

偏最小二乘法结合傅里叶变换红外漫反射光谱快速测定煤质

王家俊¹, 汪帆², 马玲¹

(1. 红河卷烟总厂 产品中心, 云南 弥勒 652300; 2. 曲靖师范学院 化学系, 云南 曲靖 655000)

摘要: 应用偏最小二乘法(PLS)结合烟煤的傅里叶变换红外漫反射光谱, 分别建立了预测发热量、挥发分和灰分的校正模型, 其相关系数分别为: 0.976 7、0.903 4 和 0.978 9; 全交互效验均方残差(RMSECV)分别为: 156.36、0.35 和 1.36; 预测范围分别为: 4 000 ~ 6 500 Cal/g、13% ~ 16%、18% ~ 40%。通过对模型进行成对 *t*-检验, 在显著性水平 0.05 的条件下, 其测定结果与标准方法对比, 两者无显著性差异。该法应用于快速分析煤质, 结果令人满意。

关键词: 偏最小二乘法; 傅里叶变换红外漫反射光谱; 煤质

中图分类号: O657.33 **文献标识码:** A **文章编号:** 1004-4957(2006)03-0060-04

Rapid Determination of Soft Coal Quality by Partial Least Squares and Fourier Transform Infrared Diffusion Reflectance Spectra

WANG Jia-Jun¹, WANG Fan², MA Ling¹

(1. Production Research Center, Honghe Cigarette General Factory, Mile 652300, China;

2. Department of Chemistry, Qujing Teacher's College, Qujing 655000, China)

Abstract: Calibration models were established for the determination of calorific value, volatile matter and ash content of soft coal by partial least squares (PLS) and Fourier transform infrared diffusion reflectance spectra. The correlation coefficients of the calibration models were 0.976 7, 0.903 4 and 0.978 9, and the root mean square errors of cross validation (RMSECV) were 156.36, 0.35 and 1.36 in the predicted range of 4 000 - 6 500 Cal/g, 13% - 16% and 18% - 40% for calorific value, volatile matter and ash content, respectively. By statistical significance test, the results of determination were compared with those of standard methods with no significant difference at 0.05 level. The method has been applied to rapid determination of calorific value, volatile matter and ash content with satisfactory results.

Key words: Partial least squares method; Fourier transform infrared diffusion reflectance spectra; Soft coal quality

发热量、挥发分和灰分等是衡量煤质的质量指标。通常, 烟煤的发热量、挥发分和灰分是按行业标准方法^[1]分别进行测定的, 操作过程繁琐, 速度慢, 且费用高。随着化学计量学多元校正方法的发展与广泛应用, 董庆年采用最小二乘法结合红外光谱成功测定了灰分^[2]; 而应用偏最小二乘法^[3](PLS)结合傅里叶变换红外漫反射光谱同时测定煤质的多个质量指标少见报道。

通过选择本地区具有代表性的 200 多个分析煤样进行实验、建模, 研究发现, 应用 PLS 把标准方法测定的发热量、挥发分和灰分等基础数据与其相应的红外漫反射光谱数据进行拟合, 表现出良好的相关性。就此, 分别建立了预测烟煤发热量、挥发分和灰分等质量指标的校正模型, 并应用于预测未进入校正集的煤样, 经成对 *t*-检验表明, 在显著水平 0.05 时, 模型的预测值与行业标准方法的测定值无显著性差异。又因本法具有操作简便、快速的优点, 应用于煤质多个质量指标的同时测定是可行的。此外, 本文还对光谱的预处理方法进行了讨论。

收稿日期: 2005-05-05; 修回日期: 2005-07-22

作者简介: 王家俊(1962-), 男, 云南弥勒人, 工程师, Tel: 0873-6196095, E-mail: honghe@vip.163.com

1 实验部分

1.1 主要仪器及标样

Nicolet NEXUS 670 FT-IR 光谱仪, 带 DTGS-KBr 检测器; 漫反射采样附件; TQ Analyst 6.2 定量分析软件(美国 Nicolet 公司)。光谱扫描范围: $4\ 000 \sim 400\ \text{cm}^{-1}$; 分辨率: $8\ \text{cm}^{-1}$; 扫描次数: 64。

