第26卷，第4期
2006年4月

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

Vol. 26, No. 4, pp620-623
April, 2006

多项式偏最小二乘法对非线性体系红外谱图的分析

张 琳[1]，张黎明[1]，李 燕[1*]，王晓斐[1,2]，胡兰萍[1,3]，王俊德[1]

1. 南京理工大学化工学院现代光谱研究室，江苏 南京 210014
2. 南京大学化学化工学院，江苏 南京 210092
3. 南通大学化学化工学院，江苏 南通 226007

摘 要 文章利用了一种非线性模型多项式偏最小二乘法(PPLS)，结合傅里叶变换红外光谱遥感技术，对大气中的五组分混合体系进行了同时分析。并与偏最小二乘法(PLS)得到的结果进行了比较，PPLS显示出较好的处理非线性数据的能力。尤其是对混合物中的苯和氯仿的预测，均方根预测误差(RMSEP)分别是0.043和0.087，用PLS预测相应的RMSEP为0.402和0.842。PPLS的这一预测精度，可以满足遥感傅里叶变换红外光谱对大气中有毒气体的实时、在线监测的需要。同时PPLS可以用较少的潜变量对变量进行解释，显示出PPLS模型的稳健性和简单化。

主题词 多项式偏最小二乘法；非线性模型；多组分分析；FTIR；大气监测
中图分类号：O657.3 **文献标识码**：A **文章编号**：1000-0593(2006)04-0620-04

引 言

作为化学计量学中最有力的工具之一的偏最小二乘法(PLS)，是一种多元统计数据分析方法，它综合了多元线性回归法(MLR)和主成分回归法(PCR)的优势，同时从自变量矩阵和因变量矩阵中提取偏最小二乘成分，可以有效地降维，并消除自变量间可能存在的复共线性关系，明显地改善了数据结果的可靠性和准确度，因此得到了日益广泛的应用[1]。但在实际应用中，由于实验数据形成的因变量X和自变量Y之间的关系，会因浓度、基线漂移等因素的影响，呈现非线性关系[2]。用线性模型PLS处理此类数据时，势必会造成较大的偏差。对于这类问题的处理，通常采用两种方法。一种是对数据进行预处理，尽可能地去除噪声部分，如多倍分散校正(MSC)[3]，正交信号处理(OSC)[4]等。但是，存在着有用信息同时被去除和模型“过拟合”等问题。另一种就是建立非线性模型[1]，如人工神经网络(ANN)[5-8]。但是ANN需要较多的训练集，另外还存在着容易“过拟合”的现象。

本文利用遥感FTIR光谱对大气中的有毒气体进行分析，采用多项式偏最小二乘法(PPLS)对红外谱图严重混叠的五组分体系进行实时、多组分同时测量。PPLS是改进的PLS非线性算法[1,9]，主要是在进行非线性迭代偏最小二乘(NIPALS)算法时，对每次提取的X和Y的得分矢量t_i和u_i实施非线性映射：$u_i = f(t_i) + h_i$，其中f为多项式函数，h_i为残差向量。从本研究的实验结果表明，PPLS可以很好地处理一些非线性组分，表现出好的预测性能。特别是对混合物中的苯和氯仿的预测。该模型在遥感FTIR的成功应用，对大气中有毒气体混合物的监测具有重要的意义。

1 实验部分

1.1 硬件

本工作采用的是Bruker EQUINOX-55遥感FTIR光谱仪，它有一个Dall-kirkham $f/4$望远镜，用于接收来自远距离的红外源信号。远距离的红外源是一个带有Dall-kirkham $f/4$准直镜的陶瓷红外源。该红外源产生的红外光，经Dall-kirkham准直光镜产生平行光，此平行光被待测气体云吸收后，由遥感FTIR光谱仪测量。检测器为液氮冷却的MCT检测器，样品扫描次数为4次，分辨率为1 cm^{-1}。所有计算均在PentiumⅣ电脑上完成。

1.2 软件

程序由Matlab6.5编写完成。PLS和PPLS算法采用NIPALS实现。采用“逐一法”(Leave-one-out)的交互验证方法，确定潜变量数。所使用数据均采用自标度化(Auto-scaling)方法处理。用均方根误差(RMSEP，root mean squared

