

# 利用可见/近红外光谱判别干枣品种

吴建虎\*, 雷俊桃, 杨 琪

(山西师范大学食品科学学院, 临汾 041000)

**摘要:** **目的** 利用可见/近红外反射光谱技术快速判别干枣的品种。 **方法** 使用光谱仪获取山西永和枣、山西板枣和新疆和田枣 3 种干枣在 345~1100 nm 波段范围内的漫反射光谱; 分别使用多元散射校正(MSC)法和一阶导数法(1<sup>st</sup>-D)和二阶导数法(2<sup>nd</sup>-D)对反射光谱进行预处理; 对预处理光谱进行主成分分析, 全交差验证法确定最佳主成分数量, 提取主成分, 结合马氏距离法和线性判别法建立品种判别模型, 建立模型过程中使用全交叉验证法确定最佳主成分数, 将模型应用于干枣的品种判别。 **结果** 可见/近红外反射光谱经过 MSC 处理后提取主成分建立品种预测模型对枣的品种判别结果最好, 利用前 4 个主成分结合马氏距离法建立的判别模型和利用前 5 个主成分结合线性判别法建立判别模型, 对于 3 个品种的枣的校正和验证判别准确率都达到了 100%。 **结论** 可见/近红外反射光谱技术可以较好地判别干枣品种, 本研究可为可见/近红外光谱技术在干枣品种和产地的快速鉴别和溯源中的应用提供一定的技术基础。

**关键词:** 干枣; 可见/近红外反射光谱; 主成分分析

## Discrimination of dry jujube cultivar varieties using visual/near infrared reflectance spectroscopy

WU Jian-Hu\*, LEI Jun-Tao, YANG Qi

(Institute of Food Science, Shanxi Normal University, Linfen 041000, China)

**ABSTRACT: Objective** To discriminate different dry jujube cultivated varieties non-destructively by visual/near infrared (VIS/NIR) reflectance spectroscopy technique. **Methods** The VIS/NIR reflectance spectrums were collected at 345~1100 nm from 3 kinds of dry jujube, including Yonghe jujube, Banzao jujube and Hetian Jujube. The spectrums were pretreated by multiplicative scattering correction (MSC), first derivative (1<sup>st</sup>-D) and second derivative (2<sup>nd</sup>-D) methods, respectively. And then, principal component analysis were performed, and the discrimination model was constructed using principal component based on Mahalanobis distance and linear distance respectively. The cross validation method was used to determinate the optimal number of principal components for constructing discrimination models. **Results** The discrimination model based on Mahalanobis distance using 4 optimal principal components, which were pretreated by MSC spectrum and the discrimination model based on linear distance using 5 optimal principal components which were pretreated by MSC spectrum had the best results. The 2 models gave good discrimination of 3 kinds of jujube with the accuracy rates of 100%. **Conclusion** The VIS/NIR reflectance spectroscopy technique is useful for discrimination of different kinds of jujube, which can provide an efficient means for the rapid and nondestructive determination of dry quality.

基金项目: 山西高校科技创新项目(2013123)

**Fund:** Supported by Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (STIP)(2013123)

\*通讯作者: 吴建虎, 博士, 副教授, 主要研究方向为食品品质安全检测技术装备。E-mail: wjhu180509@sina.com

\*Corresponding author: WU Jian-Hu, Ph.D., Shanxi Normal University, Linfen 041000, China. E-mail: wjhu180509@sina.com

**KEY WORDS:** dry jujube; visual/near infrared reflective spectrum; principal component analysis

## 1 引言

鲜枣营养丰富, 风味独特, 是深受人们喜爱的一种食品, 具有较高的经济价值, 近年的栽培面积和产量迅速提高。但是鲜枣极易腐烂, 保存期短, 对采后鲜枣进行干制成为延长枣的供应期, 提高附加值, 保证枣农收入的重要手段。干枣的产地和品种对其品质和价格有重要的影响, 目前市场上干枣的品种和产地造假日趋严重, 极大影响了枣农的经济利益。所以有效识别干枣的产地和品种对于保护枣的品牌, 维护当地枣农的经济利益具有重要的意义。

