

基于近红外光谱的煤挥发分检测

肖 锋¹, 翟阳阳², 武宝明³, 王雅圣¹, 魏 高³, 孟国营³, 翟 红³, 李 光¹

(1. 浙江大学工业控制技术国家重点实验室 智能系统与控制研究所, 浙江 杭州 310027;

2. 首都医科大学, 北京 100069;

3. 汾西矿业(集团)有限责任公司, 山西 介休 032000)

摘要:近红外光谱分析是一种新型的在线分析技术, 十分适合于例如煤的挥发分等有机物的含量测定。本研究选取 139 个煤样, 采集煤样的近红外漫反射光谱, 利用偏最小二乘算法进行回归建模。同时采用不同的光谱波段筛选和预处理方法对模型进行优化, 取得一定效果。最后针对肥煤单独进行回归分析, 模型效果进一步优化, 显示了分类建模能提高模型的预测能力。

关键词:近红外漫反射; 煤挥发分; 偏最小二乘回归; 预处理; 分类

中图分类号: O657.33 **文献标识码:** A **DOI:** 10.3969/j.issn.1001-5078.2013.08.013

Detection of coal volatile matter by near-infrared spectroscopy

XIAO Feng¹, ZHAI Yang-yang², WU Bao-ming³, WANG Ya-sheng¹,

WEI Gao³, MENG Guo-ying³, ZHAI Hong³, LI Guang¹

(1. State Key Laboratory of Industrial Control Technology, Institute of Cyber Systems and Control,

Zhejiang University, Hangzhou 310027, China; 2. Capital Medical University, Beijing 100069, China;

3. Fenxi Ming Group Ltd, Jiexiu 032000, China)

Abstract: Near infrared spectroscopy is a new on-line analysis technology, which is fit for the detection of organic substance, such as volatile matter of coal. In this paper, 139 coal samples were chosen to acquire diffuse reflectance near-infrared spectra. Then the spectra were correlated to volatile matter of coal with partial least squares. At the same time, different spectral band selecting and pre-treatment methods were used to optimize regression model. At last, the regression model was built for fat coal, and the model result was significantly improved. It shows that regression model of homogeneous set can improve the prediction ability.

Key words: diffuse reflectance near-infrared; volatile matter of coal; partial least squares regression; pre-treatments; classification

1 引言

煤的挥发分是煤样在规定条件下隔绝空气加热, 并进行水分校正后的质量损失^[1]。去掉挥发分的残渣叫焦渣。挥发分不是指煤中原有的挥发性物质, 而是指煤在严格规定条件下, 加热分解的产物。煤的挥发分主要是由水分、碳氢化合物和碳氢氧化合物组成, 物理吸附水和矿物质分解生成的二氧化碳不属于挥发分。挥发分是对煤进行分类的主要指标, 根据煤的挥发分可以大致判断其变质程度、加工

利用性质和热值的高低。实验室化学检测挥发分的方法规范性很强, 实验条件需要严格控制, 因此对检测人员的要求非常高, 同时检测时间比较长。

将一束不同波长的近红外光照射到样本表面, 有一部分光射入样本, 在样本颗粒和内部经过多次反射、折射、衍射和吸收后返回样本表面, 这样的反

作者简介:肖锋(1987-), 男, 检测技术与自动化装置专业硕士研究生。E-mail: xfl0002006@gmail.com

收稿日期: 2013-01-07

射称为漫反射。被吸收的光转化为分子能量,使近红外光谱能表征物质性质。近红外光谱的吸收区域主要为含氢基团的合频以及倍频吸收,而煤的挥发分均与含氢基团相关,近红外光谱检测煤的挥发分理论上是可行的。

