

偏最小二乘法结合傅里叶变换近红外光谱 同时测定卷烟焦油、烟碱和一氧化碳的释放量

王家俊*

梁逸曾

(云南红河卷烟厂产品中心, 弥勒 652300) (中南大学化学化工学院, 长沙 410083)

汪帆

(曲靖师范学院化学系, 曲靖 655000)

摘 要 采用多元校正技术-偏最小二乘法结合具有代表性的烟气总粒相物萃取液的傅里叶变换近红外透射光谱 220 个, 分别建立了焦油、烟碱和一氧化碳的校正模型, 并对光谱数据预处理方法和异常值诊断进行了讨论。所建的焦油、烟碱和一氧化碳模型相关系数分别为 0.9922、0.9828 和 0.9734, 均方预测残差(RMSECV) 分别为 0.3392、0.0329 和 0.3478, 预测范围分别为: 预测范围分别为 11.0 ~ 15.0 mg、1.15 ~ 1.45 mg 和 13.0 ~ 17.5 mg。通过对模型进行 t-检验, 在显著性水平大于 0.05 的条件下, 其测定结果与标准方法的测定结果对比, 两者无显著性差异。该分析方法应用于同时测定卷烟焦油、烟碱和一氧化碳的释放量, 结果令人满意。

关键词 偏最小二乘算法, 傅里叶变换近红外透射光谱, 焦油, 烟碱, 一氧化碳

1 引 言

卷烟焦油、烟碱和一氧化碳的释放量决定了卷烟制品能否上市的质量标准。按国家行业烟气分析的标准方法^[1-3]进行样品分析时, 当吸烟机抽吸完成之后, 先采用非散射红外法测定捕集烟香气相中一氧化碳含量, 继对捕集到的总粒相物进行萃取, 应用气相色谱法测定萃取液中的烟碱和水分含量, 减去烟碱和水分含量, 即得焦油含量。该过程涉及较多步骤和仪器, 分析速度慢, 特别是进行大批量测定时, 分析费用较高。采用化学计量学的多元校正技术结合傅里叶变换近红外光谱同时测定尚未见报道。

由于近红外光谱谱带较宽, 相互重叠严重, 谱带归属困难, 所以, 应用多元校正技术中的偏最小二乘法^[4], 建立了测定卷烟焦油、烟碱和一氧化碳释放量的校正模型, 并对影响建模效果的光谱数据预处理方法、奇异点和强影响点诊断等进行了讨论。此外, 校正模型采用独立校验集进行预测评价, 并通过 t-检验, 说明模型预测值与标准值方法测定结果无显著性差异。该法分析速度快, 稳定性好, 精度高, 可为卷烟质量控制提供一种可行的快速分析手段。

2 实验部分

2.1 仪器及标样制备

Antaris™傅里叶变换近红外光谱仪, InGaAs 检测器的透射分析模块(美国 Nicolet 公司), 1 mm 石英比色池; RESULT™集成软件和 TQ Analyst 6.2 定量分析软件(美国 Nicolet 公司)。光谱扫描范围: 10000 ~ 4000 cm⁻¹, 数据间隔: 4 cm⁻¹, 分辨率: 8 cm⁻¹; 扫描次数: 32。

实验中所用的 220 个标样(萃取液) 及其相应的焦油、烟碱和 CO 等基础数据均由本厂卷烟质量监督检测站按行业标准方法^[1-3]制备和测定。

2.2 实验方法

2.2.1 采集标样光谱数据 应用 RESULT™集成软件设置仪器工作参数, 编写采集光谱数据的工作流程。在恒温(25℃)恒湿(相对湿度 65%)的条件下, 开机预热 2 h 后, 调用编写好的程序, 即可采集标样的傅里叶变换近红外透射光谱数据。

2.2.2 建立校正模型 应用定量分析软件中的偏最小二乘算法把采集到的近红外透射光谱数据与标准方法测定的焦油、烟碱和 CO 等基础数据, 结合“剔一”(leave-one-out)的交互效验方法进行建模, 模型

