

近红外光谱结合化学计量学方法检测蜂蜜产地

李水芳¹, 单杨^{2*}, 朱向荣², 李忠海³

(1. 中南林业科技大学理学院, 长沙 410004; 2. 湖南省食品测试分析中心, 长沙 410025;
3. 中南林业科技大学食品学院, 长沙 410004)

摘要: 为了实现蜂蜜产地的快速判别, 应用近红外光谱结合化学计量学方法对蜂蜜产地进行了判别分析。kennard-Stone 法划分训练集和预测集。光谱用一阶导数加自归一化预处理后, 再用小波变换 (WT) 进行压缩和滤噪。结合滤波后光谱信息, 分别用径向基神经网络 (RBFNN) 和偏最小二乘-线性判别分析 (PLS-LDA) 建立了苹果蜜产地和油菜蜜产地的判别模型。对不同小波基和分解尺度进行了讨论。对苹果蜜, WT-RBFNN 模型和 WT-PLS-LDA 模型都是小波基为 db1、分解尺度为 2 时的预测精度较好, 都为 96.2%。对油菜蜜: WT-RBFNN 模型在小波基为 db4 和分解尺度为 1 时, 预测精度较好, 为 85.7%; WT-PLS-LDA 模型在小波基为 db9、分解尺度也为 1 时, 预测精度较好, 为 90.5%。研究表明: WT 结合线性的 PLS-LDA 建模比 WT 结合非线性的 RBFNN 建模更适于蜂蜜产地判别; 近红外光谱技术具有快速判别蜂蜜产地的潜力。

关键词: 近红外光谱, 小波变换, 径向基函数神经网络, 蜂蜜, 产地判别, 偏最小二乘-线性判别分析

doi: 10.3969/j.issn.1002-6819.2011.08.061

中图分类号: O43, TP3, S896.1

文献标志码: A

文章编号: 1002-6819(2011)-08-0350-05

李水芳, 单杨, 朱向荣, 等. 近红外光谱结合化学计量学方法检测蜂蜜产地 [J]. 农业工程学报, 2011, 27(8): 350-354.

Li Shuifang, Shan Yang, Zhu Xiangrong, et al. Detection of geographical origin of honey using near-infrared spectroscopy and chemometrics [J]. Transactions of the CSAE, 2011, 27(8): 350-354. (in Chinese with English abstract)

0 引言

蜂蜜是一种深受广大消费者喜爱的天然保健食品。它是蜜蜂采集植物的花蜜、分泌物或蜜露, 并与自身分泌物结合后, 经充分酿造而成的天然甜物质。蜂蜜主要成分为葡萄糖和果糖, 此外, 还含有蛋白质、氨基酸、维生素、微量元素、有机酸、色素、芳香物质的高级醇、胶质物、醇类、蜂花粉和激素等^[1-2]。蜂蜜中的各种成分及其含量随蜜种不同而异, 即使是同一蜜种的蜂蜜, 采自不同地区, 由于受气候条件、土壤性质及养蜂技术等诸多环境因素的影响, 也极少出现有完全相同成分和相同含量的分析结果^[2]。

蜂蜜品质与产地的关系密切。欧盟规定: 蜂蜜标签应标注产地; 同一植物来源而产地不同的蜂蜜, 价格会有所差异^[3]。中国在蜂蜜产地标注方面虽然没有严格规定, 但消费者在购买蜂蜜时一般都会关注其产地。加入世贸组织后, 中国对原产地的商标保护正在不断加强。作为蜂蜜生产和出口大国, 蜂蜜标签的产地标注也应引起重视。近红外光谱技术 (near-infrared spectroscopy, NIR) 已成功用于

