式和 PCA-MD 数据优化外,采用 BPANN 建模方法的建模精度均可以好于 PLS 的建模方法。

表1和表2中的相关系数值较小,最大为 0.924 1,无法满足实际分析的精度要求。主要原因是光谱测量时扫描的样品测点只有3个,太少,对整袋样品而言代表性较差。

3 结论

对于 FK2 油页岩原位样品含油率分析采用反射率、吸光度和 K-M 函数三种光谱数据形式,异常样品剔除、波长筛选及两者组合的四种数据优化方法,结合 PLS 和 BPANN 两种建模方法。

(1)不采用和采用不同的数据优化方法,两种建模方法所用建模数据库适合采用反射率或 K-M 函数两种不同光谱

数据形式。

采用 PLS 建模方法,不采用数据优化、或采用 PCA-MD 异常样品剔除、UVE 波长优化和 UVE+PCA-MD 三种数据优化处理,建模数据库的光谱数据适合采用 K-M 函数;而采用 PCA-MD+UVE 数据优化处理,建模数据库的光谱数据适合采用反射率。

采用 BPANN 建模方法,不采用数据优化和采用 UVE,建模数据库的光谱数据适合采用反射率;不采用数据优化或采用 PCA-MD 和组合数据优化,建模数据库的光谱数据适合采用 K-M 函数。

(2)两种建模方法,采用四种不同的数据优化方法,对相同数据库建模精度的影响不同。

对于 PLS 建模方法,采用 PCA-MD 和 UVE+PCA-MD 两种数据优化方法,可以提高 K-M 函数光谱数据形式数据库所建模型的分析精度,而采用 UVE 和 PCA-MD+UVE 数据优化,却降低 K-M 函数光谱数据形式数据库建模型的分析精度,对于反射率和吸光度光谱数据形式数据库无论采用何种数据优化方法均不能提高所建模型的分析精度。

对于 BPANN 建模方法,UVE、UVE 与 PCA-MD 组合三种数据优化方法,对三种数据形式的数据库的建模精度均有所提高,其中最有效的是 UVE 波长优化。

(3)除以反射率光谱数据并进行 PCA-MD 数据优化外,采用 BP 神经网络建模方法的建模精度好于 PLS 的建模方法。采用反射率光谱数据形式、UVE 波长优化的 BPANN 建模精度最高,预测相关系数 R_p 为 0.924 1、标准偏差 SEP 为 0.69%。而采用 PLS 建模方法的模型分析精度最高对应的标准偏差 SEP 为 1.09%。

致谢:吉林大学地球科学学院的董清水、柳蓉、朱建伟、孙平昌老师提供了样品、含油率分析数据,并协助了样品的光谱测量,在此对四位老师的无私帮助和大力支持表示由衷地感谢。

References

- [1] QIAN Jia-lin, YIN Liang(钱家麟, 尹 亮). Oilshale—The Alternative Energy for Petroleum(油页岩——石油的补充能源). Beijing: China Petrochemical Press(北京: 中国石化出版社), 2008, 1, 69.
- [2] The Mensuration of Oil Yield form Oilshale-Low Temperature Carbonization(油页岩含油率测定法(低温干馏法)), SH/T 0508—92.
- [3] Passey Q R. AAPG Bulletin, 1990, 74(12): 1777.
- [4] HE Jun-ling, DENG Shou-wei, CHEN Wen-long, et al(贺君玲, 邓守伟, 陈文龙, 等). Journal of Jilin University-Geoscience Edition(吉林大学学报·地球科学版), 2006, 36(6): 909.
- [5] Snyder R W, Painter P C, Conauer D C. Fuel, 1983, 62: 1205.
- [6] Alstadt K N, Katti D R, Katti K S. Spectrochimica Acta Part A, 2012, 89: 105.
- [7] XIE Fang-fang, WANG Ze, SONG Wen-li, et al(谢芳芳, 王 泽, 宋文立, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2011, 31(1): 914.
- [8] Romeo M J, Adams M J, Hind A R, et al. Journal of Near Infrared Spectroscopy, 2002, 10(3): 223.
- [9] LU Wan-zhen(陆婉珍). The Modern Analysis Technique of Near-Infrared Spectrum. 2nd ed. (现代近红外光谱分析技术, 第 2 版). Beijing: China Petrochemical Press(北京: 中国石化出版社), 2006, 1, 30, 306.
- [10] CHU Xiao-li(褚小立). Molecular Spectroscopy Analytical Technology Combined with Chemometrics and Its Applications(化学计量学方法与分子光谱分析技术). Beijing: Chemical Industry Press(北京: 化学工业出版社), 2011. 47, 61, 75, 80, 89, 259.
- [11] HAN Liang-liang, MAO Pei-sheng, WANG Xin-guo, et al(韩亮亮, 毛培胜, 王新国, 等). J. Infrared Millim. Waves(红外与毫米波学报), 2009, 28(6): 423.
- [12] WU Di, WU Hong-xi, CAI Jing-bo, et al(吴 迪, 吴洪喜, 蔡景波, 等). J. Infrared Millim. Waves(红外与毫米波学报), 2009, 28(6): 423.

Analyzing and Modeling Methods of Near Infrared Spectroscopy for In-situ Prediction of Oil Yield from Oil Shale

LIU Jie, ZHANG Fu-dong, TENG Fei, LI Jun, WANG Zhi-hong*

Instrument Science & Electrical Engineering College, Jilin University, Changchun 130026, China

Abstract In order to in-situ detect the oil yield of oil shale, based on portable near infrared spectroscopy analytical technology, with 66 rock core samples from No. 2 well drilling of Fuyu oil shale base in Jilin, the modeling and analyzing methods for in-situ detection were researched. By the developed portable spectrometer, 3 data formats (reflectance, absorbance and K-M function) spectra were acquired. With 4 different modeling data optimization methods: principal component-mahalanobis distance (PCA-MD) for eliminating abnormal samples, uninformative variables elimination (UVE) for wavelength selection and their combinations; PCA-MD+UVE and UVE+PCA-MD, 2 modeling methods: partial least square (PLS) and back propagation artificial neural network (BPANN), and the same data pre-processing, the modeling and analyzing experiment were performed to determine the optimum analysis model and method. The results show that the data format, modeling data optimization method and modeling method all affect the analysis precision of model. Results show that whether or not using the optimization method, reflectance or K-M function is the proper spectrum format of the modeling database for two modeling methods. Using two different modeling methods and four different data optimization methods, the model precisions of the same modeling database are different. For PLS modeling method, the PCA-MD and UVE+PCA-MD data optimization methods can improve the modeling precision of database using K-M function spectrum data format. For BPANN modeling method, UVE, UVE+PCA-MD and PCA-MD+UVE data optimization methods can improve the modeling precision of database using any of the 3 spectrum data formats. In addition to using the reflectance spectra and PCA-MD data optimization method, modeling precision by BPANN method is better than that by PLS method. And modeling with reflectance spectra, UVE optimization method and BPANN modeling method, the model gets the highest analysis precision, its correlation coefficient (R_p) is 0.92, and its standard error of prediction (SEP) is 0.69%.

Keywords NIR spectrum; Oil shale; Oil yield; In-situ analysis; Data format; Modeling; Data optimization

* Corresponding author

(Received May 24, 2014; accepted Jul. 29, 2014)