



李跑,谭惠珍,谢叔娥,等. 基于近红外光谱技术有监督模式识别的青皮产地溯源分析[J]. 轻工学报, 2024, 39(2): 54-59.

LI P, TAN H Z, XIE S E, et al. Traceability analysis of *Pericarpium Citri Reticulatae Viride* origin based on near infrared spectroscopy technology and supervised pattern recognition[J]. Journal of Light Industry, 2024, 39(2): 54-59. DOI: 10. 12187/2024. 02. 007

基于近红外光谱技术有监督模式识别的青皮产地溯源分析

李跑^{1,2,3}, 谭惠珍¹, 谢叔娥⁴, 苏光林¹, 董怡青¹, 唐辉³

1. 湖南农业大学 食品科学技术学院, 湖南 长沙 410128;
2. 湖南省农业科学院 湖南省农产品加工研究所, 湖南 长沙 410125;
3. 广东省粤北食药资源利用与保护重点实验室, 广东 韶关 512005;
4. 湖南省产商品质量监督检验研究院, 湖南 长沙 410007

摘要: 利用便携式近红外光谱仪采集不同产地(安徽、广东、四川)青皮外壁和内囊光谱数据, 采用单一预处理和组合预处理方法消除光谱中的多种干扰, 结合主成分分析(PCA)、簇类独立软模式分类法(SIMCA)及Fisher 线性判别分析(FLDA)等模式识别方法建立青皮产地溯源模型。结果表明, 光谱预处理可以在一定程度上消除基线漂移、背景噪声和谱峰重叠干扰, 但无法实现产地溯源。3 种模式识别方法中, PCA 无法实现青皮产地溯源; 青皮外壁和内囊原始光谱的 SIMCA 模型获得的青皮产地溯源整体鉴别率分别为 99.14% 和 98.28%; FLDA 模型获得的整体鉴别率均为 99.57%, 优于 SIMCA 模型; 经光谱预处理优化后的 SIMCA 和 FLDA 模型对青皮产地溯源的鉴别率均可达 100%, 即便便携式近红外光谱技术结合有监督模式识别方法可实现青皮产地溯源的无损分析, 可为食药同源物质产地溯源拓展新途径。

关键词: 青皮; 溯源分析; 近红外光谱技术; 有监督模式识别方法

中图分类号: TS218; O561.3 **文献标识码:** A **文章编号:** 2096-1553(2024)02-0054-06

0 引言

青皮为芸香科植物橘的干燥幼果(个青皮)或未成熟果实果皮(四花青皮)的统称, 主要成分为挥发油、黄酮类化合物、氨基酸等, 其药性苦、辛、温, 归肝、胆、胃经, 能够梳理肝胆之气, 是收录于保健食品名单的中药材^[1-4]。青皮主产于湖南、安徽、四川、广东、广西等地, 由于环境、气候等外界因素不同, 不

同产地的青皮主要成分存在差异。通常, 科研人员一般采用目测法以颜色来判断青皮的产地等, 或者借助同位素质谱、薄层色谱、代谢组学、指纹图谱等传统分析手段, 利用青皮的理化结构和化学成分差异对其进行产地溯源^[5-8]。然而, 目测法存在较多不可控的主观因素, 传统分析手段会对样品造成一定的破坏性, 且耗时较长。因此, 亟需开发一种快速、无损的青皮产地溯源分析方法^[9]。

收稿日期: 2023-03-21; **修回日期:** 2023-05-16; **出版日期:** 2024-04-15

基金项目: 国家自然科学基金项目(31601551); 湖南省自然科学基金项目(2023JJ30290); 中国博士后科学基金面上项目(2019M650187); 湖南省教育厅科学研究项目(21A0127); 2022 年湖南省研究生科研创新项目(QL20220173)

作者简介: 李跑(1989—), 男, 湖南省宁乡市人, 湖南农业大学副教授, 博士, 主要研究方向为食品分析与化学计量学。E-mail: lipao@live.cn

