

基于近红外光谱技术的枸杞产地溯源研究

汤丽华, 刘敦华*

(宁夏大学农学院, 宁夏 银川 750021)

摘要: 采用近红外光谱技术对宁夏、甘肃、青海、内蒙、河北的 8 个不同产地 40 种枸杞样品进行扫描, 在主成分分析基础上利用简易分类法(simple modeling of class analogy, SIMCA)模式识别原理分别建立模型。结果表明: 在 950~1650nm 全光谱波长范围内, 光谱经一阶导数(5 点平滑)和矢量归一化(standard normal variate, SNV)预处理后, 8 个产地模型的主成分数均为 3 时, 采用 SIMCA 模式识别法可以建立较为稳健的枸杞产地溯源模型; 在 $\alpha=5\%$ 的显著水平条件下检验模型的可靠性, 8 个产地校正集模型的识别率除青海为 80% 外, 其他产地均为 100%, 拒绝率分别为 100%、100%、97%、100%、91%、94%、97%、100%, 其验证集模型的识别率均为 100%, 拒绝率分别为 100%、100%、100%、100%、75%、88%、100%、100%。表明该方法在枸杞产地识别中具有可行性。

关键词: 近红外光谱; 鉴别; 枸杞; 产地溯源

Tracing the Geographic Origin of Chinese Wolfberry by Near Infrared Spectroscopy

TANG Li-hua, LIU Dun-hua*

(College of Agriculture, Ningxia University, Yinchuan 750021, China)

Abstract: Forty Chinese wolfberry samples from 8 different regions of China were scanned with a near infrared (NIR) spectrophotometer. The original spectra were preprocessed by the first derivative (5 point smoothing) and standard normal variate (SNV) and were subjected to principle component analysis over the wavelength range of 950 nm to 1650 nm, in which the number of principal components was selected as 3. A robust model indicating each geographic origin of Chinese wolfberry was achieved using the SIMCA (Simple Modelling of Class Analogy) pattern recognition method and its reliability at the 5% significance level was validated. The results revealed that the model indicating samples from Qinghai exhibited a recognition rate of 80% in the calibration set, the recognition rates for other 7 cultivation regions were all 100%, and the rejection rates were 100%, 100%, 97%, 100%, 91%, 94%, 97% and 100%, respectively. All the recognition rates of the 8 cultivation regions were 100% in the validation set, and the rejection rates were 100%, 100%, 100%, 100%, 75%, 88%, 100% and 100%, respectively. These results support the feasibility of applying NIR to identify the geographic origin of Chinese berry.

Key words: near infrared spectroscopy; identification; Chinese wolfberry; origin traceability

中图分类号: TS255.2

文献标识码: A

文章编号: 1002-6630(2011)22-0175-04

枸杞(*Lycium barbarum* L.)为茄科(Solanaceae)枸杞属(*Lycium*)多分枝灌木, 其果、皮、叶均可入药^[1]。枸杞具有较高的药用价值, 同时作为一种优良水土保持灌木, 广泛种植于我国川北、华北北部、内蒙古、陕西、甘肃、宁夏、青海和新疆等地^[2]。由于宁夏枸杞个大、形美、活性成分含量高、药用价值大等特点受到广大消费者的青睐。但是, 随着近年来市场上频繁

出现不同产地混杂的现象, 宁夏枸杞的品质难以保证, 而传统的枸杞鉴别方法费时、误差大。因此亟需建立一种快速、简单、准确的枸杞品质鉴定方法。

近红外光谱(near infrared spectroscopy, NIR)分析技术以其无污染、非破坏性、分析速度快等特点, 已广泛应用于食品、药品等定性、定量分析领域^[3]。目前, 近红外技术已成功应用于羊肉产地溯源^[4]、连翘产

收稿日期: 2011-06-28

基金项目: “十一五”国家科技支撑计划项目(2009BAI72B04)

作者简介: 汤丽华(1986—), 女, 硕士研究生, 研究方向为食品质量与安全。E-mail: tanglihua532@163.com

* 通信作者: 刘敦华(1964—), 男, 教授, 博士, 研究方向为食品质量与安全。E-mail: dunhualiu@126.com

地识别^[5]、中药材定性定量分析^[6]、茶叶定性和定量分析^[7]等领域,但对于枸杞产地溯源仅查到周群等^[8]采用傅里叶变换红外(fourier transform infrared spectroscopy, FTIR)光谱法并结合简易分类法(simple modeling of class analogy, SMICA)聚类分析法对不同产地的枸杞进行了聚类分析,效果较好。因此,本研究以8个不同产地的枸杞样品为研究对象,建立枸杞近红外光谱数据库和枸杞产地溯源模型,为枸杞产地鉴别提供依据。

