

不同傅里叶近红外仪器间(积分球漫反射测量)的 模型传递及误差分析

王宇恒¹, 胡文雁¹, 宋鹏飞², 束茹欣³, 杨凯³, 王萝萍², 赵龙莲¹, 李军会^{1*}

1. 中国农业大学信息与电气工程学院, 北京 100083
2. 云南中烟工业有限责任公司技术中心, 云南 昆明 653100
3. 上海烟草集团有限责任公司, 上海 200082

摘要 模型传递是近红外光谱分析技术中一个关键的共性基础技术问题, 通过在同一工作原理的两台仪器之间寻求可行的数学方法, 使得在一台仪器上建立的模型能够应用于另外一台仪器样品光谱响应的预测, 对近红外技术的实际应用具有重要意义。以 150 份烤烟作为试验样品, 以两台布鲁克公司 MPA 近红外光谱仪, 一台热电公司 Antaris 近红外光谱仪作为研究对象, 通过积分球漫反射检测技术获得光谱数据。采用一阶导数(first-order derivative, 1st Der)和标准正态变量变换(standard normal variate, SNV)对光谱数据进行处理分析, 计算不同仪器间光谱的残差值、残差一阶矩、残差信号概率密度和最大信噪比等参数, 并采用偏最小二乘法(partial least squares, PLS)建立烤烟总糖含量数学模型, 检验模型传递效果。结果表明, 一阶导数具有降低残差一阶矩, 将仪器偏差信号转换为标准高斯分布的优点, 但同时会降低信噪比。标准正态变量变换同样可以降低一阶矩, 同时可大幅度提高信噪比, 但无法将仪器偏差信号转换为标准高斯分布, 需要进一步的信号处理。一阶导数与 SNV 相结合可保留两种方法的优点, 同时在一定程度上弥补每种方法单独处理的缺点, 是一种可以消除以积分球漫反射作为光谱测量方式的因仪器厂家或型号不同、使用年限不同等原因所产生的噪声的处理方法, 可实现傅里叶型近红外光谱仪之间的模型传递效果的明显改善。

关键词 模型传递; 一阶导数; 标准变量正态变换

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)03-0964-05

引言

近红外分析技术(near infrared, NIR)是近些年发展起来的绿色分析技术, 具备简单、高效、快速等优点, 适于过程分析和在线分析, 是质量控制的理想手段, 在农业、食品、石油等领域的应用日益广泛^[1]。性能稳定可靠的近红外光谱仪是该技术的基础和前提。光谱仪是由多种部件组成的, 各台仪器中同种部件不会完全一致, 不同仪器的组装过程及使用时长存在差异; 所有这些差异均导致不同仪器间在光谱响应上有差异。这些光谱差异是影响不同仪器间模型通用的主要障碍。解决的途径之一是通过提高仪器制造水平尽可能解决仪器间的不一致性, 其二是通过数学方法处理以消除这些差异。

积分球漫反射法是近红外光谱测量方式之一, 它可以消

除镜面反射光的影响, 增加信号强度, 提高信噪比, 降低由于入射光的形状和角度等变化产生的影响, 增强仪器的稳定性并降低由于样品的不均匀和空间位置变化等对光谱测量的影响, 从而提高了测量的重复性^[2], 是本文所介绍的无标样模型传递方法得以实现的重要基础。

国外从 20 世纪 80 年代起, 就开始有关于近红外数学模型传递的研究报道。主要方法: 直接校正法(direct standardization, DS)^[3]、分段直接校正法(piecewise direct standardization, PDS)^[4]、Shenk 专利算法^[5]、FIR 法(finite impulse response, FIR)^[6]和人工神经网络方法^[7]等。

近年来, 国内外对近红外光谱的模型传递进行了较为深入的研究。Kirsten 等通过变窗宽多元散射校正(MSC)在两台近红外光谱仪间实现了直接的模型传递^[8]。Payal 等通过多条光纤让一台近红外光谱仪能够同时获得多个生物反应器中的不同浓度的葡萄糖和乳酸光谱信号并建立复合模型实现