近红外光谱技术最早 20 世纪 60 年代被用来检测谷物的水分,其快速无损方便的检测特点使它得到快速的发展,除了农作物,纺织、化工、制药和烟草等行业,都研究近红外光谱的检测方法。M. P. Fuller 和 P. R. Griffiths^[3]利用傅里叶近红外漫反射光谱检测粉末物质的性质,Fysh 和 Swinkels^[4]定性分析了煤的近红外漫反射光谱特性。日本人 Mikio Kaihara^[5]研究了煤的近红外光谱与煤的化学性质的相关性,利用偏最小二乘法对光谱进行建模分析,模型精度较差。J. M. Andres 和 M. T. Bona 分析了煤的近红外^[6-7]和中红外^[8]的光谱特性,证实近红外光谱方法检测煤有机物成分能取得较好的效果。Hongfei Cheng^[9]等人分析了高岭土、煤系高岭土的中红外和近红外光谱差异,对建模过程中光谱的波段选取有一定的指导意义。

本文利用近红外漫反射光谱对煤的挥发分进行定量分析。回归建模前对光谱进行波段筛选和数学预处理,提高光谱信噪比。最后对肥煤单独建立回归模型,进行分析。

2 方 法

2.1 样本

本实验中共挑选 139 个煤样,覆盖瘦煤、肥煤、焦煤、气煤和贫瘦煤等多个不同的煤种。煤样提供方根据国家标准——煤的工业分析方法^[2]测定了全部样本的挥发分。样本挥发分的详细参数如表 1 所示。同时,煤样送到实验室进行近红外光谱采集之前,均被粉碎到 3 mm 以下。

2.2 近红外漫反射光谱采集

实验采用德国 Bruker 公司 MATRIX-I 型工业现场级 Fourier 变换近红外光谱仪采集煤样光谱。使用 Bruker 公司提供的 OPUS 6.0 软件进行光谱仪的控制和近红外光谱的采集。处理光谱数据的预处

理操作和回归模型建立均使用 MATLAB R2007b。

采集光谱时,为了使装入样品杯的煤更好地代表该煤样的性质,需要对煤样缩分。缩分是指煤样制备时将试样分成具有代表性的几个部分,其中一份或多份用于保留。在此,利用堆锥四分法对煤样缩分^[1],即将煤样均匀堆成圆锥体,然后压成厚度均匀的圆饼状,等分成四个扇形,取相对的两个扇形部分装入样品杯作为采集光谱的煤样。对每一个煤样重复缩分和装样 15 次取平均光谱作为该煤样的最终光谱。光谱采集过程中,为了进一步减小环境造成的误差,实验室温度保持在 $(21 \pm 1)^\circ\text{C}$,相对湿度保持在 $55\% \pm 5\%$ 。

煤样光谱采集波段选择 $10000 \sim 4000 \text{ cm}^{-1}$,分辨率 8 cm^{-1} ,64 次叠加平均。光谱获取模式选为吸收模式(absorbance)。

2.3 数据预处理

实验中采集的近红外漫反射光谱包含大量环境噪声,也有物理性质的影响。在利用光谱对煤的挥发分进行回归建模时,利用光谱预处理方法可以降低噪声,以及煤样颗粒度不均匀带来的影响,提高光谱的信噪比。在本研究中,采用了波段选取方法、多元散射校正平滑、一阶求导、二阶求导和无关变量消除的光谱预处理方法。

波段选择:实验中光谱采集选择 $10000 \sim 4000 \text{ cm}^{-1}$ 波段。如果找出光谱中与挥发分显著相关的波段进行建模,可以提高建模精度。实验中首先利用全波段进行建模,然后根据光谱特征定性选择存在光谱吸收峰的波段,最后根据光谱各波段与挥发分的相关性进行波段选择。

多元散射校正:近红外漫反射吸收光谱容易受到待测样品的颗粒度等影响,造成基线偏移。多元散射校正^[12]可以有效消除散射影响导致的基线漂移,提高光谱的信噪比。多元散射校正的方法通过建立一个待测样品的“理想光谱”,以该光谱为标准对其他样品的光谱进行修正。由于“理想光谱”无法得到,实际运用中选取所有光谱的平均光谱作为“理想光谱”。

