

## 近红外光谱的不同产地柑橘无损鉴别方法

张欣欣<sup>1</sup>, 李尚科<sup>1</sup>, 李 跑<sup>1, 2\*</sup>, 单 杨<sup>2</sup>, 蒋立文<sup>1</sup>, 刘 霞<sup>1</sup>

1. 湖南农业大学食品科学技术学院食品科学与生物技术湖南省重点实验室,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

**摘要** 柑橘是世界第一大水果。不同产地的柑橘内部品质和价格有所不同,但其外观差别较小,外行人较难通过肉眼实现准确鉴别分析。DNA 标记法与仪器分析操作复杂、成本较高,且对样品具有破坏性,无法实现快速无损分析,影响了产品的二次销售。近红外光谱技术是一种快速无损的新型检测手段,可以用于不同产地农产品的鉴别分析。由于柑橘皮对光谱的干扰较大,导致现阶段柑橘产地无损鉴别研究匮乏。此外柑橘体积较大,因此需要对光谱采样点进行优化。为此,基于近红外光谱技术与化学计量学方法,提出了一种用于不同产地柑橘无损鉴别的新方法。使用近红外光谱仪得到了 120 个来自云南、湖南、广西武鸣、广西来宾的沃柑漫反射光谱数据。采用单一和组合光谱预处理方式以消除光谱中的多种干扰;采用主成分分析方法对数据进行降维处理,并以此作为输入值结合 Fisher 线性判别分析方法构建柑橘产地鉴别模型,并与主成分分析模型进行对比。此外,考察了不同光谱采样位置(赤道线 4 个采集点、果梗部以及果顶部)对结果的影响。结果表明:主成分分析方法结合优化光谱预处理的方法不能实现不同产地柑橘的准确鉴别分析,最优鉴别率仅为 5%;而采用主成分分析-Fisher 线性判别分析方法,利用赤道线 4 个点的平均光谱结合去偏置校正或多元散射校正预处理方法可实现不同产地柑橘的 100% 鉴别分析;采用主成分分析-Fisher 线性判别分析对 6 个点的平均光谱数据进行处理时,采用原始光谱便可实现不同产地柑橘的 100% 鉴别分析。为此,通过对光谱预处理方法以及光谱采集点的优化,利用主成分分析-Fisher 线性判别分析方法即可建立准确的柑橘产地鉴别模型,为不同产地柑橘的快速鉴别提供了新途径,为后续各种柑橘类水果的鉴别分析提供了参考。

**关键词** 近红外光谱; 沃柑; 无损分析; 主成分分析; 线性判别分析

**中图分类号:** O657.33    **文献标识码:** A    **DOI:** 10.3964/j.issn.1000-0593(2021)12-3695-06

## 引言

柑橘是柑、柚、橙、枳、橘(桔)等的总称,是世界第一大水果。柑橘富含水分、维生素、常量以及微量元素、矿物质、酚类、萜类等营养和生物活性物质<sup>[1]</sup>。由于气候、土壤、水分差异,不同产地柑橘在口感、质地方面有所差别,但不同产地的柑橘外观相似,表皮气味相近,非专业人士难以实现准确鉴别分析。

国内外许多学者利用化学分析与 DNA 标记等方法实现了柑橘及其副产物的品种、产地、病害的鉴别分析。Xiao 等<sup>[2]</sup>利用感官评定、气相色谱质谱联用与气相色谱-嗅觉评

判三种方法对不同产地甜橙精油的挥发性成分进行了分析。Nicolosi 等<sup>[3]</sup>利用 DNA 标记法实现了柑橘品种的鉴别分析。这些方法准确度高,但费时费力,不利于大面积推广,且对样品具有破坏性,无法实现快速无损分析,影响产品的二次销售。近年来,近红外光谱技术因其快速无损、绿色环保的特点在食品鉴别分析领域得到了广泛应用<sup>[4-6]</sup>。然而由于柑橘皮对光谱的干扰较大,导致现阶段柑橘产地无损鉴别研究匮乏。此外柑橘体积较大,因此需要对光谱采样点进行优化。

