

脐橙 VC 含量近红外光谱消噪小波优化

夏俊芳 李小昱

(华中农业大学工程技术学院, 武汉 430070)

【摘要】 运用 db1~db10 等 10 个小波基 4 尺度分解对 120 个赣州脐橙样品的近红外光谱进行消噪处理, 分别建立了各种小波消噪光谱的脐橙维生素 C(VC)含量的 PLS 模型。通过 PLS 模型预测精度比较表明: 所有 db 小波基变换都能有效消除脐橙 VC 近红外光谱噪声, PLS 模型预测精度均高于不处理光谱建立的模型预测精度。其中, 消噪效果最好的小波基是 db5, 模型预测值与实测值的相关系数 R 达到 0.942 7, 内部交叉验证均方差 R_{MSECV} 仅为 2.02 mg/(100 g)。因此, 对脐橙 VC 含量近红外光谱消噪预处理的最优小波基是 db5 小波。

关键词: 脐橙 近红外光谱 小波基 消噪

中图分类号: O657.33; S121

文献标识码: A

Wavelet Optimization for Near-infrared Spectra Denoising of Vitamin C Content of Umbilical Orange

Xia Junfang Li Xiaoyu

(Engineering and Technology Department, Huazhong Agricultural University, Wuhan 430070, China)

Abstract

By using db1~db10 (10 in total) wavelet transform with the decomposing level of 4, the near-infrared spectra (NIRS) signals obtained from 120 umbilical orange samples were denoised, the PLS models of vitamin C (VC) content were established by each wavelet base. Through comparing the prediction accuracy of PLS models, it shows that all db wavelet transform can eliminate the noise in VC NIRS effectively, the prediction accuracy of PLS models are higher than models not being treated. When the wavelet was db5, the best prediction effect was obtained, with the correlation coefficient R between the prediction and true values being 0.942 7 and the expected variance R_{MSECV} being as low as 2.02 mg/(100 g). In conclusion, the db5 is the best wavelet for NIRS denoising of VC content in umbilical orange.

Key words Umbilical orange, Near-infrared spectra, Wavelet base, Denoising

引言

维生素 C(VC)含量是衡量脐橙营养品质的重要指标之一, 运用近红外光谱(NIRS)技术可以快速、无损地对其含量进行定量分析^[1]。近红外光谱分析仪器采集的原始光谱信号不仅包含与物质化学结构相关的信息, 还包括许多其他干扰因素所产生的信号噪声, 如果直接使用原始光谱信号进行定量分析, 这些噪声的存在会影响最终分析结果的准确性。因此, 必须对原始光谱进行预处理, 以便剔除噪声, 将非信息因素降至最小^[2~4]。

前期研究表明, 小波分析能有效地消除脐橙近红外光谱中的噪声^[5]。小波分析中所用到的小波函数 $\Psi(t)$ 具有多样性。在工程应用中, 小波分析的一个十分重要的问题是小波基最优选择问题。因为同一个问题用不同的小波分析会产生不同的结果。目前, 小波基选优没有理论方法, 只能通过用小波分析方法处理信号的结果与理论结果的误差来判定小波的优劣, 由此决定最优的小波函数^[6]。

在光谱信号的小波消噪分析中, 一般使用的是离散型小波变换(DWT)。其中, 具有正交性和离散小波变换功能的 Daubechies (dbN)小波在光谱分析

中最常用^[7~8]。Daubechies小波是一个小波族,从db1~db10有10个小波基,db小波的优化选择对光谱信号的消噪效果有较大影响。本文将采用这些小波基对脐橙VC含量的近红外光谱进行消噪预处理,根据偏最小二乘法(PLS)建立的消噪光谱校正模型的精度确定最佳db小波。

1 实验材料与方法

1.1 脐橙样品

脐橙样品为江西省定南县果园多棵树上随机采摘的150个成熟脐橙,采后用塑料袋封装,冷藏于温度为5℃、相对湿度为70%的人工气候箱中,每周定时随机抽取6~8个脐橙进行近红外光谱测定,并及时测定这批脐橙的VC含量,作为建立校正和验证模型时的实测值。脐橙样品直径65~95mm;质量123~360g;VC含量化学值20.8~89.15mg/100g。将150个样品分成2组,120个为校正集,30个为验证集。

1.2 主要仪器

BRUKER FT-NIR(VECTOR33N型)近红外光谱仪,配置OPUS分析软件、Pbs检测器、镀金积分球、样品旋转器、12mm石英样品杯。测定条件:波数12000~4000cm⁻¹,数据点1960个,分辨率8cm⁻¹,扫描次数64。