表 1 煤样的挥发分参数

Tab. 1 volatile matter parameters of coal

	肥煤	焦煤	瘦煤	贫瘦煤	气煤	1/3 焦煤	全部煤样
煤样个数	31	33	25	33	4	12	139
均值/%	25.55	26.76	15.00	14.29	34.14	26.76	19.78
最小值/%	22.77	25.68	14.29	12.22	33.80	25.58	12.22
最大值/%	28.55	27.45	16.00	17.63	34.43	27.45	34.43

平滑和求导算法:平滑和求导均是常用的预处理方法,平滑可以用来降低光谱噪声,求导能提高光谱灵敏度。实验中选取 Savitzky-Golay 平滑和 Savitzky-Golay 求导的方法对光谱进行预处理。

无关变量消除:也是一种常用波段选择方法^[11]。无关变量消除基于偏最小二乘回归系数,通过光谱中混入噪声与原光谱一起作为因变量建模,比较噪声的系数和光谱的系数,认为光谱系数小于噪声系数的波段是于回归建模无关的变量,予以消除。

3 实验结果

3.1 煤近红外漫反射光谱

不同煤种采集的光谱会有一些细微的差别,如图1所示,选取样本量较多的肥煤、焦煤、瘦煤和贫瘦煤。瘦煤的近红外光谱与其他煤种的斜率相差较大,且各自谱峰的幅度和位置均有差别。从图1也可以发现光谱 1350 ~ 2400 nm (7406 ~ 4166 cm⁻¹) 波段谱峰变化显著。

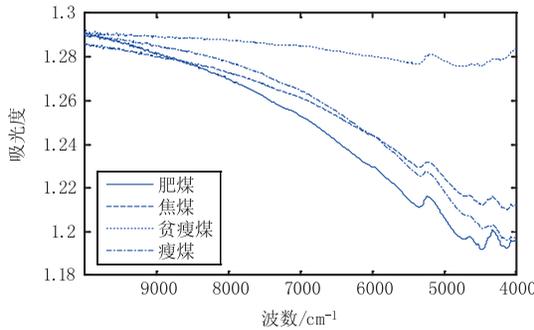


图1 4个不同煤样光谱图

Fig. 1 spectra of 4 different kinds of coal

3.2 模型建立

研究中利用的是偏最小二乘算法^[10]建立煤样光谱和挥发分的回归模型。模型好坏的标准选用内部交叉检验参数,交叉检验均方差 (root mean square error of cross validation, RMSECV) 和外部检验参数,预测均方差 (root mean square error of prediction, RMSEP)。

首先利用全部波段 (即 10000 ~ 4000 cm⁻¹ 波段) 进行建模分析,然后选取谱峰特征明显的 7406 ~ 4166 cm⁻¹ 波段建模,最后利用相关性定量进行选择波段,求取光谱各个波段与煤挥发分的相关性,如图2所示。同时对3种选取波段分别进行多元散射校正。表2所示为不同预处理得到的结果。

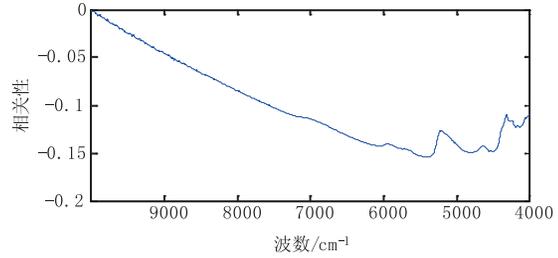


图2 光谱与煤挥发分相关性

Fig. 2 correlation between the spectra and volatile matter of coal

表2 选取波段和多元散射校正预处理的建模结果

Tab. 2 model results of different spectral ranges and MSC pre-treatment

方法		RMSECV/%	RMSEP/%
10000 ~ 4000cm ⁻¹	-	1.48	1.05
	MSC	1.58	1.48
7406 ~ 4166cm ⁻¹	-	1.03	0.89
	MSC	1.92	1.67
7251 ~ 4000cm ⁻¹	-	1.06	0.83
	MSC	1.93	1.99