建模所用分析煤样及其相应的发热量、挥发分和灰分等基础数据均按标准方法^[1]制备和测定。

1.2 实验方法

1.2.1 采集标样光谱数据 开机预热 2 h 后, 将制好的分析煤样直接装入漫反射采样附件, 即可采集分析煤样的傅里叶变换红外漫反射光谱。

1.2.2 建立校正模型 采用 Karl Norris 平滑滤波^[4]与二阶求导, 对样本光谱数据进行滤噪和基线校正处理, 应用 PLS 与对应的发热量、挥发分和灰分等基础数据关联, 结合“剔一”(Leave-one-out)的交互效验(Cross validation)方法, 剔除异常值, 逐步优化模型, 使其达到最优状态。

2 结果与讨论

2.1 光谱数据的预处理

所采集的煤烟红外漫反射光谱(见图 1a), 存在明显的噪音和基线飘移, 在通常情况下, 可采用 Savitzky-Golay 平滑滤波^[5]结合一阶或二阶求导, 预过滤噪音, 消除基线飘移的干扰。本实验采用 Karl Norris 平滑滤波, 获得了较好的效果。在实际中, 原始光谱的信号质量与样品光谱特性, 仪器参数设置和仪器硬件水平等因素有关, 在采用 Karl Norris 平滑滤波时, 窗口宽度的选择具有一定的经验性, 应根据具体情况, 选择段长(segment length)和段间距(gap between segments), 一般来说, 较大的段长会造成信号失真, 较小的段间距, 滤噪效果不佳, 应尽可能用较小的段长来建立理想的模型; 高分辨率的光谱宜选择较小的段间距, 反之, 选择较大的段间距。在本实验中, 就所使用的仪器和测定的对象, 段长不宜大于 13, 段间距在 4~6 之间比较合适。表 1 为选取不同段长和段间距结合一阶或二阶求导的建模效果比较(以灰分为例), 其中采用段长 7, 段间距 5 的 Karl Norris 平滑滤波结合二阶导数光谱(见图 1b), 用于建模, 效果理想。

