

# 脐橙总糖近红外光谱模型传递研究

胡润文, 夏俊芳\*

(华中农业大学工学院, 湖北 武汉 430070)

**摘要:**为实现近红外光谱模型在同类光谱仪器间的共享,以脐橙总糖为例,在主仪器上建立最优的偏最小二乘(PLS)近红外光谱模型,采用斜率截距(Slope/Bias)校正法和直接校正(DS)算法把主仪器上建立的模型传递到从仪器上,并探讨标准化样品个数对模型传递效果的影响。研究表明:在主仪器建立的模型经DS算法传递后,从仪器预测集的预测标准差(RMSEP)为0.448%,经Slope/Bias校正法传递后的预测标准差(RMSEP)为0.756%,DS算法更优于Slope/Bias校正法,利用直接校正法实现脐橙总糖傅里叶近红外光谱模型传递是可行的。

**关键词:**脐橙;总糖;近红外;斜率截距法;直接校正法;模型传递

## Transfer of NIRS Calibration Model for Determining Total Sugar Content in Navel Orange

HU Run-wen, XIA Jun-fang\*

(College of Engineering, Huazhong Agricultural University, Wuhan 430070, China)

**Abstract:** In order to achieve NIRS (near infrared reflectance spectroscopy) calibration model sharing among similar NIRS instruments, an optimal NIRS partial least square (PLS) regression model was established using the major instrument and transfer to the minor instrument by either slope bias correction (SBC) or direct standardization (DS). In addition, the dependence of calibration transfer on the number of standardization samples was explored. The results showed that the root mean squared errors of prediction (RMSEP) for the calibration model were 0.448% and 0.756% after standardization by DS and SBC, respectively. Thus, the DS method was better than the SBC method and could be used to successfully transfer the calibration model

**Key words:** navel orange; total sugar; near infrared; slope/bias correction; direct standardization; calibration transfer

中图分类号: O657.33

文献标识码: A

文章编号: 1002-6630(2012)03-0028-05

脐橙营养丰富,含有人体所必需的各类营养成分,糖度是决定其风味最重要的指标,如何对其进行快速准确的测定具有重要意义。近红外光谱分析技术(near infrared reflectance spectroscopy, NIRS)具有处理样品简单、分析速度快、无污染、无损等特点,已在果品成分含量的检测中得到广泛应用<sup>[1-6]</sup>。夏俊芳等<sup>[7-9]</sup>研究了基于近红外光谱的脐橙内部品质的无损检测方法,但是不同的近红外仪器间由于仪器元器件的选用、装配误差和测量环境的不同,使得相同样品的光谱数据存在一定的差异,导致一台仪器上建立的模型无法在另一台仪器上使用<sup>[10]</sup>。要解决模型在近红外仪器间的通用,最经济有效的方法是利用化学计量学算法来实现模型传递。模

型传递的本质是克服样品在不同仪器上的测量光谱的不一致性,经过传递的模型即可用于新的样品或新的测试环境<sup>[11]</sup>。

近年来,国内外学者对近红外光谱仪之间的模型传递做了许多相关研究。Alamar<sup>[12]</sup>、Greensill<sup>[13]</sup>、Roggo<sup>[14]</sup>等分别研究了苹果和橘子可溶性固形物以及甜菜糖度近红外光谱模型的传递,刘贤<sup>[15]</sup>等研究了青贮饲料的近红外光谱模型在不同类型光谱仪之间的传递,褚小立等<sup>[16]</sup>针对国产的CCD近红外光谱仪,采用普鲁克分析法解决了模型传递问题,林振兴等<sup>[17]</sup>通过分段直接校正(PDS)算法有效地解决了样品温度对近红外光谱的影响。赵龙莲<sup>[18]</sup>、李庆波<sup>[19]</sup>、陈斌<sup>[20]</sup>等分别将移动窗口支持向量回归机

收稿日期: 2011-02-24

基金项目: 华中农业大学自主科技创新基金项目(2009SC015)

