

# 结合 SNV 的分段直接标准化方法在拉曼光谱模型传递中的应用

黄承伟, 戴连奎\*, 董学锋

浙江大学工业控制技术国家重点实验室, 浙江 杭州 310027

**摘要** 为了实现拉曼光谱在不同仪器上的模型传递, 文章提出了一种改进的分段直接标准化算法。该方法先利用标准正态变换法对光谱数据进行处理, 以减少源机与目标机之间的光谱背景与强度影响, 再采用分段直接标准化算法消除不同仪器间对样本信息反应差异, 实现拉曼光谱的全谱传递。此外, 还提出了传递误差率指标用于对光谱传递效果的定量评价。针对汽油拉曼光谱的实验结果表明, 该文提出的算法具有传递样本少、传递精度高和模型稳健性强等优点, 可应用于不同拉曼光谱仪器间的模型传递。

**关键词** 分段直接标准化; 标准正态变换; 拉曼光谱; 模型传递

中图分类号: O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2011)05-1279-04

## 引言

光谱分析作为一种新型的分析检测技术, 具有无损、快速、样品制备简单及可实现在线分析等特点, 被广泛应用于食品、材料、化工、生物等领域。然而在多元校正方法实际应用中, 经常会遇到一台仪器上(称源机, master)建立的模型通常不能直接用于另一台仪器(称目标机, slave)。一般而言, 原模型“失效”的主要原因是样本变化与仪器响应函数差异所引起的测定光谱差异<sup>[1,2]</sup>。模型传递是解决上述问题的有效方法之一。

国内外对近红外光谱的模型传递进行了较为深入的研究<sup>[3-5]</sup>。目前应用最为普遍和最为成功的方法是分段直接标准化(piecewise direct standardization, PDS)方法<sup>[6]</sup>。其他一些新发展的方法, 如被运用于消除样本温度对测量光谱影响的连续分段直接标准化算法<sup>[7]</sup>, 基于小波变换基础上的直接标准化方法<sup>[8]</sup>。这些方法往往具有传递过程繁琐、光谱变量多、传递速度慢等特点。

相对于近红外光谱, 由于拉曼光谱本身的特点, 模型传递研究成果少。Olsen 等<sup>[9]</sup>采用多元信号校正(multiplicative signal correction, MSC)与波长选择的方法, 成功地在同一台仪器上, 将2005年建立的猪肉脂肪组织碘含量拉曼分析模型应用于2008年未知样本的碘含量预测。Mann 等<sup>[10]</sup>以环己烷拉曼谱图为例, 探讨了多项式拟合与函数卷积法在两台不同分辨率拉曼光谱仪器之间模型传递上的应用。Choquette

等<sup>[11,12]</sup>介绍了利用 NIST (national institute of standards and technology) 标准参考物质在激光激发下发出的荧光来校正不同拉曼光谱仪器的不同响应函数方法, 但这种标准物质价格昂贵, 并且受到光谱仪器类型的限制。

本文为了解决拉曼光谱在不同仪器上的模型传递问题, 提出了一种改进的 PDS 方法(简称 SNV-PDS 方法), 该方法先对光谱进行标准正态变换(standard normal variate, SNV), 然后采用 PDS 算法进行全谱传递。针对汽油拉曼光谱的实验结果表明, 该方法具有传递样本少、精度高和传递稳健性强等优点。

## 1 算法描述

本文在基本 PDS<sup>[6]</sup>的基础上提出的 SNV-PDS 算法流程如图 1 所示。在传递模型建立阶段, 选取若干个有代表性的传递样本, 然后在源机和目标机上分别测得这些样本的光谱  $M = \{M_{ij}, i = 1, \dots, p; j = 1, \dots, q\}$  与  $S = \{S_{ij}, i = 1, \dots, p; j = 1, \dots, q\}$ , 其中  $p$  为传递样本数,  $q$  为光谱数据点数。接着利用 SNV 对谱图矩阵  $M$  和  $S$  进行处理, 得到的谱图分别记为  $M^{snv} = \{M_{ij}^{snv}, i = 1, \dots, p; j = 1, \dots, q\}$  和  $S^{snv} = \{S_{ij}^{snv}, i = 1, \dots, p; j = 1, \dots, q\}$ 。利用  $M^{snv}$  的第  $j$  列与  $S^{snv}$  的第  $j-k$  列到第  $j+k$  列建立 PLS 回归模型, 其中  $k$  为分段半窗宽; PLS 回归系数即为转换矩阵  $F = \{F_{ij}, i = 1, \dots, 2k+1; j = 1, \dots, q\}$  的第  $j$  列。光谱矩阵  $M^{snv}$  的每一列对应有一个 PLS 模型, 共有  $q$  个模型。依次采用 PLS 算法, 便可得到完整的转换矩阵  $F$ 。

