第38卷，第8期
2018年8月

光谱学与光谱分析
Spectroscopy and Spectral Analysis

Vol. 38, No. 8, pp2407-2411
August, 2018

应用近红外光谱或化学指标判别不同产地晒红烟的相似性

马　莉[1]，李雪莹[2]，杜国荣[1]，丁　睿[3]，王允白[3]，马雁军[1*]，李军会[2*]

1. 上海烟草集团有限责任公司技术研发中心北京工作站，北京　101121
2. 中国农业大学信息与电气工程学院，北京　100083
3. 中国农业科学院烟草研究所质量安全研究中心，山东 青岛　266101

摘　要　近红外光谱采集物质信息较全面，是多种物质的复合信息，但具体特征不明显；化学指标虽能够体现物质的具体特征，但所包含的物质信息却不够全面。以2012年和2013年贵州、湖南、吉林、江西、山东、四川六个产地共115个晒红烟样品为研究对象，其近红外光谱采用一阶导数和平滑处理，测定的26个化学指标及其26个计算值(譬如糖碱比等)采用归一化处理，应用PPF投影技术对不同产地晒红烟进行相似性分析。结合近红外光谱和化学指标的分析结果表明：应用近红外光谱和化学指标分析六个产地晒红烟分布规律较为一致，表明通过近红外光谱信息和大量检测数据均可分析产地特征的相似性；对52个化学指标的方差贡献率分析得出，决定产地特征的重要指标是亚硝胺；应用近红外光谱对小产地之间进行聚类分析，得出部分小产地之间具有可替代性，结合主要内在的化学指标工业需求取向，可合理替代不同烟叶原料，降低产品加工质量的波动性。

关键词　晒红烟；近红外光谱；化学指标；相似性

中图分类号：TS42　　**文献标识码**：A　　**DOI**：10.3964/j.issn.1000-0593(2018)08-2407-05

引　言

不同产地、部位的烟叶原料采用传统的感官质量评价方法来判断烟叶原料间的相似性和差异性，用于研究烟叶间的配伍性和可替代性，具有十分重要的意义[1]。该方法使用时，判别差异不仅费时费力，往往因人为因素的不确定性造成很大差异，甚至失真。故根据可靠的分析数据，研究一种简便可靠的判别方法，用于判定烟叶原料间的相似性，寻找原料间的可替代性对于工业加工生产非常必要。近年来，近红外光谱技术不断发展，该技术成为一个简便可靠的工具已在农业[2-3]、食品[4-5]和烟草[6]等领域行业有广泛的应用[7-8]。杨凯等基于近红外光谱应用主成分及Fisher准则(PPF)的方法，对上、中、下部烟叶建立投影分析模型，结果表明上、下部烟叶的近红外光谱特征具有显著差别，中部烟叶分别与上部和下部具有一定程度重叠；目前，多数研究为应用近红外光谱建立模型对烟叶进行定性分析，但结合烟叶的化学指标用于烟叶复烤配方设计或者卷烟产品设计中不同烟叶间可替代性评价的文章鲜有报道。

我国的晒红烟种植地域广，但规模小，应用范围少，主要应用于混合型卷烟中，具有一定的使用价值。烟草产地等分析大多集中在烤烟和白肋烟上，对晒红烟产地的判别等鲜有报道。本工作以晒红烟为实验样品，应用近红外光谱和化学指标对晒红烟产地进行相似性分析，寻找烟叶原料间的可替代性，结果表明两种方法均能较好地分析不同产地晒红烟间的相似性，结合主要内在的化学指标工业需求取向，可合理替代不同烟叶原料，降低产品加工质量的波动性。该方法的应用，将对烟叶原料间的合理替代产生重要支撑作用。

1　实验部分

1.1　材料和仪器

实验样品为某企业2012年和2013年贵州、湖南、吉林、江西、山东、四川六个产地的晒红烟样品，共115个。在产地实验中，近红外光谱分析实验选取：贵州26个，湖南25个，吉林19个，山东17个，江西10个，四川9个；化学值分析实验选取：贵州26个，湖南25个，吉林22个，山东22个，江西10个，四川10个。

收稿日期：2016-10-26，**修订日期**：2017-05-10

基金项目：上海烟草集团有限责任公司科技项目(K2012-2-003Z)和国家自然科学基金项目(61144012)资助

作者简介：马　莉，1976年生，上海烟草集团有限责任公司技术研发中心北京工作站工程师　　e-mail：pclover@sohu.com

*通讯联系人　　e-mail：13366036175@189.cn；caunir@cau.edu.cn

仪器设备为 MPA 型傅里叶变换近红外光谱仪(德国 Bruker 公司)。工作参数：光谱采集范围 12 000～3 500 cm^{-1}；光谱分辨率：8 cm^{-1}；扫描次数：64 次。