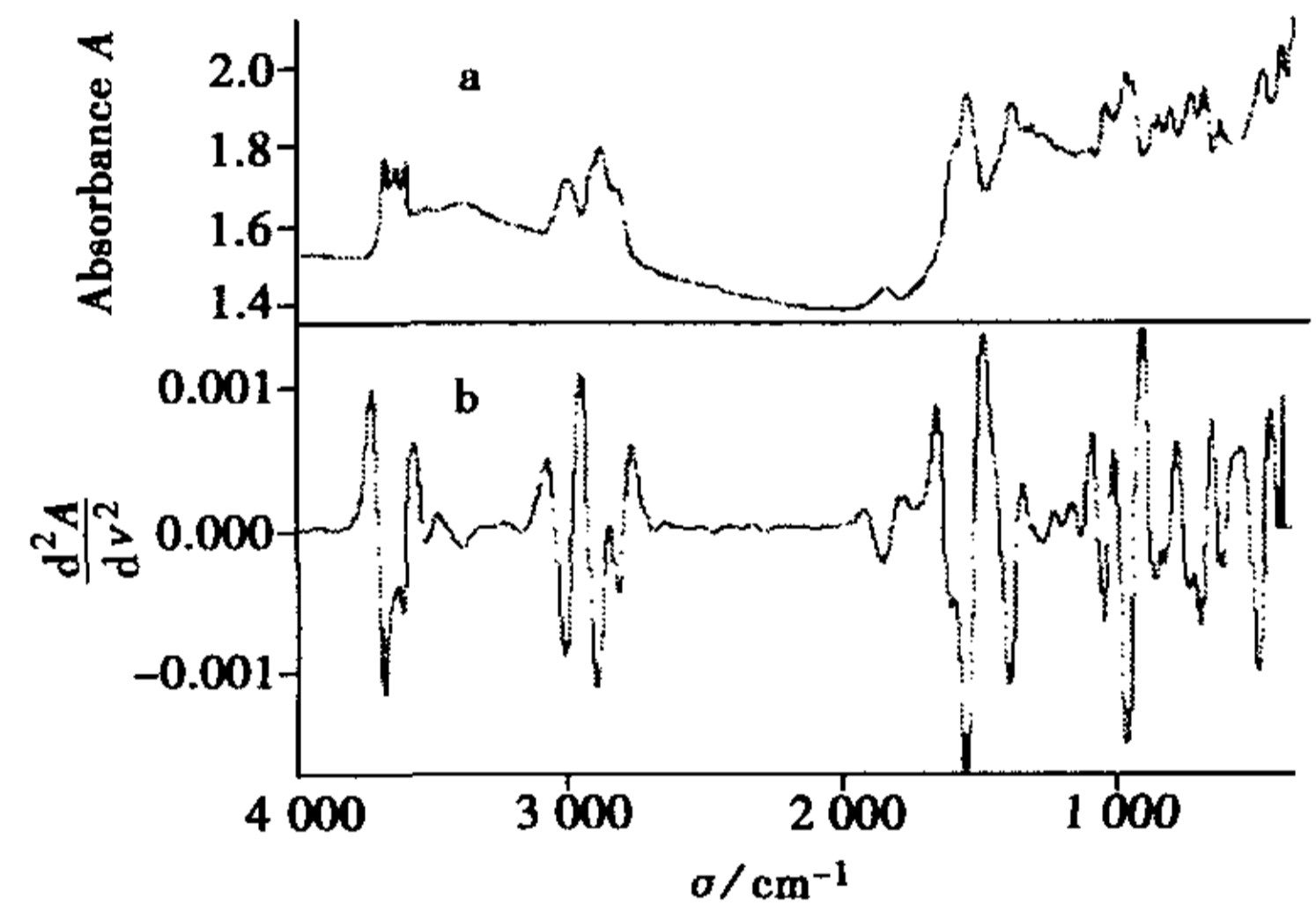


图 1 烟煤的红外漫反射光谱(a)和二阶导数光谱(b)

Fig. 1 FT-IR diffusion reflectance spectra of soft coal (a) and their 2nd derivative spectra(b)

表 1 选择不同段长和段间距平滑滤波建立的灰分校正模型的统计结果

Table 1 Statistics results for calibration models of ash with different segment length and gap between segments

Karl Norris filter		1st derivative			2nd derivative		
Segment length	Gap between segments	Factor number	Correlation coefficient	RMSECV 交互效验 均方残差	Factor number	Correlation coefficient	RMSECV 交互效验 均方残差
段长	段间距	主因子数	相关系数		主因子数	相关系数	
3	1	4	0.977 2	1.44	4	0.970 0	2.22
3	5	6	0.983 7	1.42	6	0.981 6	1.55
7	5	8	0.9824	1.41	6	0.978 9	1.36
7	7	8	0.966 4	1.37	6	0.977 7	1.35
13	5	10	0.980 2	1.37	6	0.976 2	1.35
13	7	10	0.979 8	1.35	6	0.976 6	1.35
21	5	9	0.976 9	1.36	8	0.975 4	1.35
21	7	9	0.976 4	1.39	8	0.974 9	1.39

此外, 选取建模最佳谱区也是光谱数据预处理的重要部分, 虽然定量分析软件会推荐最佳谱

区,但在模型优化过程中,常因异常值的剔除,会导致推荐的谱区发生变化。而依据光谱与预测性质表现出来的统计特征,可以准确地选取最佳谱区。现以建立灰分模型为例说明,图2a为灰分的谱区与相关系数的关系,虽然全谱区(4 000 ~ 400 cm^{-1})与灰分这一性质表现出较高的相关性,但有效信息率较高的谱区是其中大方差的谱区(3 900 ~ 3 500 cm^{-1} , 3 200 ~ 2 600 cm^{-1} , 2 100 ~ 450 cm^{-1} 等),这是因为方差光谱表明了校正集中样本光谱的吸光度在不同谱区的差异程度,见图2b,同时它也表征了光度变化的谱区与样品灰分含量变异的相关性^[6],组合这些大方差的谱区建模,可获得理想效果,见表2。