收稿日期: 2010-07-30, 修订日期: 2010-10-28

基金项目: 国家(863 计划)项目(2009A A04Z123)资助

作者简介: 黄承伟, 1985 年生, 浙江大学工业控制技术国家重点实验室硕士研究生 e-mail: oceanhew@gmail.com

\* 通讯联系人 e-mail: lk dai@ipc.zju.edu.cn

在测试阶段,先测取某一未知样本的光谱  $x^{\text{test}}$ , 并进行 SNV 处理得到光谱  $x^{\text{snv}}$ 。对于某一光谱数据点  $j$ , 选取光谱  $x^{\text{snv}}$  分段窗  $[j-k, \dots, j, \dots, j+k]$  内的数据与转换矩阵  $F$  的第  $j$  列做矩阵乘运算, 得到的结果即为转换后的光谱  $x^{\text{transferred}}$  的第  $j$  个元素。依此类推, 可得到完整的转换后光谱  $x^{\text{transferred}}$ 。

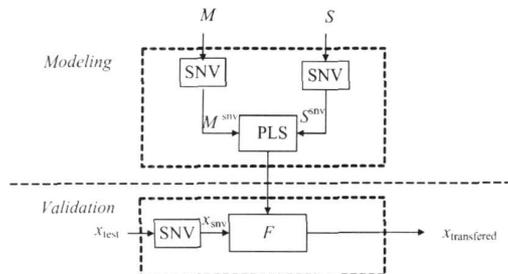


Fig 1 Schematic representation of SNV-PDS algorithm

## 2 实验部分

### 2.1 样本来源

实验数据为来源于某炼厂已知牌号的成品汽油样本, 共计 37 组数据, 其中包含了 9 个 90# 汽油、19 个 93# 汽油与 9 个 97# 汽油样本, 按顺序标记为第 1 到 37 号样本。

### 2.2 实验仪器与数据

本文实验中采用的源机和目标机都采用美国必达泰克公司(B&W Tek, Inc)的 BRM 系列激光光源, 中心波长为 785 nm, 光谱仪为美国海洋公司(Ocean Optics, Inc)的薄型

背照式 QE 65000 光纤光谱仪。不同在于, 源机使用的是必达泰克公司 BAC 100 实验室探头, 光纤长度为 2 m, 光谱仪入射狭缝为 25  $\mu\text{m}$ ; 目标机使用的是在线工业光纤探头, 光纤长度为 20 m, 光谱仪入射狭缝为 50  $\mu\text{m}$ 。

利用上述源机与目标机分别对汽油样本进行测量, 采集的拉曼光谱如图 2 所示, 两台仪器测得的数据相差 1 个数量级左右, 主要由在线探头与光纤损耗造成。

### 2.3 拉曼光谱预处理

本文选取 500~2100  $\text{cm}^{-1}$  范围内的拉曼光谱进行预处理并进行模型传递。为克服光谱噪声和光强不一致对模型传递的影响, 首先将汽油样本拉曼光谱按波数点整数插值, 然后对由源机、目标机测得的每条光谱都采用多项式平滑进行滤波, 半窗宽为 5 个波数, 再对滤波后的谱图进行标准正态变换。图 3 是所有汽油样本经预处理后的拉曼谱图。