作者简介: 胡润文(1985—),男,硕士研究生,研究方向为智能化检测技术。E-mail: hurunwen@126.com

\*通信作者: 夏俊芳(1963—),女,教授,博士,研究方向为智能化检测技术、现代农业装备设计与测控。

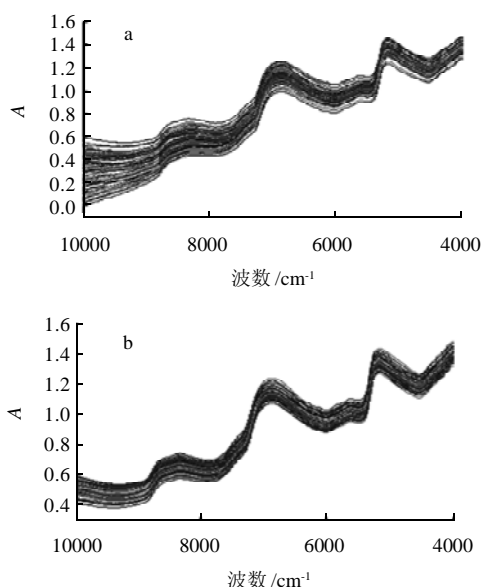
E-mail: xjf@mail.hzau.edu.cn

(SVR)方法、直接校正(DS)算法、专利算法应用于多元校正中的模型传递,均取得了不错的效果,可见经模型传递后提高了模型的预测精度。但是关于脐橙总糖近红外光谱模型在两台傅里叶光谱仪间的传递研究较少。本实验以赣南脐橙为对象,采用 Slope/Bias 校正法和 DS 算法两种方法,研究脐橙总糖含量的近红外光谱模型在两台傅里叶近红外光谱仪间的共享。

## 1 材料与方 法

### 1.1 样品与光谱的采集

实验样品从江西省定南县果园里随机采摘得到,共计 89 个。将采摘到的脐橙样品擦拭干净,分别在 BRUKER Fr-NIR(VECTOR22N 型)傅里叶近红外光谱仪和 BRUKER Fr-NIR(VECTOR33N 型)傅里叶近红外光谱仪上采集光谱(两台仪器属于德国布鲁克仪器公司生产同一系列近红外光谱仪)。设定光谱扫描范围为  $10000\sim 4000\text{cm}^{-1}$ ,镀金漫反射体作背景,分辨率为  $8\text{cm}^{-1}$ ,扫描次数为 64。将脐橙整果的横向最大直径部位放置于旋转样品杯上进行扫描,在脐橙圆周上相对  $90^\circ$  方向作 4 次光谱采集,尽量避免样品表面的斑痕,得到每个样品的 4 条普通光谱,利用近红外分析软件 OPUS5.5 将每个样品的 4 条普通光谱平均,将每个样品的平均光谱作为原始光谱数据。图 1 为脐橙样品在两台仪器上的近红外光谱图。



a. VECTOR22N 型红外光谱仪; b. VECTOR33N 型红外光谱仪。

图 1 脐橙样品近红外光谱

Fig.1 NIR spectra of navel orange samples

### 1.2 样品总糖的化学测定

样品总糖的测定采用硫酸苯酚比色法<sup>[21]</sup>。称取每个脐橙样品的质量,用组织捣碎机分别将样品捣碎,准

确称取 10.0g 样品,转移到 100mL 容量瓶中,加蒸馏水至刻度、摇匀,用滤纸过滤,取滤液 1 mL,定容到 100mL 待测,然后取 1mL 样液,1mL 6% 的苯酚,1mL 蒸馏水,5 mL 浓硫酸,摇匀、冷却后比色。最后测得的样品总糖含量的范围为 9.823%~14.7163%。

### 1.3 样品集的划分

Kennard-Stone 算法是通过计算样本之间的欧氏距离(euclidean distance, ED)来选择样品的<sup>[22]</sup>,选择样本间距离大的样品,这样所选的样品集样本张成的空间较大,覆盖范围也较大,具有代表性。本实验采用 Kennard-Stone 算法将 89 个样品划分为校正集和预测集,其中校正集包含 70 个样品,用来在主仪器上建立脐橙含糖量的近红外光谱模型,预测集包含 19 个样品,用来检验模型的优劣。计算模型传递参数的标准化样品集在校正集中选取。Kennard-Stone 算法对样品的选取过程在 Matlab 中完成。