产生后,经逐步优化,使其达到最佳状态。

3 结果与讨论

3.1 光谱数据的预处理

对采集到的原始光谱,进行必要的平滑和求导等数学预处理,过滤噪音、提高信噪比,消除基线飘移的干扰。从建模效果看,Karl Norris 导数平滑滤波^[5](karl norris derivative filter)优于 Savitzky-Golay 平滑滤波^[6];二阶求导优于一阶求导,应用二阶导数与 Karl Norris 滤波结合处理生成的导数光谱建模,可获得理想效果。但采用 Karl Norris 导数滤波时,应慎重选择段长(segment length)和段间距(gap between segments)。段长不宜大于 13,较为理想的段间距在 4~6 之间比较合适。

3.2 校正模型的优化

应用交互校验方法,对模型逐步优化,其目的是剔除奇异点(outlier)和强影响点,确定最优主因子,直到获得最佳模型。判断和剔除奇异点和强影响点的方法,一是采用 Mahalanobis 距离^[7](马氏距离),即 $MD_i = (t_i - \bar{t})S_k^{-1}(t_i - \bar{t})^T$,设定显著性水平(0.05)通过 Dixon 检验来剔除(见图 1 中垂直灰线之外的数据)。二是通过杠杆值来考察对模型的影响^[8],即采用“剔一”交互校验的方法,结合杠杆值,即 $h_i = 1/m + t_i^T(T_k^T T_k)^{-1}t_i$,与学生化残差值(studentized residual)等指标联合诊断^[9],逐步剔除这些强影响点。如建立焦油的模型中出现的两个样本 A01 和 A02,其杠杆值与学生化残差值的分布见图 2。此时模型的主因子和交互校验均方残差分别为 16 和 0.84652(见图 3a);剔除 A01 和 A02 后,重建模型的主因子和交互校验均方残差分别降为 14 和 0.78402(见图 3b)。

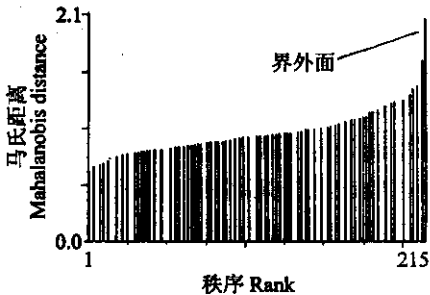


图 1 样本(光谱数据)秩序与马氏距离对应图
Fig. 1 Plot of rank versus Mahalanobis distance

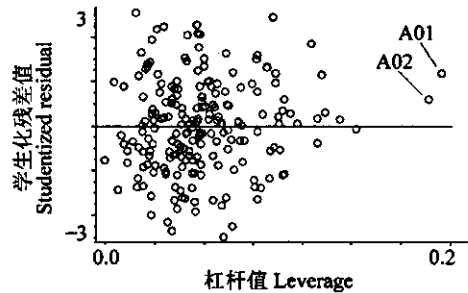


图 2 杠杆值与学生化残差值的散点图
Fig. 2 Plot of leverage versus studentized residual

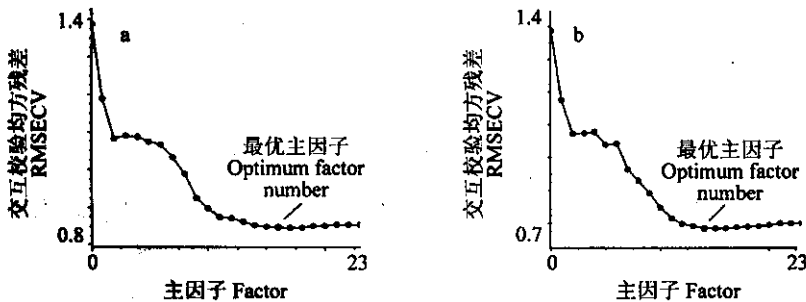


图 3 交互校验均方残差与随主因子的变化图

Fig. 3 Changes of root mean square error of cross validation (RMSECV) with factor number