从表2的结果可以看出选取波段之后 RMSECV 和 RMSEP 均明显变小,模型精度和稳定性得到改善。选择谱峰明显波段进行建模可以得到较好的结果。各个波段光谱与定标值的相关性都很低。利用相关性进行波段选择时不同的相关性阈值可以得到不同的结果,选择 0.11 时结果最好,此时对应 7251 ~ 4000 cm⁻¹ 波段,与根据光谱定性选择波段的方法得到结果类似。

多元散射校正的预处理方法没有使结果变好,选取波段之后进行多元散射校正的结果更差。Bruker 近红外光谱仪的近红外光从样品杯底部照射,取其漫反射光谱。样品表面比较平整,颗粒度不匀造成的散射影响小,因此使用多元散射校正的方法没有使结果变好。

选取 7406 ~ 4166 cm⁻¹ 波段,然后 Savitzky-Golay 平滑和求导,得到结果如表3所示。

表3 Savitzky-Golay 平滑和求导预处理建模结果

Tab. 3 model results of Savitzky-Golay smoothing and Savitzky-Golay derivative

方法	窗口宽度	多项式次数	RMSECV/%	RMSEP/%
平滑	7	2	1.05	0.88
一阶求导	7	2	1.28	1.02
二阶求导	5	2	2.53	2.63

平滑的方法可以用来消除噪声,有效信息也会被削弱。对比光谱经过平滑和未平滑的结果,可以发现结果接近。平滑没有明显优化建模效果。求导之后模型内部交叉验证结果变差,预测能力降低,尤其二阶求导之后。求导算法放大了噪声,减低信噪比,模型的稳定性和预测精度没有提高。

光谱采集和定标值测量过程中不可避免出现一些误差,误差的存在会打破光谱和定标值的对应关系,建立模型时会对模型造成比较大的影响,这样的样本点称为异常点,在建模过程中应该剔除。实验中选用学生氏残差(Studentized Residual)剔除异常点。学生氏残差是多元统计分析中常用的异常点剔除算法。计算公式如下所示:

$$ZRSEID_{e_i} = \frac{e_i}{S_e} = \frac{y_i - \hat{y}_i}{S_e}$$

其中, y_i 表示样本 i 的定标值; \hat{y}_i 表示样本 i 的预测值; S_e 为样本预测偏差的标准差。

图3所示是针对 $7406 \sim 4166 \text{ cm}^{-1}$ 波段和 $7406 \sim 4166 \text{ cm}^{-1}$ 波段 Savitzky-Golay 平滑运用学生氏残差剔除异常点的结果,学生氏残差阈值取 2.5。

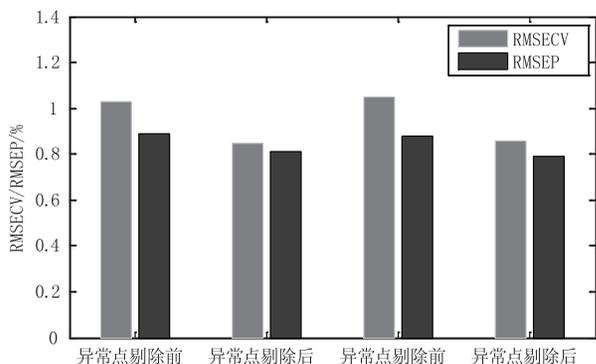


图3 学生氏残差剔除异常点结果(从左到右依次是 $7406 \sim 4166 \text{ cm}^{-1}$ 波段剔除异常点前, $7406 \sim 4166 \text{ cm}^{-1}$ 波段剔除异常点后, $7406 \sim 4166 \text{ cm}^{-1}$ 波段 Savitzky-Golay 平滑剔除异常点前, $7406 \sim 4166 \text{ cm}^{-1}$ 波段 Savitzky-Golay 平滑剔除异常点后)