1.2 方法

1.2.1 近红外光谱扫描方法

取烟末样品 25 g 左右放入样品杯，在样品上方放置压样器，自然压紧，利用积分球漫反射进行光谱测量，采集烟末样品的近红外光谱，每个样品重复装样、测量 3 次，取平均光谱作为该样品光谱数据。

1.2.2 化学指标测定方法

采用的化学指标有 12 大类，26 个检测指标，26 个比值计算指标，共计 52 个化学指标。其中 12 大类化学指标的检测分析方法和采用标准见表 1，比值计算指标的计算方法见表 2。

表 1 晒红烟化学指标采用的检测方法标准

Table 1 Test method standards for chemical index of red sun-cured tobacco

化学指标	分析方法	采用标准
总糖	连续流动法	YC/T 159—2002
还原糖	连续流动法	YC/T 159—2002
总氮	连续流动法	YC/T 161—2002
总植物碱	连续流动法	YC/T 160—2002
钾	连续流动法	YC/T 217—2007
氯	连续流动法	YC/T 162—2011
蛋白质	连续流动法	YC/T 249—2008
生物碱	气质联用	YC/T 383—2010
硝酸根	离子色谱法	YC/T 248—2008
焦油	重量法	GB/T 19609—2004
烟气中烟碱	气相色谱	GB/T 23355—2009
烟叶和烟气中亚硝胺(TSNAs)	气相色谱热能分析联用	GB/T 23228—2008

1.2.3 近红外光谱数据与化学指标处理方法

将扫描得到的近红外光谱数据进行一阶导数和 S. G. 平滑预处理[9]，52 个化学指标采用归一化处理；分别应用 PPF 投影分析方法对晒红烟的产地进行相似性分析。

PPF(projection of basing on principal component and fisher criterion)投影分析方法，是主成分分析方法和 Fisher 准则联用的方法。基于主成分分析方法得到的光谱主成分数据，该方法对类内散布矩阵的逆矩阵 $\boldsymbol{S}_w^{-1}$ 与类间散布矩阵 S_b 的乘积 $\boldsymbol{S}_w^{-1} \cdot S_b$ 进行最优投影矢量求解，遵循聚类间的距离值最大和类内的离散度最小的原则，使样本的类内距离尽量小，类间距离尽量大，从而使各类别之间达到最大程度的分离，实现正确的分类，同时也能够客观地表现类内的离散性和类间的相似性[10-11]。

依据 PPF 投影值求解各类之间相似度值的计算公式见式(1)

$$S_{pq} = 1 - \frac{D_{pq}}{D_{pp} - D_{qq}} \tag{1}$$

式中，D_{pq} 为类间投影均值的欧式距离，D_{pp} 和 D_{qq} 类内投影值的类内离散度，当 $S_{pq} < 0$ 时，两类可实现完全区分，且 S_{pq} 的绝对值越大表示差异性越大，当 $0 < S_{pq} < 1$ 时，值越大代表相似性越好。

表 2 晒红烟计算指标采用的计算方法

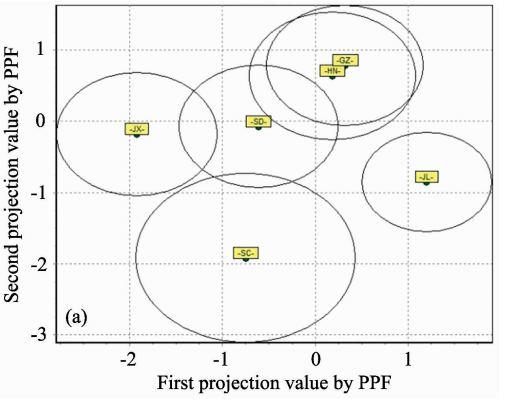
Table 2 Calculation method used in calculated index of red sun-cured tobacco

