

SIMCA 分类法与偏最小二乘法结合近红外光谱检测卷烟的内在品质

王家俊¹, 梁逸曾², 汪帆³

(1. 红河卷烟厂产品中心, 云南, 弥勒, 652300; 2. 中南大学化学化工学院, 湖南, 长沙, 410083; 3. 曲靖师范学院化学系, 云南, 曲靖, 655000)

摘要: 针对卷烟内在品质常规化学检测过程烦琐、速度慢的特点, 本文探讨了基于化学计量学方法与近红外光谱分析技术结合判别卷烟品质等级和同时检测多个品质指标的快速方法。通过采用 SIMCA 分类法与偏最小二乘法结合卷烟的傅里叶变换近红外漫反射光谱, 建立了 3 个不同级别卷烟的分类模型, 用于甄别卷烟品质真伪, 结果准确; 同时, 又建立了检测卷烟中总氮、总糖、烟碱主要化学成分含量和其主流烟气中焦油、烟碱、一氧化碳释放量等品质指标的校正模型, 其相关系数分别为: 0.9844、0.9918、0.9968、0.9505、0.8945 和 0.9625; 交互效验均方残差分别为: 0.0663、0.6682、0.0607、0.2262、0.0350 和 0.3685。另取未知样品对模型进一步验证, 经配对 *t*-检验表明, 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, 两者均无显著性差异。该法应用于卷烟品质真伪鉴别和品质指标检测, 操作简单、快速、准确。

关键词: SIMCA 分类法; 偏最小二乘法; 近红外漫反射光谱; 卷烟品质分析

中图分类号: O 657.33

文献标识码: A

文章编号: 1001-4160(2006)11-1133-1136

The quality analysis of cigarette by partial least squares and SMICA combined with fourier transform near infrared spectra

Wang Jiajun¹, Liang Yizeng² and Wang Fan³

(1. Products Research Centre, Honghe Cigarette Factory, Mile, 652300, Yunnan, China; 2. College of Chemistry and Chemical Engineering, Central South University, Changsha, 410083, Hunan, China; 3. Department of Chemistry, Qujing Techer's College, Qujing, 655000, Yunnan, China)

Abstract: The routine chemical analysis of cigarette quality was loaded down with trivial details, and slow. A fast approach based on methods of chemometrics combined with Fourier transform near infrared spectrometry, which, was discussed in this paper. The approach was intended for discrimination of cigarette quality, and simultaneous determination of many quality indexes. By using partial least squares and SMICA techniques combined with Fourier transform near Infrared spectra, the classification methods were established for discrimination of three grades cigarettes, which, was used to identify bogus cigarette with correct results. Meanwhile, the calibration models were established for determination of total nitrogen, total sugar, nicotine in cigarettes, and tar, nicotine, carbon monoxide in mainstream smoke. Correlation coefficients of the models were 0.9844, 0.9918, 0.9968, 0.9505, 0.8945 and 0.9625, and the root mean square error of cross validation (RMSECV) were 0.0663, 0.6682, 0.0607, 0.2262, 0.0350 and 0.3685, respectively. The calibration models had been applied to determination of another unknown samples. By paired-samples *t*-test, the results were compared with those of standard methods and neither of them was significant difference over 0.05 level. The approach can be used to discrimination and determination of cigarette quality, which, was simple and fast with accurate predicted results.

Key words: SMICA classification, partial least squares technique, fourier transform near infrared diffusion reflectance spectra, cigarette quality analysis

Wang JJ, Liang YZ and Wang F. The quality analysis of cigarette by partial least squares and SMICA combined with fourier transform near infrared spectra. Computers and Applied Chemistry, 2006, 23(11):1133-1136.

1 引言

一般认为, 卷烟中的总糖、总氮和烟碱等主要化学成分含量, 及其燃烧时主流烟气中的焦油、烟碱和一氧化碳等释放量可作为品质指标衡量卷烟的内在品质, 其中, 焦油、烟碱和一氧化碳的释放量又决定了卷烟能否入市的质量标准。

按国家烟草行业标准, 卷烟的内在品质检测通常分为两个部分, 一是采用连续流动法测定烟丝中的总氮、总糖、烟碱等主要化学成分含量^[1-3]; 二是通过烟气分析测定焦油、烟碱和一氧化碳的释放量^[4-6]。对于品牌规模化的生产, 按以上方法进行大批量质检, 需要较多的分析仪器, 消耗大量的化学

收稿日期: 2005-03-25; 修回日期: 2006-04-28

作者简介: 王家俊 (1962 年—), 男, 工程师。

试剂,速度慢,且分析费用高。此外,对卷烟品质的真伪甄别,通常从包装质量、评吸质量进行综合判定。

通过对我厂红河牌号3个级别卷烟的近红外漫反射光谱进行主成分分析(PCA)发现,在模式空间中,都显示出各自的特征分布,为此,采用SIMCA分类法^[7],建立了3个级别卷烟的类模型,并对未进入校正集的样品(其中包含仿冒卷烟)进行分类,获得了正确的分类结果。在进行定量分析时,因近红外光谱谱带较宽,相互重叠严重,所以,应用了多元校正方法中的偏最小二乘算法^[8,9],分别建立了检测烟丝中总糖、总氮和烟碱等主要化学成分及主流烟气中焦油、烟碱和一氧化碳释放量等品质指标的校正模型,并用于实测样品,经*t*-检验表明,在显著水平大于0.05时,校正模型的预测值与行业标准方法的测定值无显著性差异。本法应用于鉴别卷烟品质的真伪及其相应多个品质指标的同时测定,具有操作简捷、准确、无污染等优点。此外,本文对建模最佳谱区的选择进行了讨论。