海参^[4]、中药材^[5]、葡萄酒^[6]、羊肉^[7]等的产地判别。该技术用于蜂蜜理化指标的测定^[8-10]以及掺假^[1,11-12]和蜜源植物^[13-14]的判别已有较多报道, 而鲜有产地判别的报道。由于蜂蜜产地差异主要来自环境因素, 这种差异较掺假和蜜源植物的差异小, 因此产地判别比掺假和蜜源植物判别困难。小波变换 (wavelet transform, WT) 可有效消除光谱噪声、提取光谱有用信息, 是一种应用广泛的光谱数据压缩和去噪方法。该法作为一种光谱预处理用于蜂蜜产地判别未有报道。本文以 2 种单花蜜 (苹果蜜和油菜蜜) 为研究对象, 采用近红外光谱结合小波变换的非线性和线性建模方法分别对蜂蜜产地进行判别分析, 目的是检验近红外光谱结合小波变换预处理在蜂蜜产地快速判别上的可行性, 为蜂蜜产地的快速判别提供参考。

1 材料和方法

1.1 蜂蜜样本

试验所用蜂蜜样本由湖南明园蜂业有限公司提供, 于 2008 和 2009 年收集。蜂蜜从蜂农处直接购得后贮藏在 6~8℃ 冰箱中。采集光谱时, 先将蜂蜜从冰箱内取出, 放置至室温 (约 25℃), 再在 55℃ 水浴下溶解晶体, 置于约 25℃ 环境下待光谱采集。植物来源、产地及样本数见表 1。

收稿日期: 2010-11-15 修订日期: 2011-07-14

基金项目: “十一五” 国家科技支撑计划项目 (2009BADB9B07)

作者简介: 李水芳 (1968-), 女, 湖南宁乡人, 副教授, 主要从事森林食物资源的开发与利用研究。长沙 中南林业科技大学理学院, 410004。

Email: csfulishuifang@126.com

*通信作者: 单杨 (1963-), 男, 湖南攸县人, 研究员, 博士生导师, 主要从事食品安全分析与农产品加工领域研究。长沙 湖南省食品测试分析中心, 410025。

表 1 蜂蜜样本

Table 1 Honey samples

植物来源	苹果		油菜	
	山西临汾	陕西咸阳	四川资阳	湖南常德
产地				
样本数	40	40	21	41

1.2 仪器条件及光谱采集

在约 25℃ 环境下，用 Antaris II 傅里叶变换近红外光谱仪（美国 Thermo Nicolet 公司）及 TQ 软件采集和保存光谱。光谱采集条件：扫描范围 10 000~4 000 cm⁻¹，分辨率 8 cm⁻¹，增益 2，扫描次数 32 次。采用积分球透反射方式，每个样本扫描 3 次，取 3 次的平均光谱为样本光谱。

1.3 模型评价

模型的预测能力用预测总正确率（total accuracy, TA）、敏感度（sensitivity）、特异度（specificity）和 ROC（receiver operating characteristic curve）曲线下的面积为指标来评价。敏感度、特异度及总正确率 3 个指标定义如下：

$$\text{敏感度} = \frac{TP}{TP + FN}; \quad \text{特异度} = \frac{TN}{TN + FP};$$

$$\text{总正确率} = \frac{TN + TP}{TP + FN + TN + FP}$$

TP、TN、FP 和 FN 分别代表真阳性、真阴性、假阳性和假阴性。对苹果蜜，阳性和阴性分别代表山西临汾和陕西咸阳的蜂蜜；对油菜蜜，则分别表示四川资阳和湖南常德的蜂蜜。

ROC 曲线经常用来评价二分类模型性能^[15]。ROC 曲线下的面积越大，模型的分类能力越强。当 ROC 曲线下的面积为 1 时，可认为所建分类模型具有很好的分类能力；当 ROC 曲线下的面积小于 0.5 时，则所建分类模型已不具备分类能力。

1.4 分析软件

小波变换、径向基函数神经网络由 Matlab7.0 实现。PLS-LDA 代码由网络 (<http://code.google.com/p/cars2009>) 共享下载实现。Kennard-stone(KS)算法由网络提供的共享工具包下载实现^[16]。

1.5 样本集的划分

代表性样本的选择不仅可以减少建模的工作量，还可提高模型的可靠性和稳定性。Kennard-stone(KS)法^[16-17]是一种典型而有效的从样本集中选择具有代表性样本的方法。本文选用 KS 法划分训练集和预测集。80 个苹果蜜中 54 个作训练集，26 个作预测集；62 个油菜蜜中 41 个作训练集，21 个作预测集。