近红外光谱技术具有无损、快速、效率高、检测简便等突出特点,已广泛应用在石化、烟草、医药、食品等领域^[10-14]。近年来,常用近红外技术对食药产品进行真伪鉴别或者产地溯源^[15],但由于仪器本身、环境、样品特性等因素的影响,采集的光谱常出现噪声干扰、基线漂移、谱峰重叠等现象^[16-17]。为此,有研究者通过引入化学计量学方法,如去偏移(De-bias)、去趋势(Detrend, DT)、连续小波变换(Continuous Wavelet Transform, CWT)等^[18-20]进行预处理,以消除光谱中的干扰,结合模式识别方法实现不同来源样品的鉴别分析。目前近红外光谱技术应用于青皮产地溯源的研究还很少,如涂蓉玲等^[21]建立了青皮近红外光谱一致性模型,实现了青皮的真伪鉴别,但该方法需将青皮磨成粉末再进行测定,无法实现无损分析。鉴于此,本研究拟利用便携式近红外光谱仪采集不同产地青皮外壁和内囊近红外光谱数据,采用单一预处理和组合预处理方法消除光谱中的多种干扰,并结合模式识别法建立青皮产地溯源模型,以期对青皮等食药产品产地溯源分析提供新思路。

1 材料与方法

1.1 主要材料与仪器

主要材料:安徽青皮 256 份,安徽六安市丹贝尔生物科技有限公司;四川青皮 197 份,安国药源商贸有限公司;广东青皮 242 份,福曦堂药店旗舰店。青皮样品总计 695 份,将其置于干燥、室温环境 24 h,并邀请湖南农业大学食品科学技术学院相关专家对实验样品进行类别鉴定和验证。

主要仪器:i-Spec Plus 光栅型便携式近红外光谱仪(波数范围为 11 100~5900 cm^{-1} ,附有积分球漫反射附件),必达泰克光电科技(上海)有限公司。

数据处理软件:Matlab R2010b,美国 MathWorks 公司。

1.2 光谱采集

本次实验在室温下操作,将青皮置于光斑上,确保青皮完全覆盖光斑,每个样本扫描 3 次,取 3 次光谱平均值作为青皮样品的原始光谱。此外,采集了青皮样品的外壁与内囊光谱数据,分别为广东青皮

内囊、广东青皮外壁、四川青皮内囊、四川青皮外壁、安徽青皮内囊、安徽青皮外壁。

1.3 数据分析与处理

利用 Matlab R2010b 软件进行数据处理。采用 8 种单一预处理方法(De-bias、DT、标准正态变量变换(Standard Normal Variate Transformation, SNV)、多元散射校正(Multiplicative Scatter Correction, MSC)、最大最小归一化(Maximum and Minimum Normalization, Min-Max)、一阶导数(First Derivative, 1st)、二阶导数(Second Derivative, 2nd)和 CWT)和 5 种组合预处理方法(1st+DT、1st+SNV、1st+MSC、CWT+SNV 和 CWT+MSC)消除光谱中的多种干扰。为实现不同产地青皮的准确无损鉴别分析,利用 Kennard-Stone 方法将 695 个青皮原始光谱数据按约 2:1 的比例分为训练集(463 个数据)和预测集(232 个数据),采用 PCA、SIMCA 与 FLDA 模式识别方法构建青皮产地无损溯源模型。

2 结果与讨论

2.1 不同产地青皮的光谱分析

图 1 为不同产地青皮外壁和内囊的原始光谱图。由图 1 可知,不同产地青皮的外壁和内囊原始光谱具有相似的变化趋势,均在 7000 cm^{-1} 处出现明显的吸收峰,属于 R—OH 和 R—NH₂ 的近红外区主要吸收带,可能是青皮中含有的挥发油与氨基酸成分所致。此外,不同产地青皮的外壁原始光谱在 10 300 cm^{-1} 、8100 cm^{-1} 处的吸收峰强于对应内囊原始光谱,这两处吸收峰分别由 O—H 键的伸缩振动