2 实验部分

2.1 主要实验仪器及实验标样的准备

Antaris™傅里叶变换近红外光谱仪,包括InGaAs检测器的漫反射积分球、5cm石英采样杯、旋转台和TQ Analyst 6.2定量分析软件(美国Nicolet公司);Pirouette 3.11多元数据分析软件(美国Infometrix公司)。仪器的主要工作参数为光谱扫描范围:10000~4000 cm⁻¹;分辨率:8 cm⁻¹;扫描次数:72。

将不同批次的3个级别的卷烟(其中包含甲级、乙级、精品88)150个;不同产地、不同等级的陈化烤烟样品200个粉碎,过60目筛,装入密封袋,即为实验标样。烟丝中的总糖、总氮和烟碱主要化学成分含量及其卷烟相应的焦油、烟碱和一氧化碳释放量等基础数据均按行业标准方法文献^[1-6]测定。

2.2 实验方法

2.2.1 采集标样光谱数据

在恒温(25℃)恒湿(相对湿度65%)的条件下,开机预热2h后,将实验标样装入5cm石英杯中轻压至平,置于旋转台,即可采集标样的近红外漫反射光谱数据。

2.2.2 建立校正模型

选择适当的谱区,应用Pirouette软件对3个级别的150个卷烟标样(各50个)的光谱进行SIMCA分类,剔除异常点,建立3个牌号卷烟的类模型。采用TQ Analyst 6.2软件中的偏最小二乘法把采集到的烟丝和陈化烤烟的光谱数据与标准方法测定的主要化学成分含量及其相应的焦油、烟碱

和一氧化碳释放量等基础数据,结合“剔一”(Leave-one-out)的交互效验方法^[9,10]分别建立校正模型,并逐步优化,使其达到最佳状态。

3 结果与讨论

3.1 校正集样品的选择

在建立预测烟丝总糖、总氮和烟碱含量的校正模型时,由于卷烟本身的主要化学成分相对稳定,除了选择卷烟样品(150个)进入校正集外,还选入了单等级的配方烟叶(即上述的陈化烤烟样品200个),这样,既可使校正集具有较好的代表性,同时又拓宽模型的预测范围,增强了模型的适应能力。在建立预测卷烟焦油、烟碱和一氧化碳释放量的校正模型时,因为卷烟辅料(盘纸、水松纸、嘴棒等)的变化,对焦油、烟碱和一氧化碳释放量有着明显的影响,所以,选入校正集的卷烟标样,应保持相应卷烟辅料特性的相对稳定,以免引入较多的背景误差。

3.2 最佳光谱区的选择与模型的建立

虽然光谱处理软件会推荐最佳谱区,但在模型优化中常存在随奇异点(Outlier)的剔除,导致推荐的最佳谱区发生变化的现象,若采用全谱校正,虽然保留了全部信息,但也引入了更多噪声和无用信息。依据光谱特性及其与预测性质表现出来的统计特征,可以方便、准确地确定最佳谱区。

应用SIMCA分类法建立3个级别卷烟的类模型,依据平均光谱(图1),谱区与SIMCA识别能力的相关性^[11],见图2,选择8700~4120 cm⁻¹谱区分类,效果理想,分类结果见图3。而选用全谱分类则引入了弱吸收、包含较多高频噪声的谱区(10000~8700 cm⁻¹),导致分类效果变差。但值得注意的是,选择识别能力与相关性较高的过窄的谱区,虽可获得较好的分类效果,但会造成类模型的代表性损失。类与类的分离效果可用距离来衡量^[11],即 $D_{pq} = \left(\frac{s_{pq}^2 + s_{pp}^2}{s_{pp}^2 + s_{qq}^2} \right)^{1/2} - 1$,采用不同谱区分类,类与类之间的距离比较见表1。

在建立校正模型时,应基于相应的组分(以总糖为例说明),选择高相关性、大方差的谱区,分别见图4a和图4b,可获得理想的建模效果,表2为采用不同谱区建立的总糖校正模型的统计结果,其中,建模效果最佳的谱区为7250~4050 cm⁻¹。应用最佳谱区建立的总糖校正模型,其测定值(Actual)与模型预测值(Calculated)的散点图和相应的主因子数(Factor)与交互效验均方残差(RMSECV)的变化图,见图5。同样,其它组分建模的最佳谱区选择亦然,应用最佳谱区建立的校正模型的统计结果,见表3。

