第25卷，第10期
2005年10月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 25, No. 10, pp1610-1613
October, 2005

偏最小二乘法在傅里叶变换红外光谱中的应用及进展

张 琳，张黎明，李 燕*，刘丙萍，王晓斐，王俊德

南京理工大学现代光谱研究室，江苏 南京 210014

摘 要 偏最小二乘法（PLS）是一种应用非常广泛的化学计量方法，它综合了多元线性回归法（MLR）和主成分回归法（PCR）的优势，具有预测能力强和模型相对简单等优点。PLS使傅里叶变换红外光谱的应用范围不断扩大，同时算法也得到了改进和完善。文章介绍了偏最小二乘法在傅里叶变换红外光谱中的应用，对改进算法，如移动窗口PLS（MWPLS）、稳健PLS（RPLS）、加权PLS（WPLS）和非线性PLS等进行了介绍。同时，对应用PLS时数据的预处理、变量的选择、噪声的处理和非线性模型的建立进行了综述。

主题词 偏最小二乘法；傅里叶变换红外光谱；算法改进

中图分类号：O657.3 **文献标识码**：A **文章编号**：1000-0593(2005)10-1610-04

引 言

化学计量学作为化学与计算机科学、数学、统计学的接口[1]，运用计算机上实现的数学与统计方法，优化化学测量过程，并从化学测量数据(信息)中最大限度地提取有用的信息，这使得它对分析化学的发展具有重要的意义。化学计量学在分析化学的许多方面取得了成功应用[2-8]，而应用于傅里叶变换红外光谱(FTIR)中的化学计量法有经典最小二乘法(CLS)[9]、卡尔曼滤波法(KFM)[10,11]、偏最小二乘法(PLS)[12,13]、小波分析(WV)[14]以及仿生类算法人工神经网络(ANN)[15,16]、遗传算法(GA)[17]等，其中偏最小二乘法(PLS)是一种应用十分广泛的化学计量法，主要因为偏最小二乘法是在多元线性回归法(MLR)和主成分回归法(PCR)的基础上发展起来的[12,18]，它集MLR和PCR的基本功能于一体，在一个算法下可以同时实现回归建模、数据结构简化以及两组变量之间的相关分析。因此它具有以下优点[19]：(1)建模求得模型的预报残差平方和(PRESS)小，即模型的预测能力强；(2)可以很好地处理变量多，而样本少的问题；(3)模型相对简单。

FTIR是一种应用十分广泛的分析手段，具有灵敏度高、分辨本领高、速度快的特点，同时普适性强，对气、固、液样品均可进行分析，不破坏原样[20]。但面对生命、环境等学科，要求对复杂的混合物体系进行快速定性定量分析，擅长纯组分分析的FTIR遇到了挑战。化学计量学的介入在一定程度上解决了这类问题。PLS由于上述优点，是FTIR中应用最为广泛的化学计量方法[21]。PLS的基本原理，可参照文献[12，22，23]，本文将重点对PLS在FTIR中的应用和进展作一综述。

1 PLS在FTIR中的应用

顾炳和等[24]利用FTIR进行有机气体的多组分分析时，在PLS校正与预测步骤之间加入了诊断步骤，提出了利用参数SO和SA来判断待测样与分析校正样的相似性，以确保预测结果的可靠性。利用这种方法对甲苯、苯乙烯、邻二甲苯、间二甲苯和对二甲苯的混合样进行分析时，相对标准偏差RSD均小于0.5%。针对样品中含有干扰物质的情况[25]，通过在PLS算法中引入残差光谱的概念及光谱搜索，分别对含有0，1和2个干扰组分的样品进行分析，RSD没有明显的差异，这说明PLS可用于含干扰组分的样品进行分析。

Emilio等[26,27]用PLS辅助的FTIR对油漆中的乙酸丁酯、甲苯和甲基乙基酮进行了分析。同色谱分析比较，该方法测量简单、迅速、准确度和精确度高。Perez-Ponce[27]在实验中用校正阶段的3组分模型对其中的两组分进行预测时，PLS可只识别出这两种组分。显示了方法的稳健性。

文献[28-31]在利用FTIR对血液中葡萄糖的测定时，均采用了PLS进行多变量的解析。在近红外区6 600～4 250 cm^{-1}和中红外区1 200～950 cm^{-1}，Haaland[29,30]对采集的4个血液样品中的葡萄糖进行分析，平均预测误差为13 mg·L^{-1}，准确度适中。由于血液化学的特殊性，作者指出增加校正模型中样品的数目，适当提高样品的温度，会提高预测的

