

灵武长枣 VC 含量的高光谱快速检测研究

杨晓玉¹, 刘贵珊¹, 丁佳兴¹, 陈亚斌¹, 房盟盟¹, 马超², 何建国^{1*}

1. 宁夏大学农学院食品系, 宁夏 银川 750021

2. 宁夏大学物理与电子电气工程学院, 宁夏 银川 750021

摘要 采用可见-近红外高光谱成像技术结合化学计量学方法检测灵武长枣维生素 C(VC)含量, 探究一种全新的水果内部成分的快速无损检测方法。采用高效液相色谱法(HPLC)测得长枣的 VC 含量化学值, 可见-近红外高光谱成像系统采集 164 个灵武长枣 400~1 000 nm 的高光谱图像, 利用 ENVI4.8 软件提取图像的感兴趣区域(region of interest, ROI), 计算其平均光谱, 获得光谱值, 将化学值与光谱值通过 The UnscramblerX 10.4 软件建立模型。利用蒙特卡洛交叉验证法剔除异常值, 采用光谱理化值共生距离法(sample set partitioning based on joint x - y distance, SPXY)进行样本划分以提高模型的预测性能; 对光谱采用移动平滑(moving average)、中值滤波(median filter)、归一化(normalize)、基线校准(baseline)、多元散射校正(multiple scattering correction, MSC)、去趋势(detrending)和标准正态变量变换(standard normal variate, SNV)等 7 种方法进行预处理; 为进一步减少数据量, 降低维度, 提高运算速度, 使用竞争性自适应加权算法(competitive adaptive reweighted sampling, CARS)、无信息变量消除算法(uninformative variable elimination, UVE)和连续投影算法(successive projections algorithm, SPA)提取特征波长, 以期实现以少数波段代替全波段; 将全波段光谱(full spectrum, FS)以及 CARS, UVE 和 SPA 三种方法提取的特征波长分别建立偏最小二乘(partial least squares wavelength regression, PLSR)和支持向量机(support vector machine, SVM)模型, 从而确定最优的建模模型。利用蒙特卡洛交叉验证法共剔除 7 个异常样本, 采用 SPXY 法将剔除异常样本后的 157 个数据区分为校正集和预测集, 校正集中样本个数为 117, 预测集中样本个数为 40。将未经光谱预处理的建模结果与分别经过七种光谱预处理的建模结果相比, 选择未经光谱预处理的数据进行后续分析; 将未经光谱预处理的光谱值采用 CARS, UVE, SPA 方法进行提取特征波长, CARS 共优选出 406, 415, 487, 631, 636, 655, 660, 665, 670, 684, 689, 694, 723, 732, 747 和 881 nm 下的光谱变量 16 个, 利用 CARS 提取出的特征波长占总波长的 12.8%; UVE 共优选出 406, 415, 627, 631, 636, 651, 655, 660, 665, 670, 675, 679, 684, 689, 694, 699, 703, 708, 742, 747, 751, 756, 761, 766, 771, 775, 780, 785, 790, 795, 919 和 924 nm 下的 32 个特征波长, 利用 UVE 提取出的特征波长占总波长的 25.6%; SPA 共优选出 401, 665, 684 nm 三个特征波长, 利用 SPA 提取出的特征波长占总波长的 2.4%。将全波段光谱与提取出的特征波长建立 PLSR 模型和 SVM 模型, 对比模型结果显示 UVE-SVM 模型最优, 其 R^2 为 0.847 1, R_p^2 为 0.714 9, 说明 UVE 有效地对光谱进行降维, 简化了数据处理过程。本研究对高光谱成像技术在水果领域的应用进行了有益探索, 探究了一种全新的灵武长枣 VC 含量的无损检测方法, 相应建立的可见-近红外高光谱模型为其他水果成分的快速检测提供了理论基础。