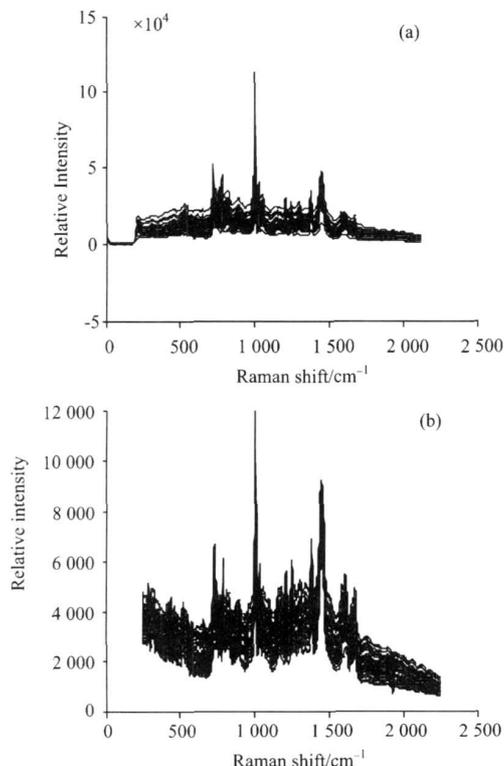


Fig 2 The original Raman spectra

- (a): Original spectra in master;  
(b): Original spectra in slave

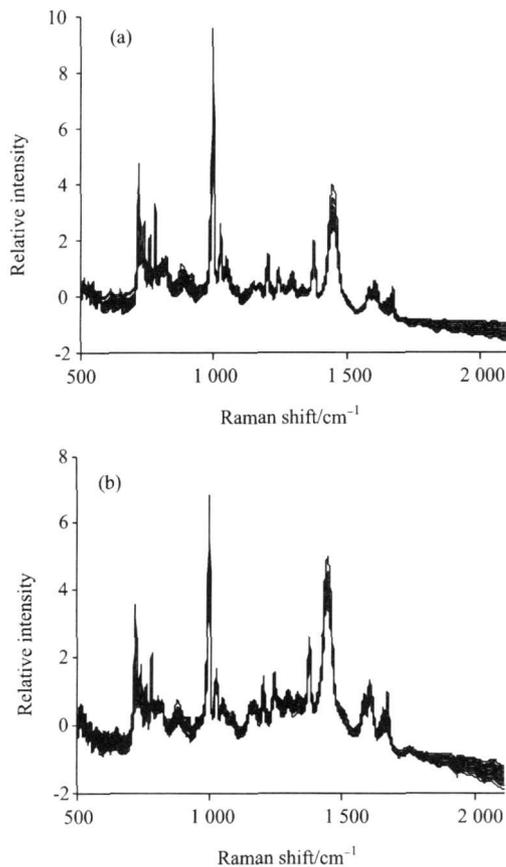


Fig 3 Pre-processed Raman spectra

- (a): Pre-processed spectra in master;  
(b): Pre-processed spectra in slave

## 3 结果与讨论

### 3.1 评价指标

对于不同光谱仪之间的模型传递, 目的是使同一样本源机的谱图与经过传递以后的目标机谱图尽可能一致。为了定量描述这种一致性, 本文提出了光谱传递误差率 (spectra transfer error rate, STER) 评价指标。某一批测试样本, 设  $\{M_{ij}, i=1, \dots, N; j=1, \dots, K\}$  为源机测得的样本光谱矩

阵,  $\{T_{ij}, i = 1, \dots, N; j = 1, \dots, K\}$  为目标机样本光谱经传递后的光谱矩阵, 其中  $N$  为测试样本数,  $K$  为传递光谱内的数据点数。

对于第  $i$  个样本的传递误差率定义为

$$STER_i = \frac{\frac{1}{K} \sum_{j=1}^K |M_{ij} - Z_{ij}|}{\frac{1}{K} \sum_{j=1}^K |Z_{ij}|} \times 100\% \quad (1)$$

其中,  $Z_{ij} = \frac{M_{ij} + T_{ij}}{2}, j = 1, \dots, K$  为第  $i$  个样本平均谱图。

对于整个测试集, 其平均传递误差率  $STER_{ave}$  和最大传递误差率  $STER_{max}$  定义为

$$STER_{ave} = \frac{1}{N} \sum_{i=1}^N STER_i, STER_{max} = \max_i(STER_i) \quad (2)$$

### 3.2 传递样本选择与传递参数确定

理想的模型传递方法应保证较少的传递样本就能得到较好的传递效果, 并且不依赖于样本的选择。对于本实验中采用的传递方法为 PDS, 主因子数最少为 1, 因此从每个类别中选取 1 个样本, 这里选第 1, 10, 30 号样本, 总共 3 个样本作为初始实验的传递样本用于建立两台仪器的传递模型, 其余样本作为测试集, 样本组成如表 1 所示。

**Table 1 The transfer set and test set for initial experiment**

Class	Sample number of transfer set	Sample number of test set
90#	1(第 1 号样本)	8
93#	1(第 10 号样本)	18
97#	1(第 30 号样本)	8
All	3	34