可见/近红外(VIS/NIR)光谱分析技术主要基于有机物分子中含 H 基团, 如 C-H、N-H、O-H 等伸缩振动的各级倍频, 以及这些基团伸缩振动与弯曲振动的频吸收, 涵盖了大多数类型有机化合物的组成和分子结构信息。不同品种和产地的农产品, 因品种、栽培模式和生长环境的不同, 其内部的质地、组分含量存在较大差异<sup>[1]</sup>, 可通过其近红外光谱反映出来。可见/近红外技术具有快速、无损和易于实现在线监测的能力<sup>[2]</sup>, 在水果<sup>[3,4]</sup>、肉类<sup>[5]</sup>、葡萄酒<sup>[6]</sup>、醋<sup>[7]</sup>以及谷物<sup>[8,9]</sup>和药材<sup>[10,11]</sup>等食品和农产品的品质评价、品种和产地的判别中得到了广泛的应用<sup>[12-14]</sup>。近年来, 可见/近红外光谱技术在鲜枣的缺陷、损伤<sup>[15,16]</sup>、成分含量检测<sup>[17-20]</sup>、酵母菌数量<sup>[21]</sup>以及品种和产地<sup>[22,23]</sup>等方面的研究都取得了较好的结果, 已有相关的研究报道。但是, 目前还鲜见针对干枣的品种判别及品质的近红外检测研究报道。

本研究以干枣为研究对象, 采集干枣样本可见/近红外反射光谱, 采用建立多元散射校正、一阶导数法和二阶导数法对反射光谱进行预处理, 采用主成分分析法结合马氏距离和线性距离判别法, 建立干枣的品种判别模型, 并比较不同数据处理方法建立判别模型对干枣品种判别准确度, 以期建立有效的干枣品种判别模型, 为快速、无损地判别干枣的品种提供理论依据。

## 2 材料与方 法

### 2.1 实验材料

实验用干枣有 3 个品种, 分别是当年产的新疆和田枣、山西骏枣和山西永和枣, 购买于临汾市尧丰农贸市场。样本购买回实验室后, 清除掉表面的污渍, 阴凉处风干, 编号, 立即测量其反射光谱。研究所使用 VIS/NIR 光谱测量系统由控制计算机、USB4000 微型便携式光纤光谱仪(美国海洋光学公司, 波长范围 345~1100 nm, 光谱分辨率 0.21 nm)、光源(14.5 V 卤素灯)和 QR600-7-VIS125BX 反射光纤探头(北京吉天公司)以及光谱仪数据采集控制软件(海洋光学 SpectraSuite 软件)组成。

### 2.2 样本 VIS/NIR 反射光谱采集

本研究采用漫反射方式采集样品光谱。光谱仪探测器曝光时间设定为 90 ms, 平滑度为 15 点(约 3 nm), 平均次数 3 次; 对每个干枣样本, 沿赤道圆周均匀取 3 处平整部位, 每处扫描 3 次, 一个样本共扫描 9 次, 取均值为该样本的反射光谱。样本的反射率计算如式(1)所示

$$R = \frac{I - D}{W - D} \times 100\% \quad (1)$$

式中:  $R$ —测定样本光谱反射率(%);  $I$ —样本反射强度;  $D$ —暗电流强度;  $W$ —参照白板( $\text{BaSO}_4$ )的反射强度。

测定光谱数据后, 每类样本都随机分为 2 组, 一组占 3/4, 用于判别模型的校正; 剩余 1/4 用于模型的验证。

### 2.3 光谱数据预处理及判别模型建立方法

光谱首先进行平滑处理, 然后分别使用多元散射校正法(multiple scattering correction, MSC), 一阶导法(first derivation, 1<sup>st</sup>-D)以及二阶导法(second derivation, 2<sup>nd</sup>-D)对光谱进行预处理; 针对预处理后光谱, 使用主成分分析, 提取特征主成分, 分别结合马氏距离法和线性判别法(LDA)建立品种判别模型。判别模型的建立过程中均使用全交叉验证法, 以确定验证模型所使用最佳主成分数和模型的稳定性。