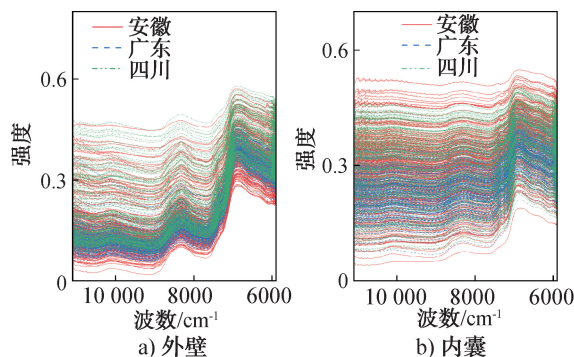


图 1 不同产地青皮外壁和内囊的原始光谱图
Fig. 1 Original spectra of the outer wall and inner capsule of *Pericarpium Citri Reticulatae Viride* from different regions

二级倍频、C—H 键的伸缩振动二级倍频引起。然而,不论外壁或内囊,不同产地的青皮原始光谱之间无明显差异,单纯使用原始光谱无法实现不同产地青皮的鉴别。此外,原始光谱中存在明显的基线漂移、背景及谱峰重叠干扰,需要借助光谱预处理方法消除这些干扰,以保证鉴别的准确性。

CWT 是一种基于小波分析的信号处理方法,也是现阶段最常用的消除原始光谱中干扰因素的预处理方法之一。图 2 为经 CWT 预处理后不同产地青皮的光谱图。由图 2 可知,相较原始光谱图,CWT 能有效消除光谱中背景噪声、基线漂移等干扰。经 CWT 预处理后,青皮外壁、内囊光谱的重叠谱峰得到了有效区分,在 $10\ 800\text{ cm}^{-1}$ 、 $10\ 300\text{ cm}^{-1}$ 、 9800 cm^{-1} 、 8600 cm^{-1} 、 8100 cm^{-1} 、 7750 cm^{-1} 、 7300 cm^{-1} 、 6650 cm^{-1} 、 6200 cm^{-1} 处存在明显的波峰或波谷。在部分波峰和波谷处,安徽、广东、四川青皮的光谱虽已呈现细微差异,但仍难以区分,如图 2a) 中 $11\ 000\sim 9100\text{ cm}^{-1}$ 波段处。此外,其他 7 种单一预处理同样较难找到不同产地青皮光谱之间的差别。总之,仅通过单一预处理的光谱无法实现青皮产地的准确溯源。

进一步采用组合预处理方法对原始光谱进行优化,考查其是否优于单一预处理方法。图 3 为经 CWT+SNV 组合预处理的不同产地青皮光谱图。由图 3 可知,CWT+SNV 组合预处理方法同样能消除基线漂移及背景干扰。然而,与 CWT 单一预处理方法相比,消除效果差异不明显,仅光谱强度有一定的提升,波峰或波谷仍出现在 $10\ 800\text{ cm}^{-1}$ 、 $10\ 300\text{ cm}^{-1}$ 、

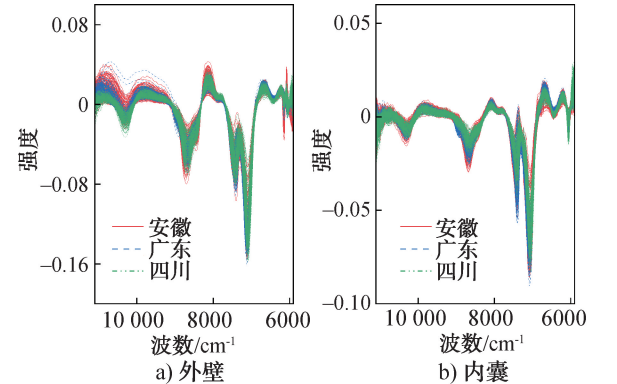


图 2 经 CWT 预处理后不同产地青皮的光谱图
Fig. 2 Spectra of *Pericarpium Citri Reticulatae Viride* from different regions after CWT pretreatment