关键词 可见-近红外; 高光谱成像技术; 维生素 C(VC); 支持向量机; 无损检测

中图分类号: TS255 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)01-0230-05

收稿日期: 2017-12-25, 修订日期: 2018-04-02

基金项目: 国家自然科学基金项目(31560481), 中央财政支持地方高校改革发展资金-食品学科建设项目(2018)资助

作者简介: 杨晓玉, 女, 1994 年生, 宁夏大学农学院食品系硕士研究生 e-mail: 675612385@qq.com

* 通讯联系人 e-mail: hejg@nxu.edu.cn

引言

灵武长枣是宁夏优势特色农产品,其品质评价一直是研究的焦点。维生素 C(VC)是人体必需营养素,也是衡量果蔬品质的关键指标^[1]。VC 含量的常规检测方法有电化学法,滴定法,分光光度法等,但滴定法易产生误差;电化学法不能进行连续检测^[2],其最大缺点是将样品破坏后才能测定 VC 含量。因此,需寻找一种快速无损测定果蔬 VC 含量的方法。

高光谱成像技术集图谱技术于一体,满足了快速无损的检测需求,成为了无损检测技术的热点,此技术广泛应用于水果的虫害缺陷^[3],自然损伤^[4],农药残留^[5]的外部检测以及可溶性固形物^[6],水分^[7],酚类物质^[8]等内部成分的检测。Gabriel A 等^[9]采用可溶性固形物和硬度指标评估高光谱对蓝莓自动分选分级的适用性,证明此技术可应用于蓝莓的分选;Zhang^[10]等利用高光谱结合支持向量机建立了不同成熟度草莓的模型,其模型的精度超过 85%;陈欣欣等^[11]对库尔勒香梨光谱信息建立 SVM 模型,完好样本和损伤样本的总识别率皆为 93.75%,表明高光谱成像技术结合 SVM 建模方法能识别损伤的库尔勒梨。

本工作以灵武长枣为研究对象,利用高光谱成像技术结合化学计量学方法建立预测模型,为可见-近红外高光谱成像技术应用于其他果蔬内部品质的评价提供参考。

1 实验部分

1.1 材料与试剂

样品采自银川大泉林场,选取 164 个无损伤的长枣,4 °C 贮藏。

VC 标品,中国食品药品检定研究院;甲醇(色谱级),偏磷酸盐(分析纯),活性炭,为天津市大贸化学试剂厂生产;水系膜(0.45 μm×50 mm),滤头(0.45 μm)。

1.2 仪器

可见-近红外高光谱成像系统(美国 Headwall Photonics 公司)和高效液相色谱仪 L-2000(日本岛津公司)。

可见-近红外高光谱成像系统由成像光谱仪(V10E-QE 型),CCD 摄像机(C8484-05G 型),光纤卤素灯(DCR III 型,150 W)和输送装置和计算机构成。

由于摄像头中暗电流的存在和光源在不同波段下强度分布不均匀,在光谱信息采集前,需进行黑白校正^[12]。如式(1)所示:

$$R(\%) = \frac{R_0 - D}{W - D} \times 100 \quad (1)$$

式(1)中, R 为校正后图像; R_0 为原始图像; D 为暗图像; W 为白板图像。

样本扫描时,曝光时间 15 ms;曝光物距 385 mm;扫描线实际长度 70 mm;电控位移台扫描速度 160 μm·s⁻¹;图像分辨率 1 004×501。

1.4 HPLC 测定 VC 含量

标准溶液的配备:称取 VC 标品 16 mg,超纯水溶解定容于 10 mL 的棕色容量瓶中,采用梯度稀释为 1/2,1/4,1/10,1/20,1/30 倍,4 °C 保存。样品待测液的制备:称取灵武长枣果肉 1.0 g,研磨,加入 2% 偏磷酸溶液 5 mL,移至 10 mL 离心管,加入 0.1 g 活性炭,4 °C 下 11 000 r·min⁻¹离心 15 min,取上清液至 10 mL 的棕色容量瓶中,2% 偏磷酸溶液定容,经 0.45 μm 滤膜过滤。液相条件:进样量 20 μL,流速 1.0 mL·min⁻¹,柱温 30 °C,检测波长 254 nm,流动相为 0.2% 偏磷酸盐溶液。

2 结果与讨论

2.1 HPLC 测定结果

图 1 为维生素 C 标品高效液相色谱图,最大吸收波长为 254 nm,出峰时间为 3.554 min。图 2 为长枣提取液的高效液相色谱出峰图,出峰时间为 3.520 min。得到线性方程为 $y=10\ 000\ 000x-473\ 152.198\ 8$,相关系数 $R^2=0.991\ 7$ 。