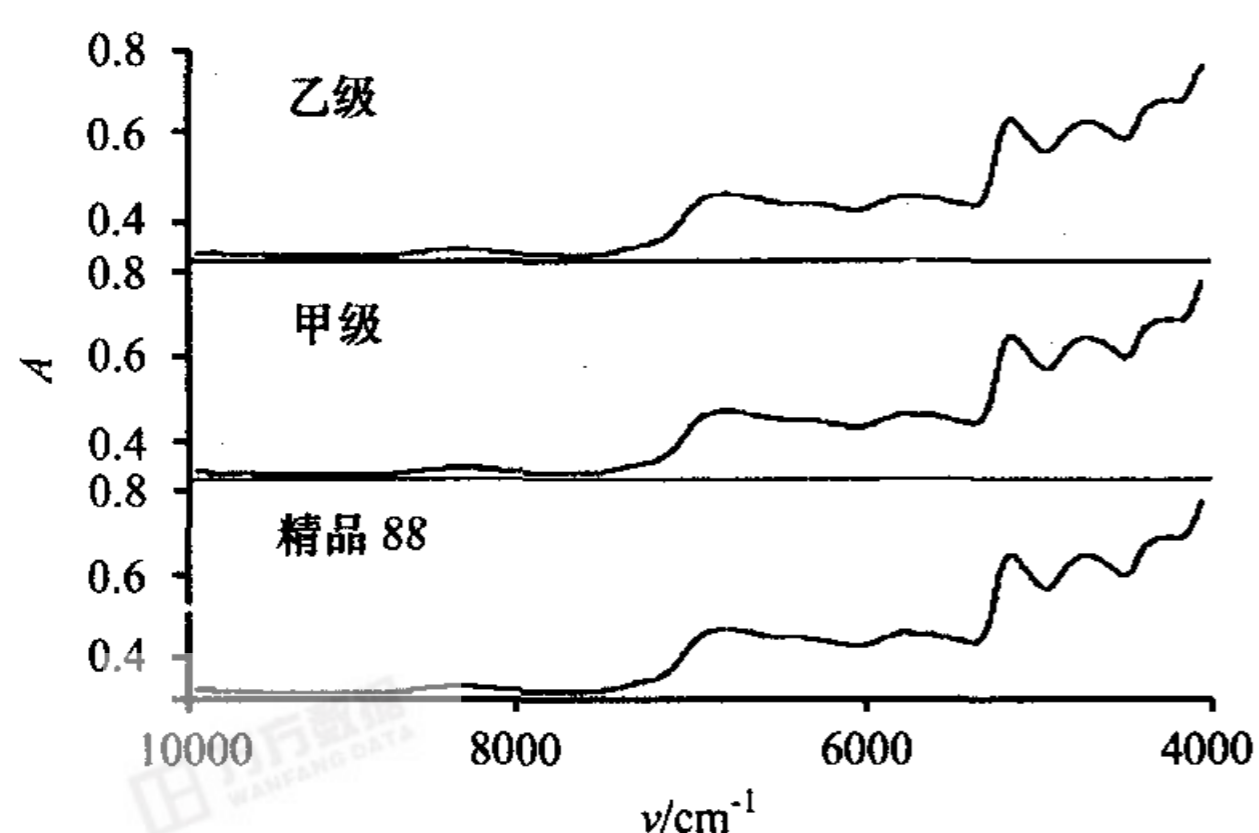


Fig. 1 NIR diffusion reflectance average spectra of three grades cigarettes.

图1 三个级别卷烟的近红外漫反射平均光谱图

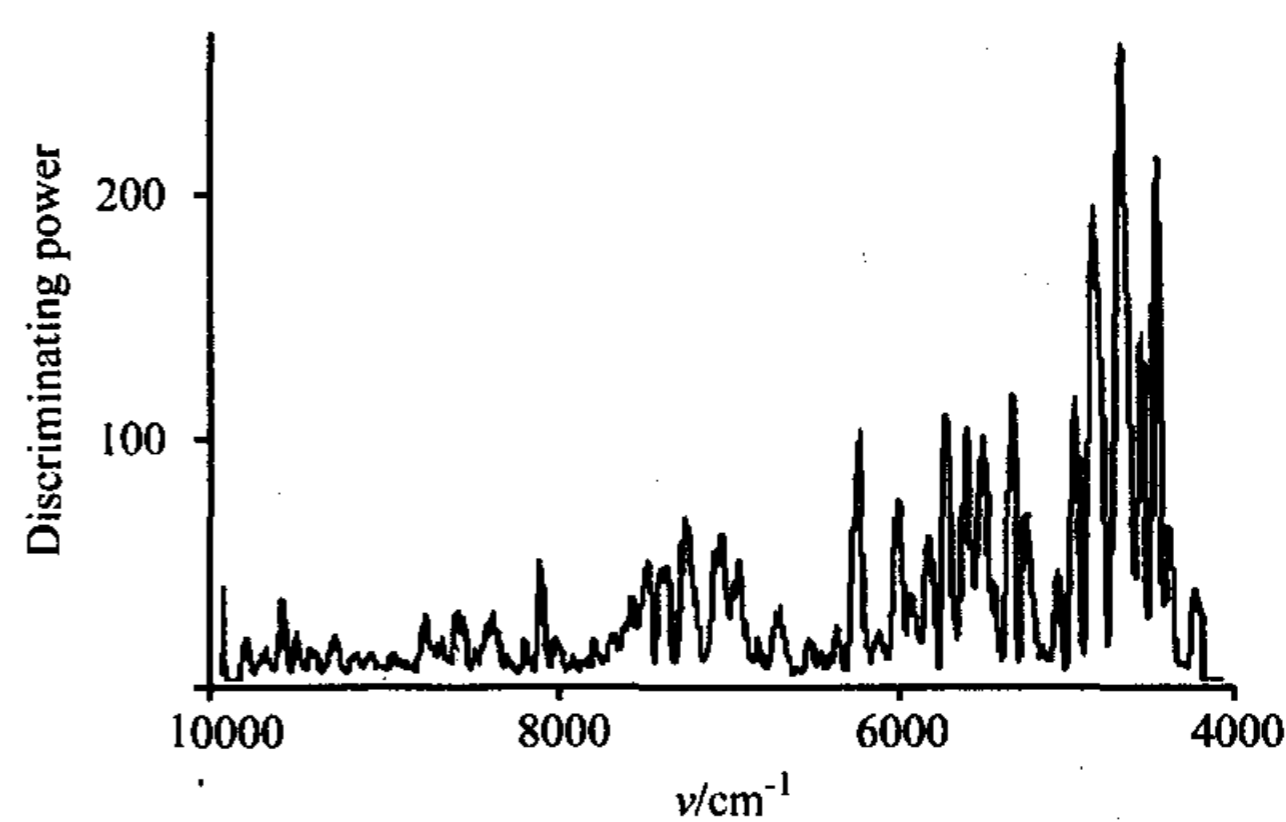


Fig. 2 Spectral region versus discriminating power from SIMCA analysis.