9800 cm^{-1} 、 8600 cm^{-1} 、 8100 cm^{-1} 、 7750 cm^{-1} 、 7300 cm^{-1} 、 6650 cm^{-1} 、 6200 cm^{-1} 处,且未出现新的波峰或波谷信息。此外,其他组合预处理同样较难找到不同产地青皮光谱之间的差别。因此,经组合预处理后的光谱仍无法实现青皮产地的准确溯源。

2.2 基于 PCA 的青皮产地溯源分析

PCA 是现阶段最常用的无监督模式识别方法之一。图 4 为不同产地青皮原始光谱的 PCA 结果。由图 4 可知,青皮外壁和内囊的置信区间均存在严重的重叠现象,无法区别青皮的产地。因此,采用 PCA 结合单一预处理方法(CWT)对原始光谱进行优化,结果见图 5。由图 5 可知,与原始光谱相比,CWT 提高了 PCA 的鉴别能力,不同产地青皮的置信区间得到一定的分离,但仍有较高的重叠。此外,研究过程中发现 PCA 结合其他预处理方法仍无法

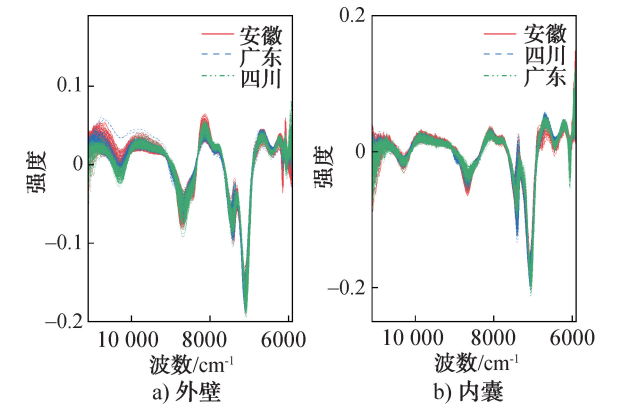


图 3 经 CWT+SNV 组合预处理后不同产地青皮的光谱图
Fig. 3 Spectra of *Pericarpium Citri Reticulatae Viride* from different regions after CWT+SNV combined pretreatment

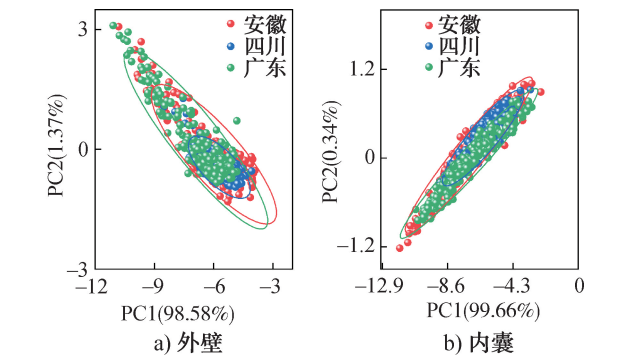


图 4 不同产地青皮原始光谱的 PCA 结果
Fig. 4 PCA results of *Pericarpium Citri Reticulatae Viride* from different regions with the original spectra

实现青皮产地的准确溯源。

2.3 基于 SIMCA 的青皮产地溯源分析

SIMCA 和 FLD 是现阶段比较常见的两种有监督模式识别方法。二者原理不同,前者通过对每一类样品分别进行 PCA 建模以实现未知来源样品的鉴别;后者在保证类内差异最小且类间差异最大的前提下将数据进行降维以实现未知样品的鉴别。图 6 为不同产地青皮光谱经预处理后的 SIMCA 结果。由图 6a)可知,青皮外壁光谱结合 SIMCA 模型的鉴别结果如下:1)青皮外壁原始光谱的整体鉴别率达到了 99.14%,安徽、广东、四川的鉴别率分别为 98.84%、98.77%、100%;2)采用合适的预处理方法消除光谱背景、噪声、基线漂移等干扰因素,可有效