2 结果与分析

2.1 光谱预处理方法的选择

近红外光谱往往包含一些与待测样本性质无关的因素引起的干扰，如样品不均匀、光散射、杂散光和仪器响应的影响，导致近红外光谱的基线漂移和光谱的不重复。适当的光谱预处理可减少或消除这些因素的影响，有效提取光谱信息。本文以 PLS-LDA 模型的预测总正确率为评价指标，选用一阶导数、二阶导数、自归一化、平滑、多元散射校正及它们的相互组合对光谱进行预处理。最终确定一阶导数与自归一化相结合为较好的光谱预处理方法。

2.2 小波变换 (WT)

WT 中小波参数（小波基、阶数和分解尺度）的选择至关重要，它直接影响 WT 后所建模型的预测能力^[18-19]。db (daubechies) 族小波基在小波中应用最广泛，具有分析近红外光谱的特性。本文选择 db 族小波基，考察不同阶数和分解尺度对模型预测能力的影响。

2.2.1 小波变换对 RBFNN 模型预测能力的影响

图 1 为选用不同阶数和分解尺度的 50 个 RBFNN 模型 (spread=2) 的预测总正确率结果。对苹果蜜，模型 2 的预测总正确率最高，为 96.2%，对应的小波系数为 389、小波基为 db1、分解尺度为 2。油菜蜜的最高预测总正确率为 85.7%，由模型 16 得到，对应的小波系数为 781、小波基为 db4、分解尺度为 1。

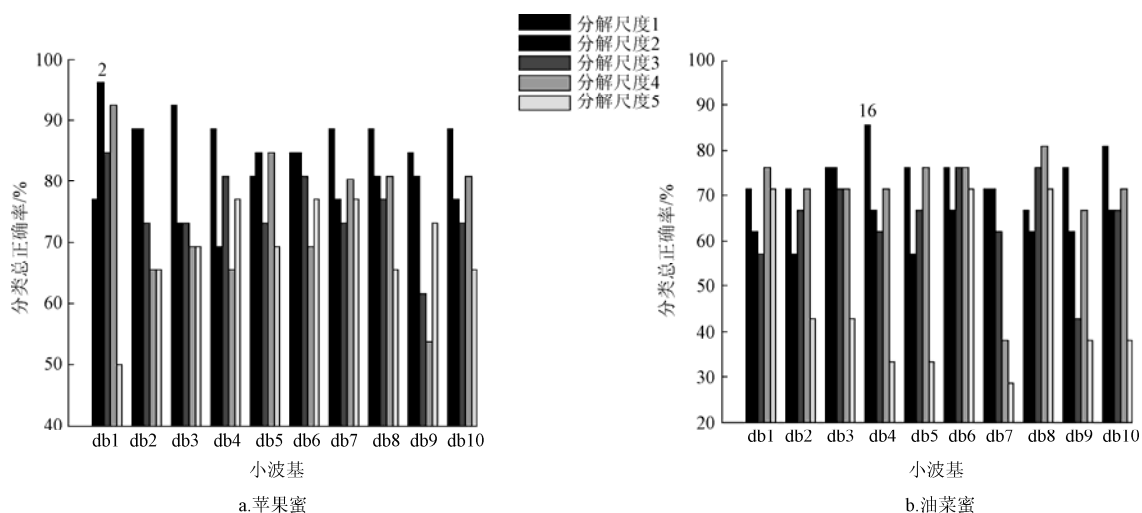


图 1 不同小波参数的 RBFNN 模型预测总正确率

Fig.1 Total accuracy rate of RBFNN models using different wavelet parameters

2.2.2 小波变换对 PLS-LAD 模型预测能力的影响

不同阶数和分解尺度的 50 个 PLS-LAD 模型的预测总正确率结果见图 2。对苹果蜜，模型 2、4、6、7、12

和 16 都得到了 96.2% 的最高预测总正确率，对应的小波系数分别为 389、98、779、391、392 和 781。为了进一步评价这 6 个模型的预测能力，对训练集进行了 10 折交