为了定量地评价模型传递的精度, 这里采用上述的平均传递误差率与最大传递误差率两个指标。平均传递误差率代表了传递方法的整体性能; 最大传递误差率代表了传递方法的最大可能风险。SNV-PDS 的可选参数仅有半窗宽, 图 4 反映了 SNV-PDS 方法的平均传递误差率与最大传递误差率随半窗宽的变化情况。综合考虑平均传递误差率与最大传递误差率指标, 当半窗宽设为 4 时, 达到了较好的传递效果, 传递模型也比较稳定。

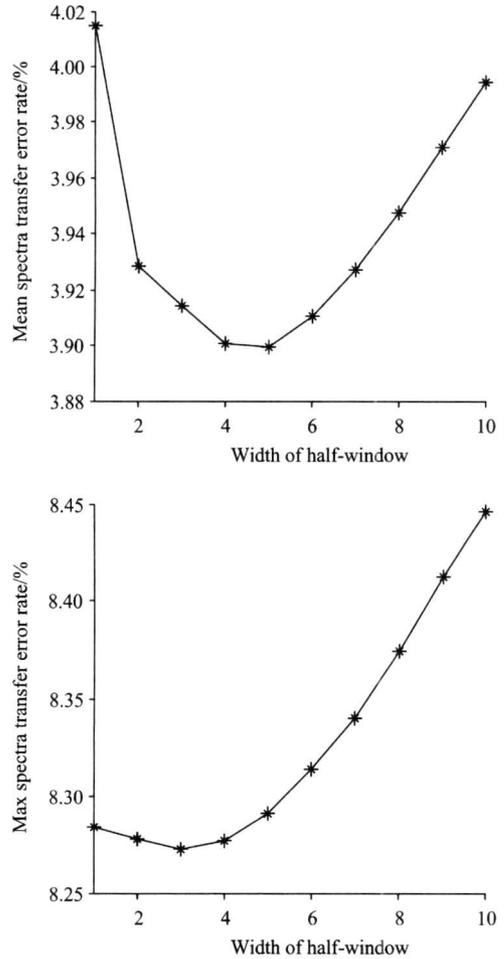
对于某一测试样本, 目标机传递前后的光谱与源机测量光谱如图 5 所示, 结果显示: 经过 PDS 传递后, 两条谱图几乎重叠在一起, 波数漂移的情况也得到了不同程度的改善。

对所有的测试样本的传递结果如表 2 所示。由于源机和目标机上汽油拉曼光谱整体存在比较大的差异, 因此不进行传递方法而直接使用 SNV 对光谱进行处理, 无法得到满意的结果; 由于仪器之间光学器件、CCD 检测器灵敏度不同, 直接使用 PDS 的传递效果也没有得到很大的改进; 而 SNV-PDS 具有最佳的模型传递性能, 其平均传递误差率与最大传递误差率分别为 4.07% 和 8.28%。