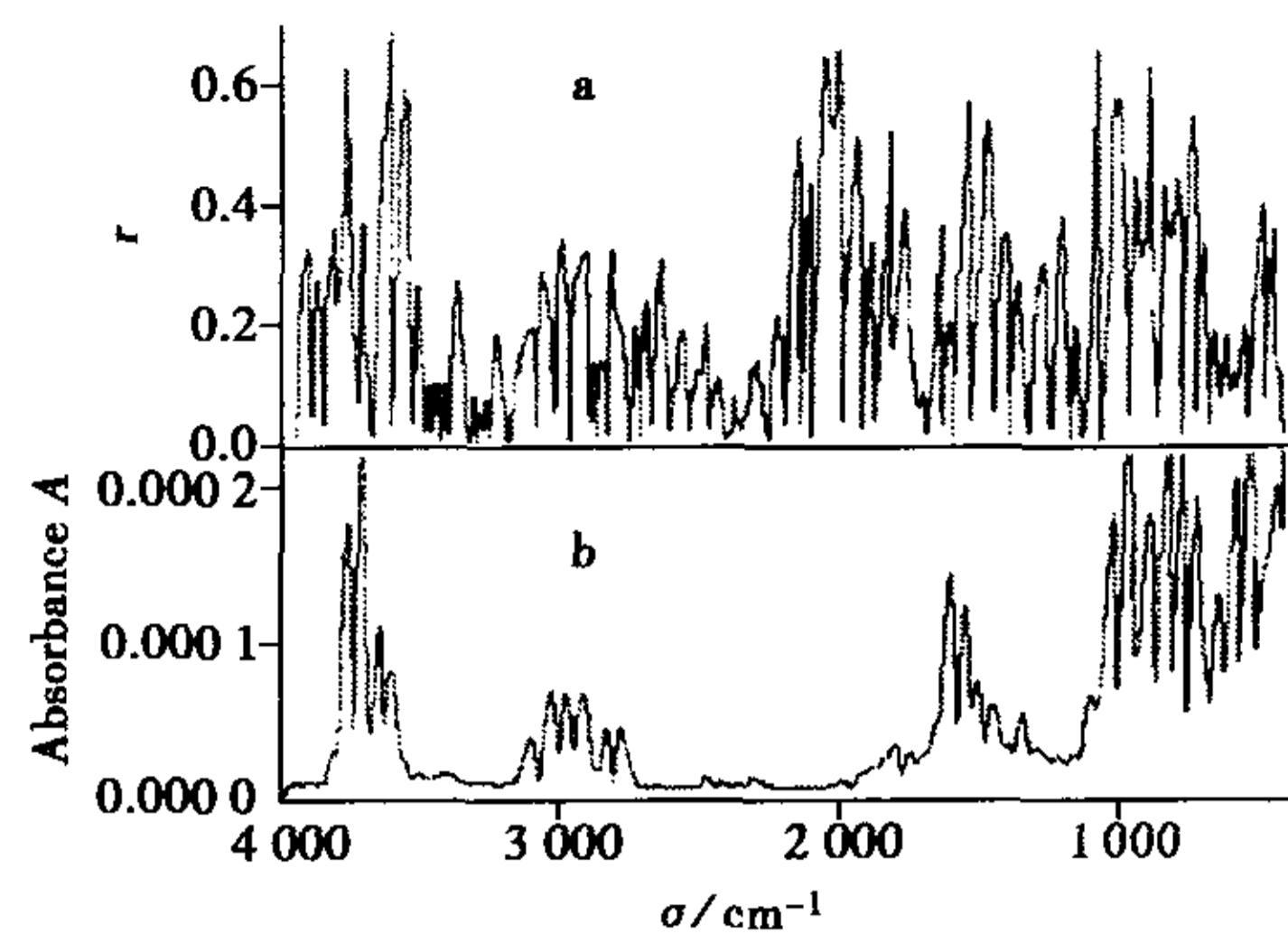


图2 灰分的谱区与相关系数 r 的关系图(a)和其方差光谱图(b)

Fig. 2 Spectral region versus correlation coefficient r for ash(a) and their variance spectrum(b)

表2 采用不同谱区建立的灰分模型的统计结果

Table 2 Statistics results for calibration models of ash with different spectral regions

Spectral regions(谱区) σ/cm^{-1}	Data processing 光谱数据处理	Correlation coefficient 相关系数	Factor number 主因子数	RMSECV 交互效验均方残差	
4 000 ~ 450	Norris filter(Segment length: 7, Gap between segments: 5) + 2nd derivative	0.972 8	4	1.45	
3 900 ~ 3 500					
3 200 ~ 2 600			0.978 9	6	1.36
2 100 ~ 450					

2.2 校正模型的建立

应用偏最小二乘法建模是一个协同过程,基于上述方法对光谱进行预处理的同时,为了获得稳健、理想的模型,一是要剔除异常值,消除对模型的影响和扰动;二是要避免“欠拟合或过拟合”未充分利用信息或引入太多的噪声,这需要采用最佳主因子来建模,最佳主因子的确定有自预测法(self prediction)、杠杆预测法(leverage prediction)和交互效验法等^[7]。本实验采用结合“剔一”的交互效验法,效果理想。图3为灰分的标准值(actual)与模型预测值(calculated)的散点图和相应的主因子(factor)与全交互效验均方残差的变化图,其中最佳主因子为6,即当交互效验均方残差(RMSECV)达到最小时的主因子。其余的发热量和挥发分等性质的最佳主因子确定亦然,采用最佳主因子建立的模型见表3。

