



引用格式:杨建云,田孟玉,裴梓烨,等. 基于近红外光谱技术的初烤烟叶致香成分含量特性快速定量评价[J]. 轻工学报,2018,33(5):60-68.

中图分类号:TS41⁺1 文献标识码:A

DOI:10.3969/j.issn.2096-1553.2018.05.008

文章编号:2096-1553(2018)05-0060-09

基于近红外光谱技术的 初烤烟叶致香成分含量特性快速定量评价

Rapid quantitative evaluation of the content of aroma components in flue-cured tobacco leaves based on near-infrared spectroscopy

杨建云,田孟玉,裴梓烨,赵蔚,马骥,李俊谷,杨乾栩

YANG Jianyun, TIAN Mengyu, PEI Ziyue, ZHAO Wei, MA Ji, LI Jungu,

YANG Qianxu

关键词:

初烤烟叶;致香成分;
近红外光谱技术;快
速定量评价;主成分
分析

云南中烟工业有限责任公司 技术中心,云南 昆明 650231

Technology Center, China Tobacco Yunnan Industrial Co., Ltd., Kunming 650231, China

Key words:

flue-cured tobacco
leaves;
aroma component;
near-infrared
spectroscopy;
rapid quantitative
evaluation; principal
component analysis

摘要:为实现对初烤烟叶致香成分含量特性的快速评价,以河南、贵州和云南K326初烤烟叶为样本,采用近红外光谱技术结合偏最小二乘法判断分析(PLS-DA)建立PLS-DA模型和定量校正模型,对不同种植环境中K326初烤烟叶致香成分含量特性进行快速定量评价,结果表明:1)3个产地初烤烟叶致香成分在PLS-DA模型中差异性明显,第1,第2主成分对模型具有较好的解释能力,对模型分类变量的解释能力为81.4%;2) β -大马酮、亚麻酸甲酯、丁基化羟基甲苯、 β -紫罗兰酮等化合物是对3个产地样品的重要区分性化合物;3)基于第1,第2主成分得分的近红外校正模型,其决定系数分别为0.9700和0.9080,具有较好的稳定性和预测准确性。

收稿日期:2018-01-24

基金项目:云南省科技厅科技计划项目(2018FD164);云南中烟工业有限责任公司科技项目(2015CP02, 2016CP03)

作者简介:杨建云(1980—),男,四川省岳池县人,云南中烟工业有限责任公司工程师,主要研究方向为卷烟产品研发。

通信作者:杨乾栩(1987—),男,重庆市人,云南中烟工业有限责任公司工程师,博士,主要研究方向为烟草原料。

Abstract: In order to realize the rapid evaluation of the content of aroma components in flue-cured tobacco leaves, the flue-cured tobacco leaves of He'nan, Guizhou and Yunnan K326 were used as samples, and the PLS-DA model and quantitative correction model were established by using near-infrared spectroscopy combined with partial least squares method. The differences in the aroma components of K326 flue-cured tobacco leaves in the planting environment were quickly and quantitatively evaluated. The results showed that: 1) The aroma components of the flue-cured tobacco leaves in the three producing areas were significantly different in the PLS-DA model, and the first and second principal components had good interpretation ability. The explanatory power of the model categorical variable was 81.4%; 2) β -Damascone, methyl linolenic acid, butylated hydroxytoluene, β -ionone and other compounds were important distinguishing compounds for the three origin samples; 3) The near-infrared correction model based on the first and second principal component scores had a coefficient of determination of 0.970 0 and 0.908 0, respectively, and had good stability and prediction accuracy.

0 引言

卷烟工业对烟叶内在化学成分,尤其是致香成分极为关注^[1],致香成分的种类和含量影响着烟叶及其制品的香气品质.致香成分含量差异的本质是致香成分代谢特性的不同,其影响因素包括烟叶品种、部位、生态区域、栽培措施、烘烤调制技术等^[2-6].各影响因素与烟草致香成分含量差异的相关性研究可揭示影响因素对致香成分代谢特性影响的强弱,同时也可发现不同条件下种植的烟叶其致香成分的关键标记物,从而为进一步科学地做好烟草种植规划、彰显卷烟风格特征奠定基础.目前,各影响因素与烟草致香成分含量差异的相关性研究一般采用气相色谱-质谱联用仪(GC-MS)结合化学计量学方法.该方法的优势在于GC-MS法定性和定量准确,化学计量学方法能有效揭示多维数据间的相关性.L. Zhang等^[3]基于初烤烟叶的GC-MS数据,结合HCA,PCA和OPLS-DA方法分析了我国不同省份烟叶致香成分的含量特性差异及其影响因素.GC-MS法可准确测定烟叶中致香成分的种类和含量,但该方法存在前处理复杂、测定时间长、易产生环境污染、成本较高等不足.近红外分析技术作为一种简便、快速、无损的检测方法已被广泛应用于对各行

