果 树 学 报 2023,40(1):155-168
Journal of Fruit Science

DOI:10.13925/j.cnki.gsxb.20210533

FTIR结合DD-SIMCA和二维相关光谱的核桃产地判别分析

王永波,李洪艳,张想芬,温卫华,杨 瑞*

(贵州大学林学院,贵阳 550025)

摘 要:【目的】建立一种基于傅里叶变换红外光谱(FTIR)的简便、快速的核桃产地鉴别方法。**【方法】**使用KBr压片法采集4类不同产地(新疆、贵州、四川、云南)的核桃仁(*n*=120)和核桃壳(*n*=80)的中红外(4000~400 cm^{-1})光谱信息,经S.G.平滑+MSC方法预处理后,进行化学计量学建模和2DCoS分析。**【结果】**各产地核桃仁中红外光谱的主要吸收峰在表征蛋白质的1649 cm^{-1}和1539 cm^{-1}处存在显著差异(VIP > 1.0)。SVM比PLS-DA模型取得了更好的识别效果,其对核桃壳和核桃仁样本的最佳产地判别正确率分别为100%和97%。DD-SIMCA法对核桃仁样本的判别灵敏度和特异性均达到100%,高于SIMCA的87%分类正确率。各产地核桃仁样本的二维同步谱图存在差异,表明该法可以用于核桃产地的分类判别。**【结论】**FTIR光谱结合DD-SIMCA化学计量学方法或2DCoS分析技术可以实现对核桃产地的高效识别。

关键词:核桃产地;傅里叶变换红外光谱;数据驱动型簇类独立软模式分类;二维相关光谱

中图分类号:S664.1 **文献标志码:**A **文章编号:**1009-9980(2023)01-0155-14

FTIR spectroscopy combined with DD-SIMCA and Two-dimensional correlation spectroscopy (2DCoS) for judgement about walnut origins

WANG Yongbo, LI Hongyan, ZHANG Xiangfen, WEN Weihua, YANG Rui*

(College of Forestry, Guizhou University, Guiyang 550025, Guizhou, China)

Abstract:【Objective】This work was to develop rapid and facile methods for the determination and classification of the walnut geographical origin based on kernel and shell samples by Fourier transform infrared spectroscopy (FTIR).【Methods】A total of 120 in-shelled walnut samples were collected from 4 areas of China, including Xinjiang, Guizhou, Sichuan and Yunnan. The walnuts were split into 120 kernel samples and 80 shell samples. After freeze-drying and oven-drying respectively, samples were numbered to maintain traceability throughout the process and were stored in a sealed bag for use. These samples were investigated under solid-state conditions in KBr pellets utilizing FTIR spectra in the range of 4000−400 cm^{-1}. Several mathematical pre-treatment methods including baseline correction, Savitsky-Golay (S.G.) smoothing, standard normal variate (SNV) and multiplicative scatter correction (MSC) were tested in the original spectral matrix. Then, the principal component analysis (PCA) was performed on the whole dataset for qualitative analysis of the spectra. Subsequently, four different classification models were applied to the FTIR data. Partial least squares discriminant analysis (PLS-DA), and support vector machine (SVM) were used to identify and categorize the spectra data from walnut kernel samples and shell samples. The class modeling techniques for soft independent modeling by class analogy (SIMCA) and data-driven soft independent modeling of class analogy (DD-SIMCA) were evaluated by using walnut kernel samples data. The DD-SIMCA was chosen because it was a powerful class-mod-

收稿日期:2022-03-17 **接受日期:**2022-08-19

基金项目:贵州省科技计划项目(黔科合支撑〔2020〕1Y011号);贵州省林业科研项目(黔林科合〔2022〕22号,黔林科合〔2020〕14号);贵州大学实验室开放项目

作者简介:王永波,男,实验师,硕士,主要从事经济林栽培生理研究。Tel:18285186282,E-mail:lukmb@163.com

*通信作者 Author for correspondence. Tel:13608507646, E-mail:yr553017@163.com

eling technique to test whether a sample was a target category or not, to assign an unknown sample to a class (target category) or not (outlier). The DD-SIMCA model was established on the PCA decomposition of each class separately. The efficiency of such chemometrics models was evaluated in terms of correct-prediction percent, sensitivity and specificity. Besides, the synchronous two-dimensional correlation spectrum (2DCoS) maps based on temperature perturbations of the walnut kernel were computed and analyzed in 1800~700 cm^{-1} wavenumber regions.【Results】The results showed that both walnut kernel and shell samples from different origins performed high similarities of the FTIR spectral profiles, especially in the position of the absorption peak. The resemblance of FTIR spectra was a result of its similar chemical composition. Separately, 14 characteristic absorption peaks were obtained from the walnut kernel average spectrum. As indicated by their FTIR spectra, their differences occurred in the region of the characteristic peaks 1750~1430 cm^{-1} and the fingerprint region (1240−600 cm^{-1}). In particular, using the variable importance in the projection (VIP) technique analyzed the peak value and provided that the main differences occurred in characterizing the absorption peaks of proteins (1649 cm^{-1} and 1539 cm^{-1}, VIP>1.0). It was difficult to classify them successfully only based on FTIR spectra. Therefore, a chemometric analysis must be conducted to detect these potential distinctions. Based on PCA and SIMCA classification performance comparison among these pretreatment methods, S.G. smoothing + MSC was superior to other methods. The samples can be clustered into four non-overlapping categories according to their collection areas. Also, the SIMCA results indicated the highest specificity and sensitivity, with a performance above 67%. In this work, all classification models used this method to pre-process spectra data. The recognition rate of the walnut kernel and shell prediction set was 73% (R^2_{cv}=0.78) and 100% (R^2_{cv}=0.95) by PLS-DA. Using the SVM pattern, 97% and 100% of prediction set samples were correctly classified in pre-treated spectra. The SIMCA model and DD-SIMCA model were selected for walnut kernel pre-treated spectra further analysis. The prediction ability of SIMCA resulted in 87% accuracy. The accurate and stable model (100% of sensitivity and 100% of specificity) was obtained with walnut kernel spectra data for the DD-SIMCA. In contrast to the four models, the DD-SIMCA model was a suitable one for walnut origin category classification. The synchronous 2DCoS constructed from the FT-IR changes in the temperature gradient range of 15−55 ℃ revealed that an autocorrelation spectrum intensity in the region of 1800−1600 cm^{-1} was obviously different. Manifestation in (1745 1745) cm^{-1} and (1650 1650) cm^{-1} showed a strong autocorrelation peak and the cross peak at (1655 1745) cm^{-1} and (1540 1745) cm^{-1}. Bands due to the combination of carboxylic acid (ester) carbonyl C = O stretching, carbonyl (amide I) C = O stretching, and protein N-H bending vibrations were varied. The 2DCoS acted out greater diversity than the one-dimensional spectrum, which can be classified.【Conclusion】Adopting a suitable spectral pre-processing (S.G. smoothing + MSC) method can significantly improve the classification and discriminatory correct-prediction percent of the model. Both walnut shells and kernels can be used to identify the origin of walnuts. The FTIR combined with DD-SIMCA and 2DCoS was potentially an ideal tool for rapidly and sensitively identifying the geographical origin of walnut samples.

Key words: Walnut origins; FTIR; DD-SIMCA; Two-dimensional correlation spectroscopy (2DCoS)

核桃(*Juglans regia* L.)是世界四大坚果之一，我国是其原产地，也是主要种植区[1]。广泛分布于新疆、云南、山东等21个省区。核桃极具经济和保健价值。现代医学研究认为，核桃仁中丰富的α-亚麻酸(ALA)、多酚、类黄酮、植物甾醇等抗氧化物质，具有很好的抗癌作用，并能有效对抗心血管疾病，对预防肥胖、糖尿病等也有很好的作用[2]。不同产地和品种的核桃营养成分存在差异，随着生活水平的

提高，人们越来越重视具有地理标志的优质核桃产品。因此对核桃产地溯源检测有实际意义。我国多以未去壳核桃销售，核桃壳（核桃内果皮）是核桃加工和消费的副产物，其含有丰富的木质素、纤维素和半纤维素以及酚酸类、黄酮类、苷类等多种活性物质，常被用于食品加工、化工、医学应用等[3-4]。核桃产地溯源研究材料大多是核桃仁，利用核桃壳进行的产地判别尚未见报道。

目前常用的产地溯源分析技术主要有同位素溯源、矿质元素指纹分析、电子鼻技术等，这些方法要求操作者有较丰富的专业知识，并且存在实验过程复杂、成本高、污染环境等缺点[5]。傅里叶变换红外光谱（Fourier transform infrared spectroscopy，FTIR）分析方法所需样品制备量少或无需样品制备，不使用危险的溶剂或试剂，具有检测步骤简便快速、特征吸收峰更易鉴别、实验重复性好等优点，已被广泛地用于化工、中药、食品等方面的定性和定量分析[6-8]。由于核桃是高度复杂的多成分系统，会导致红外振动信号的整合和重叠，这使得其直接解释困难。因此，笔者引入二维相关光谱（two-dimensional spectroscopy，2DCoS）分析技术[9-10]。与传统的FTIR技术相比，2DCoS可以放大红外光谱中微小的差异并揭示任何重叠的峰值来提高光谱分辨率[11]。

