

近红外光谱结合 SIMCA 法溯源羊肉产地的初步研究

张宁, 张德权*, 李淑荣, 李庆鹏

(中国农业科学院农产品加工研究所, 农业部农产品加工与质量控制重点开放实验室, 北京 100193)

摘要: 产地溯源是食品安全追溯制度的重要组成部分。该文采用近红外光谱结合簇类独立软模式法 (SIMCA) 建立了羊肉产地溯源模型。结果表明, 在 11995~3999 cm^{-1} 波长范围内, 光谱经 5 点平滑 (Smooth) 与多元散射校正 (MSC) 预处理, 山东济宁市、河北大厂县、内蒙临河市、宁夏银川市 4 个产地模型的主成分数分别为 5、6、5、6 时, 采用 SIMCA 模式识别方法可以建立稳健的羊肉产地溯源模型; 在 1% 的显著水平下, 4 个产地校正集模型对未知样本的识别率分别为 95%、100%、100%、100%, 拒绝率均为 100%; 其验证集模型的识别率分别为 100%、83%、100%、92%, 拒绝率均为 100%。该研究表明, 近红外光谱技术作为一种羊肉产地的溯源方法切实可行。

关键词: 近红外光谱, 模式识别, 模型, 簇类独立软模式法 (SIMCA), 产地溯源, 羊肉

中图分类号: O657.3, TS201.6

文献标识码: A

文章编号: 1002-6819(2008)-12-0309-04

张宁, 张德权, 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, 24(12): 309-312.

Zhang Ning, Zhang Dequan, Li Shurong, et al. Preliminary study on origin traceability of mutton by near infrared reflectance spectroscopy coupled with SIMCA method[J]. Transactions of the CSAE, 2008, 24(12): 309-312.(in Chinese with English abstract)

0 引言

中国是羊肉生产大国^[1], 近几年羊痒病、羊布病^[2]的出现, 严重威胁了羊肉安全。研究一种快速、准确地溯源方法, 一旦疫情发生时, 及时有效地找到源头并采取措施控制其传播速度, 对于保证羊肉安全, 保障人们身体健康有重要的现实意义。近红外光谱技术是近年来发展起来的一类新型的分析检测技术^[3-5], 具有非破坏性、快速、简单等特点, 其在农产品质量安全控制上的应用研究日益受到广泛关注^[6]。簇类独立软模式法 (SIMCA) 是 Wold 于 1976 年提出的一种基于主成分分析的特征投影显示方法, 其基本思想是先利用主成分分析的显示结果得到样本分类的基本印象, 然后分别建立各类样本的类模型, 再用这些模型对未知样本进行判别分析^[7]。采用近红外光谱技术结合 SIMCA 模式识别方法已成功应用于茶叶品种鉴别^[8]、饮料掺假识别^[9]、中草药产地判别^[10-12]等领域, 但应用于肉品产地溯源, 目前仅检索到 Xiccato 采用近红外光谱技术结合 SIMCA 模式识别方法对不同产地的鲈鱼进行了产地溯源^[13], 除此之外, 尚未发现有关的研究报道。为此, 本研究提出采用近红外光谱结合模式识别方法建立羊肉产地溯源模型, 以为肉品产地溯源提供一种新的思路。

1 材料与方法

1.1 材料

收稿日期: 2008-02-24 修订日期: 2008-05-19

基金项目: 科研院所社会公益研究专项资助项目 (2005DIA4J035)

作者简介: 张宁 (1981-), 女, 北京 中国农业科学院农产品加工研究所, 100193. Email: zhangning928@126.com

*通讯作者: 张德权 (1972-), 男, 博士, 副研究员, 主要从事肉品科学与非热力加工技术研究。北京 中国农业科学院农产品加工研究所, 100193.