业产品品质的快速分析^[7-10].在烟草领域,近红外分析技术被应用于烟草品质的定性分析和定量测定,其中定性分析应用于卷烟配方和不同批次间卷烟产品的质量稳定性评价等方面,定量测定包括快速测定烟草品质的多项化学指标,如烟碱、总氮、总糖、还原糖、多酚、淀粉、纤维素等.近红外光谱信息能从整体上反映烟草的理化性质,利用其分析初烤烟叶致香成分含量具有一定的可行性.偏最小二乘法判别分析(PLS-DA)是一种建立样本属性和预测值之间关系的有监督模式的方法,其整合了主成分分析、相关分析和判别分析,具有运算速度快、准确度高的优点,广泛应用在近红外数据分析领域^[11].

鉴于此,本研究拟以河南、贵州和云南K326初烤烟叶为样本,采用近红外光谱技术结合PLS-DA建立PLS-DA模型和定量校正模型,并对不同种植环境中K326初烤烟叶致香成分含量特性进行评价,以期快速定量评价K326致香成分差异提供参考.

1 实验材料与方法

1.1 材料与仪器

主要试剂:CH₂Cl₂(色谱纯),西陇化工股份有限公司产;萘(色谱纯),百灵威科技有限

公司产。

主要材料:收集 2014 年份的河南、贵州和云南产地 K326 品种, B2F, C3F, X2F 等级的初烤烟叶共计 300 份, 2.0 kg/份, 样品详细信息见表 1。

表 1 样品信息表

Table 1 Sample information table

产地	部位和等级	样品数量/份
河南	B2F	30
	C3F	40
	X2F	30
贵州	B2F	30
	C3F	40
	X2F	30
云南	B2F	30
	C3F	40
	X2F	30

主要仪器: Truelab2000 同时蒸馏萃取仪, 上海楚柏实验室设备有限公司产; R-3000 型旋转蒸发仪, 瑞士 Buchi 公司产; KDM 型调温电热套, 山东鄄城华鲁电热仪器有限公司产; Antaris II 型 FT-NIR 光谱仪, 美国 Thermo Fisher 公司产; Agilent GC6890N/MS5975I 气相色谱/质谱联用仪 (GC-MS), 美国 Agilent 公司产; Milli-Q 超纯水系统, 美国 Millipore 公司产。

1.2 实验方法

1.2.1 初烤烟叶样品前处理 准确称取备用样品 25.0 g, 每个样本平行 3 份, 置于恒温恒湿箱 (温度 22 °C, 湿度 60% RH) 平衡 24 h; 采用同时蒸馏萃取法, 将盛有烟叶样品的一端加入 350 mL 蒸馏水置于电热套上加热, 另一端加入 30 mL 二氯甲烷置于水浴锅中加热, 水浴锅温度为 60 °C, 蒸馏萃取 2 h; 萃取完成后, 采用旋转蒸发仪 (温度约为 20 °C, 压力为 560 MPa, 转速为 65 r/min) 将萃取液浓缩至 1.0 mL, 装入样品瓶中, 待测。

1.2.2 初烤烟叶致香物质含量测定 致香物质含量测定: 采用 GC-MS 测定上述浓缩液的致

香物质含量, 作为对应初烤烟叶中致香成分的含量, 测定条件见文献 [12]。经含量测定所得图谱用计算机谱库 (NIST05, Wiley275) 检索, 确定各谱峰归属的化合物; 采用内标法 (萘为内标), 测定各物质的相对含量。

1.2.3 初烤烟叶近红外光谱的采集 按照 DBT 497—2013《烟草及烟草制品 主要化学成分指标 近红外校正模型建立与验证导则》所述方法, 首先进行样品前处理: 将样品粉碎、过 60 目筛, 制成粒径大小均匀、水分含量相对稳定的样品 (约 60 g), 密封备用; 采用 FT-NIR 光谱仪采集样品光谱数据, 设定光谱扫描范围为 10 000 ~ 4000 cm^{-1} , 分辨率为 8 cm^{-1} , 扫描次数为 64 次, 开机预热 2 h 后, 采集样品近红外光谱。