1.3 光谱采集

在室温条件下,VECTOR33N型近红外光谱仪预热20min后,将脐橙整果按最大横向直径处置于石英样品杯口部并适当压紧,每个样品分别在相对90°部位进行4次光谱扫描,取4次光谱的平均值作为原始光谱数据。

1.4 VC含量测定

根据GB/T 6195—86《水果、蔬菜维生素C含量测定法》,采用2,6-二氯酚滴定法测定脐橙VC含量。

1.5 小波消噪原理

小波变换是将交织在一起的不同频率组成的混合信号用分辨率不同的窗口分解成对应的不相同频率的块信号,并对大小不同的频率成分采用相应的时域(或空域)取样步长,从而不断“聚焦”对象的任意微小细节,对特殊频率范围的噪声或慢背景进行滤波处理。

小波是指满足一定条件的函数 $\Psi(t)$ 通过平移和伸缩产生的一个函数族

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \Psi\left(\left|\frac{t-b}{a}\right|\right) \quad (1)$$

式中 a ——尺度参数,用于控制伸缩

b ——平移参数,用于控制位置

$\Psi(t)$ ——小波基或小波母函数

在光谱分析信号的WT处理中,一般使用的是离散型小波变换,离散小波定义 $a = a_0 m$ ($a_0 > 1$, $m \in \mathbf{Z}$), $b = nb_0 a_0 m$ ($b_0 \in \mathbf{R}$, $n \in \mathbf{Z}$), 则

$$\Psi_{m,n}(t) = a_0^{-m/2} \Psi(a_0^{-m} t - nb_0) \quad (2)$$

一般取 $a_0 = 2$, $b_0 = 1$, 此时称为二进小波。

对于等波长间隔的 k 个离散光谱数据点 x_1, x_2, \dots, x_k , 其离散二进小波变换为

$$W_X(m, n) = \sum_{i=1}^k 2^{-m/2} \Psi(2^{-m} t_i - n) x_i \quad (3)$$

式(3)说明,不同的小波基 $\Psi(t)$, 对消噪效果是非常有影响的。 W 实际是将离散信号在小波函数上投影,不同的 m, n 代表不同分辨率和不同时间域,即分解尺度水平不同,消噪效果也不同。某信号小波分解后将从原来的空间投影到小波空间,由于WT的特点,在小波空间有部分系数特别小,对信号表示没有意义。如果将小的系数去除,在重构的信号中不会丢失有意义的信息。可采用小波系数代替原始数据,取较大的小波系数组成新的数据矩阵代替原始数据矩阵,可以大大降低数据量。既能有效地消除噪声,又能提高多元校正的速度,减少模型的随机性并提高预测精度^[9]。

1.6 近红外模型评价参数

采用PLS建模,模型评价指标为模型预测值与实测值的相关系数 R 和内部交叉验证均方差 R_{MSECV} 两个指标,计算公式为

$$R^2 = \left[1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \right] \quad (4)$$

$$R_{MSECV} = \sqrt{\frac{\sum_{i=1}^N [(\hat{y}_i - y_i)/y_i]^2}{N}} \quad (5)$$

式中 N ——样品数

\hat{y}_i ——第 i 个样品的VC含量预测值

y_i ——第 i 个样品的VC含量实测值

在优化模型时,对同一样品集的VC含量,相关系数 R 越大、内部交叉验证均方差越小,则PLS算法所提取的光谱信息与VC含量的相关性越好,所建模型的预测能力和稳健性越高。

2 结果与讨论

2.1 光谱小波包消噪

首先将每条样品的近红外平均光谱在OPUS分析软件中变换成数据点的格式,再逐条导入

Matlab 6.5 软件的小波工具箱中,选择一维小波包,分别选取 db1~db10 小波对每条光谱数据进行消噪处理。消噪时将分解尺度固定为 4^[1],根据小波包的特性自动量化全局阈值,这个阈值是在置零系数的百分比和压缩的信号能量百分比之间进行折中得到的,这意味着小波包分解后所有小于阈值的高频系数将被舍去^[9]。

2.2 dbN 小波基的优化

分别将经过 db1~db10 小波 4 尺度分解消噪的脐橙近红外光谱和 VC 含量标准值导入 OPUS 软件中,设置建模波段在 7 501.7~4 245.5 cm⁻¹ 范围内^[5],用 PLS 法建立 120 个脐橙样品的 VC 含量定量校正模型。

将优化研究结果按模型精度的高低顺序排列成表 1,预测值与标准值的相关系数高、内部交叉验证均方差小的 PLS 模型精度高,则光谱预处理效果好。否则,模型预测精度低,光谱预处理效果差。