为了避免“过拟合”引入更多的噪声,同样要应用交互校验方法,通过 F-检验来确定当 PRESS(predictive residual error sum of squares)达到最小时的主因子^[10,11],即最优主因子数。所以,优化模型是一个检测奇异点、强影响点、确定最优主因子和相应交互校验均方残差 RMSECV 的协同过程。应用最优主因子建立的焦油、烟碱和 CO 等校正模型见表 1。在模型中,标准值(actual)与模型预测值(calculated)的散点图和它们相应的最优主因子与交互校验均方残差 RMSECV 的关系图分别见图 4、图 5 和图 6。

表 1 焦油、烟碱和一氧化碳等校正模型的有关数理指标

Table 1 Mathematical statistics results for calibration models of tar, nicotine and carbon monoxide

模型名称 Model name	校正集样本量 Number of samples	相关系数 Correlation coefficient	最优主因子 Factor	均方估计残差 RMSEC	交互校验均方残差 RMSECV	预测范围 (mg) Predicted range
焦油 Tar	208	0.9922	16	0.1570	0.3392	11.0 ~ 15.0
烟碱 Nicotine	191	0.9828	16	0.0158	0.0329	1.15 ~ 1.45
CO	136	0.9734	10	0.2230	0.3478	13.0 ~ 17.5

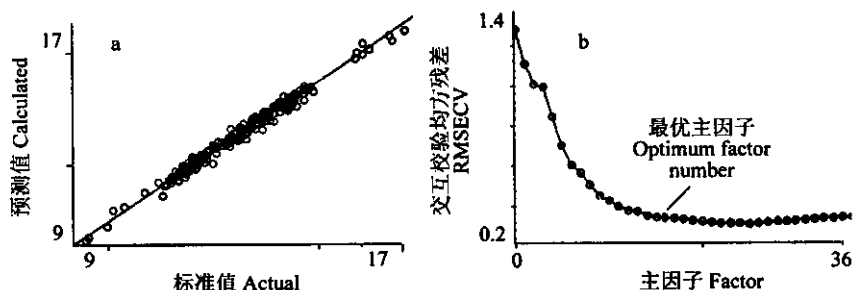


图 4 焦油的标准值与模型预测值的散点图 (a) 和交互校验均方残差与随主因子的变化图 (b)

Fig. 4 Actual value versus calculated value of tar (a) and changes of RMSECV with factor number (b)

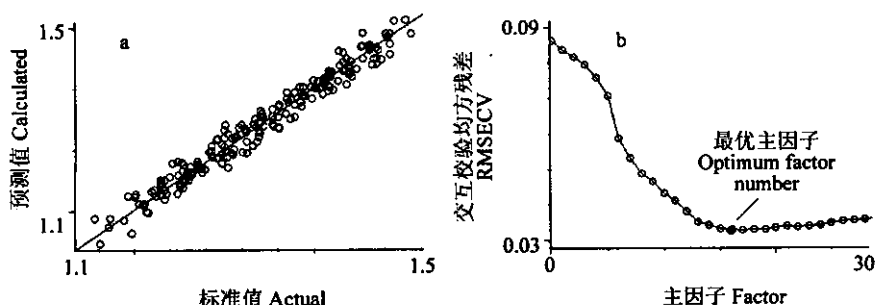


图 5 烟碱的标准值与模型预测值的散点图 (a) 和交互校验均方残差与随主因子的变化图 (b)

Fig. 5 Actual value versus calculated value of nicotine (a) and changes of RMSECV with factor number (b)