2 致香成分代谢特性的建模与评价

2.1 致香成分代谢特性 PLS-DA 模型的建立

按烟叶产地 (河南、贵州和云南) 对 300 份样品进行类别划分, 结合对应的致香成分含量, 采用 SIMCA-P+ 11.5 软件, 建立基于 K326 初烤烟叶致香成分含量的致香成分代谢特性 PLS-DA 模型, 提取能表征不同类间差异性即致香成分含量特征性分布的参数 t_1 , t_2 和 t_3 (分别表示第 1, 2, 3 主成分下每个样品对应的得分)。

2.2 定量校正模型的建立

近红外光谱的多元散射校正 MSC (multiplicative scatter correction) 可消除因颗粒分布不均匀而产生的散射影响, Norris 滤波器和二阶微分可过滤噪声, 消除基线漂移。因此本文采用 MSC + Norris (11, 3) 平滑 + 二阶微分对样品近红外光谱进行预处理, 并采用 TQ Analyst 8.6 软件完成校正。

随机挑选预处理后的 270 个样品 (河南、贵州和云南各 90 个) 的近红外光谱数据, 结合对应的 t_1 和 t_2 , 采用偏最小二乘法 (PLS) 建立表

征致香成分含量特性分布的参数 t_1 和 t_2 的近红外校正模型。

2.3 定量校正模型的评价

采用参数相关系数 (R^2)、校正均方根误差、交互验证均方根误差,基于参数预测值与实测值之间的配对 t 检验结果、验证平均误差 (\bar{e}_v)、验证标准误差和验证标准误差的标准偏差,根据未参加建模的 30 个样品的致香成分含量数据和近红外光谱数据,对近红外校正模型预测的准确性和稳定性进行外部评价。

3 结果与讨论

3.1 致香成分测定结果

采用 GC-MS 法对烟叶致香成分进行测定,共检出如下 71 种化合物:1-戊烯-3-酮、3-羟基-2-丁酮、3-甲基-1-丁醇、吡啶、己醛、面包酮、糠醛、糠醇、2-环戊烯-1,4-二酮、1-(2-呋喃基)-乙酮、丁内酯、2-吡啶甲醛、糠酸、苯甲醛、甲基糠醛、苯酚、2-戊基呋喃、2,4-庚二烯醛 A、4-吡啶甲醛、1H-吡咯-2-甲醛、2,4-庚二烯醛 B、苯甲醇、苯乙醛、1-(1H-吡咯-2-基)-乙酮、对甲基苯酚、二甲基苯酚、芳樟醇、壬醛、甲基-3-吡啶基酮、苯乙醇、氧化异佛尔酮、2,6-壬二烯醛、苯并[b]噻吩、藏花醛、胡薄荷酮、2,3-二氢苯并呋喃、吡啶、乙烯基愈疮木酚、茄酮、 β -大马酮、 β -二氢大马酮、去氢去甲基烟碱、香叶基丙酮、 β -紫罗酮、丁基化羟基甲苯、3-(1-甲基乙基)(1H)吡啶[3,4-b]吡啶、2,3'-联吡啶、二

氢猕猴桃内酯、巨豆三烯酮 A、巨豆三烯酮 B、巨豆三烯酮 C、巨豆三烯酮 D、3-氧代- α -紫罗兰醇、十四醛、肉豆蔻酸甲酯、肉豆蔻酸、降茄二酮、葱、螺岩兰草酮、新植二烯、邻苯二甲酸酯二丁酯、金合欢基丙酮 A、棕榈酸甲酯、棕榈酸、棕榈酸乙酯、异瑟模环烯醇、亚麻酸甲酯、植醇、西柏三烯二醇、金合欢基丙酮 B、茄那士酮。

3.2 PLS-DA 模型最优主成分数的选择

选择适宜的主成分数 (A) 是 PLS-DA 模型稳定性和准确性的重要因素。选择主成分数时,主成分特征值 (*Eigenvalues*) 一般需要大于 2, 同时应兼顾主成分分析对原光谱数据的累积解释能力和模型的预测准确性及精度。 R^2X 表示该主成分数下模型对光谱矩阵的解释能力, R^2X_{cum} 表示模型累积对光谱矩阵的解释能力; R^2Y 表示该主成分数下模型对分类变量的预测能力, R^2Y_{cum} 表示模型累积对分类变量累积预测能力; Q^2 是基于累计交叉有效性的模型预测精度参数, Q^2_{cum} 表示模型累积的预测精度。不同主成分下,建立 PLS-DA 模型的 R^2X , R^2X_{cum} , *Eigenvalues*, R^2Y , R^2Y_{cum} , Q^2 和 Q^2_{cum} 见表 2。