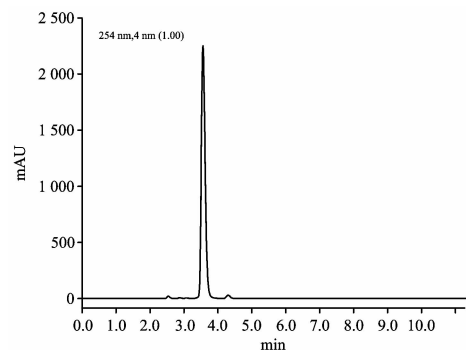


图 1 VC 高效液相色谱图

Fig. 1 High performance liquid chromatographic peak of VC

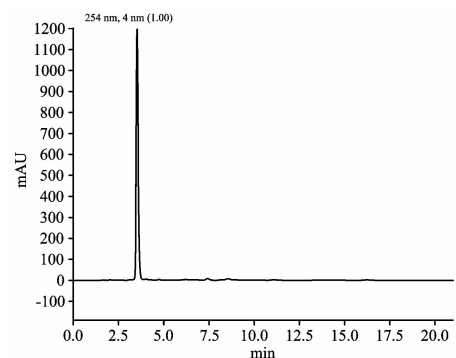


图 2 提取液高效液相色谱图

Fig. 2 High performance liquid chromatographic peak

2.2 可见-近红外光谱检测

采用可见-近红外高光谱成像系统采集长枣光谱反射曲线(见图 3),利用 ENVI 4.6 软件选取感兴趣区域。分别从每张光谱图像中提取 1 521 像素的 ROI,计算出平均反射光谱作为原始光谱。

2.3 蒙特卡洛交叉验证法剔除异常值

蒙特卡洛方法具有同时检测光谱异常值和理化参照值异

常值的优点。具体步骤: 将样本集作为校正集, 建立模型, RMSECV 最小时, 为最佳主成分数; 确定各样本预测误差的统计特征参数(均值和标准差); 以均值为横坐标, 标准差为纵坐标, 绘制散点图(见图 4)。均值和标准差平均值的 2.5 倍作为界限值, 大于界限值的为异常样本。

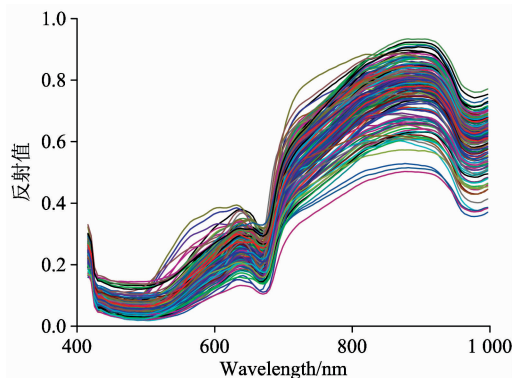


图 3 光谱反射曲线

Fig. 3 Reflectance curves of spectrum

如图 4 所示, 3, 11 样本的预测误差均值较大, 可直接判定为异常值剔除, 其余界限值以外的样本依次剔除。3, 11, 163, 69, 162, 40, 10 号样本剔除后建模 R_c^2 值由 0.465 6 上升至 0.729 3; 剔除 7, 15, 43 号样本后建模 R_c^2 值减小, 由 0.729 3 降低至 0.727 7, 说明此三个数据不是异常值。

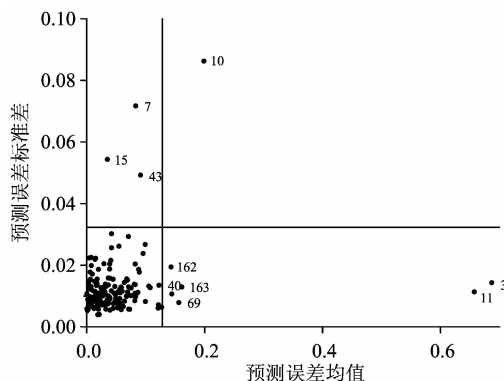


图 4 蒙特卡洛检测

Fig. 4 Monte Carlo detection

2.4 样本划分

采用 SPXY 法对剔除异常值后将样本的 VC 含量按照 3:1 的比例进行划分如表 1 所示。

表 1 长枣 VC 含量
Table 1 VC contents of long jujubes

样本集	样本数	VC 含量/(mg · g ⁻¹)			
		最大值	最小值	平均值	标准偏差
校正集	117	0.434 1	0.048 6	0.260 4	0.090 7
预测集	40	0.364 8	0.130 2	0.291 3	0.042 9