互验证,6个模型10折交互验证的总正确率分别为98.2%、94.4%、94.4%、92.6%、96.3%和94.4%,ROC曲线下的面积分别为0.962、0.934、0.951、0.940、0.944和0.948,表明模型2的预测能力最好,因此,选择小波基为db1,分解尺度为2。对油菜蜜,模型6、36、41和46都得到了

90.5%的最高预测总正确率,对应的小波系数分别为779、785、786和787。这4个模型对训练集10折交互验证的总正确率分别为82.9%、82.9%、85.4%和85.4%,但模型41的ROC曲线下的面积最大为0.863,因此其预测能力最强。模型41的小波基为db9、分解尺度为1。

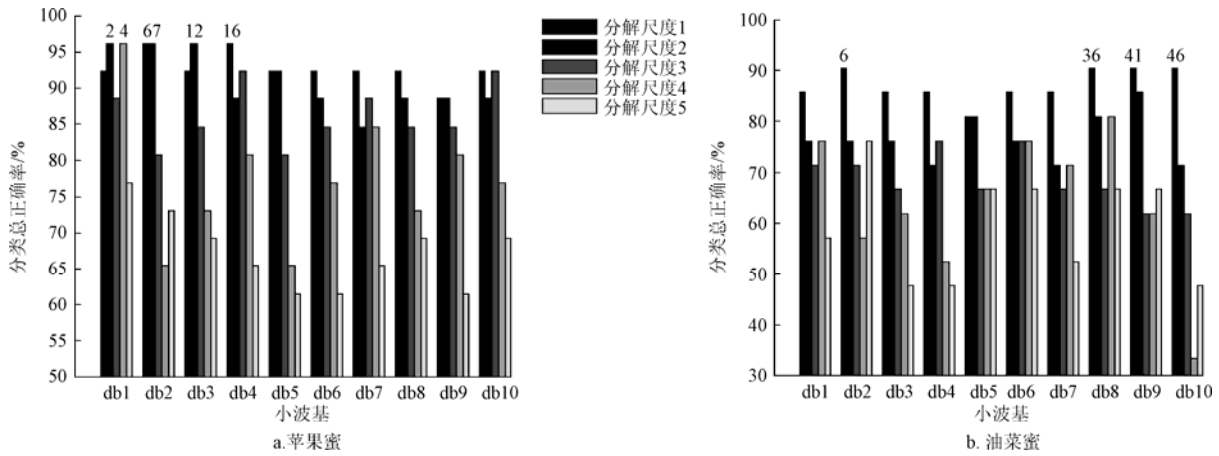


图2 不同小波参数的PLS-LDA模型预测总正确率
Fig.2 Total accuracy rate of PLS-LDA models using different wavelet parameters

2.2.3 小波变换对光谱的影响

图3为一代表性苹果蜜样本的原始光谱、一阶导数光谱、一阶段导数+自归一化处理后的光谱及一阶段导数+自归一化+WT(db1,2层分解)处理后的光谱。原始光

谱非常光滑,噪声很低。经一阶导数处理后,谱图光滑性降低,噪声增大。再经WT处理后,谱图光滑性显著增强,而光谱形状没有改变,说明噪声明显被去除,而光谱信息被保留下来。