提高鉴别率,如经 DT 单一预处理、1st+MSC 及 1st+SNV 组合预处理的青皮产地鉴别率均达 100%;3)不合适的预处理方法可能导致鉴别率降低,如经 Min-Max 预处理后,安徽青皮的鉴别率降为 90.70%,总鉴别率降为 96.57%。由图 6b)可知,青皮内囊光谱结合 SIMCA 模型的鉴别结果略逊于外壁:1)青皮内囊原始光谱的整体鉴别率为 98.28%,其中安徽、广东、四川青皮的鉴别率分别为 100%、96.30%、98.48%;2)经 1st+DT 组合预处理后实现了青皮产地的准确鉴别,鉴别率达 100%。因此,SIMCA 模型结合合适的光谱预处理方法可以实现青皮产地的准确溯源。

2.4 基于 FLDA 的青皮产地溯源分析

图 7 为不同产地青皮光谱经预处理后的 FLDA 结果。由图 7 可知,FLDA 模型结合青皮外壁、内囊原始光谱获得的整体鉴别率均达 99.57%;FLDA 模型对广东青皮的鉴别结果优于 SIMCA 模型,其原因可能是 FLDA 利用先验知识,通过数据在低维度上的投影实现了最佳的鉴别分析,而 SIMCA 是在 PCA 基础上建立起来的,无法准确实现类间差异较小的鉴别。青皮外壁和内囊的 FLDA 模型均可实现广东青皮的准确鉴别;FLDA 模型对安徽青皮内囊的鉴别结果优于 SIMCA 模型,除 1st 预处理外,FLDA 模型对安徽青皮内囊的鉴别率均为 100%,安徽青皮外壁在不同预处理条件下的 FLDA 和 SIMCA 模型鉴别结果各有优劣;四川青皮的鉴别结果略差,但鉴

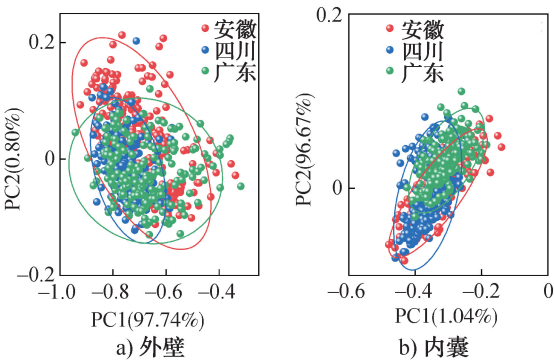


图 5 不同产地青皮光谱经 CWT 预处理的 PCA 结果

Fig. 5 PCA results of *Pericarpium Citri Reticulatae Viride* from different regions after CWT pretreatment