将建立的判别模型用于未知样本的判别, 来评定模型的验证结果。模型的校正和验证的结果用式(2)计算:

$$Q = \frac{J_1}{J} \times 100\% \quad (2)$$

式中,  $Q$ —模型的判别准确度(%),  $J$ —某品种样本的实际数量,  $J_1$ —该品种判别正确数量。数据的处理及品种判别模型的建立通过 MATLAB 7.0 软件编程实现。

## 3 结果与分析

### 3.1 样本品种及分组结果

实验采集样本及其分组情况如表 1 所示, 其中板枣 34 枚(类别号为 1)、和田枣 35 枚(类别号为 2)、永和枣 43 枚(类别号为 3)。3 类样本随机分成了 2 组, 其中校正集共 86 枚, 用于品种判别模型的校正, 其余的 26 枚用于模型判别效果的验证。

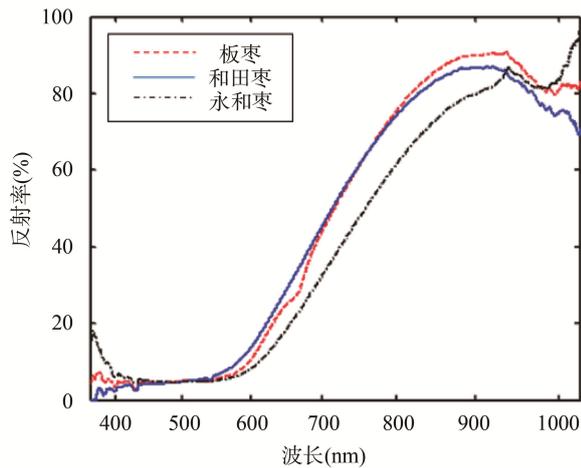
表 1 实验用样本及分组结果  
Table 1 Experimental samples and grouping results

类别号	品种	数量	校正集	验证集
1	山西板枣	34	26	8
2	和田枣	35	27	8
3	山西永和枣	43	33	10
总数		112	86	26

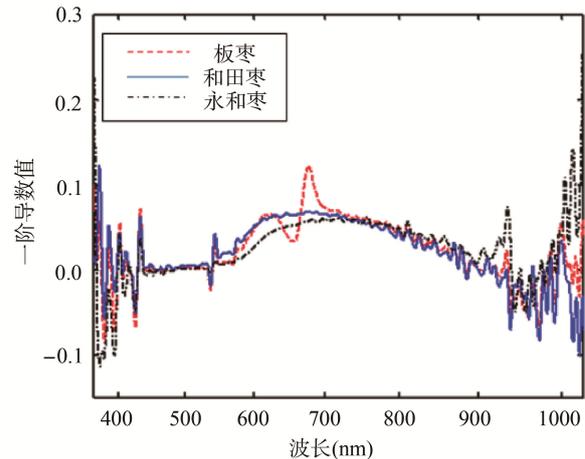
### 3.2 反射光谱及其预处理

在光谱仪的整个波段范围内, 360~1025 nm 波段外信号信噪比较低, 不具有分析价值, 因此, 本研究选择 360~1025 nm 范围作为有效波段用于进一步分析。图 1 为经过不同预处理后的平均反射光谱图。其中, 图 1a 所示为经过 MSC 处理后的 3 种枣的平均反射光谱, 从图中可以看出, 3 种枣在 360~400 nm、600~705 nm 以及 835~900