Email: dqzhang0118@126.com

试验所用材料为羊里脊肉, 分别取自山东济宁市、河北大厂县、内蒙临河市、宁夏银川市屠宰厂, 其中山东 32 条, 河北 37 条, 内蒙 39 条, 宁夏 39 条。羊里脊肉经冷冻干燥制成干品, 粉碎后过 40 目筛。

1.2 仪器与测试条件

Tensor37 近红外光谱仪 (德国 Bruker 公司), 附漫反射镀金积分球、样品旋转台和石英样品杯 (直径 42 mm)、硫化铅 (PbS) 检测器。扫描范围: 12000~4000 cm^{-1} ; 扫描次数: 64 次; 分辨率: 8 cm^{-1} ; SIMCA 软件平台 (Perkin Elmer 公司)。

1.3 数据分析

采用 SIMCA 模式识别方法, 首先针对每一类样本的光谱数据矩阵进行主成分分析, 在主成分空间下, 计算各类类内、类间的马氏距离 (以下所有距离指马氏距离), 建立判别模型, 并用识别率和拒绝率来评价模型的好坏; 其次依据该模型对未知样本进行分类, 确定未知样本类别^[14]。

$$\text{识别率} = \frac{\text{识别自身类样本个数}}{\text{该类样本的总个数}} \times 100\%$$

$$\text{拒绝率} = \frac{\text{拒绝其它类样本个数}}{\text{其它类样本的总个数}} \times 100\%$$

2 结果与分析

2.1 光谱预处理方法筛选

光谱测量中产生的噪声会掩盖有用的光谱信息, 而平滑 (Smooth) 是滤除噪声最常用的方法, 本试验以内蒙和宁夏两产地为代表, 各 39 条光谱, 其中 27 条光谱为校正集用于建立模型, 12 条光谱为验证集用于检验模型的可靠性, 在全光谱范围 11995~3999 cm^{-1} 内, 比较了 5 点 Smooth 和 9 点 Smooth 对光谱的影响, 其识别率和

拒绝率见表 1。由表可见, 5 点 Smooth 效果优于 9 点 Smooth, 原因在于平滑点数过多会造成信息失真^[15]。此外, 羊肉样本颗粒的不均匀性也会影响光的漫反射, 导致光谱波动, 为此, 本研究在平滑处理的基础上对样本的原始光谱分别进行了变量标准化 (SNV)、多元散射校正

正 (MSC) 及一阶导数 (Derivative) 等预处理, 结果表明 (表 1), MSC 处理的效果优于 SNV 和一阶导数, 从图 1 也可看出, 平均光谱进行 MSC 处理后减少了样本间因散射而引起的光谱波动。因此, 本研究最终确定选用 5 点 Smooth+ MSC 的光谱预处理方法。

表 1 光谱预处理方法对建模效果的影响
Table 1 Effects of different spectra pretreatments on modeling

产地	预处理	主成分数	校正集		验证集	
			识别率/%	拒绝率/%	识别率/%	拒绝率/%
内蒙	5 点平滑	6	100	100	100	100
	9 点平滑	6	100	100	100	100
	5 点平滑+变量标准化 (SNV)	4	100	96	100	100
	5 点平滑+多元散射校正 (MSC)	5	100	100	100	100
	5 点平滑+一阶导数	1	100	81	100	83
宁夏	5 点平滑	6	100	100	83	100
	9 点平滑	6	96	100	83	100
	5 点平滑+变量标准化 (SNV)	5	96	89	100	100
	5 点平滑+多元散射校正 (MSC)	6	100	100	100	100
	5 点平滑+一阶导数	3	100	59	100	25

注: 识别率指该类识别自身样本的能力, 如内蒙识别率=内蒙类样本个数/内蒙类样本总个数×100%; 拒绝率指该类拒绝其他类样本的能力, 如内蒙拒绝率=内蒙类拒绝宁夏类样本个数/宁夏类样本总个数×100%。