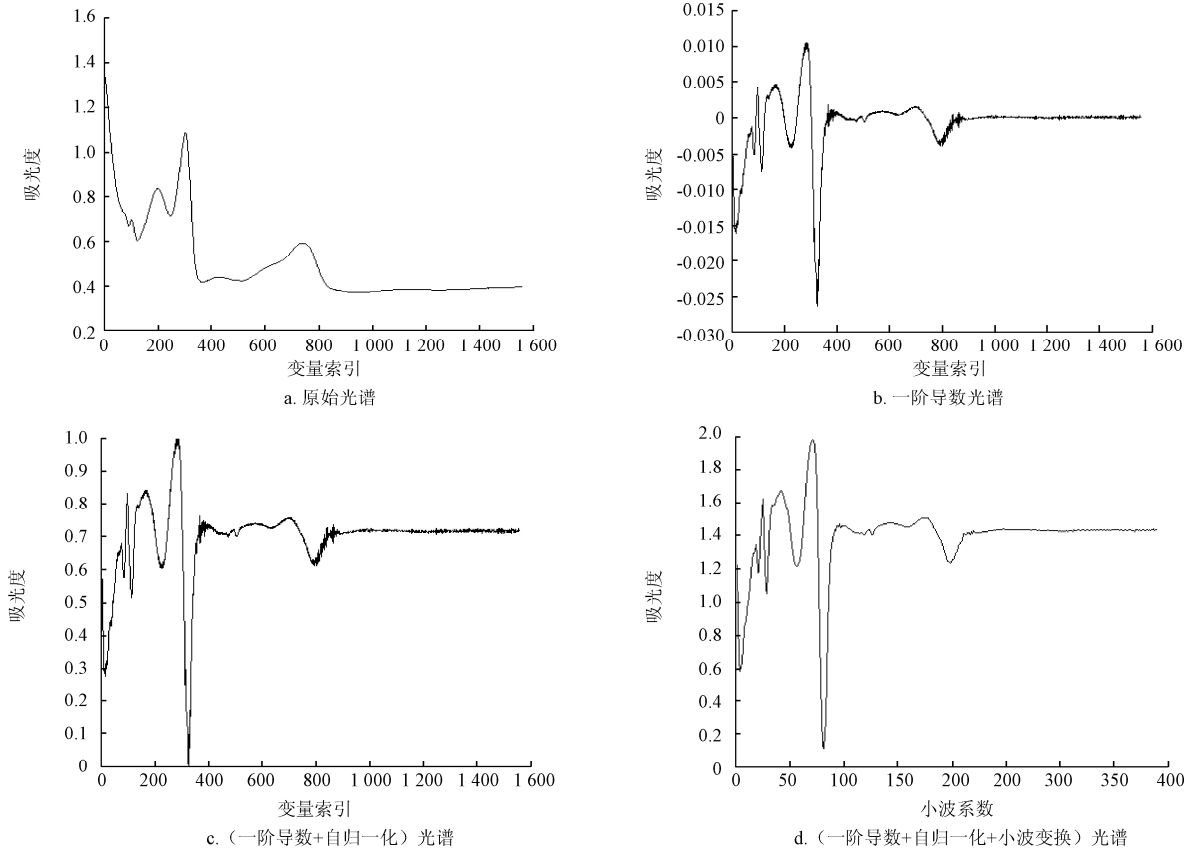


图3 苹果蜜样本近红外光谱
Fig.3 NIR spectrum of apple honey sample

2.3 产地判别

蜂蜜是一种成分复杂的天然物质,产地的差异主要受环境因素(如气候条件、土壤性质和养蜂技术等)的

影响。本文分别采用非线性和线性方法建模。RBFNN因具有很好的逼近性能和全局最优特性,并且结构简单,训练速度快,因此被选作本文的非线性建模方法。

PLS-LDA 通过 PLS 提取主成分, 再结合 LDA 进行判别分析, 比其他线性判别方法往往具有更好的模式识别能力, 因此选择该法为本文的线性建模方法。

2.3.1 RBFNN 模型

用 $\text{net=newrbe}(P, T, \text{spread})$ 函数建立 RBFNN 模型, 其中 P 是训练集光谱, T 是分类变量, 设为 1 和 -1, spread 值是需确定的参数。用模型对预测集进行预测, 计算预测集分类变量的值 Y_p ; 当 $Y_p > 0$ 时, 判定样本属于 1 这一类; $Y_p < 0$ 时, 判定样本属于 -1 这一类。步长取 0.5, 考察 spread 值在 1~10 范围内, 与之对应的 RBFNN 模型的预测总正确率。最终确定苹果蜜的 spread 值分别为 4 (RBFNN) 和 2 (WT-RBFNN), 而油菜蜜的 spread 值都为 2。RBFNN 模型结果见表 2。从表 2 可知, 苹果蜜和油菜蜜的最好预测结果都由模型 WT-RBFNN 得到, 对应的预测总正确率和 ROC 面积分别为 96.2%、0.923 (苹果蜜) 和 85.7%、0.712 (油菜蜜)。从表 2 还可看出: WT 预处理能有效提高模型的预测精度; 总的预测结果苹果蜜优于油菜蜜。