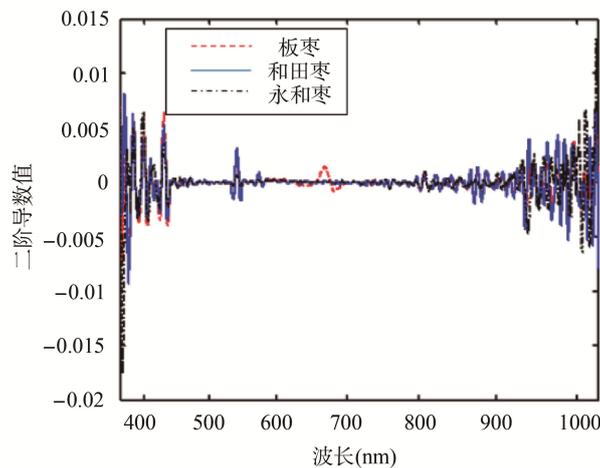
nm 之间的光谱反射率有较大的差异; 图 1b 所示为经过 1<sup>st</sup>-D 处理后 3 种干枣的平均光谱, 在 435、450、540 处以及 560~700 nm 波段范围内的光谱斜率变化的差异较显著; 图 1c 为 3 种干枣的 2<sup>nd</sup>-D 平均光谱, 3 种枣在 440~450 nm、540 nm 附近以及 700 nm 波段附近具有较大的差别。3 种不同干枣的平均反射光谱之间的差异表明可应用于其品种的鉴别。



(a) MSC 处理后平均光谱  
(a) Mean spectrum by MSC



(b) 1<sup>st</sup>-D 平均光谱  
(b) Mean spectrum by 1<sup>st</sup>-D



(c) 2<sup>nd</sup>-D 处理后平均光谱  
(c) Mean spectrum by 2<sup>nd</sup>-D

图 1 样本平均光谱曲线

Fig. 1 Mean spectrums of dry jujube samples

### 3.3 优选主成分数的确定

对经过不同预处理的验证集光谱, 进行主成分分析, 利用马氏距离法和线性距离法建立品种判别校正模型, 全交差验证法确定建立判定模型时所需的最佳主成分数。表 2 所示为不同预处理光谱在使用 2 种距离方法建立品种判别模型时所确定的最佳主成分数。从表中可知, 经过 MSC 处理的光谱, 使用马氏距离法和线性距离法建立品种判别模型时, 提取最优主成分数分别为 4 和 5, 建立的品种校正模型的校正和全交差验证的准确率都达到了 100%; 而对于一阶导数处理光谱, 2 种方法确定最佳主成分数分别为 10 个和 12 个, 模型的校正准确率分别为 98.9%和 100.0%, 线性距离法的校正模型全交差验证结果为 96.5%, 优于马氏距离法判定; 对于二阶导数处理光谱, 分别提取了 12 个和 9 个主成分, 模型校正准确率均为 96.5%, 线性距离法建立模型全交差验证结果为 94.2%, 优于马氏距离法的 84.9%的准确率。

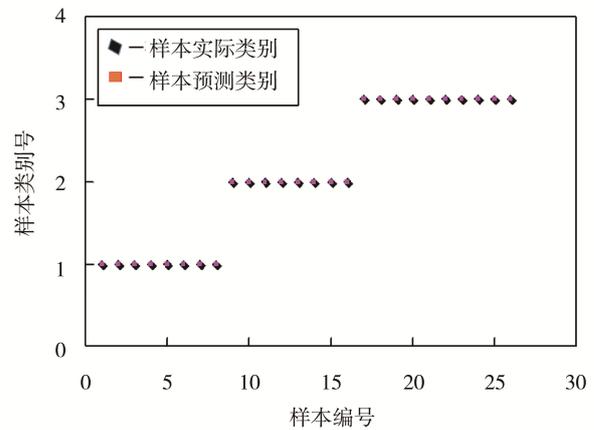
### 3.4 判别模型的验证结果

将使用 2 种方法建立的不同品种判别模型, 应用于验证集的 28 个样本, 检验模型的品种判别能力。模型的验证结果如图 2~4 所示。图 2a 和 2b 为验证集样品反射光谱经过 MSC 处理后, 分别使用马氏距离法和线性距离法建立的判别模型对验证集的预测结果。从图中可以看出, 这两种模型对属于第 1 类的 8 枚板枣(类别号 1)、第 2 类的 8 枚和田枣(类别号 2)以及第 3 类的 10 枚永和枣(类别号 3)的判别结果与实际类别完全相同, 判别准确率为 100%。