表3 发热量、挥发分和灰分等校正模型的有关数理指标

Table 3 Statistics results for calibration models of calorific value, volatile matter and ash content

Model name 模型名称	Number of samples 校正集样本量	Correlation coefficient 相关系数	Factor number 主因子数	RMSECV 交互效验均方残差	Predicted range 预测范围
Calorific value(发热量)	191	0.976 7	6	156.36	4 000 ~ 6 500 (Cal/g)
Volatile matter(挥发分)	150	0.903 4	4	0.35	13% ~ 16%
Ash content(灰分)	204	0.978 9	6	1.36	18% ~ 40%

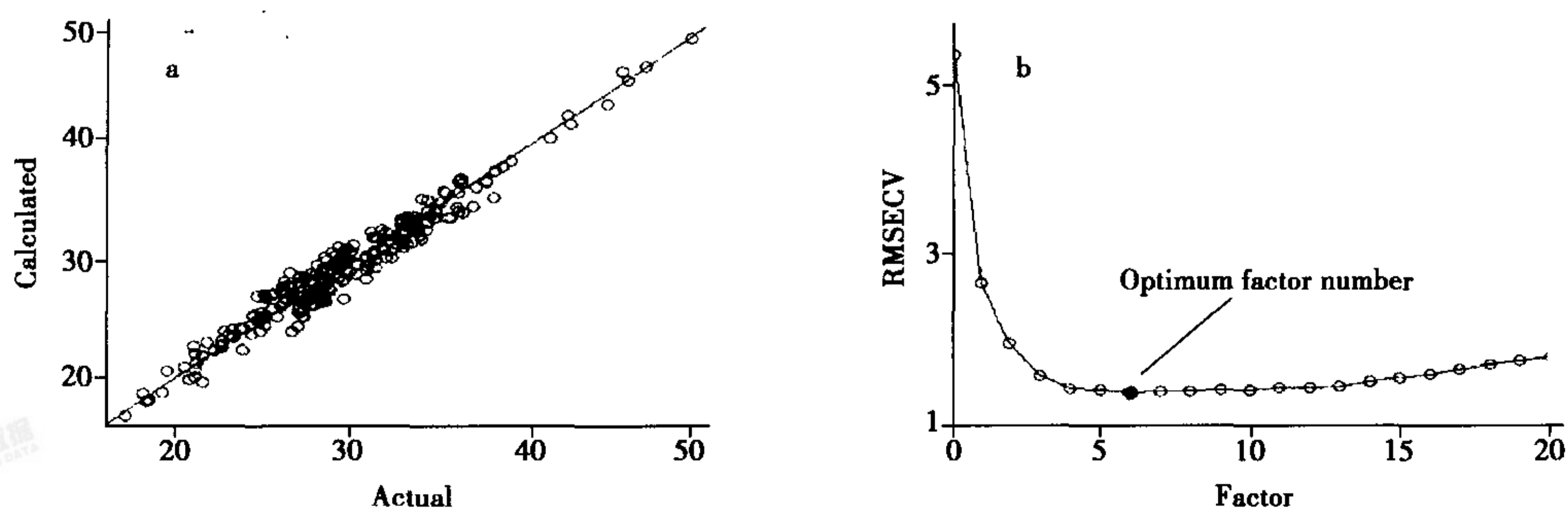


图3 灰分的标准值与模型预测值的散点图(a)和交互效验均方残差与随主因子数的变化图(b)

Fig. 3 Actual value versus calculated value of ash(a) and changes of RMSECV with factor numbers(b)

2.3 准确性试验

通常, 从以上过程基本可以判断模型的预测能力。但在建模过程中, 免不了要删除一些样本, 从而影响到校正集的代表性及模型的可靠性。为此, 另行随机抽取 16 个样本作为独立校验集进一步验证模型, 即用以上建好的模型和标准方法进行测定, 测定结果见表 4, 通过成对 t -检验显示, 在显著性水平 0.05 时, t 分布值均小于临界值 ($t(0.05, 15) = 2.131$), 说明两种测定方法不存在显著性差异, 即表明两种方法测定结果吻合。

表 4 本方法(PLS-FT-IR)与标准方法测定结果的比较($n=16$)Table 4 Comparison results of PLS-FT-IR method and standard methods($n=16$)