表 2 RBFNN 和 WT-RBFNN 模型的预测结果

Table 2 Prediction results of RBFNN and WT-RBFNN models

样本	方法	总正确率/%	敏感度/%	特异度/%	ROC 面积
苹果蜜	RBFNN	96.2	100	92.3	0.888
	WT-RBFNN	96.2	100	92.3	0.923
油菜蜜	RBFNN	81.0	100	50.0	0.608
	WT-RBFNN	85.7	100	62.5	0.712

2.3.2 PLS-LDA 模型

对 PLS-LDA 和 WT-PLS-LDA 模型的预测能力进行了比较, 结果见表 3。WT-PLS-LDA 模型获得了最好的预测结果: 苹果蜜的最高预测总正确率为 96.2%, ROC 面积为 0.923, 而油菜蜜的最高预测总正确率为 90.5%, 对应的 ROC 面积为 0.856。与 RBFNN 模型类似: 总的预测结果油菜蜜比苹果蜜要差; PLS-LDA 模型比 WT-PLS-LDA 模型差。

表 3 PLS-LDA 和 WT-PLS-LDA 模型的预测结果

Table 3 Prediction results of PLS-LDA and WT-PLS-LDA models

样本	方法	总正确率/%	敏感度/%	特异度/%	ROC 面积
苹果蜜	PLS-LDA	96.2	100	92.3	0.893
	WT-PLS-LDA	96.2	100	92.3	0.923
油菜蜜	PLS-LDA	85.7	76.9	100	0.875
	WT-PLS-LDA	90.5	84.6	100	0.856

2.3.3 模型预测结果的比较

对 2 种蜂蜜, 预测结果都是 WT-RBFNN 优于 RBFNN, WT-PLS-LDA 优于 PLS-LDA, 说明 WT 在蜂蜜产地鉴别上是一种行之有效的光谱预处理方法。WT 不仅能使数据点显著减少, 而且能有效提取原信号中的特征信息, 因此能提高模型的预测精度。对苹果蜜, 4 个模型的预测总正确率都达到 96.2%, 说明非线性和线性模型对苹果蜜产地都有很好的判别能力。对油菜蜜, PLS-LDA 模型要优于 RBFNN 模型, WT-PLS-LDA 模型要优于

WT-RBFNN 模型, 最好的预测结果由 WT-PLS-LDA 模型产生, 预测总正确率为 90.5%, 说明线性模型更适于油菜蜜的产地判别。但总体分析各模型的预测结果, 线性模型可能更适于蜂蜜产地的判别分析。同样的光谱预处理和同样的建模方法对不同植物来源的蜂蜜的产地判别结果会有较大差异, 这可能与不同植物来源的蜂蜜自身性质有关。因此, 对蜂蜜进行产地判别时, 对不同植物来源的蜂蜜分别建模, 可以获得更好的判别结果。

3 结 论

使用近红外光谱结合小波变换 (WT) 预处理和线性的偏最小二乘-线性判别 (PLS-LDA) 及非线性的径向基函数神经网络 (RBFNN) 建模方法对苹果蜜和油菜蜜的产地分别进行了判别分析。用 Kennard-stone 法划分训练集和预测集。通过对模型预测结果的分析, 可得以下结论: 1) 在蜂蜜产地快速判别上, WT 预处理光谱能有效提高所建模型的预测精度; 2) 对不同植物来源蜂蜜的产地判别, 同样的光谱预处理方法和同样的建模方法所得预测结果会存在较大差异; 3) 线性的 WT-PLS-LDA 模型比非线性的 WT-RBFNN 模型更适于蜂蜜产地判别; 4) 近红外光谱技术具有快速判别蜂蜜产地的潜力。

