

文章编号 1004-924X(2015)02-0371-07

基于近红外波长组合快速检测油页岩含油率

王智宏, 张福东, 滕 飞, 王炳建, 孙玉洋, 刘 杰*
(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快速单点测量,研究了基于近红外波长组合的油页岩含油率快速检测方法。以高岭土与机油混合的模拟样品为研究对象,利用便携式近红外光谱分析技术,研究了近红外波长组合选择方法。该方法采用自制的便携式光谱仪获得全光谱反射率数据,结合微分和多元散射校正数据处理及相关系数法进行波长初步筛选,将组合生成算法与留一交互校验多元线性回归(MLR)建模相结合,确定最佳波长组合,并用 MLR 建模对单点测量获得的最佳波长组合光谱数据进行了实验验证。结果表明:最佳波长组合为 1 644,1 720,2 210,2 260 nm,30 个建模样品的校正集和预测集的决定系数为 0.995 4 和 0.997 7,11 个验证样品的验证集的决定系数为 0.990 1。该方法为研制基于单点测量的油页岩含油率快速检测光谱仪提供了基础。

关键词:近红外光谱;油页岩;含油率;波长组合选择;留一交互校验;多元线性回归

中图分类号:O657.33;TE122 **文献标识码:**A **doi:**10.3788/OPE.20152302.0371

Rapid detection of oil yield of oil shale by combination of wavelengths in near infrared spectroscopy

WANG Zhi-hong, ZHANG Fu-dong, TENG Fei, WANG Bing-jian, SUN Yu-yang, LIU Jie*

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

* Corresponding author, E-mail: l_jie@jlu.edu.cn

Abstract: To detect rapidly the oil yield of an oil shale on a single point measurement, a detection method was proposed based on combination of wavelengths by a portable near infrared spectroscopy. By taking the kaolin mixed with oil simulating samples as the research object, how to select the combination of wavelengths by near infrared spectroscopy was performed. With proposed method, the full spectral reflectivity data were firstly acquired by the developed portable spectrometer. Then the preliminary selection for the combination of wavelengths was implemented by differential multiplicative scatter correction method and correlation coefficient method, and the optimum combination of wavelengths was determined by combination generation method and leave one cross validation Multiple Linear Regression (MLR). Finally, the data for the optimum combination of wavelengths were verified by MLR modeling. The results show that the optimum combination of wavelengths is 1 644,1 720, 2 210 and 2 260 nm, and the determination coefficients for calibration set (R_c^2), prediction set (R_p^2) of 30 modeling samples and validation (R_p^2) of 11 validation samples are 0.995 4, 0.997 7 and 0.990 1,

收稿日期:2014-06-19;修订日期:2014-08-15.

基金项目:吉林省科技发展计划资助项目(No. 20116014);国家潜在油气资源(油页岩勘探开发利用)产学研用合作创新项目(No. OSR-02-04)

respectively. The method provides a basis for design of a rapid detecting spectrometer in oil yields of oil shapes.

Key words: near-infrared spectroscopy; oil shale; oil yield; combination of wavelengths; leave one cross validation; multiple linear regression

1 引 言

油页岩是富含有机物的黏土岩和泥灰岩,主要成分为油母、水和矿物质,其中油母经过高温裂解产生页岩油,可作为石油的替代品^[1]。油页岩含油率是指油页岩中页岩油所占的质量分数,是油页岩品位评价的关键参数。油页岩含油率的常规测定方法(低温干馏法^[2],SH/T0508-92)需要在实验室进行,过程繁琐且效率低。而现有的检测方法,如利用测井数据含油率的间接评价方法^[3-4],无法应用在没有地质钻井数据的地区。便携式近红外光谱分析技术是建立在化学计量学基础上的间接评价方法,具有不破坏样品、不用试剂、不污染环境、可直接对目标进行现场及原位测定和分析速度快的特点^[5]。近红外光谱(Near Infrared Spectroscopy, NIRS)主要反映的是含氢基团 X—H(如 C—H, N—H, O—H 等)振动的倍频和合频吸收,因此 NIRS 具有丰富的结构和组成信息,非常适合用于含氢有机物质如农产品、石化产品等的物化参数测量^[5]。目前, NIRS 分析已广泛应用于农业和石化等领域。