图2 不同谱区与SIMCA类模型识别能力的关系图

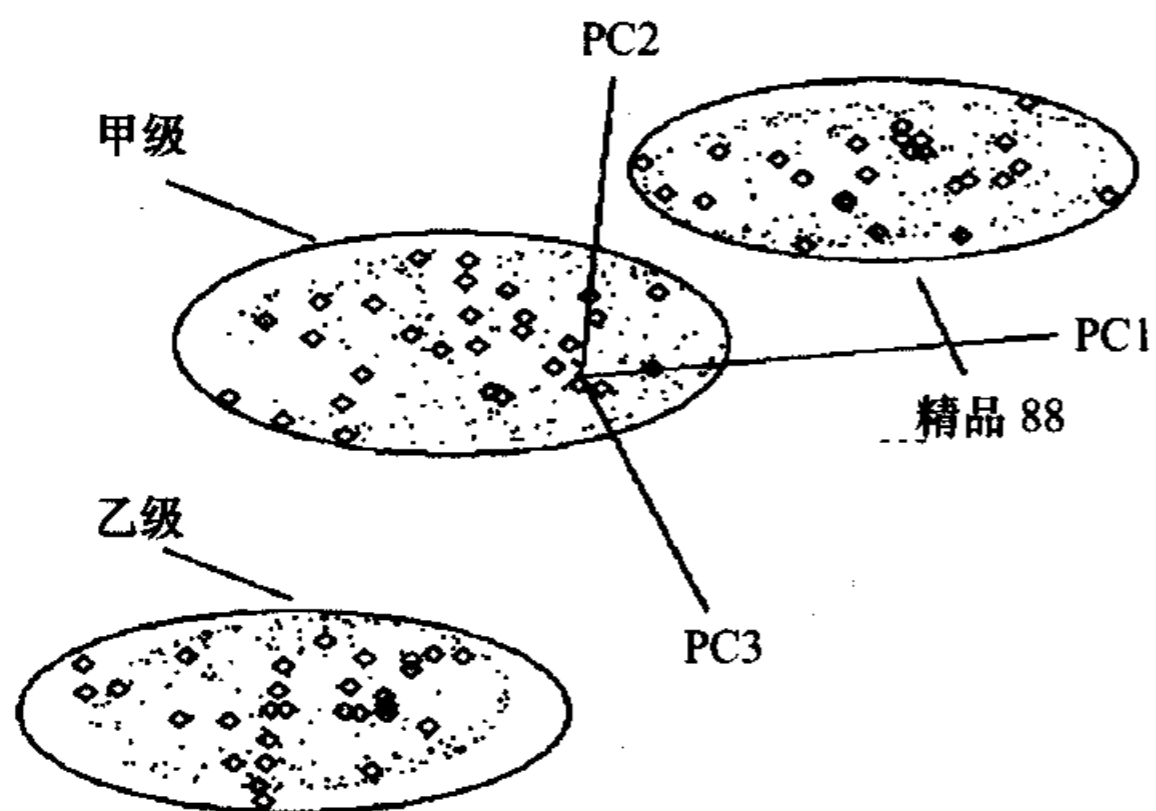


Fig. 3 Class projections plot from SIMCA analysis of three grades cigarettes.

图3 三个级别卷烟的 SIMCA 分类图

表1 采用不同谱区分类的类与类之间的距离比较

Table 1 Interclass distances with different spectral regions.

	10000 ~ 4000 cm^{-1}			8700 ~ 4120 cm^{-1}		
	乙级 Grade B	甲级 Grade A	精品 88 Superior 88	乙级 Grade B	甲级 Grade A	精品 88 Superior 88
乙级 Grade B	0.0000	3.3333	5.7917	0.0000	4.2076	8.0147
甲级 Grade A	3.3333	0.0000	2.2938	4.2076	0.0000	2.9019
精品 88 Superior 88	5.7917	2.2938	0.0000	8.0147	2.9019	0.0000