相对于 ^1H 核磁共振 (NMR) 法、荧光光谱法以及 ^{13}C 和 ^2H 同位素分析、氨基酸分析、矿质元素分析、糖类成分分析判别蜂蜜产地, 近红外光谱技术具有快速、无损、成本低、操作简单等特点, 适于现场快速分析, 具有实际应用前景, 为蜂蜜产地的快速判别提供了一种新方法。

本文仅对苹果蜜和油菜蜜两个蜜种进行了产地判别分析, 而实际中蜂蜜品种多种多样, 任何蜂蜜都不可能是完全单一花粉蜜, 且一些蜜种产地较多。因此, 采集更多品种单花蜜或多花蜜进行多产地判别分析很有必要, 今后应逐渐完善这方面的研究。

[参 考 文 献]

- [1] Downey G, Fouratier V, Kelly J D. Detection of honey adulteration by addition of fructose and glucose using near-infrared transmittance spectroscopy [J]. *Journal of Near Infrared Spectroscopy*, 2003, 11(6): 447-456.
- [2] 曹炜, 尉亚辉. 蜂产品保健原理与加工技术[M]. 北京: 化学工业出版社, 2002, 191.
- [3] Woodcock T, Downey G, Kelly J D, et al. Geographical classification of honey samples by near-infrared spectroscopy: a feasibility study[J]. *Journal of Agricultural and Food Chemistry*, 2007, 55(22): 9128-9134.
- [4] 陶琳, 武中臣, 张鹏彦, 等. 近红外光谱快速鉴定干海参产地[J]. *农业工程学报*, 2011, 27(5): 364-366.
Tao Lin, Wu Zhongchen, Zhang Pengyan, et al. Rapid identification of dried sea cucumber products from different geographical areas by near-infrared spectroscopy[J]. *Transaction of the CSAE*, 2011, 27(5): 364-366. (in Chinese with English abstract)
- [5] 刘沐华, 张学工, 周群, 等. 近红外漫反射光谱法和模式识别技术鉴别中药材产地[J]. *光谱学与光谱分析*, 2006, 26(4): 629-632.
Liu Shuhua, Zhang Xuegong, Zhou Qun, et al. Determination of geographical origins of Chinese medical herbs by NIR and