计算指标	计算方法
糖碱比	总糖/总烟碱
钾氯比	钾/氯
氮碱比	总氮/总烟碱
两糖比	还原糖/总糖
施木克值	总糖/蛋白质
碱碱比	总烟碱/挥发碱
硝酸盐 NO_3^-	阴离子百分比
五种碱和	烟碱＋降烟碱＋麦斯明＋新烟草碱＋假木贼碱
烟碱占比	烟碱/五种碱和
降烟碱占比	降烟碱/五种碱和
麦斯明占比	麦斯明/五种碱和
新烟草碱占比	新烟草碱/五种碱和
假木贼碱占比	假木贼碱/五种碱和
NNK 占比	NNK/NNK＋NAT＋NNN＋NAB
NAT 占比	NAT/NNK＋NAT＋NNN＋NAB
NNN 占比	NNN/NNK＋NAT＋NNN＋NAB
NAB 占比	NAB/NNK＋NAT＋NNN＋NAB
烟气中 NNK	转移率%/cig
烟气中 NAT	转移率%/cig
烟气中 NNN	转移率%/cig
烟气中烟碱	转移率%/cig
烟气中总粒相物	转移率%/cig
烟气中焦油	转移率%/cig

2 结果与讨论

2.1 基于近红外光谱及化学指标的晒红烟产地的相似性与替代性分析

将 1.1 中的 115 个晒红烟实验材料的近红外光谱和化学指标，按照 1.2.3 中的分析处理方法，得到 PPF 投影分析结果图见图 1，图中 GZ 代表贵州、HN 代表湖南、JL 代表吉林、SD 代表山东、JX 代表江西、SC 代表四川(注：图中圆圈表示各类的类内离散度，其半径值为类内各样品投影值标准



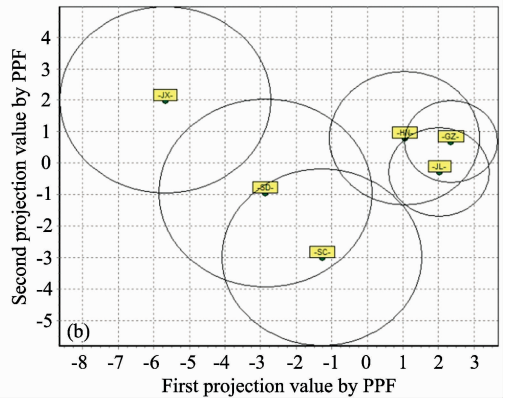


图 1　(a)产地的近红外光谱 PPF 投影分析结果；(b)产地的化学指标 PPF 投影分析结果

Fig. 1　(a) PPF projection analysis results of growing areas by near infrared spectroscopy; (b) PPF projection analysis results of growing areas by chemical index

差的 2 倍)；按照 1.2.3 中的相似度计算公式得到产地之间的相似度值见表 3。

表 3　六个产地之间的近红外光谱和化学指标相似度值

Table 3　Similarity values between near infrared spectroscopy and chemical index among six growing areas

近红外 化学值	湖南	山东	吉林	四川	江西
贵州	0.925 4 0.888 8	0.009 4 −0.229 0	−0.304 7 0.341 5	−0.413 9 −0.102 6	−0.458 9 −0.930 5
湖南		0.037 0 −0.069 0	−0.368 6 0.526 5	−0.419 3 0.087 4	−0.453 4 −0.726 4
山东			−0.487 1 −0.011 0	0.267 1 0.612 8	0.394 1 0.027 2
吉林				−0.201 8 0.274 5	−1.357 3 −1.005 6
四川					−0.161 9 −0.410 6

采用近红外光谱表征样品的判别结果图 1(a)显示，贵州晒红烟样品集与湖南晒红烟样品集有大部分交集，山东晒红烟样品集与贵州、湖南、江西、四川样品集均有较少部分交集，吉林晒红烟样品集与其他样品集完全分开，没有交集。

采用化学指标表征样品的判别结果图 1(b)显示，贵州晒红烟样品集、湖南晒红烟样品集、吉林晒红烟样品集有大部分交集，山东晒红烟样品集与江西、四川样品集均有部分交集，与湖南晒红烟样品集有较少部分交集。

综合图 1 和表 3 的分析结果，应用样品的近红外光谱或化学指标判别产地相似差异性，两个方法得到的四个产地分布规律较为一致，如贵州与湖南两个产地之间的相似性较高，具有相互替代性；吉林与江西、山东产地之间差异性较大等。

2.2　晒红烟产地差异的化学指标贡献率分析

不同的化学指标对晒红烟产地的划分有不同影响，即每个化学指标对晒红烟产地特征划分的方差贡献率有所不同，取方差贡献率大于 1%的化学指标见表 4。

表 4　晒红烟产地特征差异主要化学指标

Table 4　Difference between growing areas of red sun-cured tobacco under main chemical idexes

序号	化学成分指标	方差贡献率/%
1	亚硝胺 TSNAs	54.2
2	生物碱	10.4
3	还原糖	6.8
4	钾	5.3
5	总糖	4.1
6	总氮	4.0
7	蛋白质	3.6
8	氯	2.1
9	硝酸盐	1.9
10	总植物碱	1.2