化学计量学工具是提取红外光谱有效信息的关键数学方法，该信息与地理特征结合可实现对样本产地的快速鉴别[5-7, 12]。典型的化学计量学数据分析包括无监督的主成分分析（principal component analysis，PCA）、层次聚类分析（hierarchical cluster analysis，HCA），以及有监督的分类方法，如线性判别分析（linear discriminant analysis，LDA）、偏最小二乘判别分析（partial least squares-discriminant analysis，PLS-DA）、*k*最近邻分类算法（*k*-nearest neighbors，*k*-NN）、支持向量机（support vector machines，SVM）等[13]。簇类独立软模式分类（SIMCA）是广泛用于化学计量学中的单分类器之一。它通过对每种类别进行无监督PCA建模分析，可以将样本关联到多个类或不关联任何类[14]。Dana等[15]的研究表明，SIMCA分类与机器学习（ML）的预测模型在拉曼光谱判别蜂蜜产地和品种结果存在相关性，且2种算法分类结果正确率基本相同。最近学者提出了基于SIMCA边界构建修改的数据驱动型簇类独立软模式分类（data driven soft independent modelling of class analogy，DD-SIMCA）。DD-SIMCA通过计算每个对象的得分距离和正交距离，从而建立2个公差阈值：给定显著性水平的接受区域和异常值区域。DD-SIMCA常用于样品的掺假检测[16-18]，鲜有用于样本产地识别。

笔者旨在通过选择正确的数据预处理和适当的FTIR光谱数据分析方法，为开发一种自动量化和识别核桃产地的快速、简便的方法提供依据。

1　材料和方法

1.1　样品处理

样品是从当地种植户和经认证的销售商处采购的当年产核桃干果，其中3种分别为来自中国国家地理标志产品保护产区的贵州赫章核桃、云南大理漾濞核桃和新疆阿克苏核桃，另一种为产自于四川凉山的大凉山核桃。采集的样品密封保存于室温条件下待用，测试前剥离其核桃仁和核桃壳作为实验材料，分别重复18~38个样本。具体信息见表1。

为进一步降低样品中水分对实验结果的影响，将核桃仁样品去除种皮，用粉碎机粉碎后置于真空冷冻干燥机，冻干36 h；将核桃壳样品置于干燥箱，105 ℃条件下烘干4 h后，用粉碎机粉碎并过200目筛。干燥后的样品分别编号装入密封袋，并置于真空干燥器中储存备用。

1.2　主要仪器与试剂

傅里叶变换红外光谱仪（Frontier，美国PerkinElmer公司）、真空冷冻干燥机（LGJ-10C，北京四环

表1　样本信息

Table 1　Information of samples

产地 Origin	品种 Variety	收集时间 Year of collection	核桃仁样本数 Samples of walnut kernels	核桃壳样本数 Samples of walnut shells
新疆阿克苏 Aks, Xinjiang	温185 Wen 185	2020年12月 Dec. 2020	27	20
云南大理 Dali, Yunnan	大泡核桃 Dapao walnut	2020年11月 Nov. 2020	38	20
贵州赫章 Hezhang, Guizhou	黔核7号 Qianhe 7	2020年11月 Nov. 2020	27	22
四川凉山 Liangshan, Sichuan	大凉山核桃 Daliangshan walnut	2020年11月 Nov. 2020	28	18

公司）、压片机（DF-4B，天津港东科技公司）配13 mm免脱模压片模具、分析天平（BSA2202S，德国Sartorius公司）、玛瑙研钵、KBr（光谱纯，国药集团）。

1.3 FTIR采集和光谱数据预处理

参考黄冬晨等[19]的方法并做适当改进，实验前将光谱纯KBr放入真空烘箱105 ℃条件下烘干12 h以除去水。按质量比1∶100称取样品与KBr，分多次等比例放入玛瑙研钵中混合均匀后，放入模具压成透明薄片。

傅里叶变换红外光谱仪预热30 min，设扫描范围：4000~400 cm^{-1}，扫描次数：4次，分辨率：4 cm^{-1}。保持样品仓内无样品，扣除空气中H_2O和CO_2的干扰，扫描获取FTIR中红外透射光谱数据（ASCII格式）。实验全程环境湿度控制在45%以下，温度为室温条件。

光谱采集过程中，由于仪器、样品和测量环境的变化，原始光谱中不可避免的存在噪声、基线偏移、谱峰重叠等干扰，这就需要对图谱数据进行预处理。光谱数据的预处理：采用基线校正（baseline correct）减小样品研磨不够细和压片不够透明因素造成的红外光散射影响。采用Savitsky-Golay（S.G.）5点平滑（smooth）处理降低光谱的噪声。采用一阶导数（first derivative，1stDer）提高原光谱中的吸收峰和肩峰的识别精度。标准正态变量变换（standard normal variate transformation，SNV）和多元散射校正（multiplicative scatter correction，MSC）被用来消除样品颗粒分布不均和颗粒大小产生的、表面散射以及光程变化对光谱的影响。

1.4 二维相关光谱（2DCoS）

2DCoS是样品在系统外部扰动下获得的一系列动态光谱和通过一种交叉相关数据分析所得到的光谱。2DCoS是通过将不同产地核桃仁样品片放入带有可编程加热夹套控制器（GS20730型；Specac，英国）的样品支架中获得的。温度范围为15~55 ℃，间隔为10 ℃，在加热速度为2 ℃·min^{-1}的情况下采集的动态吸光度谱。使用OriginPro2022（OriginLab®，美国）插件（2D Correlation Spectroscopy Analysis）对采集到的动态光谱的平均光谱进行二维相关分析。

根据Noda[20]的广义二维相关光谱理论，基于温度（T）外部扰动的动态FTIR光谱[$\tilde{y}(v, T)$]由实验设定温度条件下所测光谱[$\tilde{y}(v, T)$]计算所得：

$$\tilde{y}(v,\mathrm{T})=\begin{cases} y(v,\mathrm{T})-y(v) & for\ \mathrm{T}_{min}\le \mathrm{T}\le \mathrm{T}_{max} \\ 0 & otherwise \end{cases}。\quad (1)$$

式中$y(v)$代表参考光谱，T_{min}和T_{max}为扰动变量（本实验中代表温度）所设定的最小值和最大值。通常情况下，选择平均光谱或者扰动系列中的第1个或者最后1个光谱为参考光谱，$y(v)$=0也是可以的。本实验中选择T=15 ℃时采集的FT-IR光谱作为参考光谱，因此$y(v)$=$y(v, T_{min})$=$y(v, 15\ ℃)$。

2DCoS分为同步谱Φ（v_1,v_2）和异步谱Ψ（v_1, v_2），可通过下列式（2）和式（3）计算所得：

$$\Phi(v_1,v_2)=\frac{1}{n-1}\sum_{i=1}^{n}\tilde{y}(v_1,T_i)\times\tilde{y}(v_2,T_i);\quad (2)$$

$$\Psi(v_1,v_2)=\frac{1}{n-1}\sum_{i=1}^{n}\tilde{y}(v_1,T_i)\times\Big[\sum_{j=1}^{n}M_{i,j}\times\tilde{y}(v_2,T_i)\Big]。\quad (3)$$

式中n代表收集的不同温度条件下的动态光谱的数量，M_{ij}代表被称为“Hilbert－Noda”变换矩阵的第i行和第j列元素，定义为：

$$M_{ij}=\begin{cases} 0 & if\ i=j \\ \dfrac{1}{\pi(j-1)} & otherwise \end{cases}。\quad (4)$$

1.5 化学计量学分析工具

化学计量学算法可被用于评估、区分和判别核桃的类别。核桃仁和核桃壳的FTIR数据被收集在数据矩阵x中，向量y表示核桃的产地类别。矩阵x（FTIR光谱）数据信息复杂且不含有产地等效信息，故需要通过不同的数据处理才能获得正确结果。

主成分分析法（PCA）是一种无监督学习算法，常用于数据的分类和降维[21]。本工作中，PCA被应用于评估根据不同产地核桃仁和核桃壳FTIR光谱数据的样品聚类。此外，在建立分类模型前，可用PCA检测可能的异常值。

偏最小二乘判别分析（PLS-DA）是一种广泛应用于光谱分析的统计方法。它是基于偏最小二乘回归模型，对不同处理的样本进行训练，产生训练集，并检验训练集的可信度[22]。建模过程中，各产地类别定义为数值型变量1、2、3、4，设置判别阈值为0.5。