由于光谱采集过程中存在环境、样本、操作人员等的影响,所以光谱往往存在谱峰重叠、较大背景、基线漂移等干扰。为了消除这些干扰,需要结合化学计量学方法对光谱数

收稿日期: 2020-11-15, 修订日期: 2021-02-19

基金项目: 国家自然科学基金项目(31601551, 31671931), 湖南省自然科学基金青年科学基金项目(2019JJ50240), 湖南省教育厅科学硏究项目优秀青年项目(18B118), 中国博士后科学基金面上项目(2019M650187)资助

作者简介: 张欣欣, 1997 年生, 湖南农业大学食品科学技术学院博士研究生    e-mail: zxxhunaujs@yeah.net

\* 通讯作者    e-mail: lipao@mail.nankai.edu.cn

据进行预处理<sup>[7-10]</sup>。现阶段提出了很多光谱预处理方法,如去趋势校正(DT)<sup>[11]</sup>、去偏置校正(De-bias)<sup>[12]</sup>、多元散射校正(MSC)<sup>[13]</sup>、小波变换(WT)<sup>[14-15]</sup>等,不同光谱预处理方法适用范围不同,DT 和 De-bias 用于消除光谱中存在的基线漂移, MSC 等用来消除颗粒分布不均匀及颗粒大小不同所产生的散射对光谱的影响, WT 等求导算法常用来扣除仪器背景或基线漂移对信号的影响。但实际分析光谱中往往存在多种干扰,仅用单一预处理方法无法实现对光谱的优化。为此,在处理光谱数据时通常需采用预处理组合的形式以消除多种干扰<sup>[9]</sup>。结合光谱预处理方法,可以建立准确的鉴别和定量模型。常用的鉴别算法有无监督的主成分分析(PCA)方法与有监督的 Fisher 线性判别分析(FLD),其中 Fisher 线性判别分析方法要求样本数大于变量数,需要对数据进行降维处理。本文通过结合 PCA 与 FLD 的优势,利用近红外光谱,提出了一种不同柑橘产地无损鉴别的方法。通过单一与组合预处理对获得的柑橘光谱进行预处理,利用 PCA-FLD 方法建立鉴别模型,并与 PCA 方法模型进行对比,以期为不同产地柑橘的快速无损鉴别提供一种新方法。

## 1 实验部分

### 1.1 材料与试剂

云南、湖南、广西武鸣和来宾是国内常见的沃柑主产区,且这四个产地沃柑在本地超市容易购得。它们外观上十分相似,但在价格上存在一定的差别。从本地水果超市购买新鲜的云南沃柑、湖南沃柑、广西武鸣沃柑、广西来宾沃柑各 30 个,共计 120 个,擦拭表皮,于室温下放置 12 h。

### 1.2 仪器与设备

傅里叶变换近红外光谱仪,(AntarisII,美国 Thermo Scientific 公司),采用积分球漫反射模式采集完整光谱,波数范围为 10 000~4 000 cm<sup>-1</sup>,最小间隔约为 4 cm<sup>-1</sup>,共采集 1557 个数据点。使用 MATLAB R2010b(The Mathworks, USA)软件进行数据分析与处理。

### 1.3 方法

光谱采集在室温下进行。直接将沃柑立放在近红外光谱仪光斑的中心位置,待重心平衡后,开始对样品进行扫描,利用漫反射模式对沃柑的果梗部、果顶部以及赤道线(四等分)进行光谱采集,每个样品共采集 6 个点的光谱,每个点测量 3 次,取其平均值作为原始光谱。用 Kennard-Stone(KS)分组方法将 120 个沃柑样品数据以 2:1 的比例分为 80 个校正集与 40 个验证集。此外,对每个类别样品分别进行 KS 分组计算以保证每类样品在校正集和验证集集的平衡。

为了消除仪器和环境的干扰,提高信噪比,采用 DT、De-bias、MSC、最大最小归一化(Min-Max)、标准正态变量变换(SNV)、一阶导数(1st)、二阶导数(2nd)以及连续小波变换(CWT)等预处理方法对光谱进行处理。近红外信号中存在非常明显的基线漂移干扰,求导预处理可以消除基线漂移干扰,强化谱带特征;由于柑橘样品物理性状原因,信号中存在光散射的干扰,而 MSC 和 SNV 是常用的校正散射影响的方法。因此我们采用了求导和消除光散射预处理的组合