由表 3 看出，亚硝胺 TSNAs 方差贡献率明显高于其他化学指标，达 54.2%，其次为生物碱和还原糖，分别为 10.4%和 6.8%，说明在应用化学指标分析晒红烟产地特征中，晒红烟中的亚硝胺 TSNAs 对晒红烟产地的划分起着重要的作用。

2.3　基于近红外光谱的晒红烟小产区之间的相似性分析

根据 2.1 和 2.2 分析结果，六个产地之间均有一定的相似性，现研究六个产地小产区之间的可替代性，共 20 个小产区。小产区之间基于近红外光谱 PPF 投影技术相似度值见表 5。

表 5　20 个小产区之间的近红外光谱相似度值

Table 5　Similarity values of near infrared spectroscopy among 20 different growing areas

	湖南麻阳	湖南湘西	湖南怀化	湖南辰溪	贵州望谟	贵州榕江	贵州镇远	贵州荔波	贵州铜仁	贵州天柱
湖南湘西	0.402									
湖南怀化	0.427	0.543 4								
湖南辰溪	0.395	0.808	0.358							
贵州望谟	0.423	0.027	0.126	−0.001						
贵州榕江	−0.630	−0.550	−1.000	−0.535	−0.354					
贵州镇远	−0.496	−0.602	−0.951	−0.591	−0.120	0.506				
贵州荔波	0.366	−0.081	0.097	−0.153	0.842	−0.750	−0.496			
贵州铜仁	−1.383	−1.494	−1.918	−1.617	−0.505	−0.117	0.345	−0.750		

续表 5

贵州天柱	−0.839	−1.011	−1.240	−1.048	−0.216	−0.136	0.198	−0.589	0.535	
山东沂南	−0.269	−0.821	−0.742	−0.956	0.734	−1.247	−0.772	0.464	−1.477	−0.623
山东蒙阴	−0.123	−0.321	−0.523	−0.292	0.201	0.340	0.720	−0.070	0.523	0.453
山东沂水	−0.036	−0.141	−0.415	−0.086	0.119	0.608	0.809	−0.142	0.273	0.184
四川达州	0.394	−0.339	−0.188	−0.473	0.665	−1.511	−1.130	0.630	−2.520	−1.242
四川德阳	−2.185	−2.507	−2.527	−2.786	−0.740	−2.426	−1.902	−1.216	−2.196	−0.882
四川成都	−2.936	−3.158	−3.297	−3.484	−1.334	−2.773	−2.270	−1.950	−2.469	−0.992
吉林蛟河	−0.477	−0.816	−0.443	−1.085	0.230	−2.085	−1.757	0.261	−2.767	−1.585
吉林延边	−0.828	−1.126	−0.764	−1.429	0.016	−2.374	−2.028	−0.020	−3.021	−1.755
江西石城	−0.302	−0.611	−0.734	−0.627	0.276	−0.170	0.243	−0.024	0.162	0.413
江西抚州	−0.232	−0.719	−0.876	−0.777	0.401	−0.637	−0.120	0.041	−0.871	−0.303

	山东沂南	山东蒙阴	山东沂水	四川达州	四川德阳	四川成都	吉林蛟河	吉林延边	江西石城
山东蒙阴	−0.091								
山东沂水	−0.323	0.707							
四川达州	−0.153	−0.371	−0.421						
四川德阳	−1.355	−1.015	−1.441	−2.771					
四川成都	−2.284	−1.341	−1.795	−3.861	−0.068				
吉林蛟河	−0.474	−0.957	−1.088	−0.638	−1.996	−3.041			
吉林延边	−0.725	−1.17	−1.335	−1.089	−1.916	−2.979	0.592		
江西石城	0.120	0.629	0.326	−0.449	−0.703	−1.122	−0.901	−1.087	
江西抚州	0.008	0.415	0.227	−0.509	−1.981	−2.766	−1.352	−1.675	0.416

由表 5 看出，异产区之间，贵州望谟与山东沂南、四川达州，贵州榕江与山东沂水，贵州镇远与山东蒙阴、山东沂水，贵州荔波与四川达州，山东蒙阴与贵州铜仁、江西石城，相似度值均在 0.5 以上，相似性较高，具有较好的可替代性。

2.4 讨论

近红外光谱采集物质信息较全面，是多种物质的复合信息，但具体特征不明显；化学指标虽能够体现物质的具体特征，但所包含的物质信息却不够全面，解析的是主要化学分，用于表征物质的整体信息时有缺失。结合两者的分析结果，研究不同产地晒红烟的相似性，找到之间的可替代性，既能通过近红外光谱较为全面斩获样品信息，实现产地准确分类，又能通过化学指标分析出影响样品产地划分的具体影响权重，不仅为晒红烟的科学应用提供了方法，此方法同样适用于其他类别的烟草。