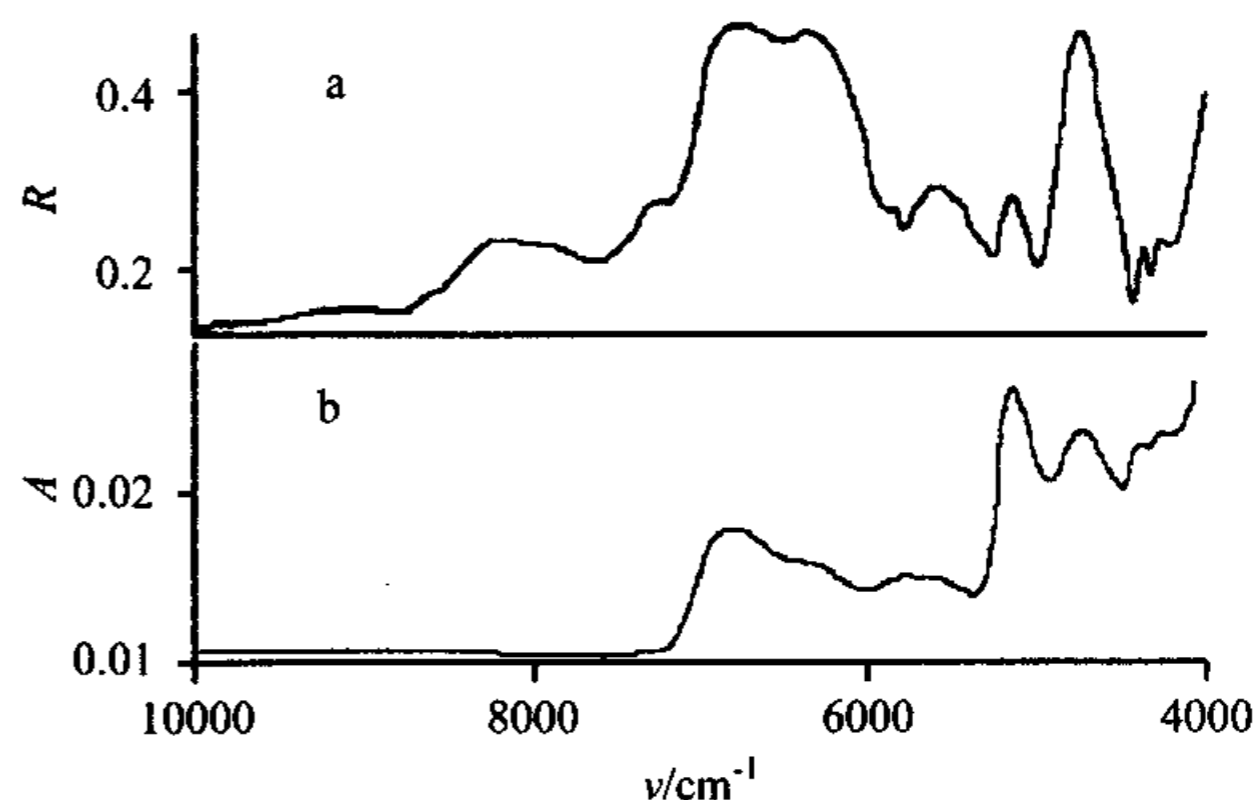


Fig. 4 Spectral region versus correlation coefficient (R) for total sugar (a) and their variance spectrum (b).

图4 总糖组分的谱区与相关系数(R)的关系图(a)和其方差光谱图(b)

3.3 样品测定与模型验证

取未进类模型的样品 37 个(乙级 12 个、甲级 11 个、精品 14 个,在每个等级中各含假烟 2 个,分别是 A01、A02, B01、B02, C01、C02),应用建好的 SIMCA 分类模型预测,均获得了正确的分类结果,尤其对假烟,该法具有良好的判别能力,分类效果见图 6。