方式以消除光谱中的多种干扰。采用 1st-DT, 1st-SNV, 1st-MSC, CWT-SNV, CWT-MSC 和 SNV-1st 等组合预处理方法对光谱进行优化。为了实现不同产地柑橘无损鉴别分析,采用 PCA 及 FLD 方法建立鉴别模型。FLD 方法要求样本数为变量数的 3~5 倍,因此采用 PCA 方法对数据进行降维处理,利用得到的主成分建立 PCA-FLD 鉴别模型。

## 2 结果与讨论

### 2.1 不同产地沃柑原始光谱特征

图 1 为不同光谱采集点的原始光谱。光谱的大致走势与其他学者采集的柑橘光谱相似,在 8 500~8 300, 7 100~6 900, 5 700~5 500 和 5 200~5 000 cm<sup>-1</sup> 四处有较为明显的波峰,分别为 CH 第三泛频带、CH 第二泛频带、CH 的第一泛频带和 CH 与 CO 组合带<sup>[16-17]</sup>,可能与柑橘中的还原糖、果胶、有机酸的近红外吸收有关。此外,原始光谱中存在明显的谱峰重叠、较大背景、基线漂移等干扰。仅通过原始光谱无法实现不同产地柑橘的鉴别。因此在建立鉴别模型前需采用预处理方法对原始光谱进行优化处理。

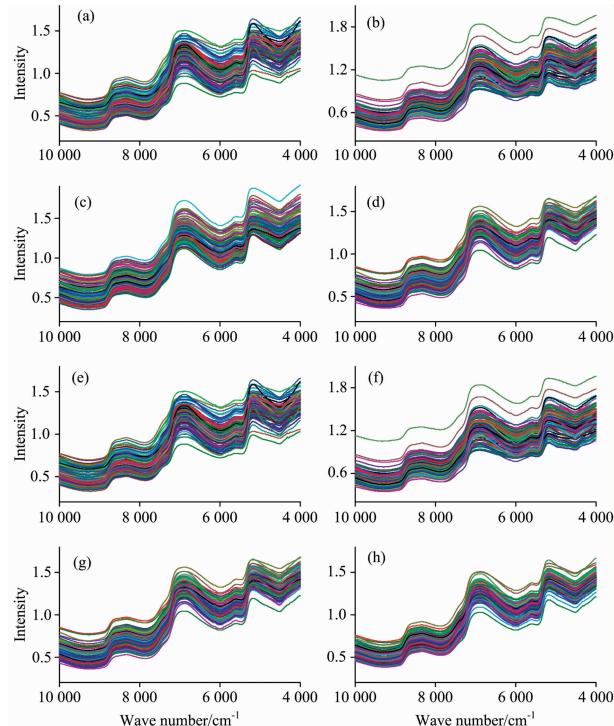


图 1 原始光谱图

(a)–(h): 赤道线 1—4、果梗部、果顶部、赤道线 4 个点平均以及 6 个点平均

Fig. 1 The raw spectra

(a)–(h): equator 1—4, top, bottom, average of 4 points, average of 6 points, respectively

### 2.2 基于单一及组合预处理和 PCA 方法的不同产地柑橘鉴别分析

采用 PCA 方法结合光谱预处理方法以建立不同产地柑橘的鉴别模型。图 2 为赤道线 4 个采集点平均光谱结合单一

预处理的 PCA 结果。其中校正集样本用实心图标，验证集样本用空心图标表示。因为前两个主成分 (PC1 和 PC2) 对柑橘光谱数据变量的累计方差贡献率在 90% 以上，因此选择 PC1 和 PC2 进行 PCA 分析。从图 2(a) 可知，不同产地柑橘数据的置信椭圆呈现交织状态，鉴别率为 0%；经 MSC, Min-Max 和 SNV 单一预处理优化后鉴别率得到了提高，但最佳鉴别率仅为 5% [图 2(b-i)]。其他光谱采集点以及 6 个采集点平均光谱得到的结果也较为类似。以上结果表明，采用单