收稿日期：2004-12-10，**修订日期**：2005-04-20
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(Njust200303)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

error of prediction)来评价模型的预测性能。

2　结果与讨论

2.1　样品

本文用遥感 FTIR 光谱对空气中人为释放的苯、甲苯、丙酮、二氯甲烷和氯仿的混合气体进行分析，实验实施细节见文献[5]。上述气体的红外谱图如图 1 所示。考虑到水和二氧化碳的影响，校正和预测的光谱区域选择为 1 360～600 cm^{-1}。

2.2　预测结果分析

对苯、甲苯、二氯甲烷、丙酮和氯仿混合物分别用 PLS 和 PPLS 进行了分析，得到的结果见表 1。

由表 1 可以看出，PPLS 的预测精度高于 PLS，这说明各个组分的浓度和吸光度间确实存在着非线性关系，且本研究中所建立的 PPLS 模型，能够较 PLS 准确地反映这种非线性关系。尤其是对苯和氯仿的预测准确度较高。但 PPLS 对甲苯、二氯甲烷和丙酮的预测准确度仍需提高。

我们同时以苯和甲苯为例，考察了它们在 PLS 和 PPLS 中得到的真实值、预测值和残差的示意图见图 2。从图 2 可以看出，图 2(a)即苯的 PPLS 模型，真实值相对于预测值呈现很好的线性，残差也在 10^{-5} 数量级。与之相比，图 2(b)，图 2(c)和图 2(d)得到的结果不够理想。

这些结果都说明，即使在同一个样品中，各个组分的因变量 X 和自变量 Y 间的关系，也不能一概而论[10]。只有能够确切反应 X 和 Y 关系的模型，才可以得到很好的预测结果。但在实际中，如果对每个组分都进行实验，以确定适合的模型，显然是不可行的。反之，如果进一步改进模型，增加模型的普适性，这将是我们未来研究的方向。

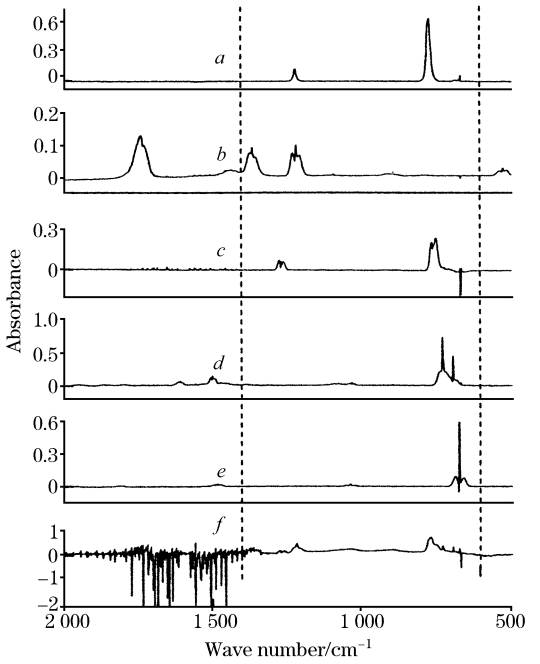

Fig. 1　FTIR spectrum of five-component system

a: Chloroform; *b*: Acetone; *c*: Methylene chloride; *d*: Toluene; *e*: Benzene; *f*: Mixture

Table 1　Results of PLS and PPLS model($\times 10^{-1}$ mg · L^{-1})