### 1.4 模型传递方法原理

#### 1.4.1 Slope/Bias 校正法

Slope/Bias 校正法通过校正主仪器上的模型来实现模型传递的。利用校正集样品光谱在主仪器上建立校正模型  $b$ ,用此模型分别预测标准化样品在两台仪器上测得的标准化样品光谱矩阵  $T_m$ 、 $T_s$ ,得到浓度值  $y_m$ ,  $y_s$ ,则:

$$y_m = T_m \times b \quad (1)$$

$$y_s = T_s \times b \quad (2)$$

用一元线性回归方程来拟合  $y_m$  和  $y_s$ ,求得此线性方程的普通最小二乘解,即得这一模型的斜率( $k$ )和截距( $l$ ):

$$k = \frac{S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}}{2S_{xy}} \quad (3)$$

$$l = \bar{y}_m - k \times \bar{y}_s \quad (4)$$

其中:

$$S_{xx} = \sum (y_{sj} - \bar{y}_s)^2, S_{yy} = \sum (y_{mj} - \bar{y}_m)^2$$

$$S_{xy} = \sum (y_s - \bar{y}_s)(y_{mj} - \bar{y}_m)^2$$

最后通过主仪器上建立的模型  $b$  和 Slope/Bias 校正项,就可以直接由从仪器上得到的光谱  $X_s$ ,计算其浓度值:

$$y = l + k(X_s \times b) \quad (5)$$

#### 1.4.2 DS 算法

DS 算法是一种全光谱校正方法,它利用转换矩阵  $F$  将从仪器上测得的未知样品光谱  $X_s$  转换为  $X_{st}$ ,  $X_{st}$  与主仪器上测得的光谱在理论上是一致的。

从校正集中选取一定量的样品作为标准样品,分别主仪器上测得光谱矩阵  $T_m$ ,从仪器上测得光谱矩阵  $T_s$ ,通过传递矩阵  $F$  将  $T_s$  转换成  $T_m$ ,使二者相匹配。其中矩阵  $F$  描述了主、从两种测定条件下光谱的差异。

由  $T_m = T_s \times F + E$ ,  $E$  为残差矩阵, 则通过最小二乘法计算可得到转换矩阵  $F = T_s^{-1} \times T_m$ , 其中  $T_s^{-1}$  为  $T_s$  的广义逆,  $F$  为  $m \times m$  矩阵( $m$  为波长点数)。对在从仪器上测得的未知样品光谱矩阵  $X_{su}$ , 利用  $F$  将  $X_{su}$  转化成与主仪器匹配的光谱矩阵  $X_{st}$ 。

$$X_{st} = F \times X_{su} \quad (6)$$

再由主仪器上建立的校正模型  $b$  计算最终结果。

$$y = X_{st} \times b \quad (7)$$

### 1.5 模型传递的评价

为了找出更优的模型传递方法, 采用主仪器上建立的最优 PLS 模型分别预测传递前后从仪器上的预测集光谱, 用相关系数( $R$ )、内部交叉验证均方差(RMSECV)、预测标准差(RMSEP)来评价模型的效果。

## 2 结果与分析

### 2.1 预测模型的建立

应用 OPUS 5.5 定量分析软件(德国布鲁克仪器公司), 将校正集样品光谱与其化学分析值结合, 建立主仪器上的脐橙糖度偏最小二乘法(PLS)校正模型。为了消除仪器状态、样品状态与测量条件对采集到的光谱的影响, 采用以下 10 种不同预处理方法对模型进行优化, 包括: 消除常数偏移量(COE)、直线差值(SLS)、矢量归一化(VN)、最小-最大归一化(MMN)、多元散射校正(MSC)、一阶导数(FD)、二阶导数(SD)、FD+SLS、FD+VN、FD+MSC。