支持向量机（SVM）是一种机器学习算法。它定义了一个决策边界，该边界通过最大化2类之间的距离来优化划分2类，用于解决模式识别和回归问题[23]。

簇类独立软模式分类（SIMCA）是基于主成分分析的一种模式识别方法。它利用先验分类知识，对每一种类别建立1个PCA模型，通过F检验设定

的分类置信区间，利用建立的模型判断未知样本的归属[24-25]。

数据驱动型簇类独立软模式分类（DD-SIMCA）是对揭示极值和异常值等特殊类型的数据进行校正分析，从而优化目标样本可接受范围的SIMCA决策阈值[26]。DD-SIMCA主要算法步骤如下：

第1步是通过PCA对训练集数据矩阵$X(I\times J)$进行分解。

$$X=TP^{t}+E \text{ 。} \quad (5)$$

式中$T=\{t_{ia}\}$为得分矩阵（$I\times A$）；$P=\{P_{ja}\}$为载荷矩阵（$J\times A$）；E=$\{e_{ij}\}$为矩阵（$I\times J$）的残差；A为主成分数。

第2步是使用PCA分解结果计算训练集样本的得分距离（SD）h_i和正交距离（OD）v_i。

$$h_i=t_i^{t}(T^{t}T)^{-1}t_i=\sum_{a=1}^{A}\frac{t_{ia}^2}{\lambda_a}\text{，}v_i=\sum_{j=1}^{J}e_{ij}^2 \text{ 。} \quad (6)$$

式中λ_a，a=1，…，A是矩阵$T^{t}T$=A=diag(λ_1，...，λ_A)的对角线元素。

样本的总距离可通过下式计算：

$$c=N_h\frac{h}{h_0}+N_v\frac{v}{v_0}\alpha\chi^2(N_h+N_v) \text{ 。} \quad (7)$$

式中h_0和v_0是比例因子，N_h和N_v是自由度数量，这些参数是先验未知的，是通过参考文献[24]中数据驱动的方法估算的。

第3步是定义目标类别的接受范围和阈值区间，考虑到Ⅰ型错误（α），接受范围被定义为：

$$c\leqslant c_{\text{crit}}(\alpha)\text{，}c_{\text{crit}}=x^2(1\text{-}\alpha,N_h+N_v) \text{ 。} \quad (8)$$

经过以上3步后，可以通过给定的α值定义的分数和正交距离，使用构建的模型对未知样本进行分类。

对于所有的分类模型，进行交叉验证。根据模型的灵敏度（sensitivity）、特异性（specificity）、识别率、校正均方根误差（RMSEC）、预测均方根误差（RMSEP）、交叉验证均方根误差（RMSECV）、决定系数（R^2）和交叉验证决定系数（R_{cv}^2）对模型的性能进行评估和优化。

灵敏度和特异性常常用于类模型（SIMCA为代表）的评价。类建模中，当为1个目标类定义1个类空间，判断样本符合与否，取决于落在类边界之内还是之外。当属于被建模类的样本，如果正确落在边界内，则称为“真阳性”（*TP*）；如果落在边界外，则称为“假阴性”（*FN*）。以此类推，与该类无关的样本，如果在边界内发现，则为“假阳性”（*FP*），如果在边界外被正确发现，则为“真阴性”（*TN*）[27]。灵敏度被定义为属于建模类样本被模型正确分类的分数：

$$\text{灵敏度（\%）}=\frac{TP}{TP+FN}\times 100\text{。} \quad (9)$$

相反，特异性是指不属于建模类样本（非目标样本）被正确拒绝的比例：

$$\text{特异性（\%）}=\frac{TN}{TN+FP}\times 100\text{。} \quad (10)$$

一个好的分类器，灵敏度和特异性应接近100%。灵敏度可用于拟合（对被建模的训练集对象）和预测（未用于建模的验证集对象）两方面的模型评价。而特异性只用于预测评价。

在本工作中，光谱数据作图采用OriginPro2022（OriginLab®，美国）软件。光谱数据预处理采用Spectrum 10（FTIR仪器自配）软件和The Unscrambler X（CAMO，挪威）。另外The Unscrambler X和MATLAB 2020b（MathWorks，美国）用于光谱数据的建模分析。MATLAB的DD-SIMCA_GUI工具从https://github.com/yzontov/dd-simca获得[28]。

2　结果和分析

2.1　光谱数据预处理

分别采用原始数据（处理1）、基线校正+平滑（处理2）、1stDer（处理3）、S.G.平滑+SNV（处理4）和S.G.平滑+MSC（处理5）等预处理方法，以4类产地核桃仁样本为研究对象，根据PCA和SIMCA结果评价各预处理方法的效果。各预处理在5%显著水平的SIMCA分类结果见表2，其中处理5的灵敏度和特异性分别在67%~100%和86%~100%之间，明显高于其他处理（1~4）灵敏度的43%~75%和特异性的21%~85%。4个产地中新疆的灵敏度和特异性均最高，说明与其他产地的样本差异较大。对比不同预处理方法对PCA和SIMCA判别结果的影响，发现S.G.平滑+MSC方法可明显提升模型的识别率。

2.2　FTIR光谱分析

2.2.1　*核桃仁FTIR光谱*　将测定的4个产地核桃仁样品的中红外光谱数据分别计算平均光谱，经基线校正+S.G.平滑处理后观察平均光谱在全谱范围内的差异。如图1所示，不同产地核桃的平均光谱在总体模式上呈现出相似的趋势，但在指纹区的2个谱段（1750~1450 cm^{-1}和1330~1100 cm^{-1}）存在较大差异，这与不同官能团的振动模式有关。所得14个

表 2 不同预处理光谱核桃仁样本的 SIMCA 判别结果

Table 2 Classification results of FTIR spectra acquired on different pretreatment of samples %

处理 Treatment	云南 Yunnan		贵州 Guizhou		新疆 Xinjiang		四川 Sichuan	
	灵敏度 Sensitivity	特异性 Specificity	灵敏度 Sensitivity	特异性 Specificity	灵敏度 Sensitivity	特异性 Specificity	灵敏度 Sensitivity	特异性 Specificity
1	50	36	43	40	65	77	46	66
2	47	27	52	21	61	88	46	45
3	51	44	57	31	72	67	48	65
4	47	47	54	29	75	88	44	85
5	100	100	67	86	90	100	90	93

