

近红外光谱法无损识别林下山参及其生长年限

卜海博¹, 聂黎行², 王丹¹, 杜红¹, 王钢力^{2*}, 李向日^{1*}

1. 北京中医药大学中药学院, 北京 100102
2. 中国食品药品检定研究院, 北京 100050

摘要 对不同生长年限的林下山参 96 支(其中,十五年生 24 支,十二年生 72 支)和园参 177 支。采集近红外光谱后,应用主成分分析-马氏距离法进行判别分析。采用原始光谱,经过预处理后,在全光谱范围内分别选择合适的主成分数,对林下山参和园参以及不同生长年限的林下山参分别建立了判别分析模型。所建立的两组模型对验证集的正确判别率均为 100%。表明该方法准确可靠、快速无损,可实际用于林下山参的质量控制。

关键词 近红外(NIR)光谱;林下山参(MCG);园参(GCG);不同生长年限;判别分析;无损识别

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2012)07-1801-05

引言

人参为我国传统名贵中药,按生长方式不同可分为“园参”、“林下山参”和“野生人参”。栽培的俗称“园参”;播种在山林野生状态下自然生长的称“林下山参”^[1]。因野生人参濒临灭绝,2008 年颁布的国家标准对野山参的定义为“自然生长于深山密林的人参(不包括野生人参)”^[2],即是药典中的林下山参,并规定野生人参禁止采挖。目前市场上野生人参罕见,林下山参已成为高档人参的主流品种,其价格为园参的几十倍^[3],不同生长年限的林下山参价格也相差悬殊。悬殊的市场价格,导致不法商人用园参冒充林下山参出售,或采取拼接等方法在生长年限上作假,使林下山参市场鱼龙混杂。

目前林下山参的鉴别方法仍依赖传统的性状和显微鉴别,不仅要求有丰富的经验,且易受主观因素影响。现代分析方法如 HPLC^[4]、LC-MS^[5]等也尝试应用,但需对样品进行复杂的前处理,成本高、费时费力,且林下山参与园参,以及不同生长时间的林下山参具有极其相似的化学成分,难以区分。林下山参作为贵重药材,送检方要求无损或微损检验,药检部门不可能对检品取样做组织、粉末及理化鉴别。因此急需建立一种快速、无损的分析方法用于林下山参的质量控制。

近红外光谱法(NIRS)是一种可用于有机物定性定量分

析的光谱分析技术,具有样品处理简单、快速、无污染、在线分析等优点,广泛用于医药、食品、石油等领域^[6]。Lee 等^[7]应用近红外光谱法成功识别了玄参产地。王钢力等应用近红外光谱法对高丽参和红参成功进行了鉴别^[8,9]。瞿海斌等采用近红外光谱漫反射光谱法和模式识别技术快速鉴别阿胶真伪。赵杰文等^[10]采用近红外光谱法成功鉴别了雪莲产地。但近红外光谱技术尚未在中药材生长年限判断方面进行应用。目前尚无应用 NIRS 识别林下山参的研究报道。

本文收集了不同产地、不同生长年限的林下山参和园参样品,采集原始光谱,结合判别分析,建立了林下山参与园参以及不同生长年限林下山参的无损鉴别模型。

1 实验部分

1.1 仪器与材料

1.1.1 仪器

Antaris 傅里叶变换近红外光谱仪(美国 Thermo Nicolet 公司,光源:卤钨灯,软件 Result 3.0 用于光谱采集,软件 TQ Analyst 7.2 用于光谱分析),配备 InGaAs 检测器,光纤附件。

1.1.2 样品

收集不同产地、不同生长年限的林下山参和园参,样品来源见表 1。所有样品经北京中医药大学李向日教授鉴定,均为正品。

收稿日期:2011-08-16, 修订日期:2011-11-15

基金项目:国家自然科学基金面上项目(81073041),高等学校博士学科点专项科研项目(20100013120010)资助

作者简介:卜海博,1985 年生,北京中医药大学中药学院硕士研究生

* 通讯联系人 e-mail: duneer@sohu.com; lixiangri@sina.com

Table 1 Details of the sample sources

样品名	产地	生长年限	编号
林下山参	辽宁宽甸	十五年	1~24
	吉林集安	十二年	25~96
园参	吉林抚松	四年	97~127
	吉林抚松	五年	128~149
	吉林集安	四年	150~204
	吉林集安	五年	205~259
	吉林通化	五年	260~273