由于 NIRS 吸收强度弱,光谱信息重叠严重^[6],并存在散射、基线平移等现象^[7],因此要从光谱信息中提取出有用信息,需要建立合适的化学模型。王智宏等研制了便携式 NIRS 分析仪,并研究了数据预处理方法对油页岩含油率 NIRS 分析的影响^[8]。李素义等将小波变换与神经网络融合建模应用到油页岩含油率的分析中^[9]。以上研究均在全波段连续光谱下进行光谱信息的处理和建模,但全光谱包含的信息量大,会影响模型的校正速度和建模效率^[10]。滤光片型和二极管阵列型近红外光谱仪采用单点测量方式,与全光谱测量方式相比,减少了建模所需的波长点数,降低了模型复杂度,提高了建模效率^[11],但波长组合的选取直接影响仪器的建模精度。尹慧敏^[12]等在滤光片确定的前提下,选取 1~8 个波长组合建立了多元线性回

归模型。但滤光片需根据经验及被测物质的 NIRS 特性选取有限的波长,其建模分析结果存在一定的局限性。申艳^[13]等采用多元散射校正和逐步回归法建立了黑土有机碳的 NIRS 定量模型,但逐步回归方法的预测精度偏低。

本文针对油页岩含油率快速单点测量的研究需要,以高岭土与机油模拟油页岩的主要成分,以二者的混合为实验对象,采用便携式 NIRS 分析技术探讨了不同预处理方法对相关系数法^[5,7]选取候选波长点的影响。利用组合生成法和留一校验多元线性回归(Multiple Linear Regression, MLR)建模计算最佳的波长组合,并验证了波长组合方法的准确性,为进一步实现油页岩含油率快速检测的单点测量或滤光片波长的选择提供了依据和方法。

2 实 验

2.1 实验样品

建模样品:产自吉林舒兰的高岭土固体经研磨粉碎,过 140 目筛,与美国壳牌公司喜力 HX5 10W-40 型号的机油用电子天平 PL203(精度为 0.001 g)按一定比重混合,得到浓度从 1%~20% 均匀分布的 30 个样品。

验证样品:用建模样品中的材料配制新浓度的样品,配制了浓度值分别为 2.3%, 3.5%, 5.5%, 8.5%, 10.8%, 13.6%, 14.8%, 16.8%, 18.2%, 19.7%, 21% 共 11 个样品。

2.2 实验仪器

实验采用自主研发的便携式 NIRS 分析仪(PISA-OS),光谱在 1 200~2 500 nm 内可选,采样间隔在 1~8 nm 可选,出射狭缝在 0.2~1.5 mm 可换;光谱分辨率为 7(0.2 mm 狭缝)~50 nm(1.5 mm 狭缝),重复性为 1 000:1(全谱平均),信噪比为 1 800:1。取样方式为积分球漫反射,参比为硫酸钡+测量烧杯;测量方式为连续扫描和单点扫描;光谱数据包括反射率、吸光度、

(Kubelka-Munk)K-M 函数。

2.3 实验数据

2.3.1 建模样品的光谱

将 30 个建模样品粉末装入测量烧杯中压实,烧杯放在仪器测量窗口,采用 1 200~2 500 nm 连续全光谱扫描方式,将烧杯旋转不同角度扫描 3 次,取平均得该样品的反射率光谱数据,图 1 为 30 个建模样品的反射率光谱。

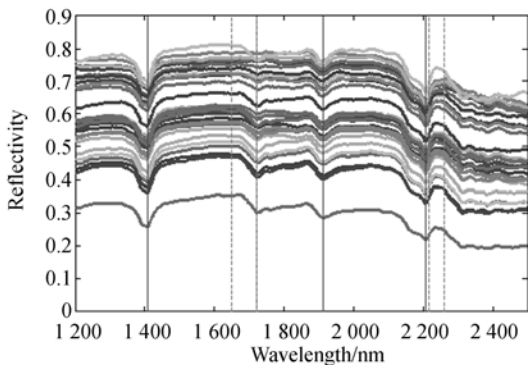


图 1 样品反射率全光谱

Fig. 1 Full reflectance spectra of samples

然后将 30 个建模样品粉末放入测量烧杯,烧杯放在仪器测量窗口,采用 $N(1\sim 9)$ 个波长点的单点扫描方式,将烧杯旋转不同角度测量 3 次,取平均得该样品的单点反射率光谱数据。