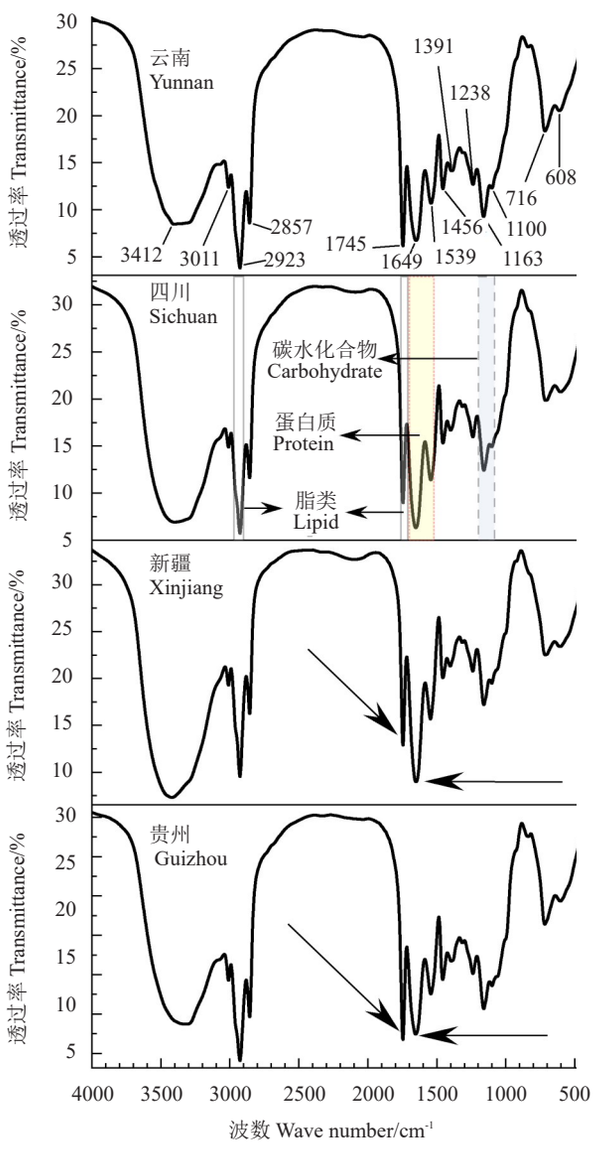


图 1 4 个产地核桃仁的平均光谱

Fig. 1 Average FTIR of the walnut kernel of four origins

主要特征吸收峰及基团归属如下：3412 cm^{-1}为液态H_2O伸缩或醇羟基O—H伸缩振动；2923 cm^{-1}和2857 cm^{-1}表现为脂肪族基团CH_2的反对称和对称伸缩振动；1745 cm^{-1}为饱和脂肪酸酯中的C＝O吸收峰；1649 cm^{-1}为酰胺Ⅰ带C＝O伸缩振动峰，氨基酸NH^{3+}振动峰，以及C—H键的伸缩振动；1539 cm^{-1}附近为酰胺Ⅱ带C—N、C＝C、N＝O的伸缩振动，以及N—H变形引起；1456 cm^{-1}和1391 cm^{-1}可能为饱和C—H和氨基酸COO吸收峰，或酰胺Ⅲ谱带；1238 cm^{-1}、1163 cm^{-1}、1100 cm^{-1}处为C—O、C—N吸收峰以及脂肪族SO_2吸收峰；716 cm^{-1}和608 cm^{-1}为S—O伸缩振动和酰胺O＝C—N吸收峰[29-30]。其中2923 cm^{-1}、2857 cm^{-1}、1745 cm^{-1}、1649 cm^{-1}、1456 cm^{-1}、1163 cm^{-1}和716 cm^{-1}处为强吸收峰，这些谱带的分配显示出核桃仁中含有丰富的蛋白质、酯类脂肪酸和碳水化合物等物质[31]。对比4个产地核桃仁的中红外光谱，箭头标示位置的表征蛋白和脂类的2个吸收峰[32]的吸收强度和相对强度存在明显差异。进一步通过变量投影重要性（variable importance in the projection VIP）算法对3100~700 cm^{-1}波数范围内主要吸收峰进行分析，筛选出统计学差异（VIP > 1.0）的吸收峰为1649 cm^{-1}和1539 cm^{-1}。这表明不同产地核桃的蛋白质和脂肪结构组成及相对含量可能存在差异。

2.2.2 核桃壳FTIR光谱 各产地核桃壳粉的FTIR透过率光谱如图2所示，3400 cm^{-1}附近的吸收峰为O—H的伸缩振动峰；2930 cm^{-1}处的特征峰与CH_2和CH_3中存在的C—H拉伸振动有关[4]；1739 cm^{-1}附近是纤维素中C＝O伸缩振动峰[33]；1610 cm^{-1}和1517 cm^{-1}处与木质素芳香环的C＝C拉伸有关，这是木质素最特征的红外吸收带[34]；1444 cm^{-1}附近的吸收峰或与CH_3和CH_2中C—H的弯曲振动有关；酚类、醇类和醚类的C—O拉伸出现在1249 cm^{-1}和1050 cm^{-1}处[35]。对比各产地核桃壳红外光谱，贵州产地的1800~1300 cm^{-1}范围光谱特征与其他3地（云南、新疆、四川）区别明显，表征为木质素和纤维素的组成和结构差异。

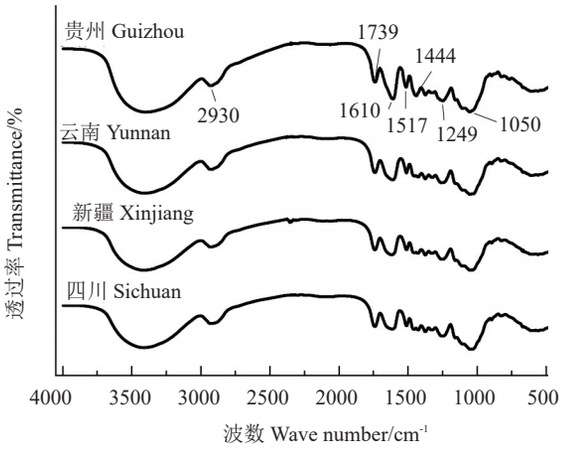


图2　4个产地核桃壳的红外光谱

Fig. 2　FTIR spectra of walnut shell

由图1和图2可知，4个产地的核桃仁和核桃壳的红外光谱，虽然存在差异，但是整体上相似，直观上并不能通过红外光谱中特异的吸收峰等特征将其区分开来，需要进行化学计量学或2DoS分析。

2.3　主成分分析（PCA）

利用经S.G.平滑+MSC预处理后核桃仁和核桃壳的全光谱数据进行PCA分析，三维得分图见图3-a~b，可以看出，不同产地的样本可明显聚为4类。图3-a中，前3个主成分PC1、PC2和PC3分别解释了66.2%、20.8%和7.7%的方差，累计贡献率为93.7%。云南产地核桃仁样本聚集更为集中且与其他3地样本的空间距离更远，说明云南核桃的化学成分组成或含量与其他3地差异较大。新疆产地样本内部聚集相对分散，编号为xj16的观测数据点处于95%置信区间外，判定其为异常值。由图3-b可知，前3主成分累计方差贡献率为91.6%，除贵州产地的核桃壳样本单独聚为一类外，其他3地样本的置信椭圆彼此间有重叠，这与核桃仁的PCA表现存在差异。

2.4　PLS-DA和SVM建模及判别

将核桃仁（n=120）和核桃壳（n=80）样本数据集分别按3∶1随机分为校正集和预测集。校正集用于

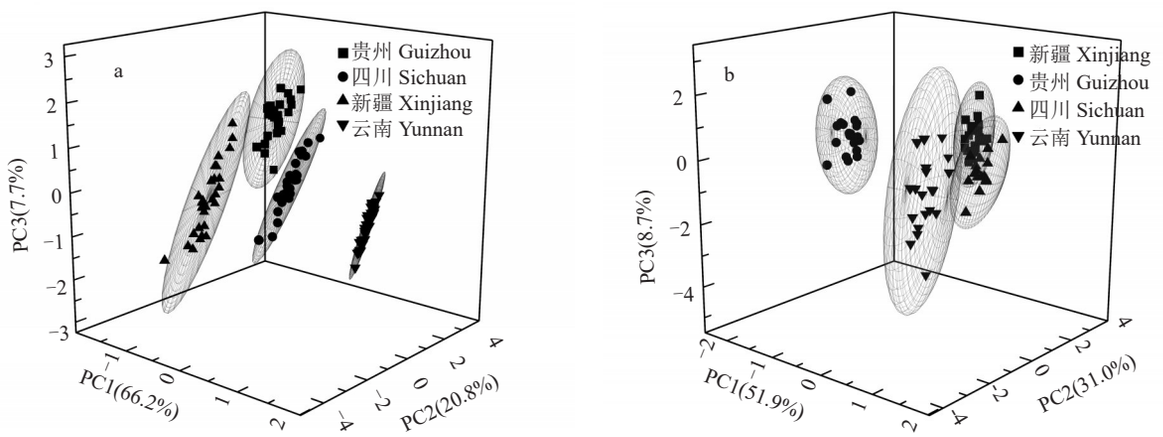


图3　不同产地的核桃仁（a）和核桃壳（b）主成分得分图

Fig. 3　3D scores PCA plot of the walnut kernel (a) and walnut shell (b)

训练和优化模型，预测集用来预测模型的泛化能力（即预测性能）评价。利用The Unscrambler X软件，通过偏最小二乘（PLSR）和支持向量机（SVM）方法，将产地信息与光谱信息分别进行数学建模分析。

2.4.1　偏最小二乘判别分析（PLS-DA）　PLS-DA作为一种常用的有监督的降维、判别分析方法，运行中一个重要的参数选择就是主成分数。本研究采用完全交叉验证的验证策略，选择最佳主成分数为5。结果表明，核桃仁校正集决定系数R^2=0.80、验证集R^2_{cv}=0.78，校正均方根误差RMSEC=0.52、交叉验证均方根误差RMSECV=0.63。校正集和验证集的决定系数相差较小，说明基于PLSR建立的核桃仁产地预测模型没有产生过拟合现象。同时较小的决定系数值，较大的均方根误差值，表明所建立的预测模型的预测效果会差强人意。另外，核桃壳样本校正集R^2=0.97、验证集R^2_{cv}=0.95，校正均方根误差RMSEC=0.19、交叉验证均方根误差RMSECV=0.26。它们对未知样本（预测集）的预测结果分别见图4-a~b，核桃仁和核桃壳样本预测集识别正确率分别为73%和100%。由此可见，通过核桃壳样本的PLS-DA模型比核桃仁样本建立的更稳健。