由表 2 可知:当模型的主成分数由 1 增加至 4 时, R^2X_{cum} 为 0.514, R^2Y_{cum} 为 0.908, Q^2_{cum} 为 0.897; 当主成分数增加到 5 时,上述 3 个参数进一步增加,但其特征值为 1.83 (< 2)。因此,所建立的 PLS-DA 模型最优的主成分数为 4。

3.3 PLS-DA 模型主成分得分分析

选择 PLS-DA 模型的主成分数为 4 时,各

表 2 不同主成分下 PLS-DA 模型的相关参数

Table 2 PLS-DA model related parameters with different principal components

主成分数	R^2X	R^2X_{cum}	<i>Eigenvalues</i>	R^2Y	R^2Y_{cum}	Q^2	Q^2_{cum}
1	0.195	0.195	12.30	0.434	0.434	0.430	0.430
2	0.153	0.348	9.63	0.379	0.814	0.658	0.805
3	0.106	0.454	6.71	0.068	0.881	0.341	0.872
4	0.059	0.514	3.74	0.027	0.908	0.195	0.897
5	0.029	0.543	1.83	0.015	0.922	0.048	0.902

样品第1,第2主成分得分(t_1, t_2)分布图如图1所示,各样品第1,第2,第3主成分得分(t_1, t_2, t_3)分布图见图2.

由图1和图2可知:由于K326烟叶产地差异,其致香成分在多维模型空间呈现一定的分布差异, t_1 和 t_2 共同作用即可实现对3个产地样品较好的区分效果,此时对模型分类变量的解释能力为81.4%; t_1, t_2 和 t_3 共同作用可实现对3个产地样品完全区分的效果;不同产地样品 t_1, t_2 与 t_3 数值之间的差异可反映出K326初烤烟叶致香成分含量在PLS-DA模型的空间差异,因此 t_1, t_2 和 t_3 可表征不同生态条件下K326致香成分含量特性.

不同地区样品对应的 t_1 和 t_2 最大值、最小值和平均值见表3,3个地区对应的 t_1 和 t_2 分布范围如图3所示.由表3和图3不难发现:3个地区样品的 t_1 值和 t_2 值具有较明显的差异,河南地区 t_1 值较低, t_2 值适中且分布较广;云南地区 t_1 值较高, t_2 值较低;贵州地区 t_1 值适中, t_2 值较高.正是由于不同地区 t_1 值, t_2 值的数据特性,即致香成分含量特性不同,3个地区K326样品具有较强的区分模式.

3.4 致香成分重要性分析

在PLS-DA模型中,不同主成分下各致香

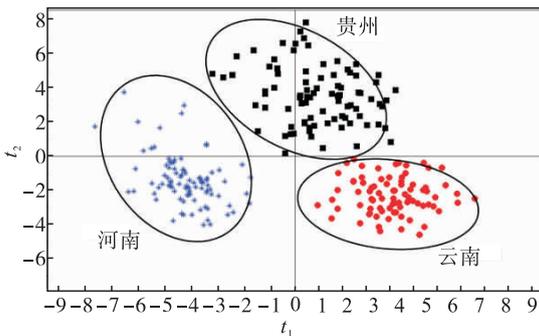


图1 不同初烤烟叶致香成分指标分类得分图 (第1,第2主成分得分)

Fig.1 Classification scores of aroma components of different flue-cured tobacco leaves (The 1st ,2nd principal components score)

成分的变量重要性投影VIP(variables importance of projection)值能够反映该物质对提高分类变量区分度的贡献率^[13].因此,分析不同主成分下各致香成分VIP值的排序,能进一步明确造成不同地区K326致香成分含量特性差异性的物质基础.第1,第2和第3主成分得分下对应的VIP值排序见表4.

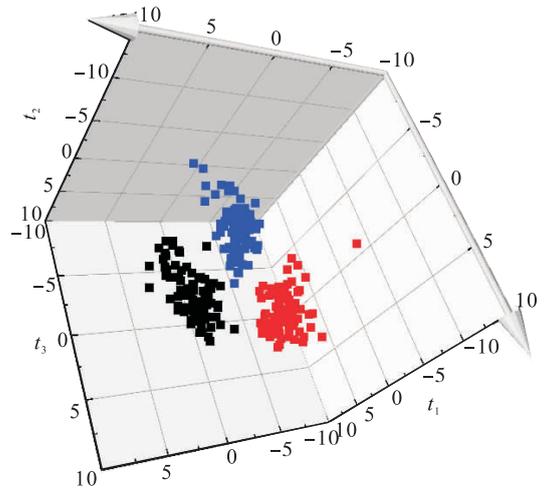


图2 不同初烤烟叶致香成分指标分类得分图 (第1,第2,第3主成分得分)