2.3.2 验证样品光谱

将 11 个验证样品粉末放入测量烧杯后置于仪器测量窗口,采用单点扫描方式,选取 N 个波长点,将烧杯旋转不同角度测量 3 次,取平均得该样品的单点反射率光谱数据。

2.3.3 含油率

建模样品与验证样品共 41 个,其含油率(即

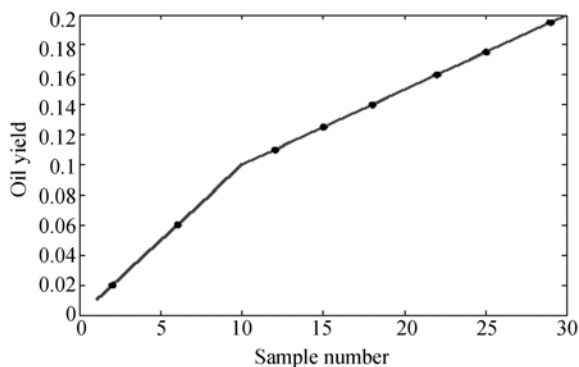


图 2 样品含油率

Fig. 2 Oil yield of samples

浓度值)由电子天平 PL203 分别量取适量的高岭土粉末与机油,两者混合均匀得到,含油率为 1%~21%。图 2 所示为 30 个建模样品的含油率。

2.4 实验方法

2.4.1 波长组合选择

2.4.1.1 预处理

对建模样品的全光谱数据进行未处理、微分和多元散射校正预处理。

2.4.1.2 候选波长组合选取

首先根据处理后的光谱数据计算出波长与含油率的相关系数 R ,按设定的阈值选取高阈值上的波峰点和低阈值下的波谷点为初始候选波长组合。依据单点扫描方式下的波长分辨率 50 nm,对初始候选波长点进行剔除,并按照高低相关性波长点 3:1 的比例手动挑出 20 个波长点作为候选波长点。

2.4.1.3 最佳波长组合确定

按波长数 $N(1\sim 9)$ 进行组合生成算法(对候选波长点进行组合生成)和留一校验 MLR 建模,计算出每次留一校验 MLR 建模的预测结果的决定系数、预测标准偏差和相应波长点下升序排列的前七组预测标准偏差,求取 1~9 个波长数留一校验 MLR 建模的最大和最小的 SEV,选取出最佳的波长组合。

$$R_N^2 = \sqrt{1 - \frac{\sum_{i=1}^n (x_{pi} - x_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}, \quad (1)$$

$$SECV_N = \sqrt{\frac{\sum_{i=1}^n (x_{pi} - x_i)^2}{n-1}}, \quad (2)$$

其中: \bar{x} 为建模数据的均值, x_i 为留一样品 i 的对应成分值, x_{pi} 为留一样品 i 模型(采用除 i 外其它 $n-1$ 个样品所建的 MLR 模型)的含油率预测值。

2.4.2 多元线性回归(MLR)建模

将 30 个建模样品中的单点光谱数据按含油率进行排序,等间隔选取中间点为预测集,其余点为校正集。30 个建模样品的单点光谱使用石英烧杯做容器消除了部分基线漂移和散射现象,因此对单点光谱数据直接进行 MLR 建模,计算得出校正集和预测集的标准偏差 SEC 和 SEP、以及决定系数^[5-6]。

$$SEC = \sqrt{\frac{\sum_{i=1}^{n_c} (x_{cpi} - x_{ci})^2}{n_c - 1}}, \quad (3)$$

$$SEP = \sqrt{\frac{\sum_{i=1}^{n_p} (x_{ppi} - x_{pci})^2}{n_p - 1}}, \quad (4)$$

$$R_c^2 = \sqrt{1 - \frac{\sum_{i=1}^{n_c} (x_{cpi} - x_{ci})^2}{\sum_{i=1}^{n_c} (x_{ci} - \bar{x}_c)^2}}, \quad (5)$$

$$R_p^2 = \sqrt{1 - \frac{\sum_{i=1}^{n_p} (x_{ppi} - x_{pi})^2}{\sum_{i=1}^{n_p} (x_{pi} - \bar{x}_p)^2}}, \quad (6)$$

其中： n_c, n_p 分别为校正集和预测集的样品个数； x_{ci}, x_{pi} 分别为校正集和预测集第 i 个样品的含油率； x_{cpi}, x_{ppi} 分别为校正集和预测集第 i 个样品的预测含油率值； \bar{x}_c, \bar{x}_p 分别为校正集和预测集样品含油率的平均值。