1.2 NIR 光谱采集

光谱采集范围为 $10\ 000\sim 4\ 000\ \text{cm}^{-1}$, 分辨率 $8\ \text{cm}^{-1}$, 扫描次数 32 次。以仪器内置背景为参比, 采用光纤附件, 分别对样品的 3 个不同部位(芦头、主根上部和主根下部)进行光谱采集, 原始光谱图见图 1。

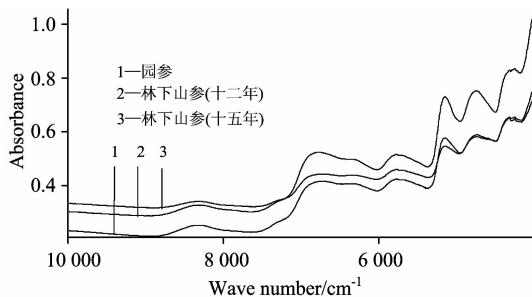


Fig. 1 The original NIR spectra of samples

2 结果分析

2.1 谱图分析

由图 1 可知, 样品基线均有一定偏移, 且所有样品的谱图极为相似, 无法直接找出特定的吸收峰加以区分。必须运用化学计量学方法, 采用软件对原始光谱进行预处理和特征信息提取后才能作出鉴别。

2.2 光谱预处理

在近红外光谱中, 谱峰重叠严重, 除样品信息外, 还含有无用信息、噪声和基线漂移。采用合适方法进行预处理, 可有效提取特征信息, 尽量消除噪声及基线漂移的影响, 减少样品表面不均匀和色差等影响, 提高模型预测的准确性。求导、平滑和多元散射校正(MSC)是常用的预处理方法。

Table 2 Discrimination results of calibration sets by different pretreatments

预处理方法	不同生长年限 林下山参		林下山参 与园参	
	校正集 数目	误判数	校正集 数目	误判数
MSC+原始光谱+SG 平滑	83	0	213	0
MSC+一阶导数+SG 平滑	83	15	213	12
MSC+二阶导数+SG 平滑	83	52	213	18
MSC+原始光谱+无平滑	83	0	213	6
MSC+一阶导数+无平滑	83	26	213	22
MSC+二阶导数+无平滑	83	54	213	42

由表 2 可知: 求导后, 误判数明显增加, 表明求导不适宜林下山参和园参样品的识别; 平滑对不同生长年限林下山参的识别没有影响, 而对林下山参和园参的识别有一定影响。确定的预处理方法为: 选择 MSC+原始光谱+无平滑识别不同生长年限的林下山参; 选择 MSC+原始光谱+SG 平滑识别林下山参与园参。

2.3 判别分析

2.3.1 相关原理

判别分析首先是收集已知类别且涵盖范围较广的样品扫描近红外图谱建立校正模型, 其次将一个未知样品在特定区域的光谱信息应用于校正模型, 来判别未知样品与标准中的哪一类最相似。

判别分析中, 主成分分析-马氏距离法^[11]是一种较为实用的方法。主成分分析的目的是将数据降维, 转换原变量, 使少数几个新变量是原变量的线性组合, 并尽可能多地表征原变量的数据特征而不丢失信息^[12]。用本方法分析样品时, 软件对标准光谱进行主成分分析, 用其结果来确定未知样品的得分值, 得分图用来计算样品到每个类别的马氏距离, 距离的值越接近于零, 该样品与标准光谱越相似。实际分析中, 当标准光谱不止一类时, 未知样品被判为马氏距离最短的那一类。

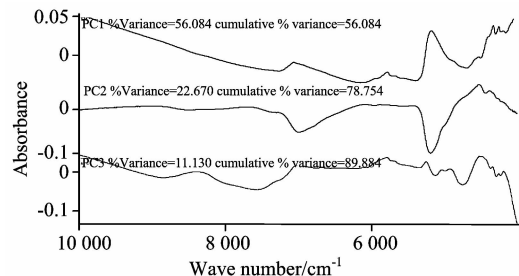


Fig. 2 Cumulative variance of the first 3 principal components of identification about MCG and GCG