No.	Calorific value/(Cal/g)			Volatile matter/%			Ash content/%		
	GB 213-87	PLS-FT-IR	Deviation	GB 212-77	PLS-FT-IR	Deviation	GB 212-77	PLS-FT-IR	Deviation
1	5 794.70	5 754.59	40.11	15.80	15.49	0.31	29.06	27.40	1.66
2	5 378.52	5 217.65	160.87	14.78	14.72	0.06	30.05	29.90	0.15
3	4 828.70	4 787.20	41.50	14.22	14.25	-0.03	31.69	30.74	0.95
4	4 842.47	4 882.32	-39.85	14.78	14.21	0.57	23.75	24.27	-0.52
5	5 418.57	5 482.80	-64.23	14.78	14.54	0.24	28.33	29.79	-1.46
6	4 924.21	5 009.62	-85.41	14.17	14.09	0.08	44.10	42.56	1.54
7	4 908.58	4 796.42	112.16	14.27	14.43	-0.16	18.28	19.25	-0.97
8	4 958.05	5 018.58	-60.53	14.71	14.45	0.26	38.75	37.03	1.72
9	4 828.70	4 787.20	41.50	14.50	14.43	0.07	33.23	32.52	0.71
10	4 908.58	4 796.42	112.16	14.90	14.72	0.18	32.86	34.07	-1.21
11	4 924.21	5 009.62	-85.41	14.39	14.57	-0.18	29.22	30.83	-1.61
12	5 240.91	5 189.85	51.06	15.37	14.92	0.45	29.43	29.37	0.06
13	5 378.52	5 217.65	160.87	13.87	14.05	-0.18	26.89	28.47	-1.58
14	5 418.57	5 482.80	-64.23	14.31	14.43	-0.12	28.16	26.92	1.24
15	5 544.84	5 483.84	61.00	14.46	14.56	-0.10	28.40	27.30	1.10
16	5 794.70	5 754.59	40.11	15.18	15.20	-0.02	24.90	25.14	-0.24
t -Value		1.248 8			1.563 4			0.317 4	

2.4 重复性试验与精度改善

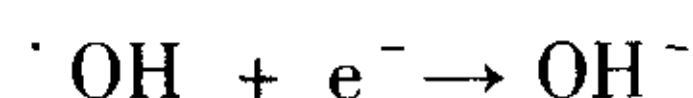
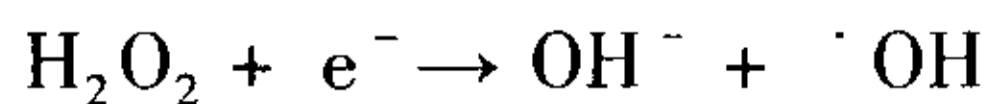
对同一样品进行 5 次测定, 结果见表 5, 与标准方法相应的重复性(发热量、挥发分和灰分的分别为: 36 Cal/g, 0.30% 和 0.5%) 比较, 该法测定挥发分有较好的测量精度, 而发热量和灰分不够理想, 这主要是由于单次采集红外光谱的重复性弱所致。但可通过重复采集(3~5 次)求平均光谱, 然后根据平均光谱预测, 这样, 可使其预测获得更好的代表性, 同时其预测精度也能够满足要求。表 5 中显示的平均值与该样本的标准方法测定值(发热量、挥发分和灰分分别为: 4 300.00 Cal/g、14.70% 和 30.13%) 吻合较好。

表 5 重复性试验($n=5$)Table 5 The repeatability test($n=5$)

No.	Calorific value/(Cal · g ⁻¹)	Volatile matter/%	Ash content/%
1	4 289.71	14.64	30.72
2	4 218.96	14.83	30.63
3	4 344.94	14.75	29.60
4	4 388.80	14.59	29.39
5	4 201.63	14.94	29.94
Mean	4 288.81	14.75	30.06
Standard deviation	80.04	0.14	0.60
RSD s_r /%	1.86	0.96	1.99

3 小 结

本实验应用偏最小二乘法结合傅里叶变换红外漫反射光谱, 建立了同时测定煤质发热量、挥发分和灰分等多个质量指标的校正模型, 通过独立校验集验证表明, 准确性好。为了克服单次采集红外光谱重复性弱的不足, 进一步提高采样代表性和预测精度, 可通过重复采集样本光谱, 并求其相应的平均光谱, 然后应用平均光谱进行预测。又因本法操作过程简捷, 用于煤质快速分析是可行的。