表 1 dbN 小波消噪预处理光谱 PLS 建模预测结果比较

Tab.1 Prediction results on PLS models of spectra denoised by different wavelet bases (dbN)

序 号	预处理 方法	优化波段 /cm ⁻¹	主成 分数	相关 系数 R	预测均方差 R_{MSECV} /mg·(100 g) ⁻¹
1	db5-4	7 501.7~7 497.9	9	0.942 7	2.02
		6 101.7~5 449.8			
2	db3-4	6 101.7~5 449.8	8	0.934 6	2.04
3	db6-4	6 101.7~5 449.8	8	0.934 4	2.08
4	db4-4	7 501.7~5 449.8	7	0.915 4	2.09
		4 601.3~4 245.5			
5	db10-4	6 101.7~5 449.8	7	0.915 1	2.08
6	db8-4	5 453.7~4 245.5	10	0.914 3	2.32
7	db2-4	7 501.7~7 497.9	7	0.906 3	2.14
		5 453.7~4 597.5			
8	db7-4	7 501.7~6 097.8	6	0.881 8	2.23
		4 601.3~4 245.5			
9	db1-4	6 101.7~5 449.8	3	0.878 5	2.49
10	db9-4	5 453.7~4 245.5	7	0.842 7	2.38
11	不处理	5 874.1~4 245.5	3	0.825 7	2.96

2.3 优化结果讨论

由表 1 可知,所有的 db 小波基都能消除脐橙 VC 含量近红外光谱噪声,所建立的 PLS 模型预测精度都高于不预处理光谱建立的 PLS 模型预测精度,说明 db 小波基消噪是脐橙 VC 含量近红外光谱的有效预处理方法。不同的 db 小波基消噪,PLS 模型的预测精度有明显差异。消噪效果好的 db 小波基依次是 db5-4、db3-4、db6-4、db4-4,其次是

db10-4、db8-4 和 db2-4,消噪效果稍差的 db 小波基是 db7-4、db1-4 和 db9-4。其中,消噪效果最好的 db 小波基是 db5-4,其 PLS 模型的相关系数 R 达到 0.942 7,内部交叉验证均方差 R_{MSECV} 仅为 2.02 mg/100 g。消噪效果最差的 db 小波基是 db9-4,PLS 模型的相关系数 R 为 0.842 7,略高于不处理的 0.825 7;内部交叉验证均方差 R_{MSECV} 高达 2.38 mg/100 g,略低于不处理的 2.96 mg/100 g。

因此,若采用 db 小波对脐橙 VC 含量近红外光谱消噪,则应优先选择 db3、db4、db5 和 db6 小波基,最好选择 db5 小波基。

2.4 优化模型验证

利用 db5 小波基消噪优化建立的 PLS 校正模型,对验证集的 30 个脐橙样品 VC 含量进行预测,考察校正模型的预测精度和稳健性,验证结果如表 2 所示。

表 2 db5 小波基 PLS 模型验证结果

Tab.2 Validation results on PLS models of spectra denoised by db5 wavelet base mg/(100 g)

序号	预测值	标准值	残差	序号	预测值	标准值	残差
1	53.851	52.89	1.96	16	48.887	50.11	-1.22
2	63.209	65.93	-1.72	17	58.401	57.97	0.43
3	73.889	73.56	0.33	18	48.121	49.05	-0.93
4	65.863	65.36	0.50	19	51.261	53.33	-3.07
5	55.102	58.14	-3.04	20	51.381	55.60	-3.21
6	65.863	65.36	0.50	21	58.197	62.24	-3.04
7	55.491	52.42	3.07	22	49.752	53.65	-3.90
8	53.376	53.17	0.20	23	47.522	43.56	2.96
9	59.635	61.05	-1.42	24	45.012	43.58	0.43
10	47.478	48.65	-1.17	25	57.405	56.97	0.44
11	58.405	53.24	3.16	26	51.161	51.30	-0.14
12	47.709	48.49	-0.78	27	58.544	56.69	1.85
13	43.657	47.68	-3.02	28	51.358	51.39	-0.04
14	59.474	58.70	0.77	29	66.128	63.61	1.51
15	55.881	52.70	3.18	30	58.465	59.74	-1.27

在没有剔除异样品的前提下,模型验证预测值与标准值的相关系数为 0.918 4,预测均方差为 0.437 mg/(100 g),残差平均值为 -0.223 mg/(100 g),预测精度较高,完全满足农产品检测精度要求,所建的 PLS 预测模型是稳健的。