图 3a 和 3b 为验证集样品反射光谱经过一阶导数处理后, 分别使用马氏距离法和线性距离法模型对验证集的预测结果。从图中可以看出, 线性距离法判别模型对 3 类枣的判别与原来完全相同, 准确率为 100%。而马氏距离判别模型的判定结果较差, 其对第 1 类的 8 个板枣的判别完全准确, 对于和田枣和永和枣的判别准确率较差, 其中 6 枚判别准确, 2 个误判为 1 类的板枣, 判别准确率为 75%; 对于 10 个永和枣, 6 个判别正确, 4 个误判为 1 类板枣, 准确率为 60%, 线性判别优于马氏距离判别。2 种方法对板枣的判别准确率为最高。

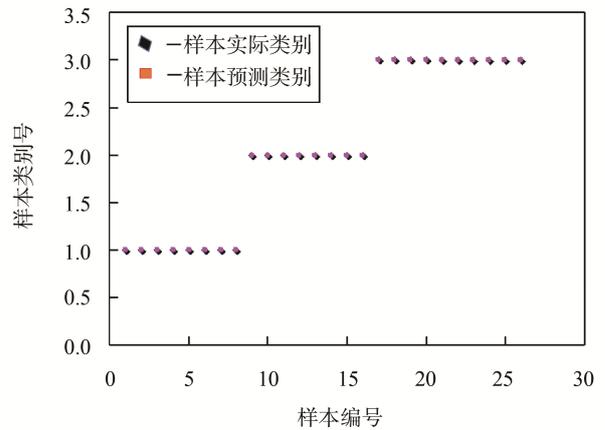
图 4a 和 4b 为验证及样品反射光谱经过二阶导数处理后, 分别使用马氏距离法和线性判别法模型对验证集的预测结果。从图中可以看出, 马氏距离判别模型对第一类的 8

个板枣的判别完全准确, 正确率为 100%, 对于第二类的和田枣, 5 个判别正确, 3 个误判为 1 类的板枣, 判别准确率为 62.5%; 对于 10 个永和枣, 9 个判别正确, 1 个误判为 1 类板枣, 准确率为 90.0%。线性判别模型判别准确率较高, 其中, 1 类有 1 个误判, 准确率为 87.5%, 2 类和 3 类的判别准确率均为 100.0%, 线性距离判别结果优于马氏距离判别法。



(a)马氏距离判别法验证结果

(a) Validation result based on Mahalanobis distance



(b)线性判别法验证结果

(b) validation result based on linear distance

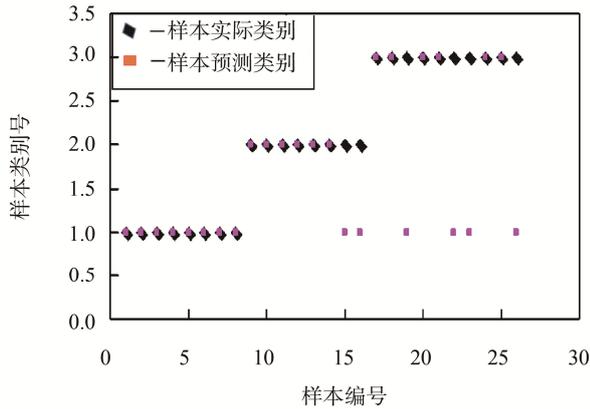
图 2 MSC 处理光谱建立判别模型的验证结果

Fig. 2 Validation results of MSC spectrum

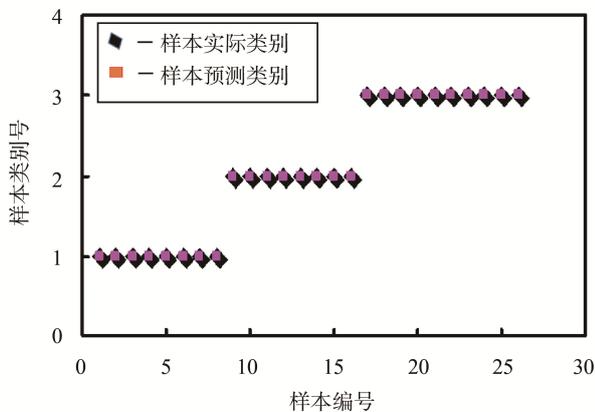
表 2 不同处理光谱最佳主成分数确定及模型校正结果

Table 2 Determination of optimal principal components and the calibration results of discrimination models