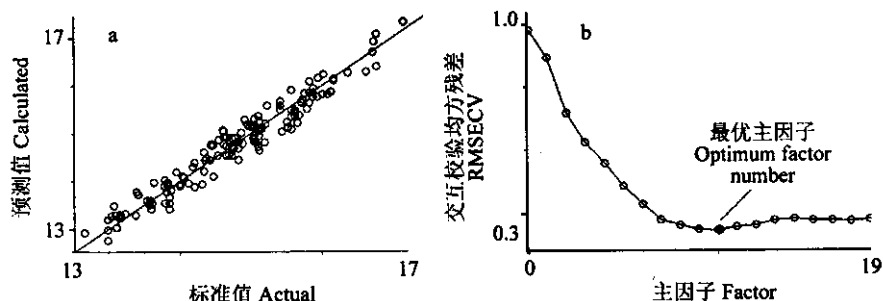


图 6 一氧化碳的标准值与模型预测值的散点图 (a) 和交互校验均方残差与随主因子的变化图 (b)

Fig. 6 Actual value versus calculated value of carbon monoxide (a) and changes of RMSECV with factor number (b)

3.3 校正模型的检验

一般来说,上述过程基本上可衡量模型预测能力的优劣。但对于复杂的多组分分析体系,在建模过程中,免不了要剔出较多的样本。此时,模型中校正集的代表性以及模型的可靠性就会受到影响,采用独立校验集验证模型是一种比较可靠、灵活的方法,即另行分批次随机抽样作为独立校验集,用建好的校正模型和标准方法进行测定,其测定结果通过配对 t-检验(由 SPSS 10.0 for windows 软件完成)进一步进行确认。统计结果显示,在显著性水平大于 0.05 时,两种测定方法不存在显著性差异,表明两种方法不存在系统误差。表 2 为两种分析方法测定结果的比较。表 2 说明,以上所建的校正模型具有良好的预测能力。此外,对同一样品进行了 6 次测定(见表 3),结果表明,该近红外光谱仪和校正模型具有良好的测量精度。

表 2 本方法(PLS-FT-NIR)与行业标准方法测定结果的比较($n=8$, mg/支)

Table 2 Comparison results of partial least squares-Fourier transform-near infrared(PLS-FT-NIR) method and standard methods($n=8$, mg per cigarette)

序号 No.	焦油 Tar			烟碱 Nicotine			CO		
	行业标准方法 YC/T29-1996 standard method (mg)	本方法 PLS-FT-NIR (mg)	偏差 Deviation (mg)	行业标准方法 YC/T29-1996 standard method (mg)	本方法 PLS-FT-NIR (mg)	偏差 Deviation (mg)	行业标准方法 YC/T29-1996 standard method (mg)	本方法 PLS-FT-NIR (mg)	偏差 Deviation (mg)
1	9.53	9.68	0.15	1.22	1.19	-0.03	13.61	13.64	0.03
2	11.43	11.32	-0.11	1.23	1.25	0.02	13.82	14.12	0.30
3	11.94	11.91	-0.03	1.29	1.25	-0.04	14.38	14.32	-0.06
4	12.36	12.43	0.07	1.30	1.33	0.03	14.87	15.19	0.32
5	12.89	12.79	-0.10	1.31	1.28	-0.03	15.04	14.85	-0.19
6	13.10	13.10	0.00	1.37	1.37	0.00	15.16	15.26	0.10
7	14.05	14.24	0.19	1.40	1.42	0.02	15.88	15.54	-0.34
8	14.68	14.50	-0.18	1.44	1.42	-0.02	16.09	15.99	-0.10

表 3 精密度实验($n=6$)

Table 3 The precision test($n=6$)

成分 Composition	平均值 Mean (mg)	标准差 Standard deviation (mg)	相对标准偏差 Relative standard deviation(%)	成分 Composition	平均值 Mean (mg)	标准差 Standard deviation (mg)	相对标准偏差 Relative standard deviation(%)
焦油 Tar	13.27	0.0259	0.20	烟碱 Nicotine	1.41	0.0024	0.17
CO	14.92	0.0320	0.21				