Table 3 Optimization of principal components in the discriminant models of MCG and GCG

主成分数	累计贡献率	误判数
28	99.963 7	1
29	99.966 0	1
30	99.967 7	0
31	99.969 5	0

2.3.2 主成分数的确定

NIR 光谱是一个高维数组, 包括许多噪声, 为充分利用光谱信息建立模型, 必须用主成分分析对光谱进行降维处理^[13]。选取主成分的个数取决于主成分的累计方差贡献率^[12], 它标志着前几个主成分概括信息之多寡。以累计贡献率和误判数筛选最佳主成分数。

以园参与林下山参(芦头部位)识别为例, 前 3 个主成分累计贡献率如图 2 所示。由表 3 可知, 选择主成分数为 30 建立模型时, 累计贡献率较大且误判数最少, 建立的林下山参和园参(芦头部位)识别模型效果最佳。

2.3.3 模型建立及验证

2.3.3.1 林下山参与园参

随机选取林下山参(L)83 个, 园参(Y)151 个作为校正集, 其余作为验证集。校正集用于建立判别模型, 验证集用于预测模型的性能。按照 2.3.2 的方法确定识别林下山参与园参的三个部位主成分数均为 30。结合 2.2 预处理方法计算, 样品明显聚为两类, 分别用验证集样品进行验证后, 均获得正确分类。三个部位判别结果依次见图 3。主根上部的验证集判定结果见表 4。判定类别与实际一致, 全部样品判断正确。其他两个部位验证后也全部判断正确。为尽量减小因样品扫描部位不同导致的误差, 使模型更加稳健, 根据性能指数确定最佳扫描部位。性能指数是用验证集验证后得到的评价模型优劣的指标。在判别分析中使用 average distance ratio 算法计算, 单位为%, 最大值 100, 值越大, 表示性能越好。依据性能指数比较, 三个部位性能指数依次为 83.8%, 85.9%, 84.1%, 确定林下山参与园参扫描的最佳部位为主根上部。

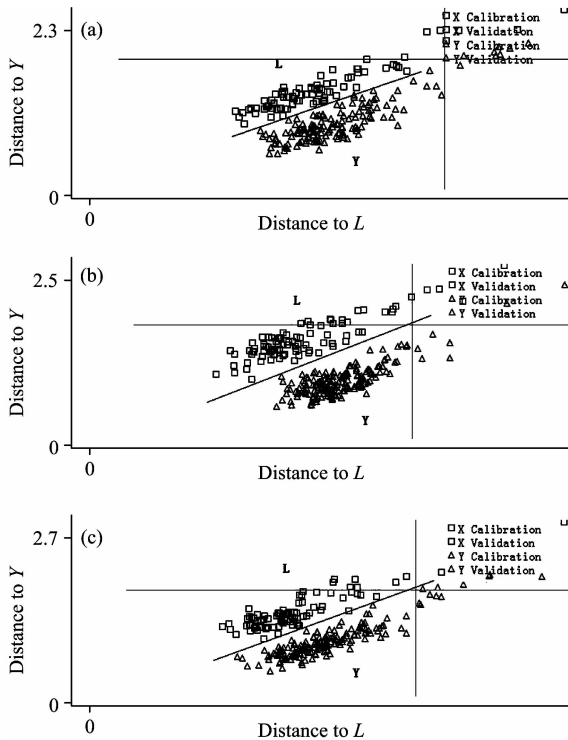


Fig. 3 Discriminant analysis results of MCG and GCG

(a): rhizome; (b): upper part of taproot;
(c): lower part of taproot

2.3.3.2 不同生长年限的林下山参

随机选取林下山参十五年生 (DL) 21 个、十二年生 (XL) 62 个作为校正集, 其余作为验证集。按照 2.3.2 的方法确定鉴别 DL 与 XL 三个部位的主成分数均为 6, 结合 2.2 预处理方法计算, 样品明显聚为两类。分别用验证集样品进行验证后, 均获得正确分类。三个部位判别结果依次见图 4。主根上部的验证集判定见表 5。判定类别与实际一致, 全部样品判断正确。其他两个部位验证后也全部判断正确。