2.4.3 验证集样品的验证模型精度

利用验证集的样品验证 2.4.2 建立的模型，得出验证集的标准偏差 SEP 和决定系数。

3 结果与分析

3.1 光谱数据预处理

对建模样品中的全光谱数据采用多元散射校正(图 3 为多元散射校正预处理后光谱)和 Savitzky-Golay 3 阶 7 点卷积平滑曲线的一阶微分(图 4 为一阶微分预处理后光谱)预处理，计算

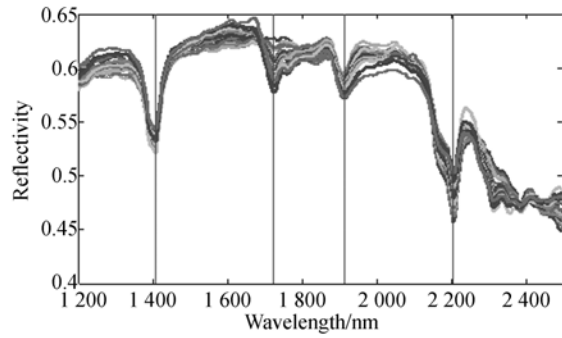


图 3 多元散射校正后全光谱

Fig. 3 Full reflectance spectra after multiple scatter correction

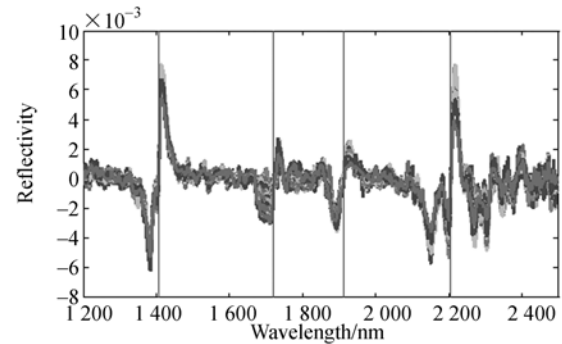
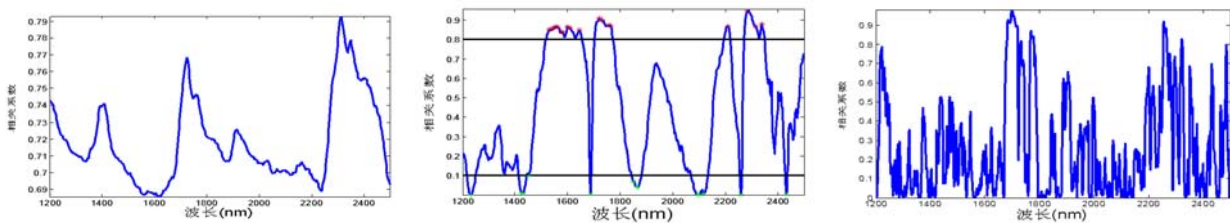


图 4 一阶微分处理后全光谱

Fig. 4 Full reflectance spectra after first derivative

得出预处理后的光谱与含油率的相关系数，如图 5 所示(彩图见期刊电子版)。

图 1 中波长 1 407, 1 722, 1 912, 2 206 nm 处有吸收峰。对原始光谱进行多元散射校正，从图 3 中可以看出，样品之间反射率的差异明显减少，消除了散射的影响。对原始光谱进行一阶微分，从图 4 可以看出，样品间反射率的差异也明显减少，消除了基线和其他背景的干扰。



(a) 未预处理 (b) 多元散射校正预处理 (c) 一阶微分预处理
 (a) No preprocessing (b) Preprocessing with multiple scatter correction (c) Preprocessing with first derivative

图 5 不同预处理的反射率数据与含油率相关系数图

Fig. 5 Correlation coefficients of reflectivity and oil yield with different preprocessing methods

图 5(a)为未处理的光谱-含油率的相关系数,从图中看出其相关系数在 0.68~0.79,相关系数的变化范围小,每个波长点所包含的样品信息量相当,无法进行波长组合的选择。图 5(b)为多元散射校正后的光谱-含油率相关系数,从图中看出其相关系数为 0~0.95,相关系数的变化范围大,而且系数峰与谷宽度明显,因此图 5(b)可进行波长组合选择。图 5(c)采用一阶微分预处理后的光谱-含油率相关系数,从图中看出虽然其相关系数变化范围大,但系数的峰与谷分辨不明显,不适合进行波长组合选择。因此,采用多元散射校正作为波长组合选取的预处理方法。