2.4.2　支持向量机（SVM）　SVM是最流行的用于样本通用分类和识别的机器学习模型之一。该模型对非线性样本具有良好的分离效果。本实验中选择

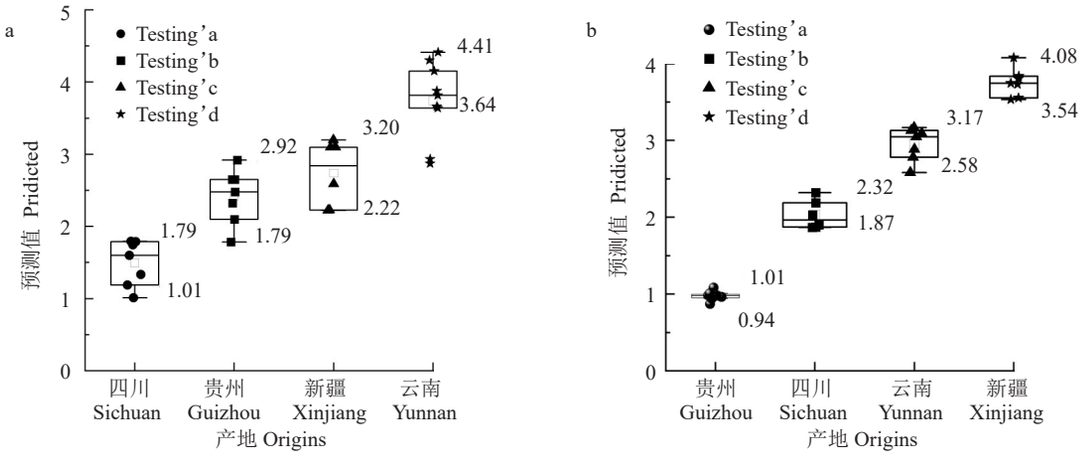


图4 基于PLS-DA的核桃仁（a）和核桃壳（b）的预测集箱点图

Fig. 4 Prediction results for the walnut kernel (a) and walnut shell (b) prediction set by PLS-DA

SVM类型：nu-SVC，Kernel类型：径向基核函数（radial basis function），Nu值：0.5，进行SVM建模和判别分析。结果表明，SVM模型在核桃仁样本校正集和验证集的识别正确率分别为100%和96%，预测集的识别正确率97%。在核桃壳样本校正集、验证集和预测集识别正确率均为100%。

比较核桃仁和核桃壳样本的PLS-DA和SVM判别结果，核桃壳样本在2种分类器中得到更高的判别正确率，说明其或更适合作为实验材料进行核桃产地鉴别。SVM比PLS-DA对核桃仁样本的判别精度提高较大。为探究核桃仁的分类潜力，对其做进一步的分析。

2.5 SIMCA和DD-SIMCA判别分析

SIMCA和DD-SIMCA属于单类分类器，该分类器用于将一个特定的目标类别对象从所有其他对象类别中区分出来。按2∶1将核桃仁样本数据集随机划分为校正集（n = 80）和独立的预测集（n = 40）。为了防止过拟合，模型中最佳主成分数分别采用杠杆率校正（leverage correction）和留一法交叉验证（leave-one-out CV）确定。

2.5.1 SIMCA判别分析 选择最佳主成分数4（sc），6（gz），5（xj）和7（yn），利用校正集分别建立各产地的PCA模型，分类结果见图5库曼图（Cooman's plot）。该图表示每个样本到模型（sc vs xj）的马氏距离，坐标轴为样品到模型的距离，坐标是计算出来的距离值（样品标准差）[36]。采用软分类标准的SIMCA，允许将样本同时分配到所调查类别中的任何一类。即当1个样本在2个类中的标准化距离都小于1时，它可以被分配到2个类中。这代表了重叠类模型的情况。此外，当1个样本的标准化距离在这个类中大于1，在另一个类中小于1时，它只能被分配到1个类。相应地，当1个样本对2个类的标准化距离都大于1时，它就被分配到2个类中的任何一个，并被视为异常值。由图5可知，各样本基本都能按照各自的组别成功分类，只有样本sc24被同时分配到四川（sc）和贵州（gz）产地。此外，样本gz8，xj16，xj14与所在样本组的距离较大。在5%显著性水平下，各产地的识别率分别是新疆（xj）93%、云南（yn）100%、四川（sc）87%和贵州（gz）79%，4类产地总体样本的测试集识别正确率为87%。

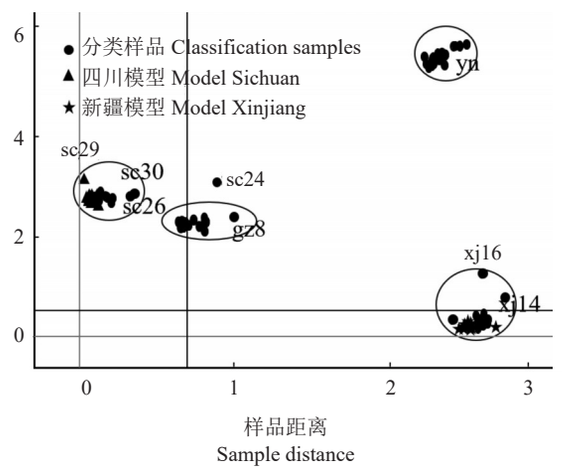


图5 各产地核桃仁样本的库曼图

Fig. 5 Cooman's plot for sc *vs* xj

图6为各产地样本与四川样本的相对类间距离。类间距描述了类之间的相似性或差异性，如果两类模型间距离大于3，则它们被区分为不同模型，且距离越大，分类越准确[37]。核桃仁样本的四川与贵州产地PCA模型间距离最小为38.38，而四川与

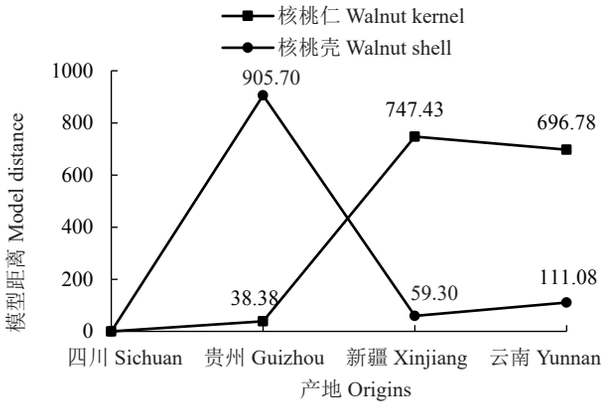


图6　各产地核桃仁样本相对四川产地的PCA模型距离

Fig. 6　PCA model distance of walnut kernel samples to Sichuan model by origin

新疆、云南产地模型的距离较远，分别为747.43和696.78。说明SIMCA方法能够成功进行核桃产地分类，其中四川与贵州两产地的核桃仁样本光谱数据比较相似，而与云南和新疆两产地的差异性较大。对比核桃壳样本各类与四川产地的类间距离，最小为59.30，最大为905.70，验证了核桃壳比核桃仁分类结果更准确。这或许与所选择不同产地品种间核桃壳差异更大有关。

2.5.2　DD-SIMCA判别分析　将随机选取的校正集和预测集样本按产地类别各分为4类，即校正集：贵州（gz；n = 19）、四川（sc；n = 20）、云南（yn；n = 26）、新疆（xj；n = 18）和预测集：g（n = 8）、s（n = 8）、y（n = 12）、x（n = 9）。在进行类建模之前，分别对总样本的校正集和预测集进行异常值检测。图7-a为校正集的DD-SIMCA接受度图（绿线内），红线为目标组样品的阈值线。两线与坐标轴组成的区域将样品分为3类：规则数据（绿点）、极值（黄点）和异常值（红点）。由图7-a可知，校正集中两极值点为xj14和sc24。图7-b中所有的校正集数据都在公差范围内（以垂直线表示）。表明基于DD-SIMCA模型的校正集不包含异常值，最终校正集和预测集中的xj14、xj16和sc24被判定为极值，这与SIMCA的检测结果基本一致。

将各产地数据分别建立DD-SIMCA模型，设参数为：校正集数据采用中心化（Centering）预处理，主成分数3，接受区域类型选择卡方分布（chi-square），并使用经典的估计方法（α = 0.01；γ = 0.01）。值得注意的是，建模过程中某些参数的设置（如主成分数过高）会导致模型不稳定，因此主成分数是通过校正集对模型反复训练，选出预测结果最好时对应的数据。