Fig.2 Classification scores of aroma components of different flue-cured tobacco leaves (The 1st ,2nd ,3rd principal components score)

表3 不同地区烟叶样品 t_1, t_2 的最大值、最小值和平均值

Table3 Maximum, minimum and average values of t_1 and t_2 in tobacco leaves samples from different regions

地区	统计量	t_1	t_2
河南	最大值	-1.794	3.845
	最小值	-7.354	-3.710
	平均值	-4.186	-1.191
贵州	最大值	4.524	7.870
	最小值	-2.878	-0.036
	平均值	0.955	3.407
云南	最大值	6.395	0.900
	最小值	-0.833	-4.614
	平均值	3.231	-2.216

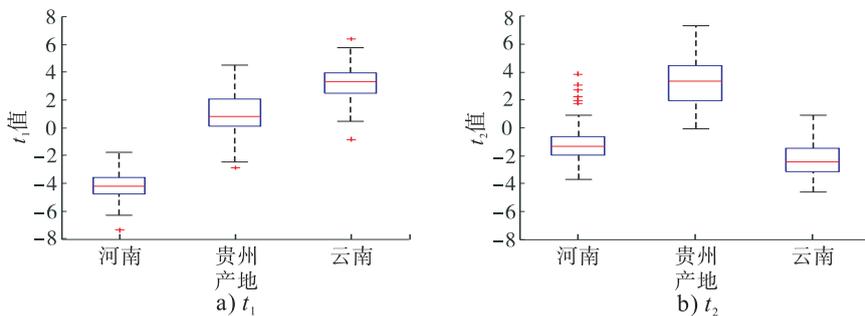


图3 不同地区烟叶样品 t_1, t_2 的最大值、最小值和平均值分布范围

Fig. 3 Maximum, minimum and average values of t_1 and t_2 in tobacco leaves samples from different regions

表4 第1,第2和第3主成分下对应的VIP值排序

Table 4 The corresponding VIP values sort of the principal components 1, 2 and 3

主成分位次	化合物名称	VIP 值	主成分位次	化合物名称	VIP 值
第1主成分	亚麻酸甲酯	1.835 21	第2主成分	胡薄荷酮	1.125 63
	香叶基丙酮	1.781 72		2-环戊烯-1,4-二酮	1.123 38
	吡啶	1.752 33		西柏三烯二醇	1.114 14
	棕榈酸乙酯	1.706 25		2,6-壬二烯醛	1.106 36
	棕榈酸甲酯	1.660 28		氧化异佛尔酮	1.080 31
	苯并[b]噻吩	1.556 05		邻苯二甲酸二丁酯	1.075 90
	β -紫罗兰酮	1.546 45		苯乙醛	1.069 29
	胡薄荷酮	1.509 73		3-氧代- α -紫罗兰醇	1.064 32
	苯乙醛	1.463 88		藏花醛	1.056 99
	4-吡啶甲醛	1.425 20		4-吡啶甲醛	1.047 65
	邻苯二甲酸二丁酯	1.380 10		1-(3-吡啶基)-乙酮	1.037 27
	金合欢基丙酮 B	1.315 72		丁内酯	1.020 40
	金合欢基丙酮 A	1.285 04		β -紫罗兰酮	1.627 79
	β -大马酮	1.272 07		丁基化羟基甲苯	1.610 48
	丁内酯	1.244 77		β -大马酮	1.506 01
	茄那土酮	1.214 11		亚麻酸甲酯	1.369 14
	去氢去甲基烟碱	1.207 84		香叶基丙酮	1.311 93
	1-(1H-吡咯-2-基)-乙酮	1.207 05		吡啶	1.257 86
	氧化异佛尔酮	1.202 08		3-甲基-2-丁烯醛	1.235 45
	西柏三烯二醇	1.182 55		棕榈酸乙酯	1.227 80
2,3'-联吡啶	1.178 52	苯甲醛	1.222 06		
藏花醛	1.178 10	β -二氢大马酮	1.222 03		
1H-吡咯-2-甲醛	1.173 86	棕榈酸甲酯	1.215 23		
β -二氢大马酮	1.060 54	巨豆三烯酮 C	1.197 62		
3-氧代- α -紫罗兰醇	1.040 89	1-(3-吡啶基)-乙酮	1.175 06		
2,3-二氢苯并呋喃	1.035 65	茄酮	1.168 14		
苯乙醇	1.021 32	吡啶	1.167 68		
β -紫罗兰酮	1.693 95	糠醛	1.162 81		
丁基化羟基甲苯	1.673 81	巨豆三烯酮 D	1.139 39		
β -大马酮	1.567 33	金合欢基丙酮 B	1.129 44		
亚麻酸甲酯	1.404 18	2-环戊烯-1,4-二酮	1.104 67		
香叶基丙酮	1.357 42	苯并[b]噻吩	1.103 28		
吡啶	1.291 36	2,6-壬二烯醛	1.099 03		
3-丁基-2-丁烯醛	1.277 94	胡薄荷酮	1.091 80		
棕榈酸乙酯	1.277 30	苯乙醇	1.075 52		
β -二氢大马酮	1.270 35	丁内酯	1.072 39		
巨豆三烯酮 C	1.226 25	西柏三烯二醇	1.071 86		
棕榈酸甲酯	1.222 43	氧化异佛尔酮	1.057 18		
苯甲醛	1.214 39	1-(1H-吡咯-2-基)-乙酮	1.054 81		
茄酮	1.211 01	邻苯二甲酸二丁酯	1.036 15		
吡啶	1.200 29	苯乙醛	1.034 84		
糠醛	1.185 84	4-吡啶甲醛	1.032 13		
巨豆三烯酮 D	1.168 96	3-氧代- α -紫罗兰醇	1.026 11		
金合欢基丙酮 B	1.167 62	藏花醛	1.022 05		
苯并[b]噻吩	1.137 65				