3.2 波长组合选择

建模样品中的全光谱经多元散射校正预处理后,利用相关系数法设定阈值,高阈值时相关系数为 0.8,低阈值时相关系数为 0.1。依据选定的阈值,图 5(b)中选取相关系数大于 0.8 的波峰位置波长点,选取相关系数小于 0.1 的波谷位置波长点。如图 5(b)所示,相关系数 0.8 处黑线以上的圆点和相关系数 0.1 以下的星号为初始候选波长点。依据 2.4.1 中的选取方法,选取 20 个候选波长点。

按波长点数 $N(N=1,2,\dots,9)$ 进行留一校验 MLR 建模,得出留一校验建模的预测结果的决定

系数和预测标准偏差 SEP。如表 1 所示,在波长点数达到 4 后 SEP 基本不变,即随着波长组合数的增加,只是在原波长组合下增加新的波长点。因此依据留一校验法判定,选定波长数 4 为最佳波长组合。由表 2 可知,4 个波长组合点下前 7 组 SEP 最小的组合中,SEP 非常接近,而且波长点数也集中在 1 644,1 720,1 866,2 210,2 260 nm。由于近红外谱带位置 C-H 基团在 1 410,1 660,2 250 nm 左右及油母质提炼出的液态烃类主要由 C—H 基团组成,而 1 886 nm 处 C—H 基团与水的吸收峰非常近,因此利用此波长点建模可能受样品水分的影响比较大。综上所述,本文选择了表 2 中第 3 行波长组合(1 644,1 720,2 210,2 260 nm),其中前 3 个波长点为高相关性点,2 260 nm 为低相关性点,选择的波长点如图 1 红色虚线所示。

3.3 多元线性回归建模结果与分析

利用 2.4.2 中的方法对建模样品中的单点光谱进行了模型分集,图 2 中黑点对应的样品为预测集,其余为校正集。图 6 为 30 个建模样品的单点光谱反射率,对它进行 MLR 建模得出校正集和预测集的预测标准偏差分别为 0.003 8 和 0.003 0、决定系数分别为 0.995 4 和为 0.997 7。

表 1 相应波长个数下模型最佳评价参数

Tab.1 Evaluation parameters of model corresponding to wavelength numbers

个数	波长/nm									SEV	R_{\max}^2	SECV
1	2 290									0.012 7	0.946 9	0.018 7
2	1 734	2 210								0.011 1	0.959 2	0.011 5
3	1 644	1 720	2 210							0.008 0	0.979 1	0.008 6
4	1 644	1 720	1 866*	2 210						0.006 9	0.984 4	0.007 2
5	1 644	1 720	1 866*	2 210	2 260*					0.006 2	0.987 3	0.006 6
6	1 644	1 720	1 866*	2 210	2 260*	2 338				0.006 2	0.987 3	0.006 4
7	1 644	1 720	1 866*	2 210	2 260*	2 290	2 338			0.006 1	0.987 7	0.006 3
8	1 548	1 644	1 720	1 866*	2 210	2 260*	2 290	2 338		0.006 2	0.987 5	0.006 2
9	1 576	1 584	1 688*	1 720	1 866*	2 210	2 260	2 290	2 432*	0.005 9	0.988 5	0.006 1

注:标有星号的波长为低相关系数点。

表 2 4 个波长组合点下前 7 组最佳波长点

Tab.2 Optimum wavelength points of the first seven groups in four wavelength point combination

序号	波长/nm				SEP	R_v^2
1	1644	1720	1866*	2210	0.006 9	0.984 4
2	1644	1724	1866*	2210	0.007 0	0.984 1
3	1644	1720	2210	2260*	0.007 1	0.983 5
4	1644	1734	1866*	2210	0.007 1	0.983 4
5	1644	1724	2210	2260*	0.007 4	0.982 0
6	1608	1720	1866*	2210	0.007 6	0.981 2
7	1644	1756	1866*	2210	0.007 6	0.981 0