一预处理结合 PCA 模型无法实现柑橘产地的鉴别分析。为了进一步消除光谱中的多重干扰，采用组合预处理优化光谱数据，并建立 PCA 鉴别模型。然而，即使结合组合预处理优化处理，赤道线 4 个点平均光谱的最佳鉴别率仅为 2.5%。此外，在 6 个点的鉴别结果中，组合预处理后的最佳鉴别分析结果也仅为 5%，可能是因为不同产地柑橘果皮干扰较大，PCA 方法无损挖掘得到隐藏在柑橘皮中的差异信息。

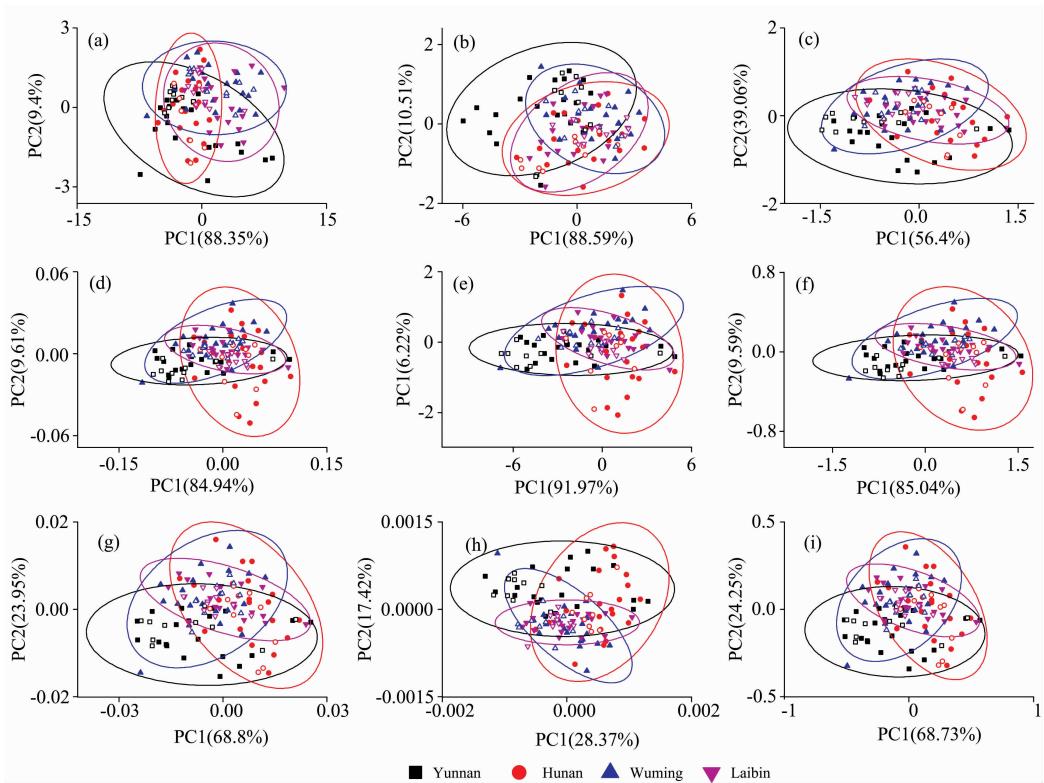


图 2 赤道线 4 个采集点平均光谱结合单一预处理的主成分分析结果

(a)-(i): 原始光谱, DT, De-bias, MSC, Min-Max, SNV, 1st, 2nd 和 CWT

Fig. 2 PCA results of 4 equator points average spectra with single pretreatment methods

(a)-(i): raw spectra, DT, De-bias, MSC, Min-Max, SNV, 1st, 2nd and CWT, respectively