1 材料与方法

1.1 材料与试剂

干果枸杞分别取自于宁夏回族自治区中宁县、南梁农场、惠农县、同心县;内蒙古自治区精河县;甘肃省靖远县、青海省、河北省巨鹿县,样品均为2009年6~9月份夏季枸杞果实。

1.2 仪器与设备

DA7200 近红外成分分析仪为瑞典波通公司产品。该仪器的波长范围:950~1650nm;扫描次数:100次/s;扫描间隔:2.0nm;噪音水平: < 20 μ A;利用软件Simplicity进行光谱数据采集。该仪器对测试环境的温度要求较高,一般室温保持在25℃,样品光谱扫描时须保持温度的稳定,同时开机需预热30min,以保证样品测定的稳定性。

1.3 方法

1.3.1 光谱采集

干果枸杞样品装入75mm样品盘中,保持样品表面平整。在全波长950~1650nm范围内进行波长连续扫描,采集样品的吸收光谱,每种样品重复扫描2次,每次扫描在电脑中编号。

1.3.2 样品分析

1.3.2.1 主成分分析法(principal component analysis, PCA)

主成分分析是一种数据降维方法,它将原变量进行转换,使少数几个新变量成为原变量的线性组合,这些新变量互相正交,尽可能多地表征原变量的数据结构特征,从而消除各变量间的相关性^[9]。主成分得分值(score)是经主成分分析降维得到的,代表了原始光谱矩阵的主要信息,反映了样本总体特征分布^[10]。

1.3.2.2 简易分类法(simple modeling of class analogy, SIMCA)

SIMCA 判别分析方法是建立在PCA基础上的一种具有监督模式识别的方法,它对训练集中每一类样本的近红外光谱数据矩阵分别进行了主成分分析,建立每一

类样本的主成分分析模型,在此基础上将未知样品与各类样本数学模型进行拟合,确定其归属^[11]。

2 结果与分析

2.1 光谱预处理

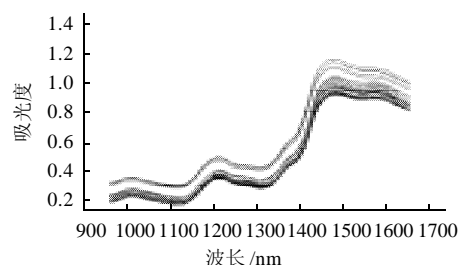


图1 8个产地枸杞的近红外光谱图

Fig.1 Original NIR spectra of Chinese wolfberry from eight regions

图1采集到的原始光谱信息量较大,光谱具有一定的代表性,达到了近红外光谱预测的要求。由于实验中枸杞样本的均匀度和粒径大小不能保证完全一致,样本的密实度也影响了光在样品中的传播,这些都对光的漫反射有一定影响^[12]。同时枸杞样品成分含量不同以及噪声等因素影响光谱存在差异,光谱之间的强度变化较明显^[13]。由于光谱信息重叠情况严重,各样品的特征信息很难从峰位、峰强等方面直观鉴别^[6]。因此,必须经过数学方法处理提取特征信息后,才能利用近红外分析技术对枸杞样品进行定性分析。在建立枸杞产地判别模型之前,为消除基线漂移、高频随机噪声等因素对样品近红外光谱的影响,提高模型的预测准确性和稳定性,需要对光谱进行预处理以获得有用信息^[14]。采用平滑处理方法在950~1650nm全光谱波长范围内,比较5、7点和9点平滑对光谱预处理的影响。同时,在平滑处理的基础上对原始光谱分别进行了一阶导数、二阶导数、多元散射校正(multiplicative scatter correction, MSC)、矢量归一化(standard normal variate, SNV)等4种预处理方法。通过对比发现5点平滑效果优于其他两种,一阶导数结合SNV的方法效果优于MSC或二阶导数预处理方法。因此,本研究最终采用一阶导数(5点平滑)+SNV的预处理方法建立模型。

2.2 波长范围筛选

波长选择一方面可以简化模型,更主要的是由于不相关或非线性变量的剔除,可以得到预测能力强、稳健性好的校正模型^[15]。图2表明,全光谱信息特征更加突出。而8个产地的枸杞光谱均存在较大差异,部分波段仅反应了不同产地枸杞光谱的部分信息,造成识别率有所下降。因此,本研究选用一阶导数(5点平滑)+