依据性能指数比较, 三个部位的性能指数依次为 92.3%, 94.8%, 93.8%, 确定不同生长年限林下山参的最佳扫描部位为主根上部。

Table 4 The validation for the identification model of MCG and GCG (upper part of taproot)

样品编号	实际类别	判定类别
5, 13, 19, 31, 39, 46, 53	L	L
61, 73, 77, 84, 87, 93	L	L
98, 103, 109, 119, 129, 135	Y	Y
144, 153, 158, 166, 173	Y	Y
177, 182, 187, 192, 198	Y	Y
207, 211, 217, 222, 229	Y	Y
233, 239, 243, 252, 269	Y	Y

Table 5 The validation result for the identification of MCG of different years (upper part of taproot)

样品编号	实际类别	判定类别
5, 13, 19	DL	DL
31, 39, 46, 53, 61	XL	XL
73, 77, 84, 87, 93	XL	XL

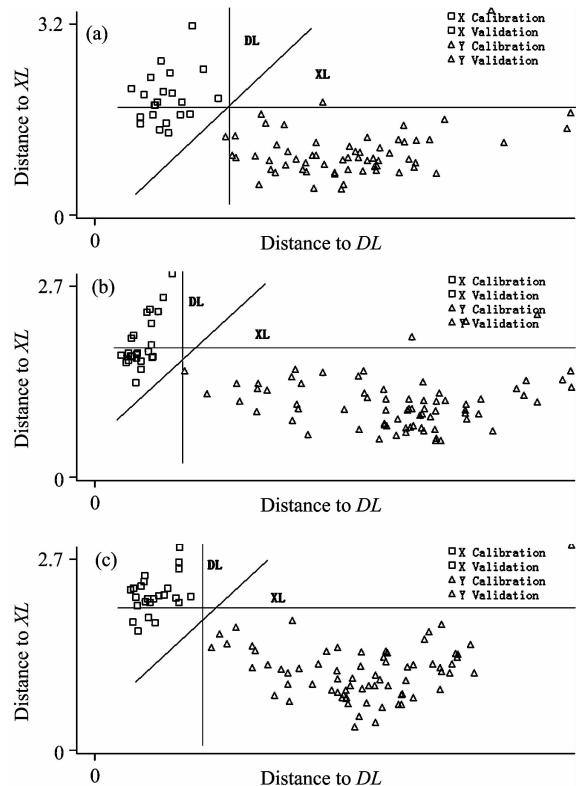


Fig. 4 The discriminant analysis results of MCG of different growth years

(a): rhizome; (b): upper part of taproot;
(c): lower part of taproot

3 结 论

采用近红外光谱漫反射分析技术, 结合主成分分析-马

氏距离法进行判别分析,建立了林下山参与园参与不同生长年限林下山参的无损鉴别模型。结果表明,运用近红外光谱法对林下山参与园参、不同生长年限的林下山参均能正确分类,并实现零误判。

以累计贡献率和误判数优选了主成分数。预处理方法为:采用原始光谱、无平滑识别不同生长年限效果最佳;而采用原始光谱、平滑建立林下山参和园参效果最佳。考虑到

10 000~4 000 cm^{-1} 范围信息量更全,有利于定性分析,因此选择全波谱范围建立模型。依据性能指数优选得到的最佳扫描部位均为主根上部。

实验收集的样品多,来源广泛,具有代表性,无需复杂前处理。该方法具有准确、快速、无损检测、成本低、无污染等优点,可实际用于林下山参的质量控制,并提高我国林下山参的市场形象,促进人参市场健康发展。