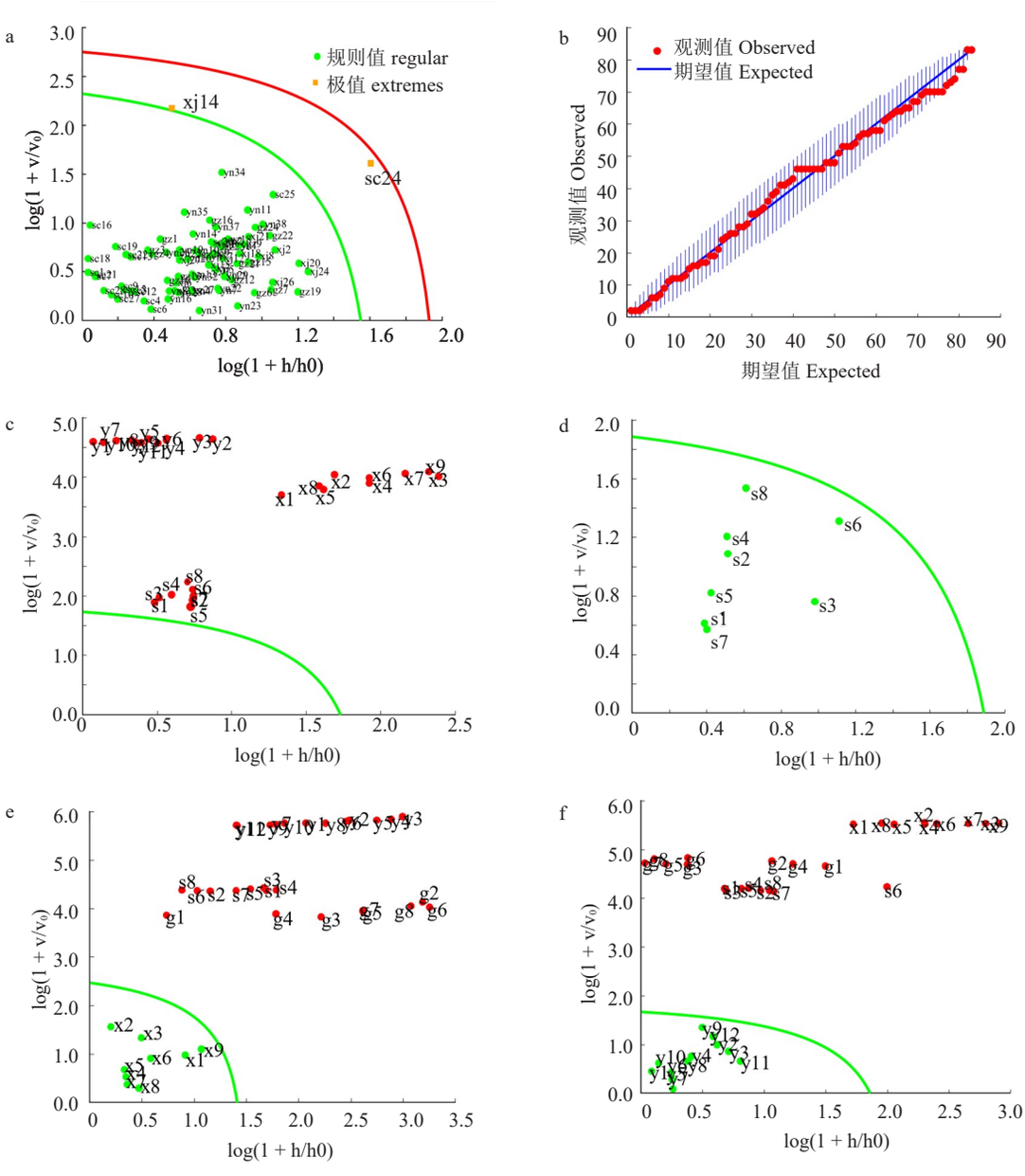
在验证阶段，使用新数据（校正集）对模型进行评估。为了测试模型对目标类或者含有外部样本的新数据集的接受和拒绝能力，分别用预测集中非目标类、目标类和所有类数据进行验证，结果见图7-c~e。图7-c和d分别表明贵州产地模型特异性与四川产地的灵敏度均为100%，图7-e和f表明云南和新疆产地模型对目标类样本和非目标类样本的特异性和灵敏度都达到100%。

2.6　二维相关红外光谱分析

比较4类产地核桃仁红外光谱在1800~700 cm^{-1}波数范围内的二维同步谱，结果如图8所示。同步2DCoS图是对角线对称的，位于对角线上的相关峰被称为自相关峰（总为正值），是由扰动引起的动态波动的自相关强度引起的。非对角线上的峰为交叉峰（正或负），是由分子间或分子内相互作用产生的官能团相对性变化的结果[38]。

由图8可知，各产地的二维同步谱图存在明显的差异，主要表现在1750~1710 cm^{-1}的脂肪酸酯、1700~1520 cm^{-1}蛋白质、1200~1080 cm^{-1}碳水化合物吸收谱带的变化。例如，图8-b和d中四川和云南产地样本2DCoS图谱在Φ（1745 1745）和Φ（1650 1650）处存在明显的强自相关峰，而贵州的Φ（1745 1745）和新疆Φ（1650 1650）处均表现为弱自相关峰（图8-a和c）。1200~1080 cm^{-1}范围自相关峰，表现为图8（a，b）和（c，d）差异较大，后者的强度明显更高。在Φ（1050 1050）处各产地均发生自相关峰，但在一维FIIR图谱中并不容易分辨，或因其在一维图谱中表现为重叠峰或肩峰，而1050 cm^{-1}归属于脂肪族酸酐（酯）的C—O伸缩振动。另外，图8（a-d）分别存在特异自相关峰，发生在Φ（1576 1576）、Φ（950 950）、Φ（1720 1720）、Φ（886 886）处。

图8-b中四川产地样本的交叉峰（1655 1745）cm^{-1}和（1540 1745）cm^{-1}为负相关与其他产地存在明显不同。它们是由羧酸（酯）C＝O伸缩振动、酰胺Ⅰ的C＝O伸缩振动和蛋白质N—H变角振动产生的。综合这些特征，可以发现各产地核桃仁样品2DCoS图谱存在明显差异，其中四川与新疆产地差异最大，这与SIMCA的类间距表现结果一致。因此可以依据2DCoS图谱进行核桃产地分类识别。



a. 校正集样本的接受度图；b. 校正集异常值验证图；c. gz 目标类模型对预测集中非目标类的接受度图；d. sc 对预测集中目标类 s 的接受度图；e & f. xj 和 yn 模型对预测集的接受度图。

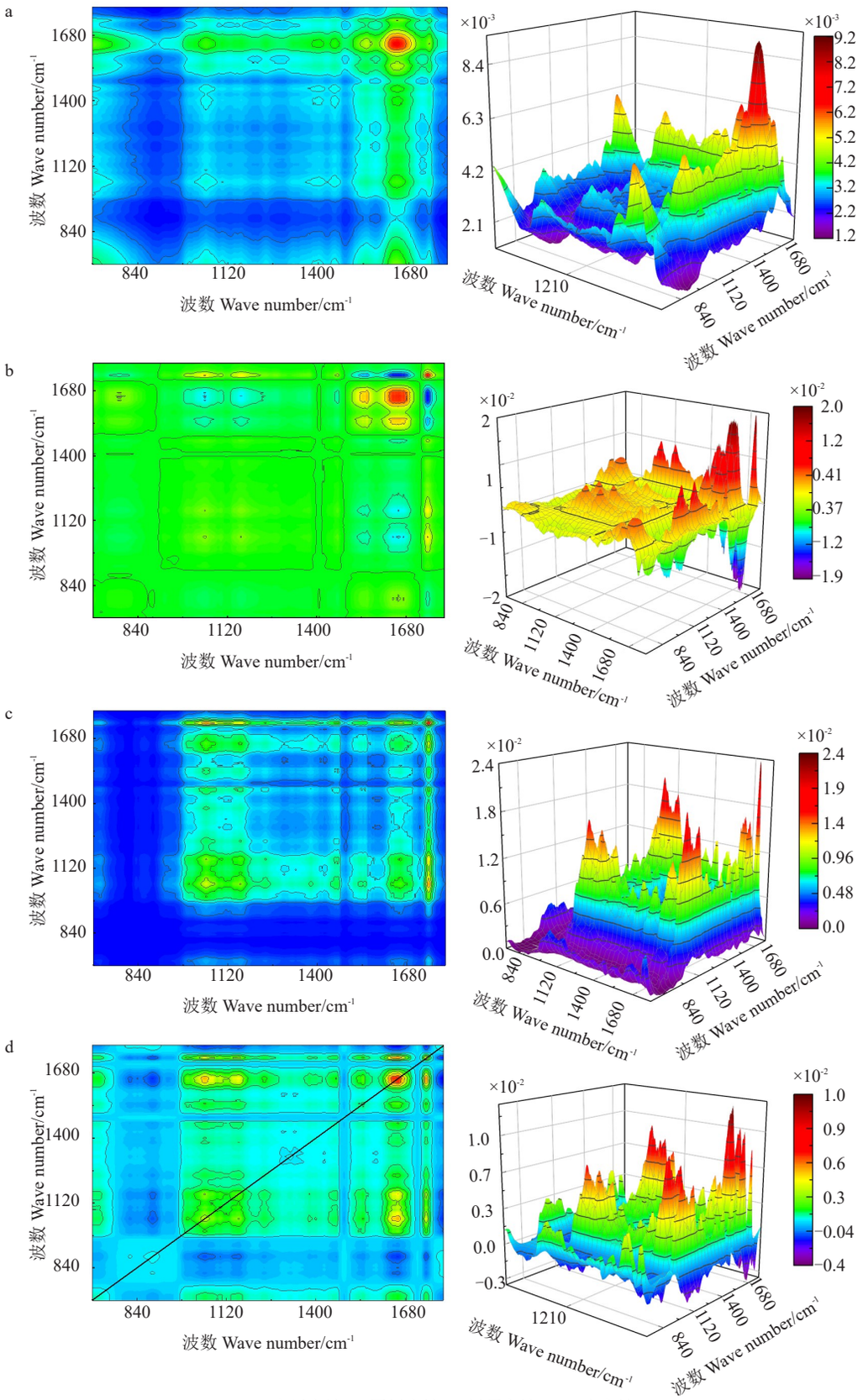
a. Acceptance area of samples by DD-SIMCA model in the training process; b. Extreme plot for the training set; c. Acceptance plot for gz samples (target class) *vs* s, x and y samples; d. sc samples *vs* s samples; e. xj samples *vs* test set; f. yn samples *vs* test set.