预处理方法	马氏距离法			线性距离判别法		
	最佳主成分数	校正准确率(%)	全交差验证(%)	最佳主成分数	校正准确率(%)	全交差验证(%)
MSC	4	100.0	100.0	5	100.0	100.0
1 <sup>st</sup> -D	10	98.9	87.2	12	100	96.5
2 <sup>nd</sup> -D	12	96.5	84.9	9	96.5	94.2



(a) 马氏距离判别法验证结果  
(a) Validation result based on Mahalanobis distance



(b) 线性判别法验证结果  
(b) validation result based on linear distance

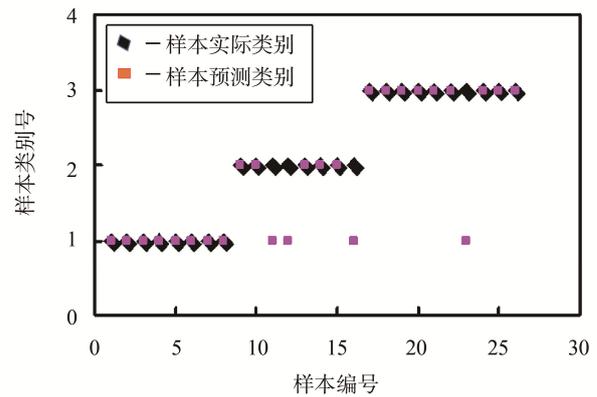
图3 1<sup>st</sup>-D 处理光谱建立判别模型的验证结果  
Fig. 3 Validation results of 1<sup>st</sup>-D spectrum

### 3.5 结果分析

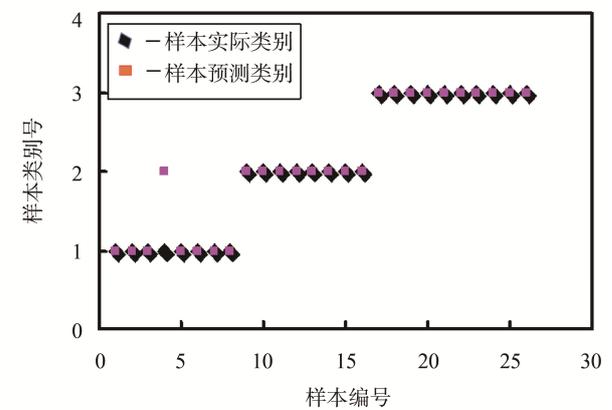
通过上述比较可以看出,反射光谱经过多元散射校正后,选取4个优选主成分使用马氏距离法建立品种判别模型,以及选取5个主成分使用线性距离法建立的品种判别模型对和田枣、永和枣和板枣的品种判别结果最好,3种枣的品种判别准确率均达到了100.0%。其余2种光谱处理方法使用线性距离判别法的验证结果较好,是可以接受的;而使用马氏距离法建立的模型的判别结果比较差。

## 4 结论

本研究利用可见/近红外反射光谱快速地判别不同品种的干枣。对反射光谱分别进行多元散射校正、一阶导数处理和二阶导数处理,对光谱进行主成分分析法,使用马氏距离法和线性距离法结合全交叉验证法,确定最优主



(a) 马氏距离判别法验证结果  
(a) Validation result based on Mahalanobis distance



(b) 线性判别法验证结果  
(b) validation result based on linear distance

图4 2<sup>nd</sup>-D 处理光谱建立判别模型的验证结果  
Fig. 4 Validation results of 2<sup>nd</sup>-D spectrum

成分数。利用主成分建立品种判别模型,实现了对不同品种干枣的快速判别。结果表明,可见/近红外反射光谱经过多元散射校正处理后选择的优化主成分组合,对3种干枣的品种判别结果最好,判别准确性达到了100%,准确性是可接受的。试验研究表明可见/近红外光谱反射技术可用于干枣品种判别,本研究可为可见/近红外光谱技术在干枣品质快速检测中的应用提供一定的技术基础。