2.5 原始光谱及光谱预处理比较

由表 2 可知, 未经光谱预处理和经过光谱预处理建立的

VC 含量预测模型中, 未处理的校正模型 RMSECV 为 0.048 1, 与经过预处理的数据相比, 值最小; R_c^2 为 0.770 7, 与预处理后的数据相比, 值最大, 说明未经预处理的光谱模型稳定, 且交互验证效果好, 故选择未经光谱预处理的数据进行后续分析。

表 2 不同预处理方法的长枣 VC 含量 PLSR 模型

Table 2 PLSR models of VC contents based on different pretreatment methods

No.	方法	R_c^2	RMSEC	RMSECV	PCs
1	未经光谱预处理	0.770 7	0.043 4	0.048 1	7
2	Moving Average	0.767 6	0.043 7	0.048 2	7
3	Median Filter	0.768 7	0.043 6	0.048 2	7
4	Normalize	0.765 4	0.043 9	0.050 9	8
5	Baseline	0.723 4	0.047 7	0.052 5	6
6	MSC	0.676 9	0.051 5	0.056 3	5
7	Detrending	0.721 8	0.047 8	0.051 7	4
8	SNV	0.681 1	0.051 5	0.056 1	7

2.6 特征波长的选取

2.6.1 竞争性自适应加权算法(CARS)

图 5(a) 为检测波长变量随采样次数的变化趋势: 采样次数增加, 所选变量个数由快到慢递减, 对应粗选与精选两个阶段; 图 5(b) 为筛选过程中交互验证错误率的变化趋势: 交互验证错误率在 0~25 时降低, 从 25 处呈现上升的趋势, 这表明选择过程剔除了与 VC 含量检测相关的关键变量; 图 5(c) 中“*”所对应的位置为 25 次 MC 采样, 此时交互验证错误率最小, 所选出的特征波长的个数最少。经 CARS 提取特征波长后, 优选出 406, 415, 487, 631, 636, 655, 660, 665, 670, 684, 689, 694, 723, 732, 747 和 881 nm 共 16 个光谱波长变量, 占总波长变量的 12.8%。

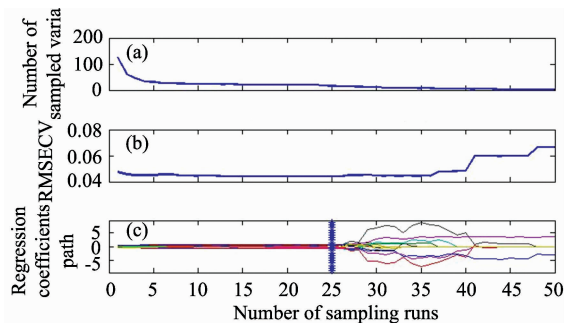


图 5 CARS 筛选过程

Fig. 5 Selection process of CARS variables

2.6.2 UVE

采用 UVE 选取特征波长时, 潜在变量数设为 1~13。潜在变量数为 7 时, RMSECV 最小, 值为 0.045 1。

图 6 中竖线两侧分别有 125 个波长变量, 两条水平虚线为变量选择阈值(30.6)。虚线外是有效信息, 其对应的为特征波长, 共提取 406, 415, 627, 631, 636, 651, 655, 660, 665, 670, 675, 679, 684, 689, 694, 699, 703, 708, 742, 747, 751, 756, 761, 766, 771, 775, 780, 785, 790, 795,

919 和 924 nm 共 32 个光谱变量, 占总波长的 25.6%。

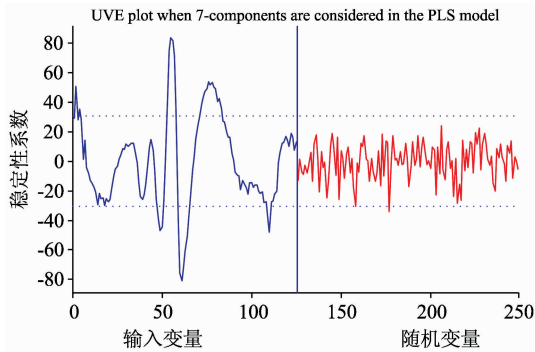


图 6 UVE-PLSR 模型的稳定性分布曲线

Fig. 6 Stability distribution curve of UVE-PLSR model

2.6.3 连续投影算法(SPA)