Fig.3 results of deleting outliers with semi-studentized residual (From left to right: $7406 \sim 4166 \text{ cm}^{-1}$ with outliers, $7406 \sim 4166 \text{ cm}^{-1}$ deleting outliers, $7406 \sim 4166 \text{ cm}^{-1}$ and Savitzky-Golay smoothing with outliers, $7406 \sim 4166 \text{ cm}^{-1}$ and Savitzky-Golay smoothing deleting outliers)

可以看出,虽然模型预测结果(RMSEP)只是稍微改善,但是 RMSECV 明显变小,学生氏残差剔除异常点可以提高模型的稳定性和预测精度。

对于选取波段 $7406 \sim 4166 \text{ cm}^{-1}$ 进行 Savitzky-Golay 平滑剔除异常点的情况进行无关变量消除, $RMSECV = 0.57$, $RMSEP = 0.78$ 。其预测集的预测

效果如图4所示。

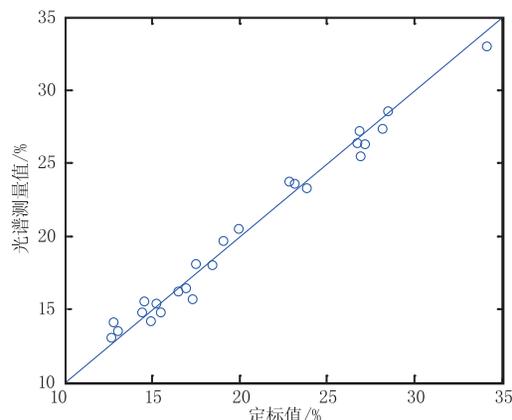


图4 挥发分预测效果图

Fig.4 prediction effect of volatile matter

3.3 分类建模探索

全体煤样一起建模时,不同煤种中粘结度,胶质层指数等物理性质上的不同会反应在光谱中,这些信息在建模时会作为噪声影响模型的结果,而一般的预处理方法无法有效的消除这些因素的影响。因此考虑如果对不同类型的煤样分类建模。相同种类的煤样物理化学性质接近,建模时能降低噪声的干扰。

本次样品中包含6种煤样,如表1所示。焦煤和瘦煤挥发分的标准差较小,覆盖范围小,气煤和1/3焦煤的样本数量又太少。综合考虑挥发分的分布范围和样本个数,选用肥煤进行分类建模探索。

进行回归建模时同样将肥煤分成预测集和校正集,均匀选取6个样本作为预测集,25个作为校正集。建模结果如表4所示。

表4 肥煤回归建模结果

Tab.4 regression model results of fat coal

方法		RMSECV/%	RMSEP/%
10000 ~ 4000 cm^{-1}	-	1.23	1.40
	UVE	0.68	1.38
7406 ~ 4166 cm^{-1}	-	0.82	0.73
	UVE	0.40	0.67
7406 ~ 4166 cm^{-1} 平滑	-	0.80	0.71
	UVE	0.43	0.68

由于求导和多元散射校正的方法对全部煤样的建模效果不好,因此没有采用。对肥煤分别进行全波段, $7406 \sim 4166 \text{ cm}^{-1}$ 和 $7406 \sim 4166 \text{ cm}^{-1}$ 平滑的分析,同时利用学生氏残差进行了异常点剔除。平滑采用移动窗口宽度7,多项式次数2的 Savitzky-Golay 平滑。其中,选取 $7406 \sim 4166 \text{ cm}^{-1}$ 波段然后运用无关变量消除后, $RMSECV$ 和 $RMSEP$ 为 0.40,

0.67。对比建模的结果可以发现对肥煤的建模精度和稳定性明显优于整体煤样。证明对煤样先进行分类然后回归建模是可行的。

4 讨论和结论

实验中,为了最大程度降低光谱采集过程中的噪声,严格控制了实验室的温度和湿度范围,尽量减小环境变化对光谱的影响。同时对煤样进行15次重复装样,降低由于煤样表面不均匀或者环境原因产生的随机误差影响。