图 7 基于 DD-SIMCA 模型的核桃产地识别结果

Fig. 7 The application of DD-SIMCA for the classification of walnut kernel samples from gz, sc, yn, and xj

3 讨 论

试验结果表明，利用核桃仁样本可以实现基于红外光谱核桃的区域识别，这与前人[12,39-40]的研究结果一致。另外利用核桃壳样本同样取得了很好识别效果，这与 Nogales-bueno 等[41]的利用近红外光谱实现对带壳核桃的品种识别的结果类似，证明核桃壳材料可用于红外光谱的定性识别。值得强调的是，影响核桃品质的因素包括不同产地的气候环境、地理条件、物候期及栽培管理措施和品种之间的差异。为取得高的材料变异性，笔者实验条件是严格选取 4 个不同物候期的核桃。实际工作中，可选取尽可能多的产地和主栽品种的核桃作为研究样本，从而建立更稳定、更准确的核桃产地判别模型。



a. 贵州产地；b. 四川产地；c. 新疆产地；d. 云南产地。

a. Guizhou; b. Sichuan; c. Xinjiang; d. Yunnan.

图 8　基于温度扰动的核桃仁样本 1800~700 cm^{-1} 波动范围的二维同步谱

Fig. 8　Temperature-dependent synchronous 2D correlation maps of the walnut kernel samples in the 1800−700 cm^{-1} wavenumber range

各分类器对核桃产地的判别效果：DD-SIMCA ＞ SVM ＞ SIMCA ＞ PLS-DA。分析PLS-DA对核桃壳数据表现良好而对核桃仁结果一般的情况，其原因或许是该分类器对类别数据不平衡或全光谱的冗余数据比较敏感。下一步可以优化信息区间，利用PCA降维数据或特征波段选择，进一步发掘模型的精度和效率。Müller等[42]使用siPLS算法将光谱分为20个区间和3个区间的组合，取得了明显优于使用所有变量的PLS最佳模型。本实验中不同分类模型都检测出极值或者异常值，它们会影响模型的预测精度和稳健性，因此可以选择剔除异常值的数据建模。

2DCoS的特点是能够有效地突出光谱细微的差异而提高光谱分辨率，事实上它也会放大数据集的噪声，而噪声会产生多余的交叉峰而严重干扰异步谱分析[43]。从光谱数据中去除非系统性噪声是后续分析前的一个重要预处理步骤。2DCoS光谱会受到材料、扰动类型和环境条件等的影响。为了尽量减少高温引起材料中蛋白质等的结构变化而影响光谱结果，本实验中温度梯度设置为15~55 ℃，连续动态光谱个数为5，或存在梯度设置不够充分从而影响二维相关光谱的表征。后续研究中可选择其他较稳定的扰动类型，以及采用机器学习算法对二维相关光谱图进行分类分析，以获得准确性和可重复性更高的结果。

4 结 论

（1）采取S.G.平滑+MSC的光谱预处理方法，能够显著地提升模型的分类和判别正确率。

（2）核桃壳和核桃仁都可以用于核桃产地的鉴别。基于核桃壳样本的PLS-DA和SVM模型判别正确率达100%，好于基于核桃仁的识别结果。

（3）核桃仁样本的FTIR光谱与DD-SIMCA结合，建立的鉴定模型能将目标类和其他产地样本分开，校正集和预测集的灵敏度和特异性均达到100%。

综上，FTIR光谱结合SVM和DD-SIMCA化学计量学方法或2DCoS分析技术可以实现对核桃产地的高效识别。

参考文献 References：

[1] TOMER V，KUMAR A，GUPTA K，SHUKLA S，RAFIQ S. Walnut[M]//NAYIK G A，GULL A. Antioxidants in vegetables and nuts - properties and health benefits. Singapore：Springer Singapore，2020.

[2] 李俊南，习学良，熊新武，杨建华，张雨思 . 核桃的营养保健功能及功能成分研究进展[J]. 中国食物与营养，2018，24(5)：60-64.
LI Junnan，XI Xueliang，XIONG Xinwu，YANG Jianhua，ZHANG Yusi. Nutritional function and functional components of walnut[J]. Food and Nutrition in China，2018，24(5)：60-64.

[3] 郑志锋 . 核桃壳树脂化基础研究[D]. 哈尔滨：东北林业大学，2006.
ZHENG Zhifeng. The basic study on resinification of walnut shell[D]. Harbin：Northeast Forestry University，2006.

[4] HSINI A，NACIRI Y，LAABD M，EL OUARDI M，AJMAL Z，LAKHMIRI R，BOUKHERROUB R，ALBOURINE A. Synthesis and characterization of arginine- doped polyaniline/walnut shell hybrid composite with superior clean- up ability for chromium (VI) from aqueous media：Equilibrium，reusability and process optimization[J]. Journal of Molecular Liquids，2020，316：113832.

[5] 梁北辰 . 基于FTIR的大米成分测量与产地溯源技术研究[D]. 哈尔滨：哈尔滨工业大学，2019.
LIANG Beichen. Research on rice composition measurement and origin traceability technology based on FTIR[D]. Harbin：Harbin Institute of Technology，2019.

[6] 高媛 . 基于FTIR光谱的宁夏中宁枸杞产地鉴别及品质快速检测[D]. 杨凌：西北农林科技大学，2016.
GAO Yuan. A rapid classification and qualification analysis of Zhongning wolfberry using FT- IR spectroscopy[D]. Yangling：Northwest A & F University，2016.

[7] 邹小波，封韬，郑开逸，石吉勇，黄晓玮，孙悦 . 利用近红外及中红外融合技术对小麦产地和烘干程度的同时鉴别[J]. 光谱学与光谱分析，2019，39(5)：1445-1450.
ZOU Xiaobo，FENG Tao，ZHENG Kaiyi，SHI Jiyong，HUANG Xiaowei，SUN Yue. Simultaneous identification of wheat origin and drying degree using near- infrared and mid- infrared fusion techniques[J]. Spectroscopy and Spectral Analysis，2019，39(5)：1445-1450.

[8] KANAKIS C D，PETRAKIS E A，KIMBARIS A C，PAPPAS C，TARANTILIS P A，POLISSIOU M G. Classification of Greek *Mentha pulegium* L. (Pennyroyal) samples，according to geographical location by fourier transform infrared spectroscopy[J]. Phytochemical Analysis，2012，23(1)：34-43.

[9] NODA I. Generalized two-dimensional correlation method applicable to infrared，raman，and other types of spectroscopy[J]. Applied Spectroscopy，1993，47(9)：1329-1336.

[10] NODA I. 2DCOS and I.：Three decades of two-dimensional correlation spectroscopy[J]. Journal of Molecular Structure，2016，1124：3-7.

[11] AZLAH M A F，CHUA L S，ABDULLAH F I，YAM M F. A fast and reliable 2D- IR spectroscopic technique for herbal leaves classification[J]. Vibrational Spectroscopy，2020，106：103014.

[12] 何勇，郑启帅，张初，岑海燕 . 基于中红外光谱和化学计量学算法鉴别核桃产地及品种[J]. 光谱学与光谱分析，2019，39(9)：2812-2817.

HE Yong，ZHENG Qishuai，ZHANG Chu，CEN Haiyan. Identification of walnut origins and varieties with mid-infrared spectroscopy analysis technique[J]. Spectroscopy and Spectral Analysis，2019，39(9)：2812-2817.

[13] BRENDEL R，SCHWOLOW S，GERHARDT N，SCHWAB J，RAU P，OEST M，ROHN S，WELLER P. MIR spectroscopy versus MALDI-ToF-MS for authenticity control of honeys from different botanical origins based on soft independent modelling by class analogy (SIMCA)：A clash of techniques?[J]. Spectrochimica Acta Part A-Molecular and Biomolecular Spectroscopy，2021，263：120225.

[14] DUCA D，MANCINI M，ROSSINI G，MENGARELLI C，PEDRETTI E F，TOSCANO G，PIZZI A. Soft independent modelling of class analogy applied to infrared spectroscopy for rapid discrimination between hardwood and softwood[J]. Energy，2016，117：251-258.

[15] MAGDAS D A，GUYON F，BERGHIAN-GROSAN C，MULLER MOLNAR C. Challenges and a step forward in honey classification based on Raman spectroscopy[J]. Food Control，2021，123：107769.