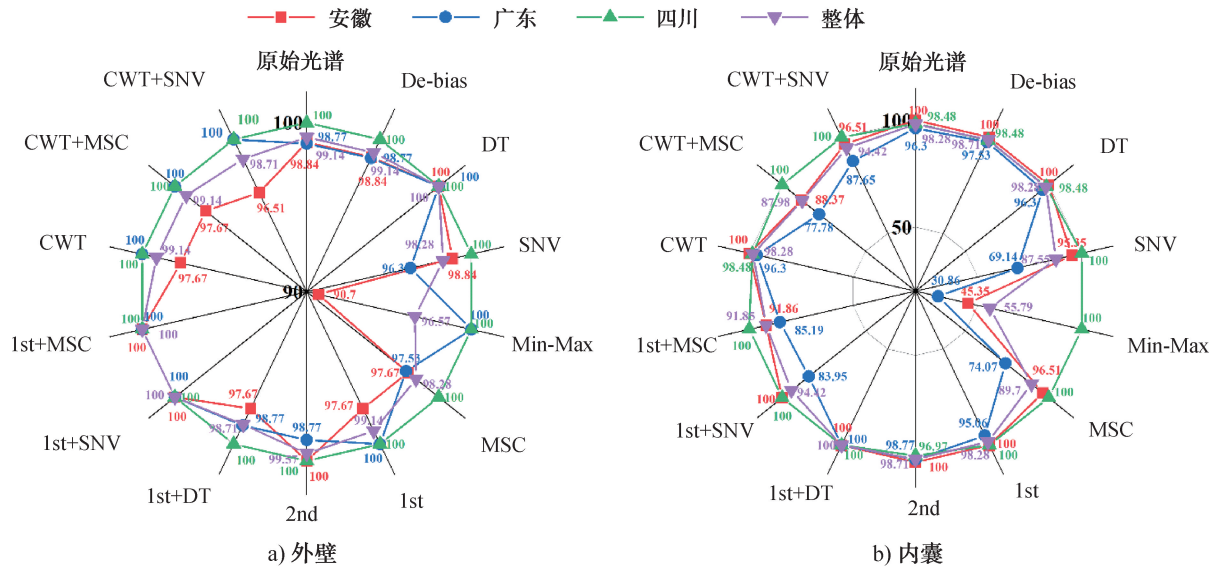


图 6 不同产地青皮光谱经预处理后的 SIMCA 结果

Fig. 6 SIMCA results of *Pericarpium Citri Reticulatae Viride* from different origins with pretreatment methods

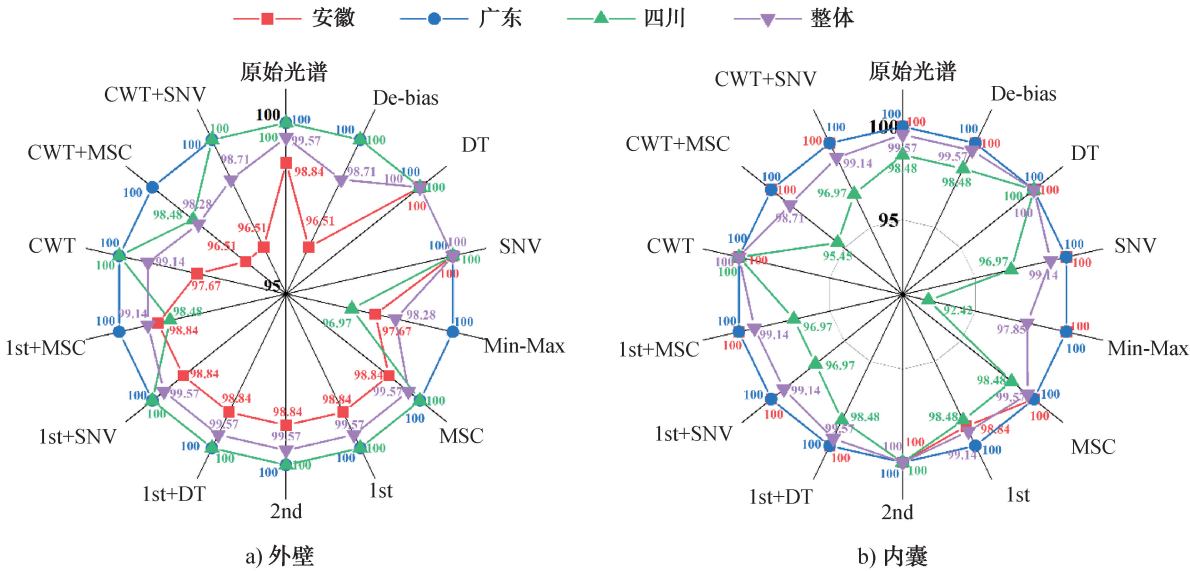


Fig. 7 FLDA results of *Pericarpium Citri Reticulatae Viride* from different origins with pretreatment methods

别率也在 92.42% 以上;结合合适的预处理方法,FLDA 模型也可实现不同产地青皮的准确鉴别,如结合 DT、SNV 预处理方法,FLDA 模型对青皮外壁的鉴别率为 100%,结合 DT、2nd、CWT 预处理方法,FLDA 模型对青皮内囊的鉴别率同样可达 100%。由此可知,FLDA 模型结合合适的光谱预处理方法可以实现不同产地青皮的准确鉴别。