表 1 不同预处理方法对模型的影响

Table 1 Effects of different data preprocessing methods on the model

预处理方法	相关系数 $R$	内部交叉验证均方差 /%	最佳主成分
NONE	0.862	0.618	10
COE	0.871	0.624	10
SLS	0.861	0.62	9
VN	0.879	0.582	9
MMN	0.883	0.572	10
MSC	0.905	0.519	9
FD	0.844	0.655	8
SD	0.771	0.777	5
FD+SLS	0.818	0.702	8
FD+VN	0.861	0.621	7
FD+MSC	0.857	0.629	7

由表 1 可知, 不同预处理对模型建立效果有影响, 经过 MSC 预处理后的模型比没有经过预处理所建的模型更好, 其所选的光谱波段范围为  $10001.5 \sim 8049.8 \text{ cm}^{-1}$  和  $5450.1 \sim 4246.7 \text{ cm}^{-1}$ , 相关系数  $R$  为 0.905, 内部交叉验证均方差为 0.519%, 最佳主成分数为 9。校正集的真实值和预测值的关系如图 2 所示。

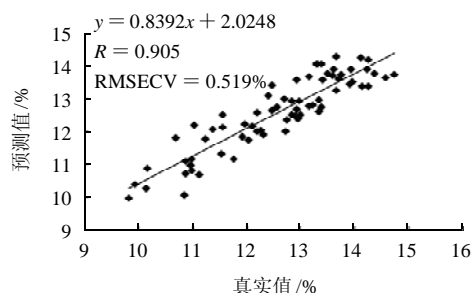


图 2 校正集中真实值与预测值散点图

Fig.2 Relationship between predicted and chemical values for the calibration set

### 2.2 模型预测结果

利用在主仪器上已建好的 PLS 模型分别预测主从仪器上 19 个预测集样品的光谱, 结果见表 2, 主仪器上建立的模型预测主仪器预测样品集的光谱, 预测值与真实值的相关系数  $R$  为 0.909, 预测标准差 RMSEP 为 0.434%, 预测效果良好, 但用此模型直接预测从仪器上预测集的光谱,  $R$  降为 0.688, RMSEP 则变为 3.675%, 预测结果产生较大差异。这是由于预测集样品光谱是在不同型号仪器上测得的, 仪器对光谱的测量影响造成图谱之间的差异较大, 主仪器上建立的模型已无法满足预测要求, 模型须经过维护或修正从仪器上的光谱才能得到满意的结果。

表 2 主仪器上模型分别预测两台仪器预测集的结果

Table 2 Predicted results for the predication sets from the major and minor instruments

目标值	主仪器	从仪器
$R$	0.909	0.688
RMSEP/%	0.434	3.675

### 2.3 标样个数对模型传递的影响

为了取得较好的模型传递结果, 本实验采用 Kennard-Stone 算法<sup>[21]</sup>依次从校正集样品中选取 10, 20, 30, …, 70 个样品作为标准化样品集, 并同时采用 Slope/Bias 校正法和 DS 算法来进行研究。Slope/Bias 校正法通过计算 Slope 和 Bias 来校正主仪器上模型的, 此过程在 Excel 中进行。DS 算法利用标准化样品求得模型传递参数, 从仪器上预测集样品通过此参数可求得传递后的光谱, DS 算法是在 Matlab 中的 PLS\_toolbox5.8 中进行的。图 3 是不同标准化样品数对两种算法的影响。标样数对 Slope/Bias 校正法影响较小, 随着样品数的增加, RMSEP 没有太大变化, 而经 DS 算法传递后测定预测集脐橙含糖量的 RMSEP 随着标样数的增加而减小, 直到标样数为全部校正集样品时, 标准化样品集包含的信息足够充分, RMSEP 减小到最小值, 模型传递效果达到最优。

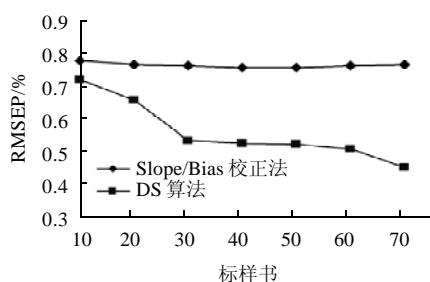


图3 预测集的RMSEP随标样数的变化

Fig.3 Change in RMSEP with the number of standardization samples

#### 2.4 Slope/Bias 校正法与 DS 算法的结果与比较

两种方法各自选取最佳标准化样品个数进行模型传递, 利用主仪器上建立的最优 PLS 模型分别预测经两种方法传递后的从仪器预测集光谱。预测集的预测值与真实值的散点图如图 4 所示。