收稿日期：2004-06-06，**修订日期**：2004-09-06

基金项目：国家自然科学基金(20175008)，教育部博士后科学基金和南京理工大学青年学者基金(Njust 200303)资助

作者简介：张 琳，女，1976年生，南京理工大学化工学院博士研究生 * 通讯联系人

准确度。最近，Lewis[31]用 PLS 方法对葡萄糖的 FTIR 单光束光谱建立了校正和预测模型，结合光纤传感系统，发展了在线、无创伤血液中葡萄糖的测定。

重水是核能反应的中子缓和剂和冷却液，建立对其连续、可靠的检测方法是十分必要的。Seung[32]用 FTIR 对重水浓度进行了分析，综合考虑灵敏性和信噪比的关系后，光程定为 0.1 mm，用 PLS 建模分析方法的标准校正误差(SEC)和标准预测误差(SEP)都有很大的改善。

油中的自由基脂肪酸(FFAs)含量是决定其质量和经济价值的主要指标，对 FFAs 的快速准确测定是一项具有工业价值的课题。传统的滴定方法耗时而繁琐，Fernando[33]对 FFAs 采用了红外光谱分析。与 CLS 校正方法相比，用 PLS 获得了更好的检测结果。Man[34, 35]用 FTIR 分析了棕榈油中茴香胺和湿度，用交互验证(cross-validation)检验方法确定了 PLS 校正模型的大小，用 2 mL 的样品在 2 min 内就可得到结果。预测标准误差满足美国油化学家协会的要求。

另外，PLS 应用于 FTIR 中解决多变量校正问题，还包括在食品行业对牛奶中蛋白质、乳糖和丙酮的测定[36]，在医药行业对扑热息通、水杨酸和咖啡因的同时检测[37]，在交通行业对摩托车尾气[38]和飞机引擎排放气体[39]的检测等。PLS 的辅助还拓展了 FTIR 在其他方面的应用，如对复杂过程的优化[40]、构效关系[41]、信号处理[42]、模式识别[43]和动力学过程检测[44]等。

2　PLS 使用策略和方法的改进

2.1　数据的预处理

用 PLS 进行多变量数据分析时，数据的预处理是重要的[18, 29, 30, 45]。预处理的方法主要有均值中心化(Mean-centering)、范围标度化(Range scaling)、自标度化(Autoscaling)和多倍分散校正(Multiplicative scattering correction)[43]等，这几种方法也可以联合使用。其中自标度化是应用最广泛的方法。Emma[45]在用 FTIR 分析己酸盐酯(EC)和二乙基丙二酸(DEM)时，对比了均值中心化、范围标度化和自标度化 3 种方法。当对 50∶50 的 EC 和 DEM 做连续 30 次的测量时，由于基线漂移等因素，收集到的原始谱图不能很好的重现。采用了范围标度化和自标度化对数据进行预处理后，谱图有了很好的重现性，而均值中心化的预处理方法，对提高谱图重现性没有明显的效果。重现性的改进可以免去背景扣除这一步骤，使得在一种仪器上建立起来的校正方法能用于其他仪器，这对工业应用有很大的意义。另外，预处理的重要意义还在于它可以使样本点的分布结构更合理，有利于计算，避免舍入误差，使变量单位一致[1]。Emma[45]认为它还可以使谱图中组分的差异最大化，避免由于浓度过高或过低使谱图差异变得模糊。

2.2　变量的选择

文献[46-48]阐述了变量的选择对多变量校正的意义，它可以去除一些不含信息的变量使模型更简单，预测性更好。另外，相对于紫外与可见光谱，红外光谱对实验条件和样品物理性能的微小扰动和变化较为敏感，由于某些波长间隔内存在非分析组分的干扰，因此变量的选择对红外光谱测试具有更重要的影响[48]。Jiang[48]提出移动窗口 PLS(MWPLS)的方法来确定合适的波长间隔。该方法在一个窗口里建立一系列 PLS 模型，然后在整个光谱区移动。根据模型的复杂性和残余量，确定适合的波长间隔以达到所需的误差水平。MW-PLS 最大优势在于：有干扰存在的情况下，模型非常稳定，而且波长的合适选择可以降低校正模型的大小。他用含不同水平噪声的两个 OP-FTIR 谱图数据和一个近红外谱图数据，验证了 MWPLS 方法对基于振动光谱的多组分分析具有良好的性能，其预测性能优于传统全光谱的 PLS。