收稿日期: 2016-03-08, 修订日期: 2017-01-05

基金项目: 国家自然科学基金项目(61144012)和上海烟草集团科技项目资助

作者简介: 王宇恒, 1992 年生, 中国农业大学信息与电气工程学院硕士研究生 e-mail: 1016726519@qq.com; yuhengw1314@gmail.com

* 通讯联系人 e-mail: caunir@cau.edu.cn

了不同仪器间的模型通用^[9]。Robert 等提出了分段正交信号校正方法处理由两台不同近红外光谱仪获得的光谱集,然后利用处理过的光谱数据建立多元校正模型实现模型传递^[10]。褚小立等通过传统标准正态变量变换与 PDS 结合的方法,很好地实现玉米样品的 PLS 校正模型在 3 台仪器之间的共享^[11]。陈增萍等将近红外光谱技术与载荷空间标准化新型模型传递方法相结合,有效消除了温度变化对近红外光谱校正模型定量分析结果的影响^[12]。

傅里叶变换光谱仪是近红外光谱仪的重要分支,具有光谱分辨率高、光通量大、通道数量多等优点,在农业生产、科学研究、环境监测、航空航天遥感等领域有着广泛的应用。本工作以傅里叶变换近红外光谱仪为实验研究对象,采用 SG-一阶导数平滑法、标准正态变量变化法以及二者结合的方法对同一样本在不同仪器所测得的光谱数据分别进行数据处理,通过对不同仪器间光谱差异信号的一阶矩,概率密度函数,最大信噪比等参数的比较,以期发现不同仪器间光谱数据的差异特征,并寻找仪器间模型通用的可行性。

1 实验部分

1.1 材料

试验样品为收集的烤烟样品共计 150 份。

1.2 设备与测试方法

美国布鲁克公司使用时间相差 10 年的 MPA 型近红外光谱检测仪两台(A, B),美国热电尼高力公司的 Antaris 型近红外光谱检测仪一台(C)。其中,设 A 为主机, B 和 C 为从机。所有近红外光谱检测仪测量工作参数一致,其中光谱采集方法:积分球漫反射测量法,采集范围:4 000~10 000 cm^{-1} ,分辨率:8 cm^{-1} ,扫描次数:64。

取适量烤烟烟末放入样品杯中,取一固定重量砝码放在样品上方,使其自然压实后进行近红外光谱扫描。

1.3 分析方法

1.3.1 一阶导数与标准正态变量变换

光谱的一阶导数是 NIR 光谱分析中常用的基线校正和光谱分辨预处理方法。导数光谱可有效地消除基线和其他背景的干扰,分辨重叠峰,提高分辨率和灵敏度。但它同时会引入噪声,降低信噪比^[13]。本研究选用的是 Savitzky-Golay 求导法,平滑点数为 15 点。

标准正态变量变换主要是用来消除因光的散射引起的光谱测量误差^[14]。对需 SNV 变换的光谱 $X_{i,k}$ 按式计算公式如式(1)所示

$$X_{i,\text{SNV}} = \frac{X_{i,k} - X_i}{\sqrt{\frac{\sum_{k=1}^m (X_{i,k} - X_i)^2}{(m-1)}}} \quad (1)$$

式(1)中, X_i 为第 i 个样品光谱的平均值(标量), $k=1, 2, \dots, m$, m 为波长点数; $i=1, 2, \dots, n$, n 为校正集样品数。

1.3.2 不同仪器光谱偏差的评价参数与方法(残差、一阶矩、信噪比)

残差(residual error)是指实际观察值与估计值(拟合值)

之间的差,蕴含了有关模型基本假设的重要信息。它应符合模型的假设条件,且具有误差的性质。残差的计算方法如式(2)所示

$$d_i = X_i - X'_i \quad (2)$$

式(2)中, X_i 为第 i 个样品的数据估计值, X'_i 为第 i 个样品的数据实际观察值。

离散随机信号的一阶矩(first-order moment)是随机信号的直流分量,而造成直流分量过大的原因包括基线漂移等系统误差,一阶矩越接近于零,代表直流分量越接近于零,即系统误差消除效果越好^[15]。