### 参考文献

- [1] 张晓焱, 苏学素, 焦必宁, 等. 农产品产地溯源技术研究进展[J]. 食品科学, 2010, 31(3): 271-278.  
Zhang XY, SU XS, Jiao BN, *et al.* Research progress in methods for geographical origin traceability of agricultural products [J]. Food Sci, 2010, 31(3): 271-278.
- [2] Prieto N, Roehe R, Lavín P, *et al.* Application of near infrared reflectance spectroscopy to predict meat and meat products quality: a review [J]. Meat

- Sci, 2009, 83(2): 175–186.
- [3] 苏学素, 张晓焱, 焦必宁, 等. 基于近红外光谱的脐橙产地溯源研究[J]. 农业工程学报, 2012, 15: 240–245.  
Su XS, Zhang XY, Jiao BN, *et al.* Determination of geographical origin of navel orange by near infrared spectroscopy [J]. Trans CSAE, 2012, 28(15): 240–245.
- [4] 张晓焱. 基于近红外光谱的柑桔产地溯源及橙汁掺假识别研究[D]. 重庆: 西南大学 2010.  
Zhang XY. Determination of geographical origin of citrus and adulteration of orange juice by near infrared spectroscopy [D]. Chongqing: Southwest University, 2010.
- [5] 孙淑敏, 郭波莉, 魏益民, 等. 近红外光谱指纹分析在羊肉产地溯源中的应用[J]. 光谱学与光谱分析, 2011, 4: 937–941.  
Sun SM, Guo BL, Wei YM, *et al.* Application of near infrared spectral fingerprint technique in lamb meat origin traceability [J]. Spectrosc Spect Anal, 2011, 4: 937–941.
- [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, S1: 374–378.  
Liu W, Zhan JC, Huang WD, *et al.* Grape wine country of origin identification method based on near infrared spectroscopy [J]. Trans CSAE, 2010, S1: 374–378.
- [7] 管晓, 刘静, 古方青, 等. 基于 NIRS 技术的食用醋品牌溯源研究[J]. 光谱学与光谱分析, 2014, 9: 2402–2406.  
Guan X, Liu J, Gu FQ, *et al.* Research on brand traceability of edible vinegar based on NIRS technology [J]. Spectrosc Spec Anal, 2014, 9: 2402–2406.
- [8] 赵海燕, 郭波莉, 魏益民. 谷物原产地溯源技术研究进展[J]. 核农学报, 2011, 4: 768–772.  
Zhao HY, Guo BL, Wei YM. Research progress on origin of grain [J]. Acta Agric Nuclear Sin, 2011, 4: 768–772.
- [9] Zhao HY, Guo BL, Wei YM, *et al.* Near infrared reflectance spectroscopy for determination of the geographical origin of wheat [J]. Food Chem, 2012
- [10] 张延莹, 张金巍, 刘岩. 近红外光谱技术鉴别三七及其伪品[J]. 中草药, 2010, 3: 364–366.  
Zhang YY, Zhang JW, Liu Y. Identification of three seven and its artifacts by near infrared spectroscopy [J]. Chin Herb Med, 2010, 3: 364–366.
- [11] 赵杰文, 蒋培, 陈全胜. 雪莲花产地鉴别的近红外光谱分析方法[J]. 农业机械学报, 2010, 8: 111–114.  
Zhao JW, Jiang P, Chen QS. Discrimination of snow lotus from different geographical origins by near infrared spectroscopy [J]. Trans CSAE, 2010, 8: 111–114.
- [12] Oliveri P, Di Egidio V, Woodcock T, *et al.* Application of class-modeling techniques to near infrared data for food authentication purposes [J]. Food Chem, 2011, 125(4): 1450–1456.
- [13] 汤丽华, 刘敦华. 基于近红外光谱技术的枸杞产地溯源研究[J]. 食品科学, 2011, 22: 175–178.  