SNV 的预处理后, 950~1650nm 全波段波长范围内建立模型, 以增加模型的稳健性。

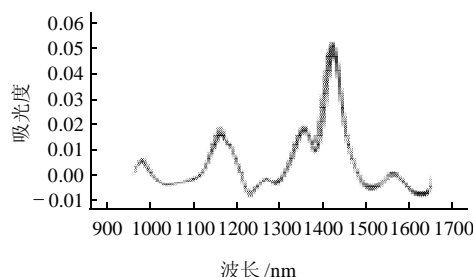
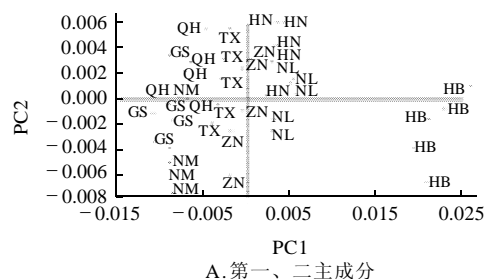


图2 预处理后的枸杞样品近红外光谱图

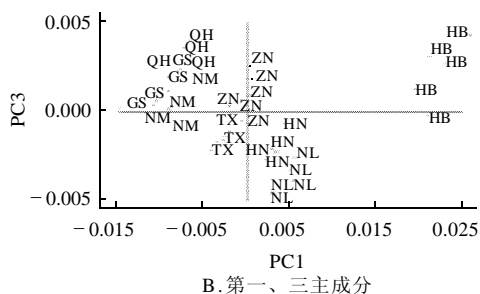
Fig.2 Preprocessed NIR spectra of Chinese wolfberry

2.3 模型的建立

对原始光谱进行一阶导数(5点平滑)+SNV 预处理后, 在全光谱波长范围内, 对8个产地枸杞样品近红外光谱进行主成分分析, 主成分数选取3。图3反映的是整个样品集中的样品点在该二维平面上的投影, 整个样品集光谱矩阵的第一主成分、第二主成分、第三主成分的方差贡献率分别为84%、10%、5%, 累计方差贡献率为99%, 表达原信息99%的变异, 因此样本在二维平面和三维立体的投影分布能够充分反映样品在多维空间的分布特征。



A. 第一、二主成分



B. 第一、三主成分

HB. 河北; NL. 南梁; HN. 惠农; ZN. 中宁; TX. 同心; QH. 青海; GS. 甘肃; NM. 内蒙。下同。

图3 8个产地枸杞样品的主成分得分图

Fig.3 Score plots of PC2 versus PC1 and PC3 versus PC1 from PCA of NIR spectra of Chinese wolfberry from eight regions

从图3可以看出, 8个产地的样品基本可以区分开, 而河北与其他7个产地样本明显地分开, 其他7个产地的样品分布较集中, 且样品有部分重叠, 不能较好地反映其他7个产地间样本的分开情况, 仅用主成分分析不能完全将产地识别。因此在主成分分析的基础上, 进行SIMCA分析。基于交互验证法, 用方差贡献率增加的显著性(F 检验), 确定中宁、南梁、同心、惠农、甘肃、青海、内蒙、河北8类模型的最佳主成分数分别为3。用SIMCA法建立的模型可以将8个产地的枸杞的基本分开来, 其中中宁、南梁、同心、惠农、甘肃、青海与内蒙的相距较近, 它们与河北的距离相对较远。由表1可以看出, 8个产地的枸杞活性成分含量有差异, 枸杞内部品质不同, 造成各产地在空间上的距离不同。产生的原因: 首先, 中宁、南梁、同心、惠农、甘肃、青海、内蒙均地处西北地区, 河北地处华北地区, 由于气候、地理位置、种植方法、土壤等外部条件的影响, 造成各地区枸杞的内部品质受不同程度的影响, 在近红外光谱上的吸收峰就存在一定的差异, 差异越大, 它们在空间的距离就越大; 其次, 采摘后的处理方式、枸杞保存方式等不同也会影响枸杞的内部品质, 造成各产地近红外光谱上的吸收峰有差异, 空间距离不同。由此可推出, 中宁、南梁、同心、惠农的枸杞内部品质较相近, 其次与甘肃、青海、内蒙有差异, 与河北枸杞内部品质相差较大。

表1 8个产地枸杞的活性成分含量

Table 1 Active ingredient contents of Chinese wolfberry from eight regions

产地	总糖 含量/%	多糖 含量/%	黄酮含量/ (mg/g)	VC含量/ (mg/100g)	甜菜碱 含量/%
中宁	48.04	6.19	0.74	13.87	0.46
南梁	47.51	3.11	0.80	25.12	0.57
惠农	49.64	5.27	0.89	20.20	0.43
同心	52.37	4.32	0.90	15.35	0.44
甘肃	51.95	5.78	0.66	27.44	0.31
青海	50.39	3.32	0.70	27.79	0.38
内蒙	49.27	3.77	0.76	30.93	0.59
河北	37.65	4.06	1.13	16.81	0.62