References

- [1] Chinese Pharmacopoeia Commission(国家药典委员会). Pharmacopoeia of the People's Republic of China, Volume I (中华人民共和国药典,一部). Beijing: China Medical Science and Technology Press(北京:中国医药科技出版社), 2010.
- [2] National Standard of the People's Republic of China(中华人民共和国国家标准). GB/T 18765—2008. Identification and Grade Quality Standards of Wild Ginseng(野山参鉴定及分等质量).
- [3] WANG Yong-ming(王永明). Research and Information of Traditional Chinese Medicine(中药研究与信息), 2003, 5(10): 25.
- [4] Shi S, Wang Y T, Li J, et al. Food Chem., 2007, 102: 664.
- [5] Nicola F. J. Chromatogr. B, 2004, 812: 119.
- [6] WEI Li-ping, WU Chun-min(魏丽萍, 吴春敏). Strait Pharmaceutical Journal(海峡药学), 2010, 41(8): 111.
- [7] Lee Dong Young, Kim Seung Hyun, Kim Young Choong, et al. Microchem. J, 2011, 5: 1.
- [8] WANG Gang-li, TIAN Jin-gai, NIE Li-xing, et al(王钢力, 田金改, 聂黎行, 等). Chinese Traditional and Herbal Drugs(中草药), 2008, 39(2): 277.
- [9] WANG Gang-li, NIE Li-xing, ZHANG Ji, et al(王钢力, 聂黎行, 张 继, 等). Chinese Traditional and Herbal Drugs(中草药), 2008, 39(3): 438.
- [10] ZHAO Jie-wen, JIANG Pei, CHEN Quan-sheng(赵杰文, 蒋 培, 陈全胜). Transactions of the Chinese Society of Agricultural Machinery(农业机械学报), 2010, 41(8): 111.
- [11] Whitfield R G, Gerger M E, Sharp R L. App. Spectrosc., 1987, 41(7): 1204.
- [12] LU Wan-zhen(陆婉珍). Modern Near Infrared Spectroscopy Analytical Technology(Second Edition)(现代近红外光谱技术, 第 2 版). Beijing: China Petrochemical Press(北京:中国石化出版社), 2007. 1.
- [13] PAN Ying, SHEN Yi, XIANG Jing-zuo, et al(潘 颖, 沈 漪, 项竞佐, 等). Chinese Journal of Antibiotics(中国抗生素杂志), 2005, 30(2): 77.

Nondestructive Identification of the Root of Mountain Cultivation Ginseng and Growth Years by Near Infrared Spectroscopy

BU Hai-bo¹, NIE Li-xing², WANG Dan¹, DU Hong¹, WANG Gang-li^{2*}, LI Xiang-ri^{1*}

1. School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 100102, China

2. National Institute for Food and Drug Control, Beijing 100050, China

Abstract Mountain cultivation ginseng (MCG) and garden cultivation ginseng (GCG) were identified by near infrared spectroscopy, so were MCG of different growth years. 96 MCG samples of different growth years, including 24 of fifteen years and 72 of ten years, and 177 GCG samples were collected. After the near infrared spectra of these samples were collected, discriminant analysis was used to distinguish MCG and GCG, so was MCG of different years. After the original spectra were pretreated, discriminant analysis models of MCG and GCG, MCG of different growth years were developed respectively with selected principal component numbers in full spectra region. The correct discrimination rate of two groups of model was both 100%. The proposed methods are accurate, fast and nondestructive, and can be applied to the quality control of MCG. It has an important significance for building market image of MCG.

Keywords Near infrared (NIR) spectroscopy; Mountain cultivation ginseng (MCG); Garden cultivation ginseng (GCG); Dif-

ferent growth years; Discriminant analysis; Non-destructive identification

* Corresponding author

(Received Aug. 16, 2011; accepted Nov. 15, 2011)

《光谱学与光谱分析》投稿简则

《光谱学与光谱分析》是由中国科协主管,中国光学学会主办,钢铁研究总院、中国科学院物理研究所、北京大学、清华大学共同承办的专业学术期刊。国内外公开发行人,从2004年起为月刊,大16开本,2012年仍为月刊,每期288页。《光谱学与光谱分析》主要报道我国光谱学与光谱分析领域内具有创新性科研成果,及时反映国内外光谱学与光谱分析的进展和动态;发现并培育人才;推动和促进光谱学与光谱分析的发展。为科教兴国服务。读者对象为从事光谱学与光谱分析的科研人员、教学人员、分析测试人员和科研管理干部。