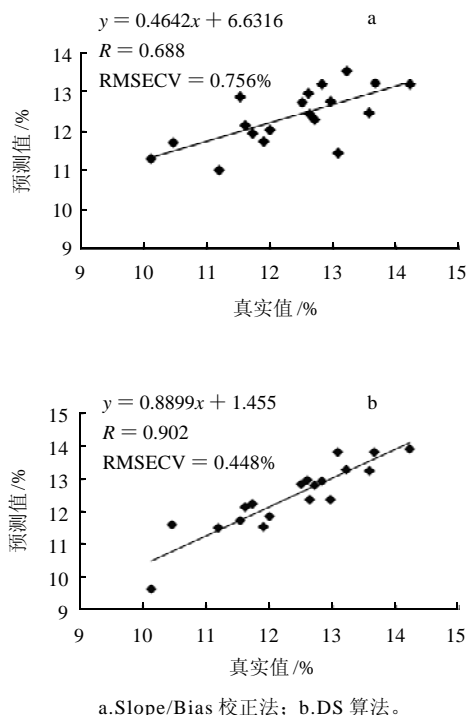


图4 经传递后真实值与预测值的散点图

Fig.4 Relationship between predicted and chemical values after transfer

由图 4 可知, 经模型传递后模型的预测能力均有所改善。Slope/Bias 校正法传递后, 相关系数  $R$  虽然没有变化, 但 RMSEP 由原来的 3.675% 减小到 0.756%, 经 DS 算法传递后的相关系数  $R$  和 RMSEP 分别为 0.902 和 0.448%, 拟合效果明显要好于 Slope/Bias 法, 其预测能力比 Slope/Bias 校正法提高了 40.74%。这可能是由于两

台仪器测量光谱时外界因素对测量结果的干扰较大, 而 Slope/Bias 校正法只是一元校正法, 仅仅对预测结果进行直线拟合校正, 对模型的修正有限, 从而没有取得较好的效果, DS 算法由于是全光谱校正法, 能消除仪器间包括测量环境的影响。因此, 在仪器结构差异及外部环境因素影响较大时, DS 算法要更好于 Slope/Bias 校正法。

### 3 结 论

3.1 建立了脐橙总糖偏最小二乘(PLS)近红外光谱模型, 并比较了不同预处理方法对模型的影响, 经多元散射校正(MSC)预处理后的模型更优, 具有更高的精度和稳定性。

3.2 比较了不同标样个数对模型传递过程的影响, 标样个数对 Slope/Bias 校正法影响较小, 但 DS 算法的效果随着标样个数的增加而变好, 当标准化样品为全部校正集样品时, 模型传递效果最优。

3.3 研究采用 Slope/Bias 校正法和 DS 算法实现脐橙糖度模型在两台傅里叶近红外光谱仪之间的传递。由于两台仪器上测得的光谱受仪器条件及环境变化引起的差异较大, 通过 Slope/Bias 校正法对预测值进行一元校正的能力有限, 而 DS 算法作为多变量标准化方法, 对仪器差异、测量环境的改变等因素引入的光谱差异的校正是可行的。经 DS 算法传递后预测值与真实值之间具有较大相关系数  $R(0.902)$  和较小预测标准差 RMSEP(0.448%), 可以实现脐橙总糖模型在两台傅里叶近红外光谱仪之间的共享。