- pattern recognition[J]. Spectroscopy and spectral analysis, 2006, 26(4): 629—632. (in Chinese with English abstract)
- [6] 刘巍, 战吉成, 黄卫东, 等. 基于近红外光谱技术的葡萄酒原产地辨识方法[J]. 农业工程学报, 2010, 26(13): 374—378.
- Liu Wei, Zhan Jicheng, Huang Weidong, et al. Application of near infrared spectroscopy and clustering analysis to classify wines from different origins[J]. Transaction of the CSAE, 2010, 26(13): 374—378. (in Chinese with English abstract)
- [7] 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]. Transaction of the CSAE, 2008, 24(12): 309—312. (in Chinese with English abstract)
- [8] Ruoff K, Luginbuhl W, Bogdanov S, et al. Quantitative determination of physical and chemical measurands in honey by near-infrared spectrometry[J]. European Food Research and Technology, 2007, 225(3/4): 415—423.
- [9] Cozzolino D, Corbella E. Determination of honey quality components by near infrared reflectance spectroscopy[J]. Journal of Apicultural Research, 2003, 42(1/2): 16—20.
- [10] Garcia-Alvarez M, Huidobro J F, Hermida M, et al. Major components of honey analysis by Near-Infrared transreflectance spectroscopy[J]. Journal of the Agricultural and Food Chemistry, 2000, 48(11): 5154—5158.
- [11] Kelly J D, Petisco C, Doweny G. Potential of near-infrared transreflectance spectroscopy to detect adulteration of Irish honey by beet invert syrup and high fructose corn syrup[J]. Journal of Near Infrared Spectroscopy. 2006, 14(2): 139—146.
- [12] Zhu X R, Li S F, Shan Y, et al. Detection of adulterations such as sweeteners materials in honey using near-infrared spectroscopy and chemometrics[J]. Journal of Food Engineer, 2010, 101(1): 92—97.
- [13] 杨燕, 聂鹏程, 杨海清, 等. 基于可见-近红外光谱技术的蜜源快速识别方法[J]. 农业工程学报, 2010, 26(3): 238—242.
- Yang Yan, Nie Pengcheng, Yang Haiqing, et al. Rapid recognition method of nectar plant based on visible-near infrared spectroscopy[J]. Transaction of the CSAE, 2010, 26(3): 238—242. (in Chinese with English abstract)
- [14] Ruoff K, Luginbuhl W, Bogdanov S, et al. Authentication of the botanical origin of honey by near-infrared spectroscopy[J]. Journal of Agricultural and Food Chemistry, 2006, 54(18): 6867—6872.
- [15] Fawcett T. An introduction to ROC analysis[J]. Pattern Recognition Letters, 2006, 27(8): 861—874.
- [16] Kennard R W, Stone L A. Computer aided design of experiments[J]. Technometrics, 1969, 11(1): 137—148.
- [17] Daszykowski M, Semeel S, Kaczmarek K, et al. TOM-CAT: A MATLAB toolbox for multivariate calibration techniques[J]. Chemometrics and Intelligent Laboratory Systems, 2007, 85(2): 269—277.
- [18] Li X L, He Y. Discriminating varieties of tea plant based on Vis/NIR spectral characteristics and using artificial neural networks[J]. Biosystems Engineering, 2008, 99(3): 313—321.
- [19] 夏俊芳, 李小昱, 李培武, 等. 基于小波变换的柑橘维生素 C 含量近红外光谱无损检测方法[J]. 农业工程学报, 2007, 23(6): 170—174.
- Xia Junfang, Li Xiaoyu, Li Peiwu, et al. Approach to nondestructive measurement of Vitamin C content of orange with near-infrared spectroscopy treated by wavelet transform[J]. Transaction of the CSAE, 2007, 23(6): 170—174. (in Chinese with English abstract)

Detection of geographical origin of honey using near-infrared spectroscopy and chemometrics

Li Shuifang¹, Shan Yang^{2*}, Zhu Xiangrong², Li Zhonghai³

(1. College of Science, Central South University of Forestry and Technology, Changsha 410004, China;

2. Hunan Food Test and Analysis Center, Changsha 410025, China;

3. College of Food Science and Technology, Central South University of Forestry & Technology, Changsha 410004, China)

Abstract: Near infrared spectroscopy combined with chemometrics methods has been used to detect the geographical origin of honey samples. The samples were divided into the training set and the test set by kennard-Stone algorithm. After being pre-treated with first derivative and autoscaling, the spectral data were compressed and de-noised using wavelet transform (WT). The radical basis function neural networks (RBFNN) and partial least squares-line discriminant analysis (PLS-LDA) were applied to develop classification models, respectively. The performances of different wavelet functions and decomposition levels were evaluated in relation to the total prediction accuracy for the test set. For apple honey samples, when wavelet function was db1 and decomposition level was 2, both WT-RBFNN model and WT-PLS-LDA model produced the largest total prediction accuracy of 96.2%. For rape honey samples, when wavelet function was db4 and decomposition level was 1, WT-RBFNN model made the largest total prediction accuracy of 85.7%; while when wavelet function was db9 and decomposition level was also 1, WT-PLS-LDA model got the largest total prediction accuracy of 90.5%; The results indicated that linear WT-PLS-LDA model was more suitable for geographical classification of honey samples than no-linear WT-RBFNN model. Near infrared spectroscopy technique have a potential for quickly detecting geographical classification of honey samples.

Key words: near-infrared spectroscopy, wavelet transform (WT), models, honey, geographical classification, radical basis function neural networks (RBFNN), partial least squares-line discriminant analysis (PLS-LDA)