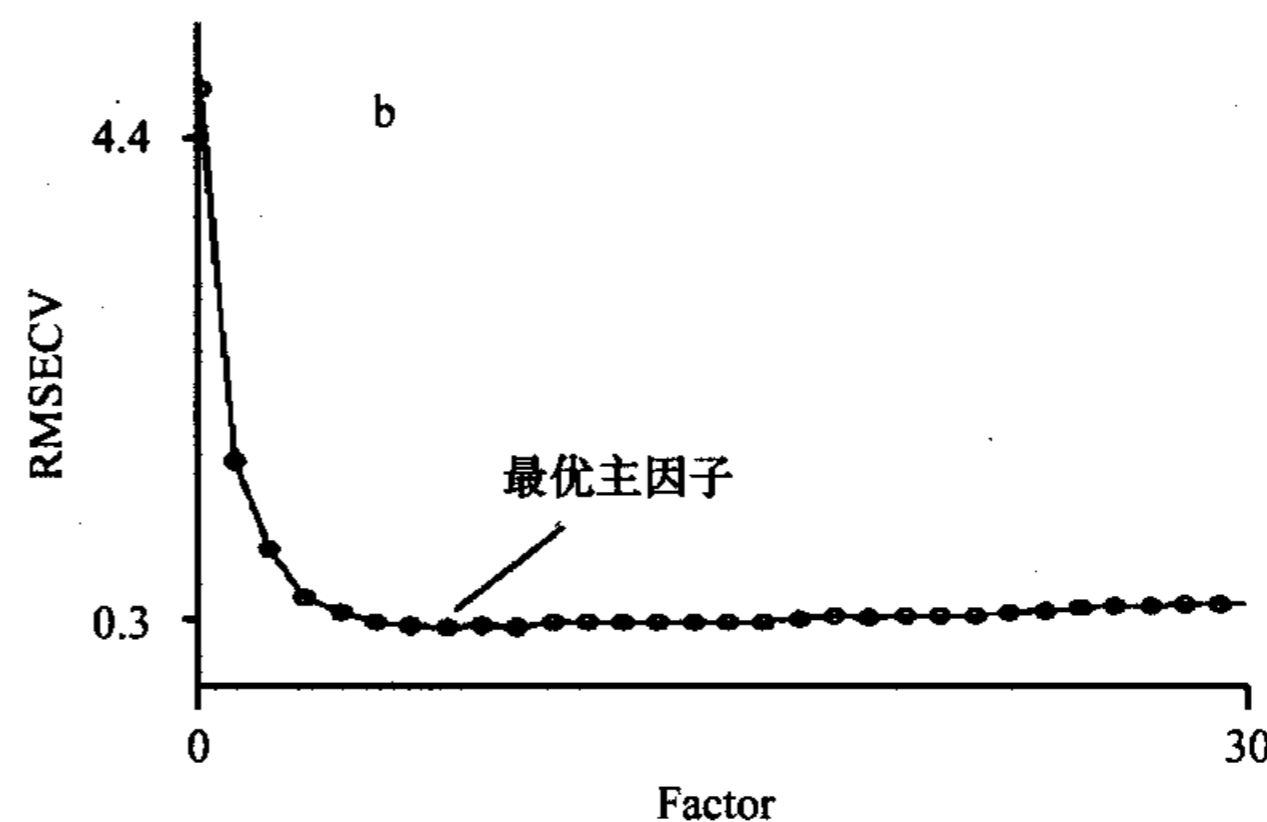
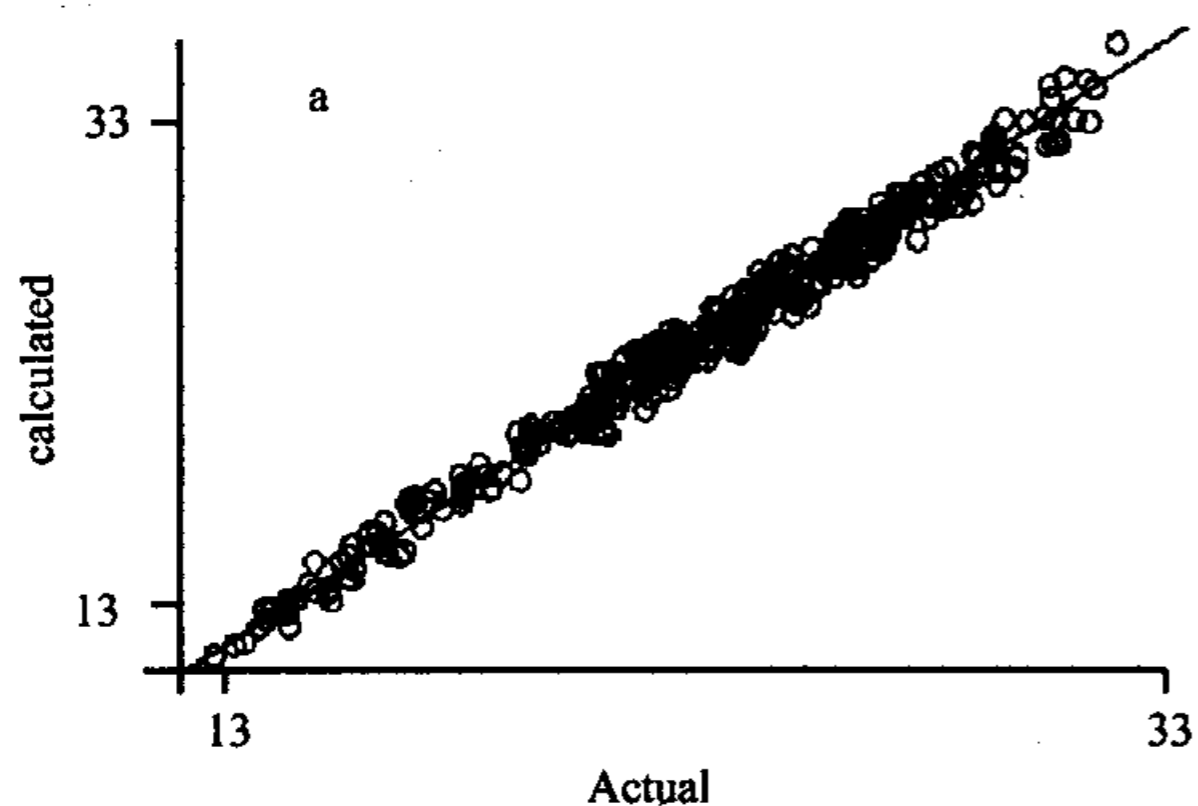


Fig. 5 Actual value versus calculated value of total sugar (a) and changes of RMSECV with factor numbers(b).

图5 总糖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的变化图(b)

表2 采用不同谱区建立的总糖校正模型的统计结果

Table 2 Statistics results for calibration models of total sugar with different spectral regions.

谱区 Spectral regions	相关系数 Correlation coefficient	最优主因子 Factor	交互效验均方残差 RMSECV
10000 ~ 4000 cm^{-1}	0.98545	7	0.7640
9000 ~ 4050 cm^{-1}	0.99273	7	0.7146
7250 ~ 4050 cm^{-1}	0.99182	7	0.6682

另行分批次随机抽取 3 个级别的样品共 8 个,用建好的校正模型和标准方法进行测定,其测定结果见表 4 和表 5,并通过配对 t -检验进一步验证,在显著性水平大于 0.05 时, $|t|$ 分布值均小于临界值 ($t_{(0.05, 7)} = 2.36$),说明两种测定方法均不存在显著性差异。由此可见,以上所建的校正模型具有较好的预测能力。