由表4可知:第1主成分下,对分类变量贡献较大的化合物有27种,第2,第3主成分下,对分类变量贡献较大的分别有30种和32种.分析第1,第2主成分下对分类变量贡献较大的物质排序可知,两者之间不同物质间的排序差异较大.分析第2,第3主成分下对分类变量贡献较大的化合物排序可知,两个主成分之间不同物质间的排序差异较小.结合不同主成分对X矩阵(致香成分含量)和Y值(不同地区分类变量)的解释和预测类能力,可知 t_1 与 t_2 间的差异是反映不同地区K326初烤烟叶含量特性差异的主要参数.因此下一步建立基于致香物质近红外光谱数据的 t_1 和 t_2 校正模型,用于快速表征K326致香成分含量特性.

为进一步明晰上述VIP值大于1的化合物与 t_1, t_2 的相关性,将上述化合物含量与 t_1, t_2 之间做Pearson相关性分析,分析结果见表5.首先需要计算化合物重要程度综合指数,公式如下:

化合物重要程度综合指数 =

$SQRT(\text{与 } t_1 \text{ 相关系数}^2 + \text{与 } t_2 \text{ 相关系数}^2)$

其中, $SQRT$ 代表开算术平方根运算,综合指数反映化合物在 t_1 和 t_2 整体上具有的相关性程度,可作为判定化合物重要程度的依据.

由表5可知,VIP值大于1的化合物与 t_1 和 t_2 值之间在0.01水平下整体均呈现较强的相关性,其中 β -大马酮、亚麻酸甲酯、丁基化羟基甲苯、 β -紫罗兰酮、棕榈酸乙酯、吡啶、香叶基丙酮、3-甲基-2-丁烯醛、2-环戊烯-1,4-二酮、糠醛的综合指数均大于0.7,是对3个产地样品的重要区分性化合物,与Y. X. Xi等^[14]研究结论基本一致.

3.5 定量校正模型的拟合结果分析

基于近红外光谱 t_1 和 t_2 的定量校正模型拟合结果分别如图4和图5所示.从图中可以看出,测定值与模型预测值整体上较为均匀地

表5 VIP值大于1的化合物与 t_1, t_2 的 Pearson 相关系数及综合指数

Table 5 Pearson correlation coefficient and composite index between compounds with VIP greater than 1 and t_1, t_2