Benzene			Toluene			Methylene chloride			Acetone			Chloroform		
Creal	Cpred		Creal	Cpred		Creal	Cpred		Creal	Cpred		Creal	Cpred	
	PPLS	PLS		PPLS	PLS		PPLS	PLS		PPLS	PLS		PPLS	PLS
1.850	1.796	2.220	2.368	2.019	2.890	2.610	2.768	2.493	0.104	0.094	0.086	2.689	2.580	3.917
3.062	3.016	2.387	2.895	3.152	2.818	2.088	1.970	2.0518	0.072	0.086	0.095	4.303	4.367	4.059
2.233	2.199	2.393	3.572	3.253	2.670	1.948	2.093	2.401	0.084	0.083	0.101	4.401	4.481	3.492
1.914	1.860	2.192	3.308	3.113	2.730	2.958	3.046	3.019	0.091	0.080	0.090	2.934	2.950	3.207
2.552	2.592	2.726	2.068	2.477	2.503	2.262	2.075	2.110	0.117	0.114	0.102	3.667	3.530	3.418
1.818	1.808	2.212	2.932	2.588	2.769	3.410	3.567	3.099	0.114	0.089	0.080	4.156	4.080	3.283
2.105	2.179	2.099	3.196	2.762	2.972	2.784	2.98	2.806	0.065	0.061	0.075	2.445	2.505	3.897
2.934	2.937	2.577	2.180	2.882	2.649	2.088	1.768	1.980	0.097	0.111	0.102	4.7922	4.873	3.774
2.392	2.427	2.768	2.256	2.273	2.431	1.914	1.906	1.953	0.128	0.121	0.113	4.401	4.319	3.332
2.711	2.736	2.171	3.572	3.515	2.928	2.784	2.809	2.534	0.084	0.081	0.084	3.6675	3.537	3.966
1.754	1.699	2.335	3.271	2.866	2.629	2.575	2.759	2.970	0.078	0.066	0.091	3.423	3.363	3.057
RMSEP	0.043	0.402		0.364	0.501		0.165	0.228		0.011	0.016		0.087	0.842

Creal : the real concentration; Cpred : the prediction concentration by PPLS or PLS

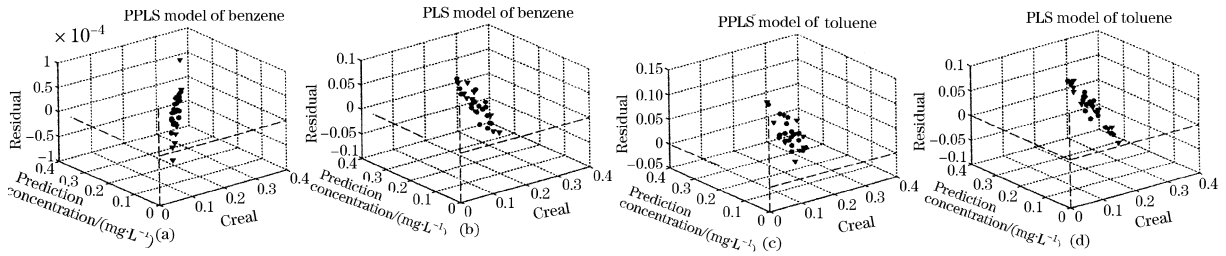

Fig. 2　Prediction concentration, real concentration and residual for benzene and toluene using PPLS and PLS

●: The result of calibration; ▼: The result of prediction

另外，表 2 总结了在 PLS 和 PPLS 模型中，每个潜变量(LV)所描述的变量。可以看出，相对于 PPLS，PLS 需要用较多的 LV 来解释变量。这是因为 PLS 在处理非线性数据时，会通过增加 LV 来提高处理非线性数据的能力[11]。而当体系中非线性较强时，增加 LV 的个数将被用于描述噪声，所以仅依靠增加 LV 个数不能提高预测的准确度。而 PPLS 可以用少的 LV 来构造模型，显示出模型的稳健性和简单化。

Table 2 Variance Captured by PLS and PPLS

LV number	PLS				PPLS			
	X-block		Y-block		X-block		Y-block	
	this LV	total	this LV	total	this LV	total	this LV	total
1	69.16	69.16	15.99	15.99	69.08	69.08	22.64	22.64
2	20.20	89.36	14.11	30.10	25.10	94.18	23.96	46.60
3	4.52	93.88	16.28	46.38	2.79	96.98	19.79	66.39
4	3.19	97.07	10.97	57.35	0.95	97.93	15.28	81.67
5	1.16	98.23	5.77	63.12	1.26	99.19	7.11	88.78
6	0.92	99.15	4.71	67.83	0.65	99.84	6.21	95.00
7	0.37	99.52	11.24	79.07	0.11	99.95	3.37	98.37
8	0.42	99.94	3.79	82.86	0.05	100	1.39	99.76
9	0.04	99.99	6.65	89.51	0	100	0.23	99.99
10	0.00	100	3.40	98.55	0	100	0.00	99.99

3 结 论

研究利用 PPLS 算法，结合遥感 FTIR 技术，对苯、甲苯、丙酮、二氯甲烷和氯仿的五组分的红外数据进行了分析。相对于 PLS，PPLS 对混合物含量的预测准确度有了提高，尤其是对苯、氯仿含量的预测，显示出很好的处理非线性数据的能力。同时，PPLS 可以用较 PLS 少的潜变量来描述变量，表现出模型的稳健性和简单化。另外，进一步改进 PPLS，提高模型的普适性，是我们下一步研究的方向。

参 考 文 献

[1] WU Xiao-hua, CHEN De-zhao(吴晓华，陈德钊). Chinese J. Anal. Chem.(分析化学), 2004, 32(4): 534.