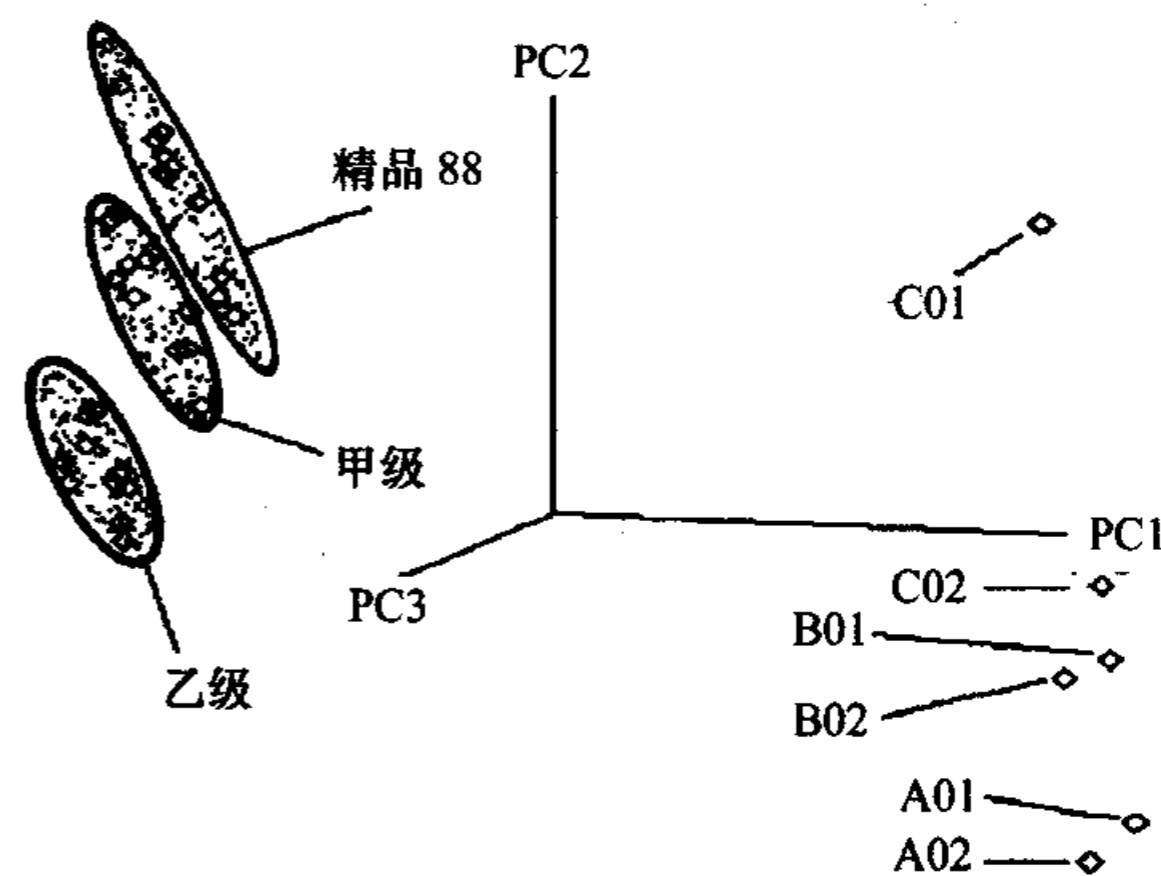


Fig. 6 Class projections plot from SIMCA analysis.

图6 实测样品的 SIMCA 分类图

表3 卷烟中总氮、总糖和烟碱及其焦油、烟碱和一氧化碳释放量等品质指标的校正模型的统计结果
Table 3 Statistics results for calibration models of total nitrogen, total sugar and nicotine in cigarettes, and tar, nicotine and carbon monoxide in mainstream smoke.

品质指标 Quality index	模型名称 Model name	相关系数 Correlation coefficient	最优主因子 Factor	交互效验 均方残差 RMSECV	预测范围 Predicted range
卷烟品质指标 Quality index in cigarettes	总氮 Total nitrogen	0.9844	7	0.0663	1.70 ~ 2.70 (%)
	总糖 Total sugar	0.9918	7	0.6682	14.0 ~ 32.0 (%)
	烟碱 Nicotine	0.9968	11	0.0607	1.40 ~ 3.80 (%)
烟气品质指标 Quality index in mainstream smoke	焦油 Tar	0.9505	3	0.2262	12.0 ~ 15.0 (毫克/支)
	烟碱 Nicotine	0.8945	5	0.0350	1.10 ~ 1.50 (毫克/支)
	一氧化碳 Carbon monoxide	0.9625	4	0.3685	12.0 ~ 17.0 (毫克/支)

表4 卷烟中总氮、总糖和烟碱等品质指标的本方法(PLS-FT-NIR)与行业标准方法测定结果比较($n=8, \%$)

Table 4 Comparison results of PLS-FT-NIR method and standard methods for total nitrogen, total sugar and nicotine in cigarette ($n=8, \%$).

序号 No.	总氮 Total nitrogen			总糖 Total sugar			烟碱 Nicotine		
	标准方法 YC/T161-2002	本方法 PLS-FT-NIR	偏差 Deviation	标准方法 YC/T159-2002	本方法 PLS-FT-NIR	偏差 Deviation	标准方法 YC/T160-2002	本方法 PLS-FT-NIR	偏差 Deviation
1	2.23	2.27	-0.04	21.80	21.23	0.57	2.20	2.06	0.14
2	2.06	2.10	-0.04	19.80	20.64	-0.84	2.20	2.25	-0.05
3	2.16	2.23	-0.07	23.10	23.21	-0.11	2.23	2.28	-0.05
4	2.02	2.02	0.00	21.80	21.79	0.01	2.20	2.07	0.13
5	2.10	2.19	-0.09	23.53	23.71	-0.18	2.04	2.19	-0.15
6	2.08	2.02	0.06	20.39	20.48	-0.09	2.03	1.96	0.07
7	2.03	1.97	0.06	21.30	20.67	0.63	2.20	2.13	0.07
8	2.31	2.27	0.04	22.70	22.20	0.50	2.29	2.29	0.00
t -分布值 t -Value	-0.481			0.352			0.563		

表5 卷烟主流烟气中焦油、烟碱和一氧化碳释放量的本方法(PLS-FT-NIR)与行业标准方法测定结果比较($n=8, \text{毫克/支}$)

Table 5 Comparison results of PLS-FT-NIR method and standard methods for tar, nicotine and carbon monoxide in mainstream smoke ($n=8, \text{mg/per cigarette}$).