化合物	与 t_1 相关系数	与 t_2 相关系数	综合指数
β -大马酮	-0.615**	0.666**	0.907
亚麻酸甲酯	0.791**	0.357**	0.868
丁基化羟基甲苯	0.000	0.853**	0.853
β -紫罗兰酮	0.526**	0.663**	0.846
棕榈酸乙酯	0.765**	0.188**	0.788
吡啶	-0.750**	0.204**	0.777
香叶基丙酮	-0.764**	-0.142*	0.777
3-甲基-2-丁烯醛	0.000	0.733**	0.733
2-环戊烯-1,4-二酮	0.000	0.729**	0.729
糠醛	0.000	0.707**	0.707
苯乙醛	-0.699**	0.029	0.700
茄酮	0.000	0.697**	0.697
西柏三烯二醇	0.541**	0.436**	0.695
氧化异佛尔酮	-0.530**	0.449**	0.695
棕榈酸甲酯	0.680**	0.065	0.683
胡薄荷酮	-0.627**	0.213**	0.662
3-氧代- α -紫罗兰酮	-0.462**	0.451**	0.646
金合欢基丙酮A	-0.638**	0.000	0.638
吲哚	0.000	0.633**	0.633
巨豆三烯酮C	0.000	0.614**	0.614
邻苯二甲酸二丁酯	-0.552**	-0.231**	0.598
4-吡啶甲醛	0.569**	0.169**	0.594
苯并[b]噻吩	0.589**	0.015	0.589
去氢去甲基烟碱	-0.579**	0.000	0.579
2,3'-联吡啶	-0.574**	0.000	0.574
β -二氢大马酮	-0.568**	0.000	0.568
金合欢基丙酮B	0.562**	0.000	0.562
巨豆三烯酮D	0.000	0.561**	0.561
苯乙醇	-0.516**	0.000	0.516
1H-吡咯-2-甲醛	0.511**	0.000	0.511
藏花醛	0.360**	0.350**	0.502
丁内酯	0.500**	-0.043	0.502
巨豆三烯酮B	0.000	0.493**	0.493
2,3-二氢苯并咪喃	-0.480**	0.000	0.480
1-(1H-吡咯-2-基)-乙酮	0.458**	0.000	0.458
茄那土酮	0.451**	0.000	0.451
2,6-壬二烯醛	0.000	0.428**	0.428
苯甲醛	0.000	-0.409**	0.409
β -大马酮	0.000	-0.214**	0.214

注:**表示在0.01水平(双侧)上显著相关;*表示在0.05水平(双侧)上显著相关;0.000表示该化合物在对应 t_1 或 t_2 上的VIP值小于1,因此不做相关性分析

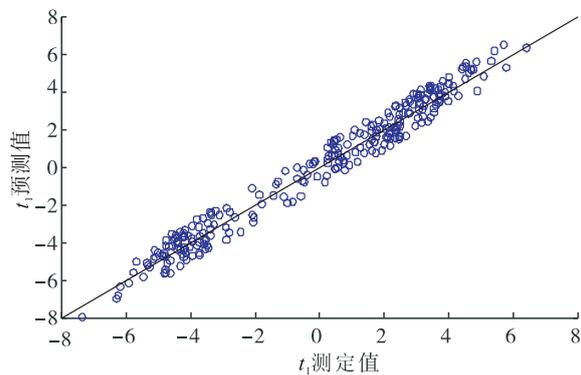


图4 t_1 测定值与模型预测值的散点图

Fig.4 Scatter plot of t_1 measured value and model predicted value

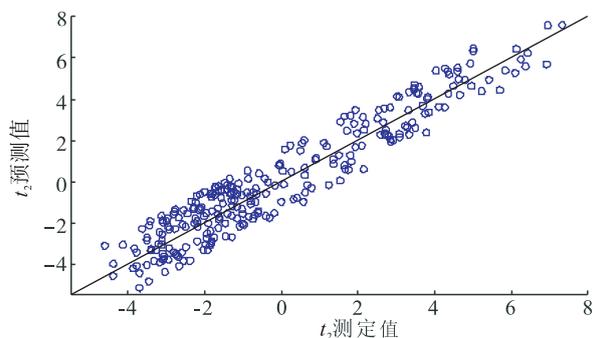


图5 t_2 测定值与模型预测值的散点图

Fig.5 Scatter plot of t_2 measured value and model predicted value

分布在直线 $y=x$ 两侧;近红外校正模型拟合效果均较为理想,对 t_1 的拟合程度要优于对 t_2 的拟合程度.

t_1 和 t_2 校正模型内部和外部评价结果如表6和表7所示.从表6可知:近红外光谱对 t_1 和

表6 t_1 和 t_2 校正模型内部评价统计量

Table 6 t_1 and t_2 correction model internal evaluation statistics

模型名称	决定系数 R^2	适宜主成分数 k	校正标准误差	交互验证均方根误差
t_1 校正模型	0.970 0	9	0.829	0.976
t_2 校正模型	0.908 0	7	1.130	1.220

表7 t_1 和 t_2 校正模型外部评价统计量

Table 7 t_1 and t_2 correction model external evaluation statistics

模型名称	验证平均误差 \bar{e}_v	验证标准误差	验证标准误差的标准偏差	t 分布值	概率 P 值
t_1 校正模型	-0.002 91	0.099 60	0.100 83	-0.183	0.856
t_2 校正模型	-0.021 11	0.126 06	0.125 86	-1.062	0.295

t_2 的决定系数分别为 0.970 0 和 0.908 0,表明近红外光谱可对 t_1 和 t_2 有较好的解释效果;且交互验证均方根误差均小于 2,表明所构建模型较为稳健,不存在过拟合现象.从表7可以看出,验证平均误差、验证标准误差均较小,且配对 t 检验结果均大于 0.05,表明内部评价结果与外部评价结果之间不存在统计学差异.以上分析表明,所建立的模型具有较好的稳定性和预测准确性.