Thomas[49, 50]认为，PLS 足以从谱图中提取所有信息而不需要进行变量选择，在随后的研究中又发现：在应用 PLS 时进行变量的选择，也可以使模型有更好的预测能力。变量的选择可以看作是一个求最佳化的问题，用 GA 算法进行 PLS 中的变量选择是一种很好的方法，Learda[46]在测定聚乙烯中添加物的浓度时，利用了 FTIR 和 PLS 算法，发现模型的预测能力和可解释性都得到了提高。

Clifford[47]首先从理论上证明了变量选择的必要性，然后提出了新的变量选择方法。该方法根据信噪比对变量进行排序，以迭代方式建立 PLS 模型，在每次循环中计算交叉有效性平均误差平方和(CVMSE)，直至对所有变量完成排序，确定最小的 CVMSE，使预测误差最小。将该方法应用于 FTIR测试葡萄糖等 3 组实验数据中，3 组结果都表明对于有大的异常值(Outlier)的数据，该方法更稳健，对于微小的噪声该方法没有明显的优势。作者同时指出使用不同的变量选择方法时，要考虑到噪声的分布。

2.3　噪声的处理

常规 PLS 校正方法包含了分析误差与噪声服从正态分布的假设，但这一假设并不总能得到满足。为此，提出了 RPLS[51]。Liu[52]利用 FTIR 分析去痛片中四组分的含量，比较了 PLS 和 RPLS 的性能，模拟数据和样品测试得出了相同的结论：在系统没有异常值时，RPLS 性能与 PLS 相当；当系统共线性很强时，RPLS 具有优势。对 RPLS 方法的研究有助于拓宽其实际应用范围。

在体系中噪声确定的情况下，可采用 WPLS 的方法。WPLS 的基本思路就是对不同的误差项 e_i 加不同的权重，这样可以保证拟合的精确度。Haaland[21]用真实和模拟 FTIR 数据对 WPLS 算法进行了验证。结果表明，在其他条件相同的情况下，WPLS 算法比未加权的 PLS，预测的精确度提高了 9 个百分点。

对 FTIR 信号建立 PLS 预测模型时，Douglas[53]开展了一项用 Savistsky-Golay(SG)对潜变量平滑处理的研究，即 PoLiSh。其基本思想是把噪声，从重要的潜变量中移至次重要的潜变量中，进行迭代。在每一步平滑的迭代过程中，用 Durbin Waston(DW)标准来评估 PLS 模型中，每个潜变量中噪声的水平。Douglas 用含不同噪声水平的模拟 FTIR 信号，进行 PoLiSh 处理时发现，噪声水平高于 10%～20%时，模型的预测能力提高，相对于传统的 PLS，该模型更稳健。Douglas 认为这项技术也可以用于建立二维的 PLS 模型。

2.4　非线性 PLS 的建立

通常的 PLS 是线性模型，为将 PLS 拓至非线性的情况，Wold[54, 55] 先后以多项式和样条函数形成内部关系。Emma[45] 在 FTIR 对 EC 和 DEM 的混合体系进行定量分析时，对 EC 和 DEM 分别建立了线性校正模型和多项式模型，预测平均偏差分别是 4%～14%和 3%～9%。Yang[56] 对三氯甲烷、二氯甲烷和一氯甲烷的混合体系，分别应用 ANN、传统线性 PLS、多项式 PLS 和样条函数 PLS 算法，ANN 得出了更好的分析结果。李燕[57] 在用 FTIR 对谱图严重重叠的五组分 1,3-丁二烯、邻二甲苯、氯苯和丙烯醛的混合体系，进行多组分同时定性定量分析时，比较了 CLS，KFM，PLS 和 ANN 的分析效果。用 RSD 和平均相对偏差(MRE)评定四种方法，结果表明 PLS 最优。结论的不同在于各研究体系中的线性和非线性特性的不同。

Yang[56] 同时比较了 ANN、传统线性 PLS、多项式 PLS、样条函数 PLS 的计算时间和模型容易使用程度。计算时间排序为：线性 PLS≈多项式 PLS＜ANN＜样条 PLS，模型容易使用程度为：线性 PLS≈多项式 PLS＞样条 PLS＞ANN。因此可根据不同的使用要求以及体系的不同特性，选择合适的化学计量模型。

随着对 PLS 研究的深入，新的改进方法不断出现，Wold[58] 还提出隐含非线性潜变量回归(INLR)作为 PLS 的一种简化非线性形式，Bro[59] 将 PLS 扩充至高维(Multiway)的情形，提出了高维 PLS。

3 结 论

综上所述，PLS 是一种非常有效的化学计量学工具。采用 PLS 校正方法使得 FTIR 的应用领域越来越广泛，具有快速、准确、便捷和安全等优势。同时，FTIR 在不同领域的应用特性，也促进了 PLS 方法的改进和完善。随着 FTIR 应用范畴的不断拓展以及 PLS 的不断改进，二者一定可以互相促进，相得益彰。

参 考 文 献

[1] Otto M(M 奥托). Chemometrics(化学计量法). Beijing: Science Press(北京：科学出版社), 2003. 1.