Tang LH, Liu DH. Tracing the geographic origin of Chinese wolfberry by near infrared spectroscopy [J]. Food Sci, 2011, 22: 175–178.
- [14] 马冬红, 王锡昌, 刘利平, 等. 近红外光谱技术在食品产地溯源中的研究进展[J]. 光谱学与光谱分析, 2011, 4: 877–881.  
Ma DH, Wang XC, Liu LP, *et al.* Current progress in food geographical origin traceability by near infrared spectroscopy technology [J]. Spectrosc Spect Anal, 2011, 4: 877–881.
- [15] 王斌, 张淑娟, 赵春慧. 壶瓶枣轻微损伤可见/近红外光谱动态判别模型研究[J]. 现代食品科技, 2015, 31(4): 258–263.  
Wang B, Zhang SJ, Zhao CH. Dynamic discriminate models for the detection of subtle bruising in huping jujube constructed based on their visible/near-infrared spectral data [J]. Mod Food Sci Technol, 2015, 31(4): 258–263.
- [16] Wu LG, He JG, Liu GS, *et al.* Detection of common defects on jujube using Vis-NIR and NIR hyperspectral imaging [J]. Postharv Biol Technol, 2016, 112:134–142.
- [17] 彭海根, 罗华平, 彭云发, 等. 近红外技术对新疆冬枣糖度无损检测的研究[J]. 农机化研究, 2014, 11: 146–150.  
Peng HG, Luo HP, Peng YF, *et al.* Research on the non-destructive testing technology of near-infrared in sugar content of dongzao jujube in south of Xinjiang [J]. J Agric Mech Res, 2014, 11: 146–150.
- [18] 张淑娟, 张海红, 赵艳茹, 等. 鲜枣可溶性固形物可见/近红外光谱检测建模方法比较[J]. 农业机械学报, 2012, 43(3): 108–112.  
Zhang SJ, Zhang HH, Zhan YR, *et al.* Comparison of modeling methods of fresh jujube soluble solids measurement by NIR spectroscopy [J]. Trans CSAE, 2012, 43(3): 108–112.
- [19] 罗雪宁, 倪明航, 孔维楠, 等. 二维相关红外光谱技术对不同冷藏温度的新疆骏枣品质分析研究[J]. 现代食品科技, 2016, 2: 171–175.  
Luo XN, Ni MH, Kong WN, *et al.* Quality analysis of jun jujube from southern xinjiang stored at different cold temperatures by two-dimensional correlation infrared spectroscopy [J]. Mod Food Sci Technol, 2016, 2:171–175.
- [20] Guo Y, Ni YN, Kokot S, Evaluation of chemical components and properties of the jujube fruit using near infrared spectroscopy and chemometrics, Spectrochimica [J]. Acta Part A: Mol Biomol Spect, 2016, 153(15): 79–86.
- [21] 胡耀华, 刘聪, 何勇. 近红外光谱检测鲜枣酵母菌的动力学模型[J]. 光谱学与光谱分析, 2014, 34(4): 922–926.  
Hu YH, Liu C, He Y. Detection of kinetic model of fresh jujube yeast by near infrared spectroscopy [J]. Spectrosc Spect Anal, 2014, 34(4): 922–926.
- [22] 张淑娟, 王凤花, 介邓飞, 等. 鲜枣品种和可溶性固形物的近红外光谱无损检测[J]. 农业机械学报, 2009, 40(4): 139–142.  
Zhang SJ, Wang FH, Jie DF, *et al.* Detection of the fresh jujube varieties and SSC by NIR spectroscopy [J]. Trans CSAE, 2009, 40(4): 139–142 .
- [23] 陈璐, 谷晓红, 王文博, 等. 近红外光谱技术识别沾化和陕西冬枣产地的研究[J]. 山东农业科学, 2016, 3: 133–136.  
Chen L, Gu XH, Wang WB, *et al.* Identification of geographical origins of ziziphus jujuba mill.in Zhanhua and Shaanxi by near infrared spectroscopy [J]. Shandong Agric Sci, 2016, 3: 133–136.

(责任编辑: 金延秋)

## 作者简介



吴建虎, 副教授, 主要研究方向为食品品质快速检测技术。  
E-mail: 287718596@qq.com