### 2.3 基于 PCA-FLD 的不同产地柑橘鉴别分析

FLD 是一种有效的有监督分类方法，常用于寻找目标类之间的最优边界。为了使样本数达到变量数的 3~5 倍，我们采用 PCA 方法对数据进行降维处理。图 3 为随着主成分 (PC) 数量增加的累积方差贡献率。可以看到，随着 PC 数量的增加，其方差贡献率迅速增加最终趋于 100%。除 2nd 预处理方法以外，前 30 个主成分基本包含所有信息 (>99.99%)，且样本数刚好达到了变量数的 3~5 倍。因此我们采用 PC 数为 30 的 PCA-FLD 方法对数据进行降维处理。

利用 PCA-FLD 结合单一与组合预处理优化以实现不同产品柑橘无损鉴别，并考察不同光谱采集部位对结果的影响。表 1 为不同预处理方法后的 PCA-FLD 结果。由表中可以看出，与 PCA 分析模型相比，PCA-FLD 模型鉴别率得到了显著提高。利用 PCA-FLD 方法分析 4 个点平均光谱数据

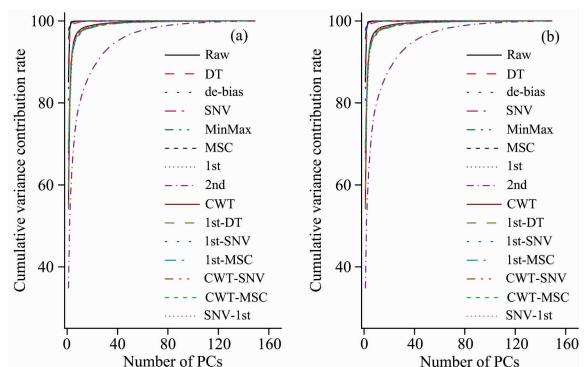


图 3 不同部位光谱分析的累积方差贡献率

(a): 赤道线 4 个点；(b): 赤道线 4 个点+果梗部+果顶部

Fig. 3 Cumulative variance contribution rates at different positions

(a): 4 points on equator; (b): 4 points on equator+stem+top

表 1 基于不同预处理的 PCA-FLD 模型鉴别准确率

Table 1 Identification accuracies by PCA-FLD with different pretreatment methods

Spectral acquisition site	Pretreatment	Identification accuracies/%				
		Yunnan	Hunan	Wuming	Laibin	All
4 points on equator	Raw spectra	90	100	100	100	97.5
	DT	90	100	100	100	97.5
	De-bias	100	100	100	100	100
	SNV	90	100	100	90	95
	Min-Max	90	100	100	90	95
	MSC	100	100	100	100	100
	1st	80	100	100	100	95
	2nd	70	80	90	30	67.5
	CWT	80	100	100	100	95
	1st-DT	80	100	100	100	95
	1st-SNV	80	100	100	100	95
	1st-MSC	90	100	100	100	97.5
4 points on equator and the top and bottom of orange	CWT-SNV	90	100	100	100	97.5
	CWT-MSC	80	100	100	100	95
	Raw spectra	100	100	100	100	100
	DT	100	100	100	100	100
	De Bias	100	100	100	100	100
	SNV	100	100	100	100	100
	Min-Max	100	100	100	100	100
	MSC	100	100	100	100	100
	1st	100	100	110	100	100
	2nd	100	80	80	80	100
	CWT	100	100	100	90	85
	1st-DT	80	100	100	90	97.5