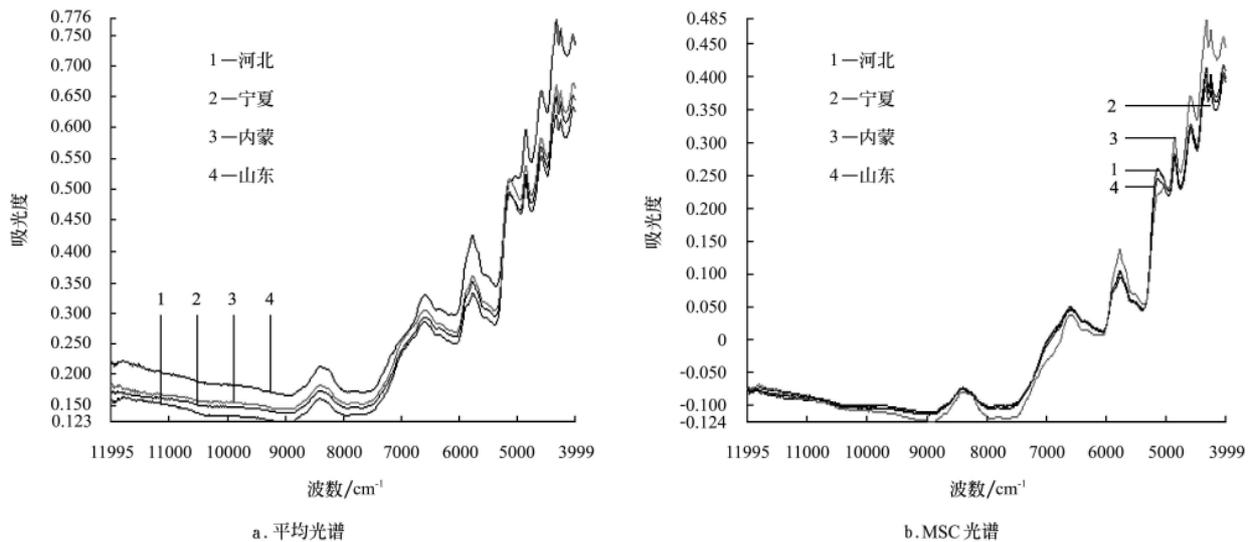


图 1 4 个产地羊肉的平均光谱图及多元散射校正 (MSC) 处理图

Fig.1 Mean spectra and Multiplicative Scatter Correction(MSC) spectra of mutton from four regions

2.2 波长范围筛选

为确定最佳的光谱范围, 分别对内蒙、宁夏两产地的全光谱、分段光谱进行筛选。各取 39 条光谱, 其中 27 条光谱为校正集用于建立模型, 12 条光谱为验证集用于检验模型的可靠性, 结果表明 (表 2), 由全光谱建立的模型优于部分波段建模的效果, 其识别率和拒绝率均达到 100%。从图 2 也可看出, 在水分吸收敏感区 5155 cm^{-1} 处和脂肪特征峰 5767 、 4329 cm^{-1} 处及蛋白特征峰 6670 、 4650 cm^{-1} 处, 4 个产地的羊肉光谱均存较大差异, 而部分波段仅反应了不同产地羊肉光谱的部分信息, 造成识别率下降, 为此, 本研究选用全波段光谱进行建模, 增加模型的稳健性。

表 2 不同波长范围对建模效果的影响

Table 2 Effects of different spectra ranges on modeling

产地	波数/ cm^{-1}	主成分数	校正集		验证集	
			识别率/%	拒绝率/%	识别率/%	拒绝率/%
内蒙	11995~3999	5	100	100	100	100
	5300~4900	6	100	100	100	100
	5900~5500+4500~3999	7	96	100	0	100
	6800~6300+5000~4500	7	100	100	92	100
宁夏	11995~3999	6	100	100	100	100
	5300~4900	6	100	100	92	100
	5900~5500+4500~3999	6	100	100	92	100
	6800~6300+5000~4500	6	100	100	92	100

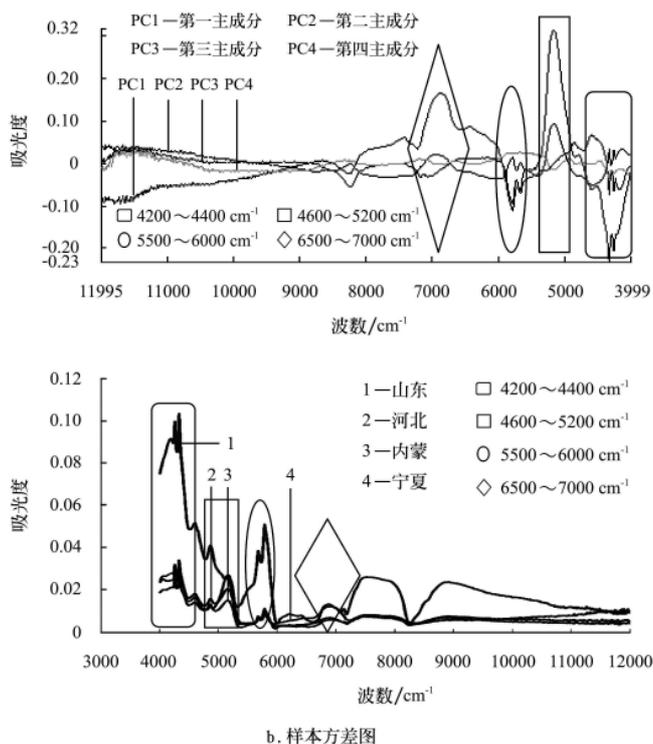


图2 样本主成分图与方差图

Fig.2 Weight spectra for the first four Principal Components (PCs) and variance plot of sample