3 结 论

应用近红外光谱和化学指标分析晒红烟六个产地分布规律较为一致，表明通过近红外光谱信息或者大量检测数据均可用来判别各产地样品间的相似性；52 个化学指标的方差贡献率分析结果得出，决定晒红烟产地特征的重要指标是亚硝胺，其次是生物碱；应用近红外光谱对小产区样品间进行相似性分析，可得出部分产区样品间具有可替代性。结合主要内在的化学指标工业需求取向，可合理替代不同产地的烟叶原料，降低产品加工质量的波动性，该方法同样适用于其他类别的烟草。

References

[1] DU Wen, YI Jian-hua, TAN Xin-liang, et al(杜 文，易建华，谭新良，等). Acta Tabacaria Sinica(中国烟草学报), 2009, 15(5): 1.

[2] Peerapattana J, Shinzawa H, Otsuka K, et al. J. Near Infrared Spectroscopy, 2013, 21(3): 195.

[3] Liu F, Tang X. Bioengineered, 2015, 6(3): 166.

[4] Graham S F, Haughey S A, Ervin R M, et al. Food Chemistry, 2012, 132(3): 1614.

[5] Xu H, Qi B, Sun T, et al. Journal of Food Engineering, 2012, 109(1): 142.

[6] XIE Juan, LUO Jian-qun, YAO He-ming, et al(谢 娟，罗建群，姚鹤鸣，等). Tobacco Science & Technology(烟草科技), 2008, (7): 42.

[7] Xu F, Yu J, Tesso T, et al. Applied Energy, 2013, 104: 801.

[8] Haughey S A, Graham S F, Cancouët E, et al. Food Chemistry, 2013, 136(3): 1557.

[9] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al(严衍禄，赵龙莲，韩东海，等). Basis and Application of Near-Infrared Spectroscopy(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京：中国轻工业出版社), 2005.

[10] Jolliffe I T. Principal Component Analysis, Springer Series in Statistics, 2nd ed. Springer, NY, USA, 2002.

[11] WEN Ya-dong, WANG Yi, WANG Neng-ru, et al(温亚东，王 毅，王能如，等). Acta Tabacaria Sinica(中国烟草学报), 2009, 15(5): 6.

Analyzing the Similarity of Different Growing Areas in Red Sun-Cured Tobacco by Using Near Infrared Spectroscopy or Chemical Data

MA Li[1], LI Xue-ying[2], DU Guo-rong[1], DING Rui[3], WANG Yun-bai[3], MA Yan-jun[1*], LI Jun-hui[2*]

1. Beijing Working Station, Technical Center of Shanghai Tobacco Group Corp. Ltd., Beijing 101121, China
2. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China
3. Quality and Safety Research Center, Institute of Tobacco Research of Chinese Academy of Agricultural Sciences, Qingdao 266101, China

Abstract Near infrared spectroscopy is a comprehensive information collection material, composite information of various substances, but the specific feature is not obvious; the chemical index can reflect the specific characteristics of the material, but the material information is not comprehensive enough. In this paper, total of 115 red sun-cured tobacco samples in 2012 and 2013 were tested, which were from Guizhou, Hunan, Jilin, Jiangxi, Shandong, Sichuan. After near infrared spectroscopy was processed by the first derivative and SG smooth and 26 items of chemical data and 26 items of calculated data (for example, ratio of sugar to nicotine, etc.) were normalized, the PPF (Projection of Basing on Principal Component and Fisher Criterion) projection method was used to analyze the similarity and substitution of growing areas in red sun-cured tobacco. Combined with the analysis of near infrared spectroscopy and chemical indicators, the results showed that the similarity analysis result by using near infrared spectroscopy is in basic agreement with the result by using chemical data, which means two methods can both analyze the similarity of growing areas. Nitrosamines would determine the growing areas as an important indicator by analyzing the variance contribution rate of 52 test data. The similarity of small areas was analyzed by used near infrared spectroscopy and the result was that the part of different small areas can be replaced. Application of near infrared spectroscopy technology can quickly and accurately analyze the similarity and substitution of different growing areas. Combining chemical data can be used to analyze its intrinsic characteristics, which can play an important role in the scientific exploitation of tobacco raw materials.

Keywords Red sun-cured tobacco; Near infrared spectroscopy; Chemical data; Similarity

(Received Oct. 26, 2016; accepted May 10, 2017)

* Corresponding authors