对于离散随机变量来说,一阶矩的表示如式(3)所示

$$M_n = \sum_n x f(x) = \frac{1}{n} \sum_1^n x_i \quad (3)$$

式(3)中, x_i 为第 i 个离散随机变量, n 为离散随机变量的总数目。

信噪比(signal-to-noise ratio)是描述信号中有效成分与噪声成分的比例关系参数。本实验采用最大信噪比计算方法,即样品的最大额定信号与所有波长下的平均噪声的比值。其计算公式如式(4)和式(5)所示

$$\text{SNR}_i = \sqrt{\frac{x_{i,\text{max}}^2}{d_i^2}} \quad (4)$$

$$d_i = \frac{\sum_{j=1}^{j=n} d_{i,j}}{n} \quad (5)$$

式(4)和式(5)中, $x_{i,\text{max}}$ 为第 i 个样品的额定最大信号, d_i 为第 i 个样品在所有波长下的平均噪声, $d_{i,j}$ 为第 i 个样品在第 j 个波长点下的噪声, n 为波长点数。

高斯噪声是指幅度起伏遵从高斯分布的噪声。高斯噪声通过线性系统后,仍是高斯噪声。将近红外光谱仪仪器之间的差别经化学计量学方法转化为高斯分布后,可避免进一步的线性校正。其一维概率密度函数可用数学表达式如式(6)所示

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-a)^2}{2\sigma^2}\right] \quad (6)$$

式(6)中, a 为噪声的数学期望值,也就是均值, σ^2 为噪声的方差。

本实验中,将主机测量的光谱数据作为估计值(拟合值),将 A 仪器重新装样以及在 B 和 C 两台仪器测量得到的光谱数据作为三组实际观察值,计算残差值。对不同仪器测得的光谱数据进行一阶导数、SNV、和两种方法结合的处理,并计算残差值的一阶矩。其中主机重新装样测量的残差值作为标准对比。将主机的光谱数据中的最大值作为额定最大信号,将两台从机与主机的光谱数据的残差值作为噪声成分。

1.3.3 不同仪器间模型通用性的验证与评价方法

本实验采用偏最小二乘法作为建模方法,评价模型通用性的参数为决定系数(R^2)和预测标准偏差(square error of prediction, SEP)^[2]。