2.3 模型的建立

在全波段范围内，对山东（32 个样本）、河北（37 个样本）、内蒙（39 个样本）、宁夏（39 个样本）共计 147 个羊肉样本的近红外光谱进行主成分分解，结果表明，主成分图中的特征区域 $4200 \sim 4400 \text{ cm}^{-1}$ 、 $4600 \sim 5200 \text{ cm}^{-1}$ 、 $5500 \sim 6000 \text{ cm}^{-1}$ 和 $6500 \sim 7000 \text{ cm}^{-1}$ 与样本方差图中的特征峰位置基本一致（图 2），说明可以用主成分来代替原始数据。再以所有光谱数据的前 3 个主成分得分作图发现（图 3），3 个主成分得分图不能直观地反映出各个产地样本的分开情况，表明仅用主成分分析不能将产地识别。光谱相关分析表明，宁夏和山东、河北、内蒙的相关系数均较大，分别达到 0.9678、0.9940、0.9989，是导致不同产地样本相互重叠的主要原因。为此，本研究在主成分分解的基础上，从山东、河北、内蒙、宁夏 4

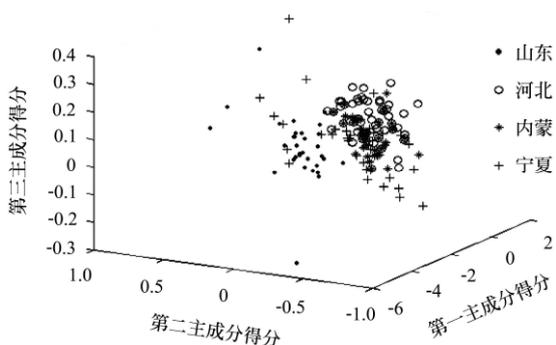


图 3 4 个产地羊肉样本的前 3 个主成分得分三维投影图
Fig.3 Score three-dimensional cluster plot using top three PCs of mutton from four regions

类样本中各随机挑选 20、25、27 和 27 个样品作为校正集，对每类羊肉样本光谱分别进行 SIMCA 分析，剩下 48 个样本作为验证集，检验模型的可靠性。基于 SIMCA 软件，确定山东、河北、内蒙和宁夏 4 类模型的最佳主成分分数分别为 5、6、5、6，在主成分空间下，计算出各个类样本类间的距离（表 3）及各类临界距离，分别为 1.46、1.4、1.37、1.38。由表 3 可以看出，任意两个产地类间距离均大于两类的各自临界距离，表明类类间有明显的类界限，说明可以将该 SIMCA 模型用于鉴别不同来源的羊肉产地。

表 3 4 个产地羊肉样本的类间距离

Table 3 Mahalanobis distances between classes in Soft Independent Modeling of Class Analogy (SIMCA) of mutton from four regions

	内蒙	宁夏	山东	河北
内蒙	1	3.90	9.30	13.59
宁夏		1	7.04	10.07
山东			1	11.53

2.4 校正和验证的结果

用山东、河北、内蒙、宁夏 4 类中各 20、25、27 和 27 个样本作为校正集，建立 SIMCA 模型，剩下 48 个样本作为验证集，检验模型的可靠性，结果见表 4。由表可见，在 1% 的显著水平下，山东、河北、内蒙、宁夏校正集模型的识别率分别为 95%、100%、100%、100%，拒绝率全部为 100%；用建立的模型对验证集样本进行检验，其识别率分别为 100%、83%、100%、100%，拒绝率全部为 100%。

表 4 SIMCA 产地模型校正与验证的结果

Table 4 Results of calibration and validation set by SIMCA

产地	主成分数	校正集		验证集	
		识别率/%	拒绝率/%	识别率/%	拒绝率/%
山东	5	95	100	100	100
河北	6	100	100	83	100
内蒙	5	100	100	100	100
宁夏	6	100	100	100	100