利用 SPA 选择特征波长时, 指定波长数 N 为 3~15, 共优选出 401, 665, 684 nm 三个光谱波长变量, 占总波长的 2.4%。

2.7 模型的建立

2.7.1 PLSR 模型

建立全光谱和特征波长的 PLSR 模型, 对其建模效果进行分析。由表 3 可知, CARS-PLSR 模型与 FS-PLSR 模型相比, R^2 值, R_p^2 值接近, 与其他模型相比, RMSEC, RMAECV 均为最小, 分别为 0.042 9 和 0.046 6, 具有较好的校正性能和交互验证性能, 其优选出 16 个特征波长, 可有效地对光谱数据进行降维。

2.7.2 SVM 建模

建立不同提取特征波长方法的 SVM 模型(见表 4), UVE-SVM 模型的 RMSECV 值为 0.048 9, R_p^2 值为 0.714 9, 较其他模型最优。

表 3 不同特征波长下的 VC 含量 PLSR 模型

Table 3 PLSR models of VC content based on different wavelengths

建模方法	特征波长提取方法	波长数	校正集		预测集		RMSECV
			R^2	RMSEC	R_p^2	RMSEP	
PLSR	FS	125	0.770 7	0.043 4	0.321 0	0.042 9	0.048 1
	CARS	16	0.776 2	0.042 9	0.336 1	0.043 0	0.046 6
	UVE	32	0.776 5	0.044 3	0.343 7	0.043 3	0.047 4
	SPA	3	0.692 3	0.050 3	0.390 9	0.037 6	0.051 6

表 4 不同特征波长下的 VC 含量 SVM 模型

Table 4 SVM models of VC content based on different wavelengths

建模方法	特征波长提取方法	C	ϵ	γ	R^2	RMSECV	R_p^2
SVM	FS	100	0.1	0.01	0.857 1	0.053 2	0.531 6
	CARS	100	0.1	0.01	0.749 6	0.051 1	0.608 0
	UVE	10	0.1	0.1	0.847 1	0.048 9	0.714 9
	SPA	100	0.1	0.1	0.734 6	0.050 7	0.707 4

3 结 论

利用 VIS/NIR 可见-近红外高光谱技术对灵武长枣 VC 含量进行预测, 主要结论如下:

(1)将灵武长枣的光谱信息经过蒙特卡洛交叉验证法剔除 7 个异常值后进行不同方法的光谱预处理。结果表明, 不

经过光谱预处理的建模效果较好。

(2)利用 CARS, UVE 和 SPA 进行降维处理, 提取特征波长个数为 16, 32, 3, 占原光谱数据的 12.8%, 25.6% 和 2.4%。

(3)对提取的特征波长建立 PLSR 和 SVM 模型。结果表明, UVE-SVM 建立的 VC 含量模型最优, 其 R^2 为 0.847 1, R_p^2 为 0.714 9, RMSECV 为 0.048 9。

References

- [1] CHEN Chen, LU Xiao-xiang, ZHANG Peng, et al(陈 辰, 鲁晓翔, 张 鹏, 等). Food & Machinery(食品与机械), 2015, 31(5): 70.
- [2] LI Ye, YIN Li-hui, GAO Shang, et al(李 野, 尹利辉, 高 尚, 等). Chin. J. Pharm. Anal. (药物分析杂志), 2016, 36(5): 756.
- [3] Liu Guishan, He Jianguo, Wang Songli. International Journal of Food Properties, 2016, 19: 41.
- [4] SUN Shi-peng, PENG Jun, LI Rui, et al. Food Science(食品科学), 2017, 38(2): 301.
- [5] ZHAO Man-tong, LI Bo-cheng, ZHOU Yao, et al(赵曼彤, 李伯承, 周 瑶, 等). Optical Technique(光学技术), 2016, 42(5): 408.
- [6] Polder G A, Leiva-Valenzuela L, Lu R F. Journal of Food Engineering, 2013, 115: 91.
- [7] Pu Yuanyuan, Sun Dawen. Innovative Food Science and Emerging Technologies, 2016, 33: 348.