2 结果与讨论

2.1 不同仪器间的光谱偏差分析

选取 150 份烤烟样品其中的 1 份，绘出的在不同处理情况下不同厂家近红外光谱仪所测光谱数据之间的残差分布图

和残差概率密度图，具体情况如图 1 所示。

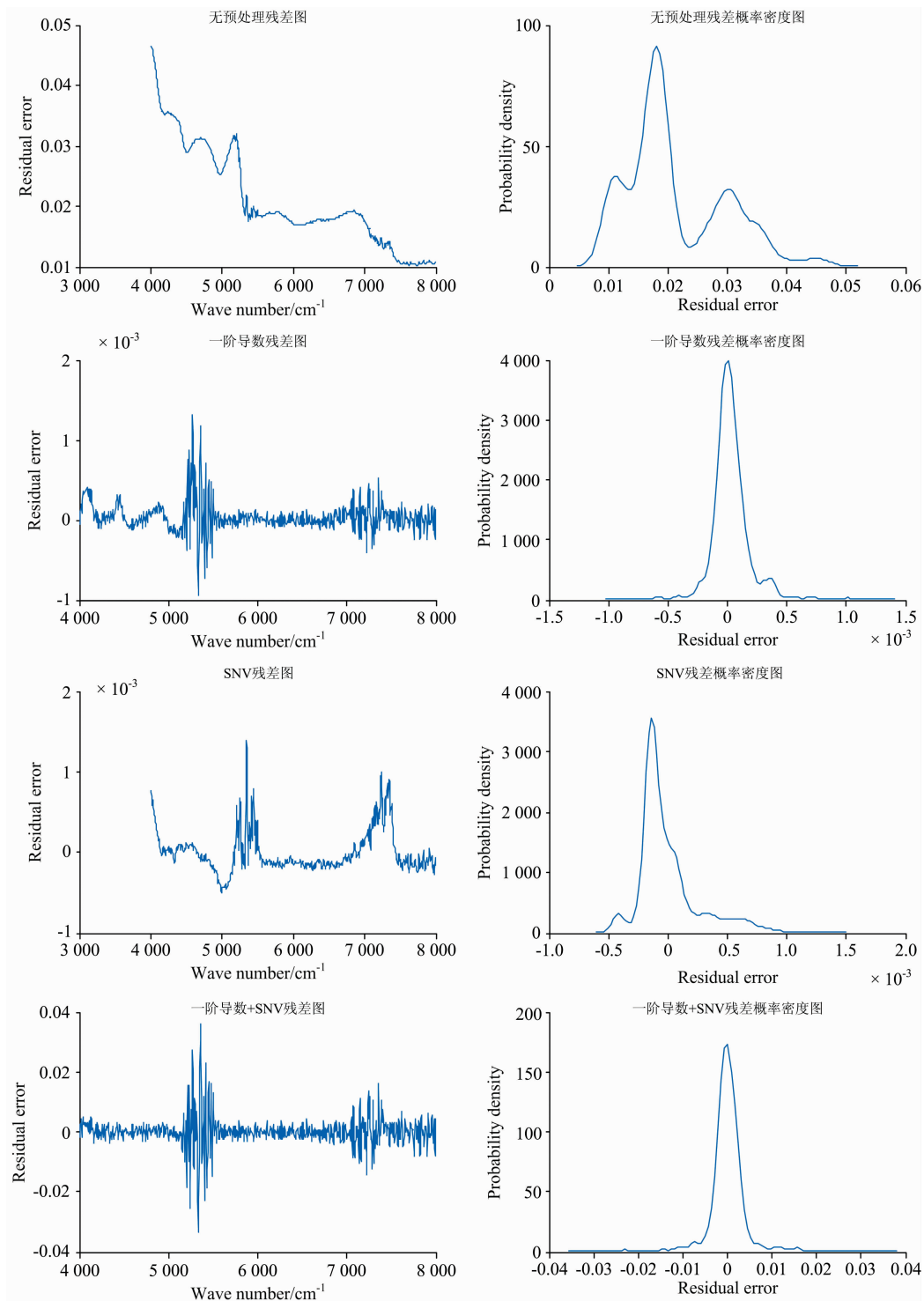


图 1 不同处理情况下的残差图和概率密度图

Fig. 1 The residual error and it's probability density with different process

由图 1 的结果可以看出未处理的残差与原始光谱分布接近，仍含有基线漂移和其他的背景干扰，包含了较大的系统误差，概率密度分布含有多个峰值；SNV 处理后有所改善，

但噪声幅度变化较大，概率密度分布峰值减少至两个；经过导数和两种方法结合处理后的残差分布接近随机噪声，同时概率密度分布为高斯分布。

选取 150 份烤烟样品其中的 10 份, 计算在不同处理情况下的残差一阶矩和信噪比, 并计算平均值, (其中, A-A 表示主机重新装样测量, A-B 表示相同厂家仪器, A-C 表示不同厂家仪器)。其中, 不同处理情况下的残差一阶矩见表 1(a), 不同处理情况下的最大信噪比见表 1(b)。

表 1(a) 不同处理情况下的残差一阶矩
Table 1(a) First-order moment of residual error with different process

处理方法	A-A	A-B	A-C
无处理	-4.723×10^{-3}	1.453×10^{-2}	3.696×10^{-2}
一阶导数	-5.029×10^{-6}	3.881×10^{-5}	3.393×10^{-5}
SNV	-3.566×10^{-9}	-1.156×10^{-9}	-1.927×10^{-10}
一阶导数+SNV	5.685×10^{-9}	2.794×10^{-9}	4.817×10^{-10}

表 1(b) 不同处理情况下的最大信噪比
Table 1(b) Maximal signal to noise ratio with different process

处理方法	A-A	A-B	A-C
无处理	195.8	43.63	17.92
一阶导数	62.67	51.43	47.70
SNV	189.7	382.7	257.9
一阶导数+SNV	59.69	70.62	60.52