3 结论

本文利用便携式近红外光谱仪采集不同产地(安徽、广东、四川)青皮外壁和内囊光谱数据,通过单一预处理和组合预处理方法消除光谱中存在的基线漂移、背景噪声和谱峰重叠干扰,结合 PCA、SIMCA、FLDA 模式识别法分别建立青皮产地溯源模型并进行对比分析,得出如下结论:无监督的 PCA 模式识别法结合光谱预处理无法实现青皮产地的准确溯源,而经光谱预处理优化后的有监督 SIMCA 和 FLDA 模型对青皮产地溯源的总鉴别率均可达 100%,即便便携式近红外光谱技术结合有监督模式识别方法可以实现青皮产地的溯源分析。本研究为食药同源产品产地溯源拓展了新途径,有效保障了食药同源物质的“道地性”和“特色性”。

参考文献:

[1] 何民友,叶瑞平,段志文,等. 青皮及其炮制品 UPLC 指

纹图谱与多成分含量测定研究[J]. 中药材,2022,45(10):2418-2424.

[2] 刘龙焯,王星宇,李林楠,等. 青皮与陈皮化学成分及分析方法研究进展[J]. 中国中药杂志,2022,47(11):2866-2879.

[3] 任桂林,柯刚,郑琴,等. 介孔二氧化硅 350FCP 固化陈皮、青皮挥发油研究[J]. 中草药,2021,52(5):1323-1334.

[4] WANG Y, CUI W L, YANG C G, et al. Comparison of Gepingpi and Sihuaqingpi based on ultra-high-performance liquid chromatography-tandem mass spectrometry combined with multivariate statistics, network pharmacology analysis, and molecular docking[J]. Journal of Separation Science, 2022, 45:4079-4098.

[5] 余代鑫,郭盛,杨健,等. 稳定同位素技术在中药材产地溯源中的应用与展望[J]. 中国中药杂志,2022,47(4):862-871.

[6] 牛露露,刘磊,李瑶瑶,等. 柑橘属同源药对青皮-枳壳的薄层色谱鉴别[J]. 中成药,2022,44(2):667-671.

[7] 高薇,王秀,李见春,等. 基于 HPLC 指纹图谱与多元统计分析的太平猴魁产地判别方法[J]. 轻工学报,2022,37(1):33-39.

[8] 靳贝贝,裴香萍,梁惠珍. 青皮药材的 HPLC 指纹图谱建立及聚类分析和主成分分析[J]. 中国药房,2018,29(24):3336-3339.

[9] KANG K S, OLIKAGU C, LEE T, et al. Sulfenyl chlorides: An alternative monomer feedstock from elemental sulfur for polymer synthesis[J]. Journal of the American Chemical Society, 2022, 144:23044-23052.

[10] 杨建云,田孟玉,裴梓烨,等. 基于近红外光谱技术的初烤烟叶致香成分含量特性快速定量评价[J]. 轻工学报,2018,33(5):60-68.