3 结论

本文以河南、贵州和云南 2014 年的 K326 初烤烟叶为样本,将近红外光谱技术与 PLS-DA 相结合,建立了 PLS-DA 模型和定量校正模型,用于评价不同种植环境 K326 初烤烟叶致香成分的含量特性.结果表明:3 个产地初烤烟叶致香成分在 PLS-DA 模型中差异性明显,第 1,第 2 主成分对模型具有较好的解释能力,对模型分类变量的解释能力为 81.4%,第 1,第 2 主成分得分可作为样品致香成分的特性参数; β -大马酮、亚麻酸甲酯、丁基化羟基甲苯、 β -紫罗兰酮等化合物是对 3 个产地样品的重要区分性化合物;基于第 1、第 2 主成分得分的近红外校正模型,其决定系数分别为 0.970 0 和 0.908 0,具有较好的稳定性和预测准确性.

在后续的研究中,可进一步拓宽 K326 致香成分含量特性差异性的样品,从校正模型的波长变量筛选和建模方法优化方面,进一步提

高 PLS-DA 模型和 t_1, t_2 校正模型的稳定性和预测准确性.

参考文献:

- [1] 邵丽, 晋艳. 生态条件对不同烤烟品种烟叶产质量的影响[J]. 烟草科技, 2002(10): 40.
- [2] 常寿荣, 罗华元, 王玉, 等. 云南烤烟种植海拔与致香成分的相关性分析[J]. 中国烟草科学, 2009, 30(3): 37.
- [3] ZHANG L, WANG X Y, GUO J Z, et al. Metabolic profiling of Chinese tobacco leaf of different geographical origins by GC-MS [J]. Journal of Agricultural and Food Chemistry, 2013, 61: 2597.
- [4] YANG C, WU W, WU S C, et al. Aroma types of flue-cured tobacco in China: Spatial distribution and association with climatic factors [J]. Theoretical and Applied Climatology, 2014, 115 (3/4): 541.
- [5] ZHU X L, GAO Y, CHEN Z Y, et al. Development of a chromatographic fingerprint of tobacco flavor by use of GC and GC-MS [J]. Chromatographic, 2009, 69(7/8): 745.
- [6] 王玉, 王保兴, 武怡, 等. 卷烟挥发性成分的聚类分析[J]. 烟草科技, 2007(2): 48.
- [7] 陈贵平, 童佩瑾, 耿金培, 等. 近红外光谱结合膜富集技术测定饮料中微量邻苯二甲酸二异辛酯的含量[J]. 分析测试学报, 2012, 31(5): 605.
- [8] 刘波平, 荣茵, 邓泽元, 等. 基于 PLS - 自组织竞争神经网络近红外光谱技术对鲜乳和掺假乳的检测方法研究[J]. 分析测试学报, 2008, 27(11): 1147.
- [9] 褚小立, 许育鹏, 陆婉珍. 支持向量回归建立成品汽油通用近红外校正模型的研究[J]. 分析测试学报, 2008, 27(6): 619.
- [10] CHU X, YUAN H F, LU W I. Research and applications of near infrared spectroscopy in China in recent years [J]. Analytical Instrumentation, 2006, 22(2): 1.
- [11] 王惠文. 偏最小二乘回归方法及其应用 [M]. 北京: 国防工业出版社, 1999.
- [12] 杨乾栩, 冯斌, 吴长伟, 等. 基于大孔树脂静态吸附的烟草增香组分制备及应用 [J]. 烟草科技, 2014(10): 39.
- [13] OUSSAMA A, ELABADI F, PLATIKANOV S, et al. Detection of olive oil adulteration using FT-IR spectroscopy and PLS with variable importance of projection (VIP) scores [J]. Journal of the American Oil Chemists Society, 2012, 89(10): 1807.
- [14] XI Y X, SONG J Z, LI F, et al. Different types flue-cured tobacco aroma precursors and variance analysis of aroma components content [J]. Journal of Zhejiang Agricultural Sciences, 2011 (2): 355.