由表 1 的结果可以看出, 无处理情况下残差一阶矩量级为 $10^{-2} \sim 10^{-3}$, 不同厂家仪器间差距较大; 同台仪器重新装样的最大信噪比远远大于不同台仪器, 同厂家仪器的最大信噪比是不同厂家型号仪器的两倍左右。一阶导数处理后残差一阶矩数量级为 $10^{-5} \sim 10^{-6}$; 信噪比虽然数量级差距减小, 但相同仪器的最大信噪比仍高于不同仪器; SNV 处理后残差一阶矩数量级为 $10^{-8} \sim 10^{-10}$; 不同仪器的最大信噪比远高于同台仪器, 同时信噪比值得到大幅度提升。一阶导数与 SNV 结合处理后残差一阶矩数量级为 $10^{-8} \sim 10^{-9}$; 不同仪器的信噪比与相同仪器的信噪比基本持平, 并相比于未处理和导数有所提高。说明导数和 SNV 对消除随机信号的直流分量, 减小系统偏差有着明显的效果, 其中单独使用 SNV 和同时使用两种方法的情况效果更好, 消除效果与同台仪器重新装样效果持平。

2.2 偏最小二乘法模型在不同仪器间通用性验证

建模检验比例为 1:1, 在 150 个烟草粉末实验样品中, 选取 75 个作为建模集, 其他 75 个作为检验集, 对实验样品的总糖含量进行建模预测, (其中总糖含量的测定方法是流

动分析法), 得到预测决定系数(R^2)和预测标准偏差(SEP)结果见表 2。

由表 2 的结果可以看出无处理情况下, 不同仪器间的模型传递效果不理想, 其中相同厂家的不同仪器预测决定系数下降约 4%, 不同厂家仪器预测决定系数相差 8%, 且标准偏差相差较大。分别进行一阶导数和 SNV 处理后, 预测决定系数有显著提高, 预测标准偏差差别缩小。其中, 一阶导数处理后不同厂家仪器之间仍相差 3 个百分点, 标准偏差差别有所减小。SNV 处理后不同厂家仪器决定系数差别为 1%, 标准偏差的差别进一步减小。而一阶导数与 SNV 结合处理后, 效果优于任一方法单独处理, 预测决定系数相差在一个百分点以内, 同时 R^2 均在 95% 以上, 标准偏差值均维持在 1 左右, 可以看出, 在处理后 R^2 显著提高, 标准偏差值显著降低, 同时不同仪器间差值在 0.1 以内, 相比于单独处理有明显减小。

表 2 总糖模型预测决定系数及标准偏差(R^2 /SEP)
Table 2 Decision coefficient and standard deviation of total sugarmodel

建模	检验	无处理	一阶导数	SNV	一阶导数+SNV
A	A	93.84/1.24	93.62/1.26	94.03/1.22	95.37/1.07
A	B	89.41/1.62	92.72/1.35	94.23/1.20	95.30/1.08
A	C	85.21/1.92	90.57/1.53	95.63/1.04	96.15/0.98

3 结 论

通过对同台仪器不同测试时间、同厂家仪器、不同厂家仪器相同样本测量数据的残差、一阶矩、最大信噪比、概率密度分布图以及偏最小二乘模型的统计分析, 发现一阶导数具有降低一阶矩, 将仪器偏差转换为标准高斯分布以及提高模型预测准确度的优点; 但会降低信噪比, 不同厂家仪器模型通用效果不够理想。SNV 同样可以降低一阶矩, 并可大幅度提高信噪比, 在不同厂家仪器间作用明显; 但无法将仪器偏差转换为标准高斯分布, 需要进一步的光谱信号处理方法消除高斯噪声外的系统偏差。

综上所述, 一阶导数与 SNV 相结合的方法是一种可以消除以积分球漫反射作为光谱测量方式的因仪器厂家或型号不同、使用年限不同等原因所产生的噪声以及因测量导致的基线漂移等背景干扰的处理方法, 可实现傅里叶型近红外光谱仪之间的模型传递效果的明显改善。

References

- [1] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al(严衍禄, 赵龙莲, 韩东海, 等). Foundation and Application of NIR Spectra Analysis(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005. 13.
- [2] YAN Yan-lu(严衍禄). The Principle, Technique and Application of NIR Spectra Analysis(近红外光谱分析的原理、技术与应用). Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2013. 35.
- [3] JI Na-yu, HAN Dong-hai(吉纳玉, 韩东海). Journal of Food Safety & Quality(食品安全质量检测学报), 2014(3).
- [4] Wang Yongdong, Velramp David J, Kowalski Bruce R. Anal. Chem., 1991, 63 2750.