[11] YUAN L B, MENG X R, XIN K H, et al. A comparative

- study on classification of edible vegetable oils by infrared, near infrared and fluorescence spectroscopy combined with chemometrics[J]. *Spectrochimica Acta Part A (Molecular and Biomolecular Spectroscopy)*, 2023, 288: 122120.
- [12] PHILLIPS V Z, CANOY R J, PAIK S H, et al. Functional near-infrared spectroscopy as a personalized digital healthcare tool for brain monitoring[J]. *Journal of Clinical Neurology (Seoul, Korea)*, 2023, 19(2): 115–124.
- [13] 帖金鑫, 何文苗, 李石头, 等. 基于级联分类的复烤片烟产地预测方法研究[J]. *轻工学报*, 2022, 37(4): 94–99.
- [14] OHASHI A, NISHIO T, SAITO A, et al. Baseline drift vector of multiple points on body surface using a near-infrared camera[J]. *Physical and Engineering Sciences in Medicine*, 2022, 45(1): 143–155.
- [15] 吴木兰, 宋萧萧, 崔武卫, 等. 基于近红外光谱技术对南阳豌豆产地鉴别研究[J]. *光谱学与光谱分析*, 2023, 43(4): 1095–1102.
- [16] BIAN X H, WANG K Y, TAN E, et al. A selective ensemble preprocessing strategy for near-infrared spectral quantitative analysis of complex samples [J]. *Chemometrics and Intelligent Laboratory Systems*, 2020, 197: 103916.
- [17] HAN L, SUN Y, WANG Y, et al. Ultra-high resolution near-infrared spectrum by wavelet packet transform revealing the hydrogen bond interactions [J]. *Spectrochimica Acta Part A (Molecular and Biomolecular Spectroscopy)*, 2023, 289: 122233.
- [18] WANG Y Q, LIU G M, HU L P, et al. Prediction of anthocyanidins content in purple Chinese cabbage based on visible/near infrared spectroscopy[J]. *Foods*, 2023, 12(9): 1922.
- [19] HONG Y S, CHEN Y Y, CHEN S C, et al. Improving spectral estimation of soil inorganic carbon in urban and suburban areas by coupling continuous wavelet transform with geographical stratification[J]. *Geoderma*, 2023, 430: 116284.
- [20] HAN X, TAN Z, HUANG Z X, et al. Nondestructive detection of tri-closan in antibacterial hand soaps using digitally labelled Raman spectroscopy [J]. *Analytical Methods*, 2017, 9(24): 3720–3726.
- [21] 涂蓉玲, 王建伟. 近红外光谱法青皮药材真伪鉴别研究[J]. *中国民族民间医药*, 2014, 23(24): 15–16.

Traceability analysis of *Pericarpium Citri Reticulatae Viride* origin based on near infrared spectroscopy technology and supervised pattern recognition

LI Pao^{1,2,3}, TAN Huizhen¹, XIE Shue⁴, SU Guanglin¹, DONG Yiqing¹, TANG Hui³

1. College of Food Science and Technology, Hunan Agricultural University, Changsha 410128, China;

2. Hunan Agricultural Product Processing Institute, Hunan Academy of Agricultural Sciences, Changsha 410125, China;

3. Guangdong Provincial Key Laboratory of Utilization and Conservation of Food and Medicinal Resources in Northern Region, Shaoguan 512005, China;

4. Hunan Provincial Institute of Quality Supervision and Inspection of Product, Changsha 410007, China

Abstract: Spectral data of the outer wall and inner capsule of *Pericarpium Citri Reticulatae Viride* from different regions (Anhui, Guangdong, and Sichuan) were collected with portable near infrared spectrometer. Multiple interferences in the spectra were eliminated using single and combined pretreatment methods. Pattern recognition methods such as principal component analysis (PCA), cluster independent soft pattern classification (SIMCA) and Fisher linear discriminant analysis (FLDA) were used to establish traceability models of *Pericarpium Citri Reticulatae Viride* origin. The results showed that spectral pretreatment could eliminate the interferences of baseline drift, background, and spectral peak overlap to a certain extent. However, the traceability analysis of the origin couldn't be achieved. Among the three pattern recognition methods, accurate traceability analysis of *Pericarpium Citri Reticulatae Virid* couldn't be achieved with PCA method. The whole identification rate of SIMCA model with the outer wall and inner capsule original spectra were 99.14% and 98.28%, respectively. The whole identification rates of FLDA models were both 99.57%, better than that of SIMCA models. The identification rates of SIMCA and FLDA models with appropriate spectral pretreatment were 100%. Therefore, portable near infrared spectroscopy technology combined with the supervised pattern recognition methods can achieve non-destructive traceability analysis of *Pericarpium Citri Reticulatae Viride* origin, expanding a new way for the traceability analysis of food and drug homologous substances origin.

Key words: *Pericarpium Citri Reticulatae Viride*; traceability analysis; near infrared spectroscopy technology; supervised pattern recognition