注:标有星号的波长为低相关系数点。

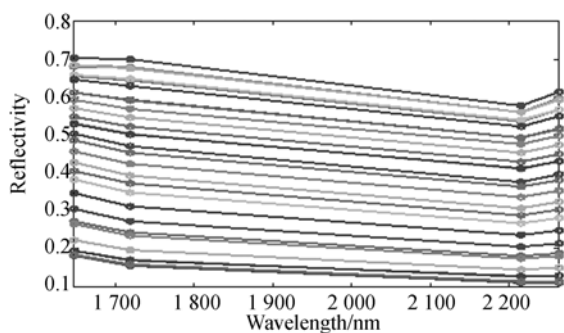


图 6 样品反射率单点光谱

Fig. 6 Partial point reflectance spectra of samples

MLR 建模得到的回归函数为:

$$y = 0.1824 + 1.7953x_1 - 2.6614x_2 + 2.2601x_3 - 1.4177x_4 \quad (7)$$

3.4 模型验证

利用 11 个验证集样品预测模型, 得出验证集的预测标准偏差 SEP 为 0.0065 和 0.9901。图 7 为建模样品的单点光谱反射率与含油率的关系

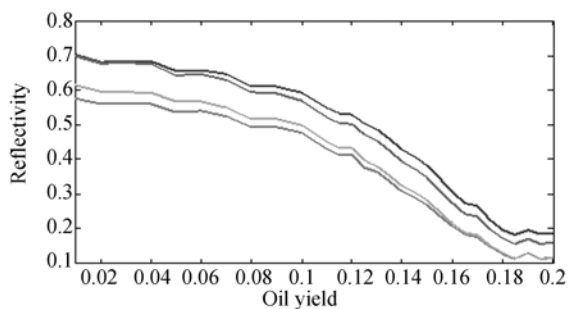


图 7 单点光谱反射率与含油率关系

Fig. 7 Relationship between oil yield and reflectance

系, 可看出含油率与反射率呈线性关系。图 8 所示为预测含油率和实际含油率的关系, 其中圆点为预测含油率值。

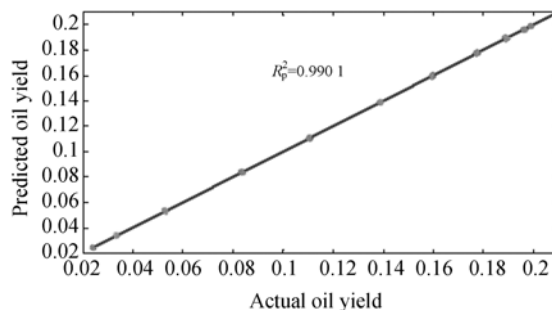


图 8 实际的含油率与预测的含有率关系

Fig. 8 Relationship between actual and predicted oil yields

4 结 论

本文结合微分和多元散射校正数据处理与相关系数法进行了波长初步筛选; 用组合生成算法与留一交互校验多元线性回归(MLR)建模, 确定了最佳波长组合; 通过单点测量光谱获得了最佳波长组合的光谱数据, 并进行了 MLR 建模验证实验。实验结果表明, 建模样品与验证样品的决定系数均达到了 0.99 以上。本文以高岭土与机油混合样品为对象, 初步实现了基于近红外波长组合的油页岩含油率的快速检测, 为进一步实现基于单点测量的油页岩含油率快速检测光谱仪的研制奠定了基础。

参考文献:

- [1] YOUNGQUIST W. Shale oil-the elusive energy [Z]. *Hubbert Center Newsletter*, 1988, 4: 1-8.
- [2] 中华人民共和国石油化工行业标准. 油页岩含油率测定法——低温干馏法[S]. 1992. SH/T 0508-92. The Mensuration of Oil Yield form Oilshale-Low Temperature Carbonization[S]. 1992. (in Chinese)
- [3] PASSEY Q R, CREANEY S, KULLA J B, et al.. A practical model for organic richness from porosity and resistivity logs[J]. *AAPG Bulletin*, 1990, 74 (12): 1777-1794.
- [4] 贺君玲, 邓守伟, 陈文龙, 等. 利用测井技术评价松辽盆地南部油页岩[J]. *吉林大学学报: 地球科学版*, 2006, 36(6): 909-914. HE J L, DENG SH W, CHEN W L, et al.. Evaluation of oil shale in the Southern Songliao basin using logging technique[J]. *Journal of Jilin University: Earth Science Edition*, 2006, 36(9): 909-914. (in Chinese)
- [5] 褚小立. 化学计量学方法与分子光谱分析技术[M]. 北京化学工业出版社, 2011: 47, 61-68, 75-76, 80, 89, 259. ZHU X L. *Molecular Spectroscopy Analytical Technology Combined with Chemometrics and its Application*