致谢 标样及其基础性数据均由云南红河卷烟厂质量监督检测站提供,特此致谢。

References

- 1 Edited by The State Tobacco Monopoly Administration of P. R. C.(国家烟草专卖局编), YC/T29-1996 *Cigarettes-Determination of Total Particulate Matter and Tar Using a Routine Analytical Smoking Machine*(YC/T29-1996 卷烟用常规分析用吸烟机测定总粒相物和焦油). Beijing(北京):The State Tobacco Monopoly Administration of P. R. C.(国家烟草专卖局发布), 1996
- 2 Edited by the State Tobacco Monopoly Administration of P. R. C.(国家烟草专卖局编), YC/T8-93 *Cigarettes-Determination of Moisture and Nicotine in Total Particulate Matter of Smoke-Gas Chromatography Method*(YC/T8-93 卷烟烟气总粒相物水分和烟碱的测定气相色谱法). Beijing(北京):The State Tobacco Monopoly Administration of P. R. C.(国家烟草专卖局发布), 1993
- 3 Edited by the State Tobacco Monopoly Administration of P. R. C.(国家烟草专卖局编). YC/T30-1996 *Cigarettes-Determination of Carbon Monoxide in Gas Phase of Smoke-NDIR Method*(YC/T30-1996 卷烟烟气气相中一氧化碳的测定非散射红外法). Beijing(北京):The State Tobacco Monopoly Administration of P. R. C.(国家烟草专卖局发布), 1996
- 4 Liang Yizeng(梁逸曾). *White, Grey and Black Multicomponent System and Their Chemometric Algorithms*(白灰黑复杂多组分分析体系及其化学计量学算法). Changsha(长沙):Hunan Publishing House of Science and Technology(湖南科技出版社), 1996 :32~36
- 5 Edited by Thermo Nicolet Corporation. *TQ Analyst User's Guid*, Madison :Thermo Nicolet Corporation, 2003 :153~159

- 6 Savitzky A, Golay M J E. *Anal. Chem.*, **1964**, 36 : 1627
- 7 Edited by Infometrix Inc. . *Multivariate Data Analysis*, Woodinville : Infometrix Inc. , **2003** : 5 , 24
- 8 Weisberg S. *Applied Linear Regression*, New York : John Wiley & Sons , **1985** , Chapter 5
- 9 Cook R D , Weisberg S. *Residuals and Influence in Regression*, New York : Chapman and Hall , **1982** , Chapter 2
- 10 Haaland D M , Thomas E V. *Anal. Chem.* , **1988** ,(60) : 1193 ~ 1202
- 11 Osten D W. *J. Chemometrics* , **1988** ,(2) : 39

Simultaneous Determination of Tar , Nicotine and Carbon Monoxide in Smoking by Partial Least Squares and Fourier Transform Near Infrared Transmission Spectrometry

Wang Jiajun *

(Products Research Centre , Honghe Cigarette Factory , Mile 652300)

Liang Yizeng

(College of Chemistry and Chemical Engineering , Central South University , Changsha 410083)

Wang Fan

(Department of Chemistry , Qujing Teacher's College Qujing 655000)

Abstract The contents of tar , nicotine and carbon monoxide in smoking is the standard of quality for coming into the market. Determination of these contents plays an important role in their quality control. The calibration models were established for determination of tar , nicotine and carbon monoxide by partial least squares and Fourier transform near infrared transmittance spectrometry of 220 representative extractions from total particulate matter of smoking. Spectral data preprocessing and outliers diagnostics were also discussed. Correlation coefficients of the models were 0.9922 , 0.9828 and 0.9734 , and the root mean square error of cross validation (RMSECV) were 0.3392 , 0.0329 and 0.3478 in the predicted range of 11.0 ~ 15.0 mg , 1.15 ~ 1.45 mg and 13.0 ~ 17.5 mg for tar , nicotine and carbon monoxide , 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 simultaneous determination of tar , nicotine and carbon monoxide in smoking with satisfactory results.

Keywords Partial least squares method , Fourier transform near infrared transmission spectrometry , tar , nicotine , carbon monoxide

(Received 15 May 2004 ; accepted 26 August 2004)