- [5] Shenk J S, Wes terhaus M O. U. S. Patent: 4866644, Sept. 12, 1989.
- [6] Blank Thomas B, Brown Steven D. Anal. Chem. , 1996, 68(17): 2987.
- [7] Goodacre Royston, Timmins Eadaoin M, Jones Alun, et al. Analytica Chimica Acta, 1997, 348(1-3): 511.
- [8] Kirsten E K, Robert E M, Susan L R. Chemometrics and Intelligent Laboratory Systems, 2008, 92: 33.
- [9] Roychoudhury P, O'Kennedy R, McNeil B, et al. Analytica Chimica Acta, 2007, 590(1): 110.
- [10] Tan H W, Brown S D. Journal of Chemometrics, 2002, 16(5): 228.
- [11] YANG Hui-hua, ZHANG Xiao-feng, CHU Xiao-li(杨辉华, 张晓峰, 褚小立). Chinese Journal of Analytical Chemistry(分析化学), 2014, (9): 1229.
- [12] SHI Xin-zhen, WANG Zhi-guo, CHEN Zeng-ping(史新珍, 王志国, 陈增平). Chinese Journal of Analytical Chemistry(分析化学), 2014, (11): 1673.
- [13] LIANG Yi-zeng, YU Ru-qin(梁逸曾, 俞汝勤). Analytical Chemistry Manual (10)—Stoichiometry(分析化学手册(10)—化学计量学). Beijing: Chemical Industry Press(北京: 化工出版社), 2001.
- [14] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立, 袁洪福, 陆婉珍). Progress in Chemistry(化学进展), 2004, 16(4): 528.
- [15] Papoulis A. Probability, Random Variables, Stochastic Processes, 2nd ed. New York: McGraw-Hill, 1984.

Model Transfer Between Different Fourier Instruments and the Analysis of Error

WANG Yu-heng¹, HU Wen-yan¹, SONG Peng-fei², SHU Ru-xin³, YANG Kai³, WANG Luo-ping², ZHAO Long-lian¹, LI Jun-hui^{1*}

1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Technology Center of China Tobacco Yunnan Industry Co., Ltd., Kunming 653100, China

3. Technology Center of Shanghai Tobacco (Group) Corporation, Shanghai 200082, China

Abstract Model transfer is a key common technical problem in the near infrared spectral analysis technology. By seeking feasible mathematical methods between the two instruments that have the same working principle, we can make the model which was set up on one instrument be applied to another one. In this paper, with 150 flue-cured tobaccos as test samples, with two Bruker MPA near infrared spectrometer and one Thermo Antaris near infrared spectrometer as the research object. We obtained spectral data by integrating sphere diffuse reflectance. Processed and analyzed the spectral data by using the first derivative and standard normal variate (SNV) transformation, and calculated the value of residual error between different instruments, first moment, signal probability density and maximum signal to noise ratio(SNR) and so on. Also, we established mathematical model of total sugar content by partial least squares (PLS) to test the effect of model transfer. The results showed that first derivative could reduce the first moment and transfer the deviation between different instruments into the standard Gaussian distribution, but at the same time, it could also put SNR down. SNV could also reduce the first moment and even could do better than first derivative, and it could increase SNR significantly, but SNR could not transfer the deviation into standard Gaussian distribution, which would need other ways to make up for it. The combination method of the first derivative and SNV can retain the advantages of both, and make up for the disadvantages of individual treatment, and it can settle the model transfer problem caused by different instrument types and different using time between different instruments that work in the principle of Fourier, which is based on the integrating sphere diffuse reflectance. This method is an ideal model transfer method without the prototype.

Keywords Model transfer; First-order derivative; Standard normal variate

(Received Mar. 8, 2016; accepted Jan. 5, 2017)

* Corresponding author