2.4 模型的校正和检验

在PCA的基础上, 用8个产地样品中的32个组成校正集, 建立每类样品的SIMCA模型, 剩下的8个样品作为验证集, 验证模型的预测能力和准确性, 验证结果如表2所示。在 $\alpha=5\%$ 的显著水平条件下检验模型的可靠性, 在对校正集样品进行预测时, 8类模型对本类样品的识别率中中宁、南梁、同心、惠农、甘肃、内蒙、河北模型的识别率均为100%, 而青海模型的识别率为80%; 在拒绝非本类样品时, 中宁、南

梁、同心、河北模型的拒绝率均为 100%，惠农、甘肃、青海、内蒙模型的拒绝率分别为 97%、91%、94%、97%。在对验证集样品进行预测时，8 个模型对样品的识别率均为 100%，中宁、南梁、同心、惠农、内蒙、河北模型的拒绝率均为 100%，甘肃、青海模型的拒绝率为 75%、88%。由于宁夏、甘肃、青海、内蒙等地样本距离较近，枸杞生长环境差异小，投影点之间有相互交叉的区域，产生识别错误的结果。说明可以用该 SIMCA 模型鉴别不同产地来源的枸杞，所建立的产地识别模型基本能够达到正确识别 8 个不同产地枸杞样品的目的。

表 2 SIMCA 产地模型校正与预测效果

Table 2 Results of calibration and validation for models indicating different cultivation regions

产地	主成分数	校正集		验证集	
		识别率/%	拒绝率/%	识别率/%	拒绝率/%
中宁	3	100	100	100	100
南梁	3	100	100	100	100
惠农	3	100	97	100	100
同心	3	100	100	100	100
甘肃	3	100	91	100	75
青海	3	80	94	100	88
内蒙	3	100	97	100	100
河北	3	100	100	100	100

3 结 论

本研究采用近红外光谱方法判别枸杞产地，在主成分分析的基础上利用 SIMCA 模式识别原理建立模型，模型基本能够正确识别 8 个产地的枸杞，其他产地识别

时拒绝率达 91% 以上。综上所述，近红外光谱技术鉴别枸杞产地具有可行性且效果较好，为枸杞产地识别提供了技术依据，具有广阔的应用前景。

参考文献：

- [1] 杨薇, 谭自力. 枸杞的价值及栽培技术[J]. 河北农业科技, 2008(6): 35-36.
- [2] 郑国琦, 胡正海. 宁夏枸杞的生物学和化学成分的研究进展[J]. 中草药, 2008(5): 796-780.
- [3] 张晓瑜, 王庭欣, 谢飞. 近红外光谱技术鉴别地理标志产品黄骅冬枣[J]. 食品工业科技, 2010(11): 111-113.
- [4] 张宁, 张德全, 李淑荣, 等. 近红外光谱结合 SIMCA 法溯源羊肉产地的初步研究[J]. 农业工程学报, 2008(12): 309-312.
- [5] 张晓慧. 基于近红外光谱的连翘有效成分分析与产地鉴别技术研究[D]. 洛阳: 河南科技大学, 2008.
- [6] 孙丽英. 用于中药材定性定量分析的近红外指纹图谱研究[D]. 武汉: 中南民族大学, 2008.
- [7] 牛智有, 林新. 茶叶定性和定量近红外光谱分析方法研究[J]. 光谱学与光谱分析, 2009, 29(5): 2417-2420.
- [8] 周群, 孙素琴, 梁曦云. 枸杞产地的红外指纹图谱与聚类分析法研究[J]. 光谱学与光谱分析, 2003, 23(3): 509-511.
- [9] 占茉莉. 茶叶产地及铅污染源溯源技术研究[D]. 无锡: 江南大学, 2008.
- [10] 李彦周, 闵顺耕, 刘霞. 主成分分析在近红外定量分析校正集样本优选中的应用[J]. 分析化学研究简报, 2007, 35(9): 1331-1334.
- [11] 高文骥, 袁洪福, 邱藤, 等. 原位实时近红外光谱研究核壳乳液聚合过程[J]. 高等学校化学学报, 2009, 30(7): 1293-1299.
- [12] 陈全胜, 赵杰文, 张海东, 等. SIMCA 模式识别方法在近红外光谱识别茶叶中的应用[J]. 食品科学, 2006, 27(4): 186-189.
- [13] 姜伟. 基于近红外技术的苎麻化学成分快速定量分析研究[D]. 青岛: 青岛大学, 2009.
- [14] 张晓炎. 基于近红外光谱的柑桔产地溯源及橙汁掺假识别研究[D]. 重庆: 西南大学, 2010.
- [15] 褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. 化学进展, 2004, 16(4): 529-531.