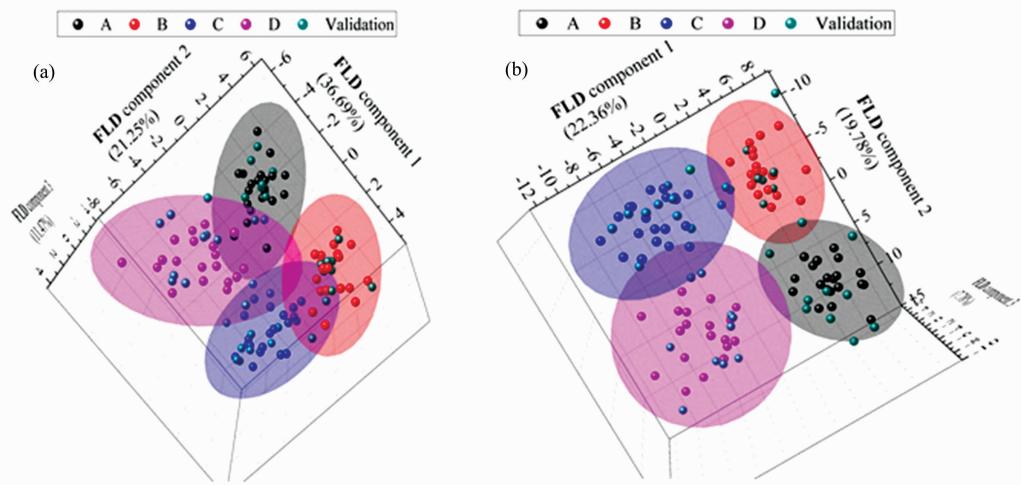


图 4 De-bias 预处理的 PCA-FLD 结果

(a): 赤道线 4 点平均光谱数据; (b): 6 点平均光谱数据

Fig. 4 PCA-FLD results with De-bias method

(a): 4 points average spectra; (b): 6 points average spectra

时, 无需结合光谱预处理即可达到 97.5% 的鉴别准确率; 结合 De-bias 或 MSC 预处理可获得 100% 的鉴别准确率; 最低

的鉴别率结果为经过 2nd 预处理后的 67.5%, 可能原因是 2nd 预处理的累计方差贡献率小于 99%。采用 6 个点平均光

谱数据时，无需结合预处理方法，便可实现不同产地柑橘的 100% 鉴别；除 CWT 外，其他单一预处理优化后的 PCA-FLD 模型鉴别结果均达到了 100%；经组合预处理优化的模型鉴别结果均超过 90%。图 4 为原始光谱经过 De-bias 预处理后的 PCA-FLD 结果，表明 PCA-FLD 模型可以实现所有产地沃柑的鉴别分析。

### 3 结 论

基于近红外光谱结合化学计量学方法，建立了不同产地柑橘的快速无损的鉴别方法。在不破坏沃柑样品的情况下，获得了沃柑赤道线及其果梗部、果顶部共 6 个位置的光谱数据。用光谱预处理方法对光谱进行优化处理，并利用 PCA 与 PCA-FLD 模式识别方法建立鉴别模型，同时对柑橘的光谱

采集位置进行了优化。结果表明：仅通过预处理和光谱采集位置的优化，PCA 方法都不能实现不同产地柑橘的鉴别分析，最高鉴别率仅为 5%；采用 PCA-FLD 方法建立的模型鉴别结果显著优于 PCA 方法，采用 4 个点平均光谱获得的鉴别率可达到 97.5%，结合 De-bias 或 MSC 预处理可以实现不同产地柑橘 100% 的鉴别；当采用 6 个点平均光谱数据时，无需预处理即可实现对不同产地柑橘的 100% 鉴别。PCA 与 PCA-FLD 的结果有很大差别，主要原因是 PCA 为无监督的模式识别方法，而采用有监督模式识别的 FLD 方法对不同产地沃柑 6 点平均光谱进行处理可实现 100% 的聚类分析，因为该方法需提供类别的先验知识，在处理分类问题时有更好的降维与分类效果。本实验为不同柑橘产地的无损鉴别提供了一个参考，在今后的研究中，将对其他柑橘水果进行进一步分析，以建立适用性更强的鉴别模型。