3 结论

本研究利用近红外光谱结合 SIMCA 模式识别方法对 4 个产地羊肉分别建立了类模型，模型基本能正确识别 4 类产地羊肉，对其他产地识别时，拒绝率达到 100%。结论充分表明，近红外光谱结合 SIMCA 模式识别方法在羊肉产地溯源中具有可行性，为解决羊肉安全问题提供了一种新的思路。

[参考文献]

[1] 中华人民共和国农业部. 2006 年牛羊肉市场形势分析 [R]. 20070-03-09.
 [2] 中华人民共和国卫生部. 关于加强布鲁氏菌病防治工作的通知 [R]. 2007-12-04.
 [3] 金钦汉. 从 2000 年匹茨堡会议看分析化学和分析仪器发

- 展的一些动向[J]. 现代科学仪器, 2000, (3): 14—16.
- [4] Peres B, Barlet N, Loiseau G, et al. Review of the current methods of analytical traceability allowing determination of the origin of foodstuffs[J]. Food control, 2007, 18(3): 228—235.
- [5] Luykx D M A M, van Ruth S M. An overview of analytical methods for determining the geographical origin of food products[J]. Food Chemistry, 2008, 107(2): 897—911.
- [6] 陆婉珍, 袁洪福, 徐广通, 等. 现代近红外光谱分析[M]. 北京: 中国石化出版社, 2000.
- [7] Wold S. Pattern recognition by means of disjoint principle components models[J]. Pattern Recognition, 1976, 8(3): 127—139.
- [8] 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. 食品科学, 2006, 27 (4): 186—189.
- [9] Pontes M J C, Santos S R B, Araújo M C U, et al. Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectrometry[J]. Food Research International, 2006, 39: 182—189.
- [10] Cho C H, Woo Y A, Kim H J, et al. Rapid qualitative and quantitative evaluation of deer antler(*Cervus elaphus*) using near infrared reflectance spectroscopy[J]. Microchem, 2001, 68(2-3): 189—195.
- [11] Woo Y A, Cho J H, Kim H J, et al. Classification of cultivation area of ginseng by near infrared spectroscopy and ICP-AES[J]. Microchem Journal, 2002, 73(3): 299—306.
- [12] Woo Y A, Kim H J, Ze K R, et al. Near-infrared(NIR) spectroscopy for the non-destructive and fast determination of geographical origin of *Angelicae gigantis Radix*[J]. Journal of Pharmaceutical and Biomedical Analysis, 2005, 36(5): 955—959.
- [13] Xiccato G, Trocino A, Tulli F, et al. Prediction of chemical composition and origin identification of European sea bass (*Dicentrarchus labrax* L.) by near infrared reflectance spectroscopy(NIRS)[J]. Food Chemistry, 2004, 86(2): 275—281.
- [14] 严衍禄. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005.
- [15] 许禄. 化学计量学方法[M]. 北京: 科学出版社, 1997.

Preliminary study on origin traceability of mutton by near infrared reflectance spectroscopy coupled with SIMCA method

Zhang Ning, Zhang Dequan^{*}, Li Shurong, Li Qingpeng

(Institute of Agro-Food Science and Technology, Chinese Academy of Agricultural Sciences, Key Laboratory of Agricultural Product Processing and Quality Control, Ministry of Agriculture, Beijing 100193, China)

Abstract: Origin traceability is an important part of food safety traceability system. The mutton origin traceability model was developed by near infrared reflectance (NIR) spectroscopy coupled with soft independent modeling of class analogy (SIMCA). The results showed that the model of mutton origin traceability could be established very well when original NIR spectra was preprocessed by five point smooth and Multiplicative Scatter Correction (MSC) in the spectra region between 11995 and 3999 cm^{-1} and the optimal number of principal component was 5, 6, 5, 6 for the origin traceability model of mutton from Jining city of Shandong Province, Hebei Province, Linhe city of inner mongolia and rinchuan city of Ningxia region, respectively. Under the 1% significant level, the recognition rate of calibration set for four models was 95%, 100%, 100% and 100%, respectively. The recognition rate of validation set was 100%, 83%, 100% and 100%, respectively and the rejection rate of four models were all 100%. It indicates that the technique of NIR was a feasible way for tracing the origin of mutton.

Key words: near infrared spectroscopy, pattern recognition, models, soft independent modeling of class analogy (SIMCA), origin traceability, mutton