[2] Wold S, Antti H, Lindgren F, et al. Chemom. Intell. Lab. Syst., 1998, 44(1-2): 175.

[3] Andersson A Claus. Chemom. Intell. Lab. Syst., 1999, 47(1): 51.

[4] Yee G N, Coghill G G. Chemom. Intell. Lab. Syst., 2003, 67(1): 145.

[5] Li Yan, Wang Junde, Chen Zuoru, et al. Anal. Lett., 2001, 34(12): 2203.

[6] LI Yan, WANG Jun-de(李 燕，王俊德). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(6): 1104.

[7] LI Yan, WANG Jun-de, WANG Lian-jun (李 燕，王俊德，王连军). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2000, 20(4): 477.

[8] 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.

[9] Wold S. Chemom. Intell. Lab. Syst., 1989, 7(1-2): 53.

[10] Emma S H, Anthon D W, Stephen J H. Anal. Chim. Acta, 1997, 337(1): 191.

[11] Yang Husheng, Griffith P R, Tate J D. Anal. Chim. Acta, 2003, 489(2): 125.

Multi-Component Analysis of FTIR Spectra of Non-Linear System Using Polynomial Partial Least Squares Method

ZHANG Lin[1], ZHANG Li-ming[1], LI Yan[1*], WANG Xiao-fei[1,2], HU Lan-ping[1,3], WANG Jun-de[1]

1. Laboratory of Advanced Spectroscopy, Nanjing University of Science and Technology, Nanjing 210014, China
2. Department of Chemistry, Nanjing University, Nanjing 210092, China
3. Department of Chemistry, Nantong University, Nantong 226007, China

Abstract A non-linear algorithm, polynomial PLS was applied to the simultaneous analysis of OP-FTIR spectra of a five-component system whose FTIR spectra were seriously overlapped. The results were compared with the one obtained from PLS. PPLS yielded good performance, especially for the prediction of benzene and chloroform. RMSEP(root mean squared error of prediction) of benzene and chloroform in PPLS model were 0.043 and 0.087 and the corresponding values in PLS were 0.402 and 0.842, respectively. Meanwhile, variance was accounted by PPLS with fewer latent variables, which indicates the simplicity and robustness of the model. The successful application of PPLS to non-linear system was meaningful for the use of remote sensing FTIR in air monitoring.

Keywords Polynomial PLS; Non-linear system; Multi-component analysis; FTIR; Air monitoring

(Received Dec. 10, 2004; accepted Apr. 20, 2005)

* Corresponding author

“第四届国际华夏学者分析化学研讨会”会议通知

The Fourth International Symposium of Worldwide Chinese Scholars on Analytical Chemistry (ISWCSAC 2006)

第四届国际华夏学者分析化学研讨会(ISWCSAC 2006)，原名国际华裔学者分析化学研讨会，经中华人民共和国外交部和中国科学技术协会批准，将于 2006 年 9 月 22～26 日在大连召开。

此次会议的宗旨是进一步加强海内外华裔分析化学学者的相互了解与合作研究。会议期间将同时举办分析化学新产品、新技术展示会。充分展示分析化学、生命科学和实验室设备的最新产品和技术成果。我们诚挚期待您的参加。

详情访问 http://iswcsac2006.dicp.ac.cn.

征文内容：

会议将涵盖包括的领域有：原子光谱和分子光谱、核磁共振、电化学、色谱、质谱、分析仪器及应用

联系方式：

联系人：张丽华　张维冰

地　址：大连市中山路 457 号 中国科学院大连化学物理研究所

邮　编：116023　　**电话/传真：**+86-411-84379779

E-mail：iswcsac2006@dicp.ac.cn