[2] Witjes H, Simonetti A W, Buyden L. Anal. Chem., 2001, 73(19): 548A.

[3] Wold J P, Kvaal K. Appl. Spectroscopy, 2000, 54(6): 900.

[4] Boyswork M, Obando L, Booksh K. Proc. SPIE, 1999, 3856: 308.

[5] Kalman E, Lofvebdahl A, Winquist F. Anal. Chim. Acta., 2000, 403(1～2): 31.

[6] Van Rhee A, Stocker J P, Creech C. J. Comb. Chem., 2001, 3(3): 267.

[7] Linusson A, Gottfries J, Lindgren F. J. Med. Chem., 2000, 43(7): 1320.

[8] Lavine B K, Workman J Jr. Anal. Chem., 2002, 74(12): 2763.

[9] Haaland D M, Easterling R G, Vopick D A. Appl. Spectroscopy, 1985, 39(1): 73.

[10] Brown S D. Anal. Chim. Acta, 1986, 181: 1.

[11] Monfre S L, Brown S D. Appl. Spectroscopy, 1992, 46(11): 1711.

[12] Geladi P, Kowalski B! R. Anal. Chim. Acta, 1986, 185(1): 19.

[13] Natasa Smola, Uros Urkb. Anal. Chim. Acta, 2000, 410(1～2): 203.

[14] LIU Fang(刘 芳). Research on the Technology of FTIR Spectra Analysis about Toxic Organic Compounds in the Atmosphere and the Establishment of Diffusion Models in the Indoor Air [Ph. D. Thesis]. Nanjing(南京), Nanjing University of Sci. and Tech.(南京理工大学), 2003.

[15] Yu Ruqin, Jiang Jianhui. Chemom. Intell. Lab. Syst., 1999, 45(1～2): 191.

[16] LI Yan, SUN Xiu-yun, WANG Jun-de(李 燕，孙秀云，王俊德). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2000, 20(6): 773.

[17] Liu Fang, Wang Junde. Spectroscopy. Lett., 2001, 34(1): 13.

[18] Geladi P, Kowalsk B R. Anal. Chim. Acta, 1986, 185(1): 1.

[19] ZHU Er-yi, YANG Peng-yuan(朱尔一，杨芃原). Chemometrics and Its Application(化学计量学技术及应用). Beijing: Science Press(北京：科学出版社), 2003. 92.

[20] Wang Junde(王俊德). The Application of Remote Sensing in FTIR(遥感技术在傅里叶变换红外光谱中的应用), in Modern Fourier Transform Infrared Spectroscopy and its Application(Vol. 1)(近代傅里叶变换红外光谱技术及应用). Wu Jinguang Ed(吴瑾光). Beijing: Scientifical and Technical Documents Publishing House((北京：科学文献出版社), 1994. 442.

[21] Haaland D M, Jones H D T. AIP. Conf. Proc., 430(Fourier Transform Spectroscopy), 1998, 253.

[22] Fuller M P, Ritter G L, Drapper C S. Appl. Spectroscopy, 1998, 42(2): 217.

[23] LI Yan(李 燕). Temporally and Spatially Extension of Analytical Chemistry [Ph. D. Thesis]. Nanjing, (南京), Nanjing University of Sci. and Tech.(南京理工大学), 2003.

[24] Gu Binghe, Wang Junde. Spectroscopy. Lett., 1998, 31(5): 1053.

[25] Gu Binghe, Wang Lianjun, Wang Junde. Spectroscopy Lett., 1998, 31(7): 1451.

[26] Emilio L A, Garrigues S, Miguel G. Analyst, 1998, 123(6): 1247.

[27] Perez-Ponce A, Rambla F J, Garrigues J M. Analyst, 1998, 123(6): 1253.

[28] Janatsch G, Kruse Jarrres J D, Marbach R. Anal. Chem., 1989, 61(18): 2016.