数据分析过程中,选取偏最小二乘算法对光谱进行回归分析。同时,尝试不同的预处理方法,如平滑、求导和波段选取,以及针对颗粒物品的多元散射校正对光谱进行处理。分析证实在对煤样光谱与挥发分含量进行回归建模时,波段选取能改善模型结果,通过选择含有效信息更多的波段可以提高光谱的信噪比。进行学生氏残差剔除异常点之后建模结果明显改善。利用无关变量消除再次挑选与挥发分相关性大的波段进行建模,进一步改善了模型结果。最后对分类建模的方法进行了尝试,选取肥煤进行类似的回归建模分析,结果显示单独对肥煤建模的效果优于全体煤样。可以先对煤样分类然后回归建模的方法。

挥发分主要由水分、碳氢氧化物和碳氢化合物组成,而近红外光谱对含氢基团物质检测能力较好。理论上证实了煤的近红外光谱能包含挥发分信息,而近红外光谱方法用于检测煤的灰分和硫分时,因为灰分主要由无机物构成,而硫分则包含了有机硫和无机硫,导致近红外光谱对煤灰分和硫分的检测精度不如对挥发分的检测精度。

近红外漫反射光谱方法检测煤挥发分含量,无损快速的检测特征让其具有良好的运用前景。需要增加样本数量以提高模型的稳定性和精度,同时运用分类建模的方法进一步提高检测精度。

参考文献:

- [1] GB/T 3715-2007. 煤质及煤分析有关术语[S]. 北京:中国标准出版社,2007.
- [2] GB/T 212-2008, 煤的工业分析方法[S]. 北京:中国

标准出版社,2008.

- [3] M P Fuller, P R Griffiths. Diffuse reflectance measurements by infrared Fourier transform spectrometry[J]. Analytical Chemistry, 1978, 50(13): 1906-1910.
- [4] A Fysh, A J Swinkels. Near-infrared diffuse reflectance spectroscopy of coal[J]. Applied Spectroscopy, 1985, 39(2): 354-357.
- [5] Mikio Kaihara, Toshihiro Sato, et al. Application of near infrared spectroscopy to rapid analysis of coals[J]. Spectroscopy Letters: An International Journal for Rapid Communication, 2002, 35(3): 369-376.
- [6] M T Bona, J M Andres. ASTM clustering for improving coal analysis by near-infrared spectroscopy[J]. Talanta, 2006, 70: 711-719.
- [7] M T Bona, J M Andres. Coal analysis by diffuse reflectance near-infrared spectroscopy: Hierarchical cluster and linear discriminant analysis[J]. Talanta, 2007, 72(4): 1423-1431.
- [8] M T Bona, J M Andres. Application of chemometric tools for coal classification and multivariate calibration by transmission and drift mid-infrared spectroscopy[J]. Analytica Chimica Acta, 2008, 624: 68-78.
- [9] Cheng Hongfei, Zhang Jinshan, et al. A spectroscopic comparison of selected Chinese kaolinite, coal bearing kaolinite and halloysite: a mid-infrared and near-infrared study[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2010, 77(4): 856-861.
- [10] S Wold. PLS-regression; a basic tool of chemometrics[J]. Chemometrics and Intelligent Laboratory Systems, 2001, 58: 109-130.
- [11] Shao Xueguang, Wang Fang, Cheng Da, et al. A method for near-infrared spectral calibration of complex plant wavelet transform and elimination of uninformative variables[J]. Analytical and Bioanalytical Chemistry, 2004, 378(5): 1382-1387.
- [12] Inge S. Helland, Tormod Nas, Tomas Isaksson. Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data[J]. Chemometrics and Intelligent Laboratory Systems, 1995, 25: 233-241.