### 参考文献:

- [1] 应义斌, 刘燕德, 傅霞萍, 等. 苹果有效酸度的近红外漫反射无损检测[J]. 农业机械学报, 2004, 35(6): 124-126.
- [2] XIE Lijuan, HUANG Haibo, YING Yibin, et al. Effect of wavelet denoising techniques on the determination of navel orange sugar content with near-infrared spectra[C]//Optics for Natural Resources, Agriculture, and Foods. Boston: Yud-Ren Chen, 2006.
- [3] LIU Yande, SUN Xudong, ZHOU Jianmin, et al. Linear and nonlinear multivariate regressions for determination sugar content of intact Gannan navel orange by Vis-NIR diffuse reflectance spectroscopy[J]. Mathematical and Computer Modelling, 2010, 51(11): 1438-1443.
- [4] TEWARI J C, DIXIT V, CHO B K, et al. Determination of origin and sugars of citrus fruits using genetic algorithm, correspondence analysis and partial least square combined with fiber optic NIR spectroscopy[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2008, 71(3): 1119-1127.
- [5] 马兰, 夏俊芳, 张战锋, 等. 基于小波变换的番茄总糖近红外无损检测[J]. 农业工程学报, 2009, 25(10): 350-354.
- [6] NISHIZAWA T, MORI Y, FUKUSHIMA S, et al. Non-destructive analysis of soluble sugar components in strawberry fruits using near-infrared spectroscopy[J]. Japan Nippon Shokuhin Kagaku Kogaku Kaishi, 2009,

- 56(4): 229-235.
- [7] 夏俊芳. 基于近红外光谱的贮藏脐橙品质无损检测方法研究[D]. 武汉: 华中农业大学, 2007.
- [8] 夏俊芳, 李小昱, 李培武, 等. 基于小波变换的柑橘维生素C含量近红外光谱无损检测方法[J]. 农业工程学报, 2007, 23(6): 170-174.
- [9] 夏俊芳, 李培武, 李小昱, 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.
- [10] 严衍禄, 赵龙莲, 韩东海, 等. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005: 199-200.
- [11] 褚小立, 袁洪福, 陆婉珍. 光谱多元校正中的模型传递[J]. 光谱学与光谱分析, 2001, 21(6): 881-885.
- [12] ALAMAR M C, BOBELYN E, LAMMERTYN J, et al. Calibration transfer between NIR diode array and FT-NIR spectrophotometers for measuring the soluble solids contents of apple[J]. Postharvest Biology and Technology, 2007, 45(1): 38-45.
- [13] GREENSILL C V, WALSH K B. Calibration transfer between miniature photodiode array-based spectrometers in the near infrared assessment of mandarin soluble solids content[J]. Journal of Near Infrared Spectroscopy, 2002, 10(1): 27-35.
- [14] ROGGO Y, DUPONCHEL L, NOE B, et al. Sucrose content determination of sugar beets by near infrared reflectance spectroscopy. Comparison of calibration methods and calibration transfer[J]. Journal of Near Infrared Spectroscopy, 2002, 10 (2): 137-150.
- [15] 刘贤, 董苏晓, 韩鲁佳, 等. 青贮饲料近红外光谱分析模型转移研究[J]. 农业机械学报, 2009, 40(5): 153-157.
- [16] 褚小立, 袁洪福, 陆婉珍, 等. 普鲁克分析用于近红外光谱仪的分析模型传递[J]. 分析化学, 2002, 30(1): 114-119.
- [17] 林振兴, 邬蓓蕾, 王豪, 等. PDS用于不同温度下的近红外光谱模型传递研究[J]. 分析测试学报, 2008, 27(12): 1330-1333.
- [18] 赵龙莲, 李军会, 张文娟, 等. 基于SVR的傅里叶变换型近红外光谱仪间数学模型传递的研究[J]. 光谱学与光谱分析, 2008, 28(10): 2299-2303.
- [19] 李庆波, 张广军, 徐可欣. DS算法在近红外光谱多元校正模型传递中的应用[J]. 光谱学与光谱分析, 2007, 27(5): 873-876.
- [20] 陈斌, 王豪. 专利算法在白葡萄酒精度近红外光谱分析模型转移中的应用[J]. 红外技术, 2006, 28(4): 245-248.
- [21] 于瑞涛, 朱鹏程, 陶燕铎, 等. 苯酚硫酸法测定迷果芹多糖的含量[J]. 分析实验室, 2008, 12(27): 222-224.
- [22] FILHO H A D, GALVAO R K H, ARAUJO M C U, et al. A strategy for selecting calibration samples for multivariate modelling[J]. Chemometrics and Intelligent Laboratory Systems, 2004, 72: 83-91.