即溶解氧分子首先经由超氧阴离子($\text{O}_2^{\cdot-}$)在 -0.2 V 被还原为 H_2O_2 , H_2O_2 随后经由羟基自由基($\cdot\text{OH}$) 在 -1.00 V 被还原为 H_2O 。由于肉桂酸的极谱还原峰出现在 -1.50 V , 比 H_2O_2 的还原峰峰电位更负, 因此, 是由 $\text{O}_2^{\cdot-}$ 产生的 H_2O_2 再生的 $\cdot\text{OH}$ 在滴汞电极上还原催化了肉桂酸极谱还原峰峰电流, 导致其还原峰峰电流增加。具体的还原过程为:

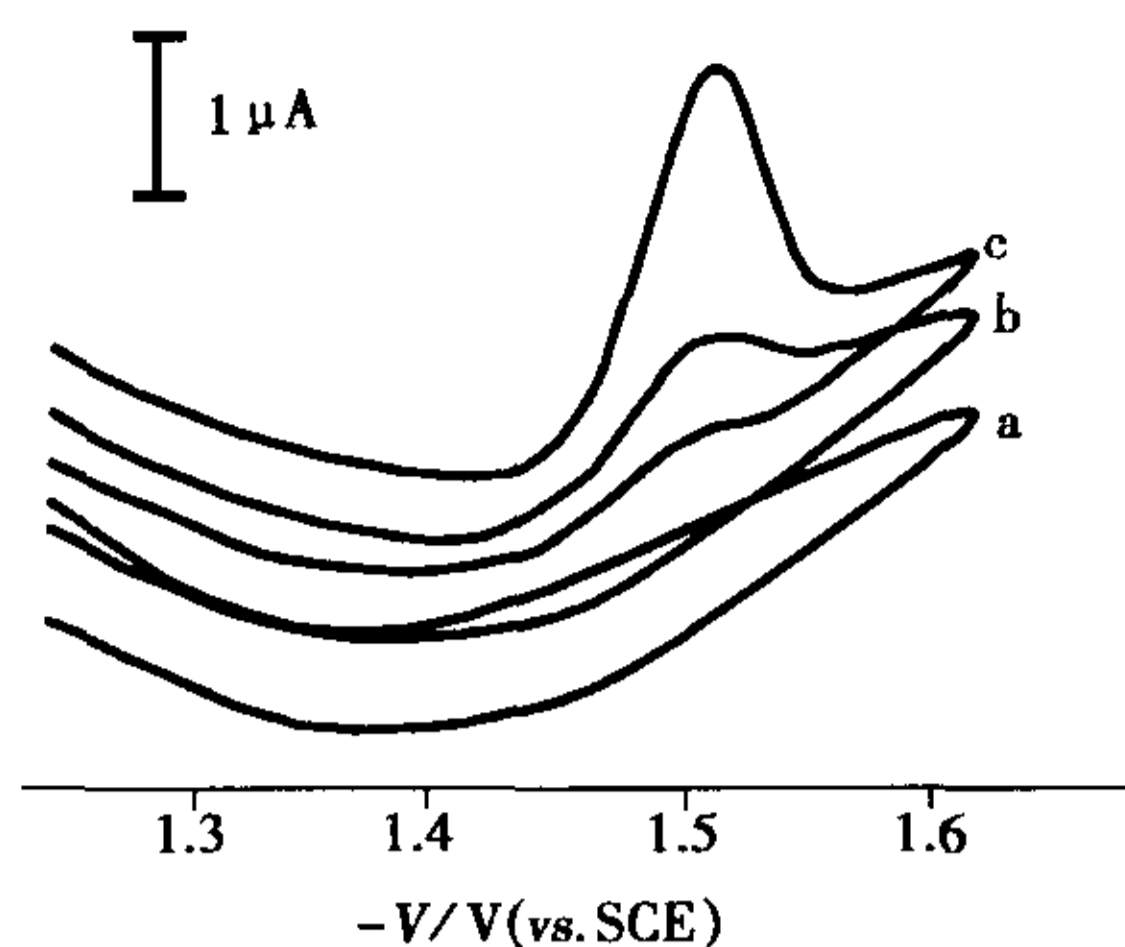
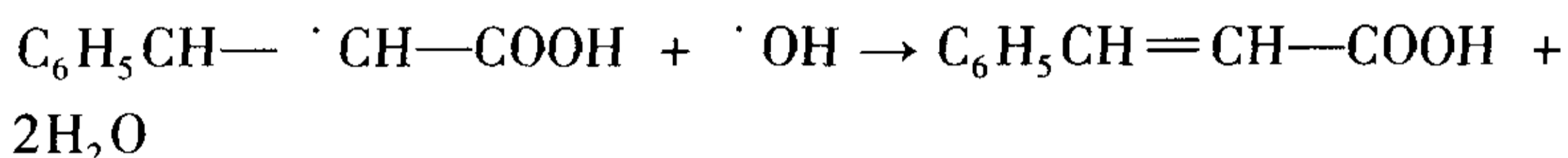
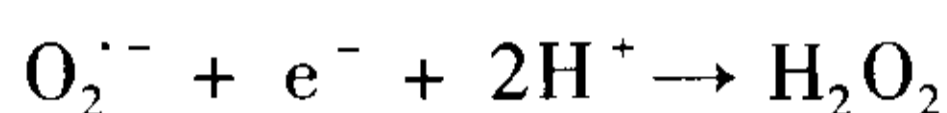
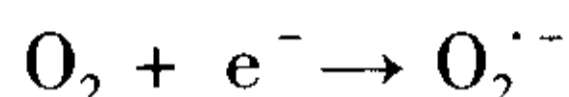