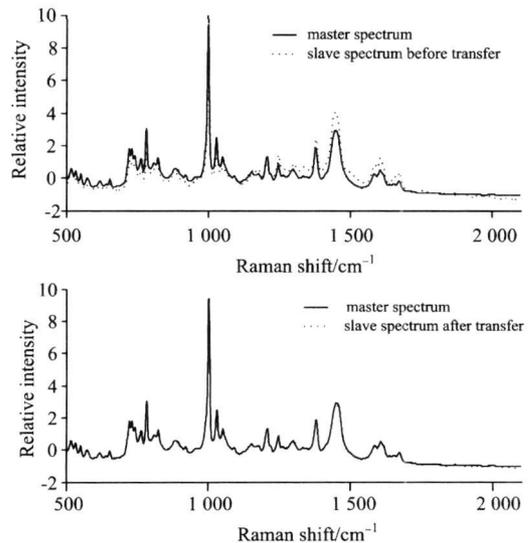
### 3.3 不同传递样本对结果影响

为了展示 SNV-PDS 方法的稳健性, 这里首先选取了部分不同的传递样本组合。每组均采用三个传递样本建立传递

模型, 然后对其他 34 个测试样本进行传递。从表 3 可以看到, 传递样本如果包含至少两种牌号的汽油, 就可以得到比较好的传递结果。



**Fig. 4 Influence of SNV-PDS half-window width on spectra transfer error rates**



**Fig. 5 Raman spectra transfer result using SNV-PDS**

**Table 2 Comparison of results between three calibration model transfer methods**

传递方法	平均传递误差率/%	最大传递误差率/%
SNV	22.47	28.74
PDS	13.73	36.78
SNV+PDS	4.07	8.28

**Table 3 Comparison of results based on different transfer samples**

传递样本组成	传递样本标号	平均传递误差率/%	最大传递误差率/%
90#, 93#, 97#	1, 10, 30	4.07	8.28
90#, 93#, 97#	1, 15, 35	4.24	8.24
90#, 93#, 97#	7, 25, 36	4.60	9.41
90#, 93#, 97#	3, 13, 34	4.15	6.98
93#, 97#	10, 20, 32	4.16	7.91
90#, 97#	1, 32, 36	4.69	8.86
90#, 93#	4, 15, 25	5.15	9.29
90#	4, 5, 8	11.40	21.00
93#	12, 20, 26	6.20	11.23
97#	29, 34, 37	7.01	13.55

## 4 结论

本文成功地将结合标准正态变化与分段直接标准化的 SNV-PDS 模型传递方法应用于汽油拉曼光谱的模型传递, 并提出了衡量拉曼光谱传递效果的定量指标。SNV-PDS 法有效地扣除了两台拉曼光谱仪器的光谱差异, 具有传递样本少、精度高和传递稳健性高等优点, 可应用于不同拉曼光谱仪器间的模型传递。

## References

- [1] Feudale R N, Woody N A, Tan H W, et al. *Chemometrics and Intelligent Laboratory Systems*, 2002, 64: 181.
- [2] Quang H N, Jouan M, Dao N Q. *Analytical Chimica Acta*, 1999, 379: 159.
- [3] CHU Xiaoli, YUAN Hongfu, LU Wanzhen (褚小立, 袁洪福, 陆婉珍). *Spectroscopy and Spectral Analysis (光谱学与光谱分析)*, 2001, 21(6): 881.
- [4] ZHANG Xuebo, FENG Yan-chun, HU Chang-qin (张学博, 冯艳春, 胡昌勤). *Chinese Journal of Pharmaceutical Analysis (药物分析杂志)*, 2009, 29(9): 1390.
- [5] Pereira C F, Pimentel M F, Galvão R K H, et al. *Analytical Chimica Acta*, 2008, 611: 41.
- [6] Wang Y D, Veltkamp D J, Kowalski B R. *Anal. Chem.*, 1991, 63: 2750.
- [7] Børring H K, Boelens H F M, DE Noord O E, et al. *Applied Spectroscopy*, 2001, 55(4): 458.
- [8] Tan H W, Brown S D. *Journal of Chemometrics*, 2001, 15(8): 647.
- [9] Olsen E F, Baustad C, Egelandsdal B, et al. *Meat Science*, 2010, 85(1): 1.
- [10] Mann C K, Vickers T J. *Applied Spectroscopy*, 1999, 53(7): 856.
- [11] Choquette S J, Etz E S, Hurst W S, et al. *Applied Spectroscopy*, 2007, 61(2): 117.
- [12] Etz E S, Choquette S J, Hurst W S. *Microchimica Acta*, 2005, 149(3-4): 175.

# The Application of Piecewise Direct Standardization with SNV in Calibration Transfer of Raman Spectra

HUANG Cheng-wei, DAI Lian-kui\*, DONG Xue-feng

State Key Lab of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China

**Abstract** To implement calibration transfer between Raman spectrometers, an improved piecewise direct standardization (PDS) is proposed in the present paper. Standard normal variate (SNV) is firstly introduced to reduce the influence of spectral background and intensity corresponding to the master spectrometer and the slave spectrometer; then PDS algorithm is used to eliminate the differences between Raman spectra for a specific sample. Moreover, a new quantitative criterion, called transfer error rate, is proposed to evaluate the performance of calibration model transfer. This improved PDS is applied to Raman spectral analysis of gasoline. The result shows that the proposed algorithm not only needs a small quantity of transfer samples, but also obtains high transfer accuracy and strong model robustness.

**Keywords** Piecewise direct standardization; Standard normal variate; Raman spectrum; Calibration model transfer

\* Corresponding author

(Received Jul. 30, 2010; accepted Oct. 28, 2010)