[29] Haaland D M, Robinson M R, Koepp G W. Appl. Spectroscopy, 1992, 46(10): 1575.
[30] Ward J K, Haaland D M, Robinson M R. Appl. Spectroscopy, 1992, 46(10):959.
[31] Lewis C, Mcnichols R, Gowda A. Appl. Spectroscopy, 2000, 54(10): 1453.
[32] Seung Y C, Jaebum C, Hoeil C. Vibrational Spectroscopy, 2003, 31(1): 251.
[33] Fernando A I, Jose M G, Salvador G. Anal. Chim. Acta, 2003, 489(1): 59.
[34] Man Y B Che, Setiowaty G. J. Am. Oil. Chem. Soc., 1999, 76(2): 243.
[35] Man Y B Che, Mirghani M E. J. Am. Oil. Chem. Soc., 2000,77(6): 631.
[36] Luinge H J, Hop E ,Lutz E T G. Anal. Chim. Acta, 1993, 284(2): 419.
[37] Bouhsain Z, Garrigues S, Miguel G. Analyst, 1996, 121(12): 1935.
[38] Wang Junde, Bian Haiyan, Chen Zuoru. Spectrosc. Lett., 1988, 21(6): 935.
[39] Andrade J M, Carrigues S, Miguel G. Anal. Chim. Acta, 2003, 482(1): 115.
[40] Wold S. J. Chemom., 1996, 10(5～6): 463.
[41] Alifrangis L, Hjorth C, Inge T. J. Med. Chem., 2000, 43(1): 103.
[42] Norgaard L, Saudland A, Wager J. Appl. Spectroscopy, 2000, 54(3): 413.
[43] Liang Yiceng, Yu Ruqin,(梁逸曾, 俞汝勤). Handbook of Analytical Chemistry(分析化学手册), Vol. 10, Chemometrics(第十分册, 化学计量学). Beijing: Chemical Industry Press(北京: 化学工业出版社), 2000. 369.
[44] Yan Bing, Yan HongBin. J. Comb. Chem., 2001, 3(1): 78.
[45] Emma S H, Anthon D W, Stephen J H. Anal. Chim. Acta, 1997, 337(1): 191.
[46] Leardu R, Seasholtz M B, Pell R J. Anal. Chim. Acta, 2002, 461(2): 189.
[47] Clifford H S, Michael J M, Marchel J G. Anal. Chem., 1998, 70(1): 35.
[48] Jiang Jianhui, Berry R J, Siesler H W. Anal. Chem., 2002, 74(14): 3555.
[49] Thomas E V, Haaland D M. Anal. Chem., 1990, 62(15): 1091.
[50] Thomas E V. Anal. Chem., 1994, 66(15): 795.
[51] Yu Ruqin(俞汝勤). Research on Chemometrics Methodology(化学计量学基础与方法学研究), in Advances in Analytical Chemistry(分析化学新进展). Wang Erkang Ed. (汪尔康). Beijing: Science Press(北京: 科学出版社), 2002. 379.
[52] Liu Shiqing, Wang Weiwen. Chemom. Intell. Lab. Syst., 1999, 45(1): 131.
[53] Douglas N R, Antonio B, Ivonne D. Anal. Chim. Acta, 2001, 446(1～2): 281.
[54] Wold S. Chemom. Intell. Lab. Syst., 1989, 7(1～2): 53.
[55] Wold S. Chemom. Intell. Lab. Syst., 1992, 14(1～3): 71.
[56] Yang Husheng, Griffith P R, Tate J D. Anal. Chim. Acta, 2003, 489(1): 125.
[57] Li Yan, Wang Junde, Yuan Weiqun. J. Enviorn. Sci. Health, 2000, A35(9):1673.
[58] Berglund A, Wold S. J. Chemom., 1997, 11(2): 141.
[59] Bro B. J. Chemom., 1996, 10(1): 47.

Application and Improvement of Partial-Least-Squares in Fourier Transform Infrared Spectroscopy

ZHANG Lin, ZHANG Li-ming, LI Yan*, LIU Bing-ping, WANG Xiao-fei, WANG Jun-de
Laboratory of Advanced Spectroscopy, Nanjing University of Science and Technology, Nanjing 210014, China

Abstract Partial least squares(PLS) algorithm is an effective chemometric tool. It takes the advantages of multipal linear regression (MLR) and principal component regression (PCR), which makes Fourier transform infrared spectrometry (FTIR) more powerful and useful. Accompanied with increasing use of FTIR, the algorithm is modified and corrected under different circumstances. The applications of PLS to FTIR were mentioned. Improved algorithms were presented, such as moving windows PLS(MWPLS), robust PLS (RPLS), weighted PLS(WPLS), and non-linear PLS. Data pre-processing, selection of variable, noise elimination and non-linear model of PLS were introduced.

Keywords PLS; FTIR; Improved algorithms

(Received Jun. 6, 2004; accepted Sep. 6, 2004)

* Corresponding author