[16] GERÔNIMO D M，OLIVEIRA S C D，SOARES F L F，PERALTA-ZAMORA P，NAGATA N. Determination of main raw material source in bar soaps using mid-infrared spectroscopy combined with classification tools[J]. Microchemical Journal，2021，164：106029.

[17] PINTO F G，MAHMUD I，RUBIO V Y，MÁQUINA A D V，FURTADO DURANS A F，NETO W B，GARRETT T J. Data-driven soft independent modeling of class analogy in paper spray ionization mass spectrometry-based metabolomics for rapid detection of prostate cancer[J]. Analytical Chemistry，2022，94(4)：1925-1931.

[18] SALIH M A M. 牛奶掺假的近红外光谱分析技术研究[D]. 北京：中国农业科学院，2018.
SALIH M A M. Study on detection of adulteration in raw milk by near-infrared spectrometers[D]. Beijing：Chinese Academy of Agricultural Sciences，2018.

[19] 黄冬晨，杨天鸣，付海燕，楚霞，吴海龙 . 基于 PLSDA 对中红外透射法压片过程中溴化钾与样品质量配比的研究[J]. 计算机与应用化学，2013，30(3)：241-245.
HUANG Dongchen，YANG Tianming，FU Haiyan，CHU Xia，WU Hailong. The application of PLSDA to investigate the choice of the tabletting process in the mid-infrared transmission method of KBr and the ratio of the quality of the sample[J]. Computers and Applied Chemistry，2013，30(3)：241-245.

[20] NODA I. Two-dimensional correlation analysis of unevenly spaced spectral data[J]. Applied Spectroscopy，2003，57(8)：1049-1051.

[21] 林海明，杜子芳 . 主成分分析综合评价应该注意的问题[J]. 统计研究，2013，30(8)：25-31.
LIN Haiming，DU Zifang. Some problems in comprehensive evaluation in the principal component analysis[J]. Statistical Research，2013，30(8)：25-31.

[22] LEE L C，LIONG C，JEMAIN A A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data：A review of contemporary practice strategies and knowledge gaps[J]. Analyst，2018，143(15)：3526-3539.

[23] 苑江浩，常青，赵会义，唐芳 . 基于 SVM 的粮食霉变预测分类方法研究[J]. 中国粮油学报，2021，36(9)：138-144.
YUAN Jianghao，CHANG Qing，ZHAO Huiyi，TANG Fang. Research on prediction and classification method of rain mildew based on SVM[J]. Journal of the Chinese Cereals and Oils Association，2021，36(9)：138-144.

[24] POMERANTSEV A L，RODIONOVA O Y. Concept and role of extreme objects in PCA/SIMCA[J]. Journal of Chemometrics，2014，28(5)：429-438.

[25] 郝勇，孙旭东，高荣杰，潘媛媛，刘燕德 . 基于可见/近红外光谱与 SIMCA 和 PLS-DA 的脐橙品种识别[J]. 农业工程学报，2010，26(12)：373-377.
HAO Yong，SUN Xudong，GAO Rongjie，PAN Yuanyuan，LIU Yande. Application of visible and near-infrared spectroscopy to identification of navel orange varieties using SIMCA and PLS-DA methods[J]. Transactions of the Chinese Society of Agricultural Engineering，2010，26(12)：373-377.

[26] NEVES M D G，POPPI R J. Authentication and identification of adulterants in virgin coconut oil using ATR/FTIR in tandem with DD-SIMCA one class modeling[J]. Talanta，2020，219：121338.

[27] OLIVERI P，DOWNEY G. Multivariate class modeling for the verification of food-authenticity claims[J]. Trends in Analytical Chemistry，2012，35：74-86.

[28] ZONTOV Y V，RODIONOVA O Y，KUCHERYAYSKIY S V，POMERANTSEV A L. DD-SIMCA - A MATLAB GUI tool for data driven SIMCA approach[J]. Chemometrics and Intelligent Laboratory Systems，2017，167：23-28.

[29] COSTA N，AMARAL S，ALVIM R，NOGUEIRA M，SCHWANNINGER M，RODRIGUES J. Assessment of resin formulations and determination of the formaldehyde to urea molar ratio by near- and mid-infrared spectroscopy and multivariate data analysis[J]. Journal of Applied Polymer Science，2013，128(1)：498-508.

[30] YU W J，SHI C L，LIU Y M，DAI M，SHI X M. Sublethal thermal treatment of vibrio parahaemolyticus and its determination by micro. fourier transform infrared spectroscopy[J]. Chinese Journal of Analytical Chemistry，2013，41(10)：1470-1476.

[31] MECOZZI M，PIETROLETTI M. Chemical composition and surfactant characteristics of marine foams investigated by means of UV-vis，FTIR and FTNIR spectroscopy[J]. Environmental Science and Pollution Research，2016，23(22)：22418-22432.

[32] 杨静怡，陶兴月，夏玉芳，谢钊俊，丁晓霞 . 不同贮藏条件下核桃种子的红外光谱特征分析[J]. 果树学报，2015，32(5)：950-958.

YANG Jingyi，TAO Xingyue，XIA Yufang，XIE Zhaojun，DING Xiaoxia. Infrared spectra of walnut seed under different storage conditions[J]. Journal of Fruit Science，2015，32(5)：950-958.

[33] 胡爱华，邢世岩，巩其亮 . 基于 FTIR 的针阔叶材木质素和纤维素特性[J]. 东北林业大学学报，2009，37(9)：79-81.
HU Aihua，XING Shiyan，GONG Qiliang. Lignin and cellulose characteristics of different woods based on FTIR[J]. Journal of Northeast Forestry University，2009，37(9)：79-81.

[34] MAHESHWARAN M V，HYNESS N R J，SENTHAMARAIKANNAN P，SARAVANAKUMAR S S，SANJAY M R. Characterization of natural cellulosic fiber from *Epipremnum aureum* stem[J]. Journal of Natural Fibers，2018，15(6)：789-798.

[35] CAO J S，LIN J X，FANG F，ZHANG M T，HU Z R. A new absorbent by modifying walnut shell for the removal of anionic dye：kinetic and thermodynamic studies[J]. Bioresource Technology，2014，163：199-205.

[36] HASHIM D M，MAN Y B C，NORAKASHA R，SHUHAIMI M，SALMAH Y，SYAHARIZA Z A. Potential use of fourier transform infrared spectroscopy for differentiation of bovine and porcine gelatins[J]. Food Chemistry，2010，118(3)：856-860.

[37] BALAN B，DHAULANIYA A S，JAMWAL R，YADAV A，KELLY S，CANNAVAN A，SINGH D K. Rapid detection and quantification of sucrose adulteration in cow milk using attenuated total reflectance-fourier transform infrared spectroscopy coupled with multivariate analysis[J]. Spectrochimica Acta Part A：Molecular and Biomolecular Spectroscopy，2020，240：118628.

[38] LIU Z M，YANG S B，WANG Y Z，ZHANG J Y. Discrimination of the fruits of amomum tsao-ko according to geographical origin by 2DCOS image with RGB and resnet image analysis techniques[J]. Microchemical Journal，2021，169：106545.

[39] 焦俊，圣阳，马鑫，李绍辰，滕燕，李春，蒋玲 . 基于近红外和远红外光谱信息融合的核桃品种鉴别方法[J]. 农业工程，2021，11(12)：35-41.
JIAO Jun，SHENG Yang，MA Xin，LI Shaochen，TENG Yan，LI Chun，JIANG Ling. Identification of walnut cultivars based on near-infrared and far-infrared spectral information fusion[J]. Agricultural Engineering，2021，11(12)：35-41.

[40] GU X Z，ZHANG L，LI L T，MA N，TU K，SONG L J，PAN L Q. Multisource fingerprinting for region identification of walnuts in Xinjiang combined with chemometrics[J]. Journal of Food Process Engineering，2018，41(4)：e12687.

[41] NOGALES-BUENO J，FELIZ L，BACA-BOCANEGRA B，HERNÁNDEZ-HIERRO J M，HEREDIA F J，BARROSO J M，RATO A E. Comparative study on the use of three different near infrared spectroscopy recording methodologies for varietal discrimination of walnuts[J]. Talanta，2020，206：120189.

[42] MÜLLER A L H，PICOLOTO R S，MELLO P D A，FERRÃO M F，DOS SANTOS M D F P，GUIMARÃES R C L，MÜLLER E I，FLORES E M M. Total sulfur determination in residues of crude oil distillation using FT-IR/ATR and variable selection methods[J]. Spectrochimica Acta Part A：Molecular and Biomolecular Spectroscopy，2012，89：82-87.

[43] PARK Y，NODA I，JUNG Y M. Smooth factor analysis (SFA) to effectively remove high levels of noise from spectral data sets[J]. Applied Spectroscopy，2018，72(5):765-775.