### References

- [1] Farag M A, Abib B, Ayad L, et al. Food Chemistry, 2020, 331: 127306.
- [2] Xiao Z B, Ma S T, Niu Y W, et al. Flavour and Fragrance Journal, 2016, 31(1): 41.
- [3] Nicolosi E, Deng Z N, Gentile A, et al. Theoretical and Applied Genetics, 2000, 100(8): 1155.
- [4] Liu P, Wang J, Li Q, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2019, 206: 23.
- [5] Li P, Zhang X X, Li S K, et al. Sensors, 2020, 20(6): 1586.
- [6] LI Shang-ke, DONG Yi-qing, LI Pao, et al(李尚科, 董怡青, 李跑, 等). Journal of Chinese Institute of Food Science and Technology (中国食品学报), 2020, 20(4): 240.
- [7] Han X, Huang Z X, Chen X D, et al. Fuel, 2017, 207: 146.
- [8] Han X, Tan Z, Huang Z X, et al. Analytical Methods, 2017, 9: 3720.
- [9] Li J Y, Yu M, Li S K, et al. Food Science & Nutrition, 2021, 9(8): 4176.
- [10] Bian X H, Wang K Y, Tan E X, et al. Chemometrics and Intelligent Laboratory Systems, 2020, 197(2): 103916.
- [11] Barnes R J, Dhanoa M S, Lister J S. Applied Spectroscopy, 1989, 43(5): 772.
- [12] LIU Yan-de, YING Yi-bin, FU Xia-ping(刘燕德, 应义斌, 傅霞萍). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(11): 1793.
- [13] Helland S I, Naes T, Isaksson T. Chemometrics and Intelligent Laboratory Systems, 1995, 29(2): 233.
- [14] Shao X G, Leung K M, Chau F T. Accounts of Chemical Research, 2003, 36(4): 276.
- [15] Liu Y, Cai W S, Shao X G. Chemometrics and Intelligent Laboratory Systems, 2013, 125: 11.
- [16] Liu Y, Sun X, Ouyang A, et al. Food Science and Technology, 2010, 43(4): 602.
- [17] Li P, Li S K, Du G R, et al. Food Science & Nutrition, 2020, 8(5): 2543.

# A Nondestructive Identification Method of Producing Regions of Citrus Based on Near Infrared Spectroscopy

ZHANG Xin-xin<sup>1</sup>, LI Shang-ke<sup>1</sup>, LI Pao<sup>1, 2\*</sup>, SHAN Yang<sup>2</sup>, JIANG Li-wen<sup>1</sup>, LIU Xia<sup>1</sup>

1. College of Food Science and Technology, Hunan Provincial Key Laboratory of Food Science and Biotechnology, Hunan Agricultural University, Changsha 410128, China

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

**Abstract** Citrus is one of the popular fruits in the world. There are differences in internal quality and price of different-regions citrus. However, the appearance differences are small, and it is difficult for laypeople to identify by appearance. Methods such as DNA labeling and instrumental analysis are complex in operation and destructive to samples, which cannot achieve rapid and non-destructive analysis, affecting the secondary sales of products. Near-infrared spectroscopy is a fast and nondestructive detection method that can be used to identify different-regions agricultural products. Due to the large interference of citrus peel on the spectra, there is a lack of nondestructive identification of citrus origin. Besides, citrus is large. Therefore, it is necessary to optimize the spectral collection points. This paper proposed a new method for nondestructive identification of different-regions citrus based on near infrared spectroscopy and chemometrics. The diffuse reflectance spectra of 120 fertile oranges from Yunnan, Hunan, Wuming and Laibin of Guangxi were obtained by near-infrared spectroscopy. Single and combined spectral pretreatment was used to eliminate multiple interferences in the spectra. The principal component analysis method was used to reduce the data dimension, which was used as the input value. Combining with Fisher linear discriminant analysis method, the citrus origin identification model was obtained and compared with the principal component analysis model. In addition, the effects of different spectral collection locations (4 collection points along the equator, top and bottom) on the results were investigated. The results showed that the principal component analysis method combined with the optimized spectral pretreatment method could not accurately identify different-regions citrus, and the best identification accuracy was 5%. When the principal component analysis-Fisher linear discriminant analysis was used, the average spectra of 4 collection points along the equator combined with De-bias correction or multivariate scattering correction pretreatment method could achieve 100% identification analysis of different-regions citrus. Furthermore, the average spectra of 6 collection points combined with raw data could achieve 100% identification analysis of different-regions citrus. Therefore, by optimizing the spectral pretreatment methods and spectral collection points, the accurate identification model of different-regions citrus can be established by using principal component analysis-Fisher linear discriminant analysis method, which provides a new method for rapid identification of different-regions citrus.

**Keywords** Near infrared spectroscopy; Fertile orange; Non-destructive analysis; Principal component analysis; Fisher linear discriminant analysis

(Received Nov. 15, 2020; accepted Feb. 19, 2021)

\* Corresponding author