图4 肉桂酸的循环伏安图

Fig. 4 Cyclic voltammogram of cinnamic acid

a. 1.0 mol/L HAc - NaAc buffer ($\text{pH } 4.1$, without deaeration); b. a + $8.0 \times 10^{-5}\text{ mol/L}$ cinnamic acid (after deaeration for 20 min); c. a + $8.0 \times 10^{-5}\text{ mol/L}$ cinnamic acid (without deaeration); Potential scan rate $v = 250\text{ mV/s}$

参考文献:

- [1] 徐向阳, 张 蕙, 陈 希. HPLC 法测定脉络宁口服液中肉桂酸的含量[J]. 江苏药学与临床研究, 2003, 11(1): 14 - 16.
- [2] 刘传华, 陈沪宁, 刘师莲, 等. BR - HPLC 测定桂龙咳喘宁胶囊中肉桂酸的含量[J]. 中国中药杂志, 2004, 29(3): 276 - 277.
- [3] 张 玲, 单卫华. 双波长薄层扫描法测定肉桂中肉桂酸的含量[J]. 药物分析杂志, 1997, 17(6): 408 - 410.
- [4] 翟海云, 徐健君, 陈缙光, 等. 毛细管电泳高频电导法快速测定桂枝中的桂皮酸[J]. 中草药, 2005, 36(1): 109 - 111.
- [5] 李丽敏. 线性扫描极谱法测定微量肉桂酸[J]. 理化检验: 化学分册, 2002, 38(6): 295 - 298.
- [6] 亢晓峰, 过 玮, 赵 川, 等. H_2O_2 - 四丁基溴化铵 - 醋酸盐体系中肉桂酸的吸附平行催化波研究[J]. 中国科学, B, 2000, 30(4): 317 - 322.
- [7] CHEVALET J, ROUELLE F, GOERST L, et al. Electrogenation and some properties of the superoxide ion in aqueous solutions[J]. J Electroanal Chem, 1972, 39(1): 201 - 216.

(上接第 63 页)

参考文献:

- [1] 北京煤化学研究所. 煤炭试验方法标准及其说明[S]. 北京: 中国标准出版社, 1992: 362 - 392.
- [2] 吴瑾光. 近代傅里叶变换红外光谱技术及应用[M]. 第二卷. 北京: 科学技术文献出版社, 1994: 76 - 78.
- [3] 梁逸曾. 白灰黑复杂多组分分析体系及其化学计量学算法[M]. 长沙: 湖南科技出版社, 1996: 32 - 36.
- [4] NORRIS K H, WILLIAMS P C. Optimization of mathematical treatments of raw near-infrared signal in the measurement of protein in Hard Red Spring Wheat[J]. Cereal Chem, 1984, 61: 158 - 160.
- [5] SAVITZKY A, GOLAY M J E. Smoothing and differentiation by simplified least squares procedures[J]. Anal Chem, 1964, 36: 1627 - 1639.
- [6] Edited by Thermo Nicolet Corporation. TQ analyst user's guid[M]. Madison: Thermo Nicolet Corporation, 2003: 303 - 306.
- [7] 许 禄. 化学计量学: 一些重要方法的原理及应用[M]. 北京: 科学出版社, 2004: 20 - 23.