序号 No.	焦油 Tar			烟碱 Nicotine			一氧化碳 Carbon monoxide		
	标准方法 YC/T29-1996	本方法 PLS-FT-NIR	偏差 Deviation	标准方法 YC/T156-2001	本方法 PLS-FT-NIR	偏差 Deviation	标准方法 YC/T30-1996	本方法 PLS-FT-NIR	偏差 Deviation
1	12.75	12.79	-0.04	1.29	1.30	-0.01	14.63	14.92	-0.29
2	13.01	13.19	-0.18	1.30	1.32	-0.02	15.47	13.82	0.10
3	13.22	13.04	0.18	1.34	1.31	0.03	13.54	13.65	-0.11
4	13.34	13.37	-0.03	1.36	1.35	0.01	15.20	15.26	-0.06
5	13.89	14.04	-0.15	1.39	1.38	0.01	15.72	15.69	0.03
6	14.09	14.05	0.04	1.40	1.39	0.01	14.02	13.97	0.05
7	14.31	14.50	-0.19	1.43	1.40	0.03	14.93	14.86	0.07
8	14.51	14.35	0.16	1.46	1.45	0.01	13.92	15.65	-0.18
t -分布值 t -Value	-0.511			1.433			-0.151		

References

- 1 Edited by The State Tobacco Monopoly Administration of PRC. YC/T161-2002 Tobacco and tobacco products-Determination of total nitrogen-Continuous flow Method. Beijing: The State Tobacco Monopoly Administration of PRC, 2002.
- 2 Edited by The State Tobacco Monopoly Administration of PRC. YC/T159-2002 Tobacco and tobacco products-Determination of water soluble sugars - Continuous flow Method. Beijing: The State Tobacco Monopoly Administration of PRC, 2002.
- 3 Edited by The State Tobacco Monopoly Administration of PRC. YC/T160-2002 Tobacco and Tobacco Products-Determination of Total Alkaloids-Continuous Flow Method. Beijing: The State Tobacco Monopoly Administration of PRC, 2002.
- 4 Edited by The State Tobacco Monopoly Administration of PRC. YC/T29-1996 Cigarettes-Determination of Total Particulate Matter and Tar Using a Routine Analytical Smoking Machine. Beijing: The State Tobacco Monopoly Administration of PRC, 1996.
- 5 Edited by The State Tobacco Monopoly Administration of PRC. YC/T156-2001 Cigarettes-Determination of Nicotine in Total Particulate Matter of Smoke - Gas Chromatography Method. Beijing: The State Tobacco Monopoly Administration of PRC, 2001.
- 6 Edited by The State Tobacco Monopoly Administration of PRC. YC/T30-1996 Cigarettes-Determination of Carbon Monoxide in Gas Phase of Smoke -NDIR Method. Beijing: The State Tobacco Monopoly Administration of PRC, 1996.
- 7 Wold S. Pattern recognition by means of disjoint principal components models. Pattern Recognition, 1976, (8):127-139.
- 8 Liang YiZeng. White, Grey and Black Multicomponent System and their Chemometric Algorithms. Changsha: Hunan Publishing House of Science and Technology, 1996:32-36.
- 9 Haaland DM, and Thomas EV. Partial least squares methods for spectral analysis. Anal Chem, 1988, (60):1193-1202.
- 10 Osten DW. Selection of optimal regression model via cross validation. J Chemometrics, 1988, (2):39.
- 11 Edited by Infometrix Inc. Multivariate Data Analysis. Woodinville: Infometrix Inc, 2003, 6.14-6.15.

附中文参考文献

- 1 国家烟草专卖局编. YC/T161-2002 烟草及烟草制品总氮的测定连续流动法. 北京: 国家烟草专卖局发布, 2002.
- 2 国家烟草专卖局编. YC/T159-2002 烟草及烟草制品水溶性糖的测定连续流动法. 北京: 国家烟草专卖局发布, 2002.
- 3 国家烟草专卖局编. YC/T160-2002 烟草及烟草制品总植物碱的测定连续流动法. 北京: 国家烟草专卖局发布, 2002.
- 4 国家烟草专卖局编. YC/T29-1996 卷烟用常规分析用吸烟机测定总颗粒物及焦油. 北京: 国家烟草专卖局发布, 1996.
- 5 国家烟草专卖局编. YC/T156-2001 卷烟烟气总颗粒物烟碱的测定气相色谱法. 北京: 国家烟草专卖局发布, 2001.
- 6 国家烟草专卖局编. YC/T30-1996 卷烟烟气气相中一氧化碳的测定非散射红外法. 北京: 国家烟草专卖局发布, 1996.
- 8 梁逸曾. 白灰黑复杂多组分分析体系及其化学计量学算法. 长沙: 湖南科技出版社, 1996:32-36.