- [8] Zhang Ni, Liu Xu, Jin Xiaoduo, et al. Food Chemistry, 2017, 237: 811.
- [9] Gabriel A, Leiva-Valenzuela, Renfu Lu. Innovative Food Science and Emerging Technologies, 2014, 24: 2.
- [10] Zhang Chu, Guo Chentong, Liu Fei. Journal of Food Engineering, 2016, 179: 11.
- [11] CHEN Xin-xin, GUO Chen-tong, ZHANG Chu, et al(陈欣欣, 郭辰彤, 张 初, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2017, 37(1): 150.
- [12] Zou Xiaobo, Xu Yiwei, Chen Wu, et al. Mod. Food Sci. Technol., 2015, 32(9): 263.

A Rapid Evaluation of VC Content on Lingwu Long Jujube Using Hyperspectral Technique

YANG Xiao-yu¹, LIU Gui-shan¹, DING Jia-xing¹, CHEN Ya-bin¹, FANG Meng-meng¹, MA Chao², HE Jian-guo^{1*}

1. School of Agriculture Department of Food, Ningxia University, Yinchuan 750021, China

2. School of Physics and Electronic-Electrical Engineering, Ningxia University, Yinchuan 750021, China

Abstract In this paper, Lingwu Long Jujube VC content was regarded as the research object, and a combination of hyperspectral imaging technique with chemometrics method was used to explore a rapid and nondestructive detecting method for fruit internal components. Vitamin C content of Long jujube was measured by high performance liquid chromatography (HPLC). A total of 164 Lingwu long jujubes of hyperspectral images in region of 400~1 000 nm were acquired. Then spectral curves were obtained by ENVI 4.8 software from the region of interest (ROI). The models were built for chemical value and spectral data by UnscramblerX 10.4 software. Outliers were to be eliminated by Monte Carlo cross validation method; Samples division was set partitioning based on joint X-Y distance(SPXY) method to improve the prediction performance of the model; The spectral's pretreatment was analyzed, such as Moving Average, Median Filter, Normalize, Baseline, multiple scatter correction (MSC), Detrending and standard normal variate (SNV) and so on; To reduce the amount and dimension of data, the feature wavelengths were extracted by competitive adaptive weighting algorithm (CARS), uninformative variable elimination (UVE) and continuous feeding Shadow algorithm (SPA) ; Compared to the models of full spectrum (FS) and the feature wavelengths extracted by CARS and UVE of PLSR and SVM built, the optimal model was determined. A total of 7 abnormal samples were eliminated using Monte Carlo cross validation method. After eliminating abnormal sample data, the samples were divided into calibration set and prediction set by SPXY method, and calibration samples is 117, and prediction samples is 40. The spectral pretreatment were studied by the 7 methods. The results showed that the model effect without spectral pretreatment was the best, and its R_c was 0.8779, and RMSECV was 0.0481; Without a preprocessing method by CARS, UVE and SPA method to reduce the dimensions, a total of 16 feature wavelengths were selected by CARS, which were 415, 487, 406, 631 636, 655, 660, 665, 670, 684, 689, 694, 723, 732, 747 and 881 nm. A total of 32 feature wavelengths were selected by UVE, which were 415, 406, 627, 631, 636, 651, 655, 660, 665, 670, 675. 679, 684, 689, 694, 699, 703, 708, 742, 747, 751, 756, 761, 766, 771, 775, 780, 785, 790, 795, 919 and 924 nm. A total of 3 feature wavelengths were selected by SPA, which were 401, 665 and 684 nm. Comparing models of the full band spectrum with the models of extracted characteristic wavelengths of PLSR and SVM, the UVE-SVM model is the best, and its R_c^2 is 0.847 1 and R_p^2 is 0.714 9, which indicates that UVE effectively reduces the dimension of the spectrum and simplifies the data processing. This study explores the application of hyperspectral imaging technology in the field of fruit, explores a new method for nondestructive testing of Lingwu Long Jujube VC content, provides a theoretical basis for visible and near infrared hyperspectral model established for the rapid detection of other components of fruit.

Keywords Visible near infrared; Hyperspectral imaging; Vitaman C(VC); Support vector machine; Non-destructive detection

(Received Dec. 25, 2017; accepted Apr. 2, 2018)

* Corresponding author