- tions[M]. Beijing:Chemical Industry Press, 2011:47, 61-68,75-76,80,89,259. (in Chinese)
- [6] 谷筱玉,徐可欣,汪熾,等. 波长选择算法在近红外光谱学中药有效成分测量中的应用[J]. 光谱学与光谱分析, 2011, 31(1): 91-94.
- GU X Y, XU K X, WANG Y, *et al.*. Application of wavelength selection algorithm to measure the effective component of chinese medicine based on near-infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2011, 31(1): 91-94. (in Chinese)
- [7] 陆婉珍. 现代近红外光谱分析技术(第二版)[M]. 中国石化出版社,2006:1-11,30-31, 306-334.
- LU W ZH. *The Modern Analysis Technique of Near-Infrared Spectrum* [M]. 2nd Ed.. Beijing: China Petrochemical Press, 2006: 1-11, 30-31, 306-334. (in Chinese)
- [8] 王智宏,刘杰,王静茹,等. 数据预处理方法对油页岩含油率近红外光谱分析的影响[J]. 吉林大学学报:工学版,2013,43(4):1017-1022.
- WANG ZH H, LIU J, WANG J R, *et al.*. Influence of data preprocessing on NIRS analysis for oil yield of oil shale[J]. *Journal of University: Engineering and Technology Edition*, 2013, 43(4): 1017-1022. (in Chinese)
- [9] 李素义,嵇艳鞠,刘伟宇,等. 小波变换与神经网络融合法在油页岩近红外光谱分析中的应用[J]. 光谱学与光谱分析,2013,13(4):968-971.
- LI S Y, JI Y J, LIU W Y, *et al.*. Application of wavelet transform and neural network in the near-infrared spectrum analysis of oil shale[J]. *Spectroscopy and Spectral Analysis*, 2013,13(4):968-971. (in Chinese)
- [10] 介邓飞,谢丽娟,饶秀勤,等. 近红外光谱变量筛选提高西瓜糖度预测模型精度[J]. 农业工程学报, 2013,29(12):264-270.
- JIE D F, XIE L J, RAO X Q, *et al.*. Improving accuracy of prediction model for soluble solids content of watermelon by variable selection based on near-infrared spectroscopy[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, 29(12):264-270. (in Chinese)
- [11] 石吉勇,邹小波,赵杰文,等. BiPLS结合模拟退火算法的近红外光谱特征波长选择研究[J]. 红外与毫米波学报,2011,30(5):458-462.
- SHI J Y, ZOU X B, ZHAO J W, *et al.*. Selection of wavelength for strawberry NIR spectroscopy based on BIPLS combined with SAA[J]. *J. Infrared Millim. Waves*, 2011,30(5):458-462. (in Chinese)
- [12] 尹慧敏. 基于波长组合的多元线性回归在近红外光谱波长优选中的应用[J]. 红外, 2010, 31(10): 31-34.
- YIN H M. Application of multiple linear regression based on wavelength combinations in near-infrared spectral wavelength optimization [J]. *Infrared*, 2010,31(10):31-34. (in Chinese)
- [13] 申艳,张晓平,梁爱珍,等. 多元散射校正和逐步回归法建立黑土有机碳近红外光谱定量模型[J]. 农业系统科学与综合研究,2010,26(2):175-179.
- SHEN Y, ZHANG X P, LIANG A ZH, *et al.*. Multiplicative scatter correction and stepwise regression to build NIRS model for analysis of soil organic carbon content in black soil [J]. *System Sciences and Comprehensive Studies in Agriculture*, 2010, 26(2):175-179. (in Chinese)

作者简介:



王智宏(1968—),女,吉林长春人,教授,博士生导师,1995年于吉林工业大学(现吉林大学)获得硕士学位,2005年于吉林大学获得博士学位,主要从事近红外光谱及应用技术的研究。E-mail:zhwang@jlu.edu.cn

通讯作者:



刘杰(1969—),女,副教授,1991年于吉林工业大学获得学士学位,1994年于中科院长春光机所获得硕士学位,主要从事光电传感仪器方面的研究。Email: l_jie@jlu.edu.cn