3 结束语

小波包变换是脐橙 VC 近红外光谱消噪预处理的有效方法。所有的 db 小波基变换都能消除脐橙 VC 近红外光谱噪声,所建立的 PLS 模型预测精度

均高于不处理光谱建立的模型预测精度。根据 PLS 模型预测精度高低,各小波基消噪效果的顺序依次为 db5、db3、db6、db4、db10、db8、db2、db7、db1、db9。消噪效果较好的小波基是 db5、db3、db6,其中,消噪

效果最好的小波基是 db5,其 PLS 模型的相关系数 R 达到 0.9427、内部交叉验证均方差 R_{MSECV} 仅为 2.02 mg/(100 g)。因此,对脐橙 VC 含量近红外光谱消噪预处理的最优小波基是 db5 小波。

参 考 文 献

- 夏俊芳,李小昱,李培武,等. 基于小波变换的柑橘维生素 C 含量近红外无损检测方法[J]. 农业工程学报, 2007, 23(6):170~174.
Xia Junfang, Li Xiaoyu, Li Peiwu, et al. Approach to nondestructive measurement of vitamin C content of orange with near-infrared spectroscopy treated by wavelet transform[J]. Transactions of the Chinese Society of Agricultural Engineering, 2007, 23(6): 170~174. (in Chinese)
- 袁洪福,陆婉珍. 现代光谱分析中常用的化学计量学方法[J]. 现代科学仪器, 1998(5):6~9.
Yuan Hongfu, Lu Wanzhen. Popular chemometrics used in spectrometric analysis[J]. Modern Scientific Instruments, 1998(5): 6~9. (in Chinese)
- 陈华才,陈星旦. 近红外光谱在药物领域的应用与研究进展[J]. 中国计量学院学报, 2003, 14(4):261~267.
Chen Huacai, Chen Xingdan. Applications and researches of near-infrared spectroscopy in the pharmaceutical filed [J]. Journal of China Institute of Metrology, 2003, 14(4): 261~267. (in Chinese)
- 彭玉华. 小波变换与工程应用[M]. 北京:科学出版社, 2002.
- 夏俊芳,李小昱,李培武,等. 不同预处理对近红外光谱检测脐橙 VC 含量的影响[J]. 农业机械学报, 2007, 38(6): 107~111.
Xia Junfang, Li Xiaoyu, Li Peiwu, et al. The effect of different pretreatment method to nondestructive measure vitamin C content of umbilical orange with near-infrared spectroscopy [J]. Transactions of the Chinese Society for Agricultural Machinery, 2007, 38(6): 107~111. (in Chinese)
- 陈斌. 基于小波变换的方便面含油率近红外光谱检测技术[J]. 农业机械学报, 2001, 32(6):74~76.
Chen Bin. Improvement on predicting precision of oil content in instant noodles with near-infrared spectroscopy treated by wavelet transformation [J]. Transactions of the Chinese Society for Agricultural Machinery, 2001, 32(6): 74~76. (in Chinese)
- 田高友,袁洪福,刘慧颖,等. 小波变换用于近红外光谱性质分析[J]. 分析化学, 2004, 32(9):1125~1130.
Tian Gaoyou, Yuan Hongfu, Liu Huiying, et al. Application of wavelet transform on the near infrared analysis[J]. Chinese Journal of Analytical Chemistry, 2004, 32(9): 1125~1130. (in Chinese)
- 田高友,袁洪福,刘慧颖,等. 小波变换在近红外光谱分析中的应用进展[J]. 光谱学与光谱分析, 2003, 23(6):1111~1114.
Tian Gaoyou, Yuan Hongfu, Liu Huiying, et al. The application of wavelet transform in near infrared spectroscopy [J]. Spectroscopy and Spectral Analysis, 2003, 23(6): 1111~1114. (in Chinese)
- 夏俊芳,李小昱,李培武,等. 小波变换在脐橙 VC 含量近红外光谱预测中的应用[J]. 中国农业科学, 2007, 40(8): 1760~1766.
Xia Junfang, Li Xiaoyu, Li Peiwu, et al. Application of wavelet transformation in umbilical orange vitamin C content prediction with near-infrared spectroscopy[J]. Scientia Agricultural Sinica, 2007, 40(8):1760~1766. (in Chinese)
- 林新,牛智有,马爱丽,等. 绿茶游离氨基酸总量近红外光谱定量分析模型优化[J]. 农业机械学报, 2008, 39(10):144~147.
Lin Xin, Niu Zhiyou, Ma Aili, et al. Optimization on quantitative analysis models of free amino acids content in green tea by near infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2008, 39(10): 144~147. (in Chinese)