栏目设置和要求

1. 研究报告 要求具有创新性的研究成果,一般文章以8000字(包括图表、参考文献、作者姓名、单位和中文、英文摘要,下同)为宜。
2. 研究简报 要求在前人研究的基础上有重大改进或阶段性研究成果,一般不超过5000字。
3. 评述与进展 要求评述国内外本专业的发展前沿和进展动态,一般不超过10000字。
4. 新仪器装置 要求介绍新型光谱仪器的研制、开发、使用性能和应用,一般不超过5000字。
5. 来稿摘登 要求测试手段及方法有改进并有应用交流价值,一般以3000~4000字为宜。

稿件要求

1. 投稿者请经本刊编委(或历届编委)一人或本专业知名专家推荐,并附单位保密审查意见及作者署名顺序,主要作者介绍。文章有重大经济效益或有创新者,请说明,同时注明受国家级基金或国家自然科学基金资助情况。
2. 来稿要观点明确、数据真实可靠、层次分明、言简意明、重点突出。来稿必须是网上在线投稿(含各种符号和外文字母大写、小写、正体、斜体;希腊字母、拉丁字母;上角、下角标位置应标清楚)。中文摘要以300字为宜,英文摘要以2000字符(相当于300个英文单词)为宜;另附关键词。要求来稿应达到“齐、清、定”,中文、英文文字通顺,方可接受送审。
3. 为了进一步统一和完善投稿方式、缩短论文发表周期,本刊在**2007年7月1日以后**,不再接收以邮寄方式或e-mail方式的投稿,**只收网上在线投稿**。严禁“一稿多投”,对侵权、抄袭、剽窃等学术不端行为,一经发现,取消三年投稿资格。
4. 文中插图要求完整,图中坐标、线条、单位、符号、图注等应标注准确、完整。如作者特殊要求需出彩色插图者,必须在投稿时事先加以说明,并承担另加的彩印费用。图幅大小:单栏图7.5cm(宽)×6cm(高);双栏图:14cm(宽)×6cm(高);图中数字、图题、表题全部用中文、英文对照,图中数字、中文、英文全用6号字(电子文档中除实物图外,曲线图尽可能用Matlab, Excel, Visio 或 Origin 等软件制作,稿件中图片的原图随电子版修改稿一同发送到本刊的修改稿专用邮箱)。
5. 文中出现的单位必须按“中华人民共和国计量标准”及有关GB标准规定缮写。物理量符号一律用斜体,单位符号和词头用正体字母。
6. 名词术语,请参照全国科学技术名词规定缮写。
7. 参考文献,采用顺序编码制,只列主要文献;以15~20条为宜。内部资料、私人通讯、未经公开发表的一律不能引用。日文、俄文等非英文文献,请用英文表述;中文文献和中文图书采用中、英文对照表述,文献缮写格式请参照本刊。
8. 请在投稿第一页左下角写明投稿联系人的**电话和两个e-mail**,以便及时联系。

稿件处理

1. 自收到稿件之日起,一个月内作者会收到编辑部的稿件处理意见。请根据录用通知中所提出的要求认真修改,希望修改稿在30天内寄回编辑部,并作为作者最终定稿(当作者接到校样时,以此修改稿为准进行校对,请勿再做大的改动),若二个月内编辑部没收到修改稿,将视为**自行撤稿处理**。
2. 有重大创新并有基金资助者可优先发表;不录用的稿件,编辑部将尽快通知作者,底稿一律不退,请自留底稿。
3. 来稿一经发表将酌致稿酬并送样刊2册,当期封面1份。
4. 遵照“中华人民共和国著作权法”,投稿作者须明确表示,该文版权(含各种媒体的版权)授权给光谱学与光谱分析期刊社。国内外各大文献检索系统摘录本刊刊出的论文;凡不同意被检索刊物无稿酬摘引者,请在投稿时事先声明,否则,本刊一律认为已获作者授权认可。

5. 修改稿请寄:100081 北京市海淀区学院南路76号,《光谱学与光谱分析》期刊社(收)

电话:010-62182998 或 62181070 传真:010-62181070

e-mail: chngpxygpfx@vip.sina.com; 修改稿专用邮箱: gp2008@vip.sina.com 网址: <http://www.gpxygpfx.com>