光谱预处理方法选择研究

第五鹏瑶，卞希慧*，王姿方，刘 巍

天津工业大学省部共建分离膜与膜过程国家重点实验室，环境与化学工程学院，天津 300387

摘 要 复杂样品光谱信号往往会受到杂散光、噪声、基线漂移等因素的干扰，从而影响最终的定性定量分析结果，因此通常需要在建模前对原始光谱进行预处理。目前已有的光谱预处理方法包括很多种，如何寻找合适的预处理方法是很棘手的问题。一种途径是观察光谱信号特点选择预处理方法(visual inspection)，另一种途径是根据建模性能的优劣反过来选择预处理方法(trial-and-error strategy)。前者无需建模，更具有解释性，但是有时会由于选择者主观的因素导致错误的结果；后者无需观察光谱特点，但需要考察大量的预处理方法，对大数据集比较费时。因此需要探讨哪种选择方式更科学与合理。本研究采用9组数据，通过对10种预处理方法的120种排列组合来探讨预处理的必要性及预处理方法的选择。首先，优化偏最小二乘(PLS)的因子数及一阶导数、二阶导数、SG平滑的窗口参数，连续小波变换(CWT)的小波函数和分解尺度。然后把无预处理及一阶导数、二阶导数、CWT、多元散射校正(MSC)、标准正态变量(SNV)、SG平滑、中心化、Pareto尺度化、最大最小归一化、标准化10种预处理方法按照背景校正、散射校正、平滑和尺度化的顺序进行排列组合，得到120种预处理及其组合方法。最后对不同数据及相同数据的不同组分分别进行120种预处理，分析光谱信号特点及预处理后PLS建模的预测均方根误差值(RMSEP)。结果表明，相比观察光谱信号特点，根据光谱与预测组分的建模效果可以更为准确地选择最佳预处理方法。对于多数数据，采用合适的预处理方法可以提高建模效果；对于不同的数据集，因为其数据集信息和复杂性不同，所以其最佳预处理方法也不同；对于相同数据集，即使光谱相同，但不同组分的预处理方法也不相同。因此，不存在普适性的最佳预处理方法，最佳预处理方法除了与光谱有关，还与预测组分有关。通过对已有预处理方法按照预处理目的进行分类再排列组合是选择最佳预处理方法的一种有效途径。

关键词 预处理方法；复杂样品；偏最小二乘；参数优化；方法选择

中图分类号：O433.4 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2019)09-2800-07

引 言

光谱分析因其分析速度快、操作简便、不需纯样品等优点在工业、医疗、农业方面的复杂样品分析中得到广泛应用[1-2]，但光谱数据易受到杂散光、噪声、基线漂移等因素的干扰，从而影响建模效果[2,3]。因此，需要在建模前对光谱数据进行预处理。

目前已有的光谱预处理方法包括很多种，根据预处理的效果可以分为基线校正、散射校正、平滑处理和尺度缩放四类[4-5]，每一类又包括多种预处理方法。基线校正包括一阶导数、二阶导数、连续小波变换(continuous wavelet transform，CWT)等；散射校正包括多元散射校正(multiplicative scatter correction，MSC)、标准正态变量(standard normal variate，SNV)等；平滑处理包括SG平滑等；尺度缩放包括中心化、Pareto尺度化、最大最小归一化、标准化等。其中，导数处理和CWT主要是扣除仪器背景或漂移对信号的影响；MSC和SNV用来消除由于颗粒分布不均匀及颗粒大小不同所产生的散射对光谱的影响；SG平滑能够非常有效的提高谱图信噪比，降低随机噪声的影响；中心化、Pareto尺度化、最大最小归一化、标准化可以消除尺度差异过大带来的不良影响。

对复杂样品的光谱数据进行分析时，虽然多种预处理方法可以被用于数据的预处理，但如何找到一种最佳的预处理方法是很重要的工作。一种是根据光谱信号的特征直接选择预处理方法，虽然无需建立模型，但需要选择者的经验。Zhu

收稿日期：2018-08-02，**修订日期**：2018-12-18

基金项目：国家自然科学基金项目(21405110)和天津市教委科研计划项目(2018KJ200)资助

作者简介：第五鹏瑶，1993年生，天津工业大学环境与化学工程学院硕士研究生 e-mail：diwupengyao@163.com

*通讯联系人 e-mail：bianxihui@163.com

等[6]通过比较 SNV、一阶导数、二阶导数对 400 个蛹的动态光谱预处理效果，选择了二阶导数的预处理方法结合建模方法实现了对不同品种的活蛹进行性别判断。另一种是根据建模效果选择预处理方法，这样能够选择最优的方法，但由于目前存在很多预处理方法，对大数据集是很耗时的过程。Qiao 等[8]采用 6 种预处理方法对原始光谱进行处理，应用偏最小二乘(partial least squares，PLS)方法建立预测模型确定土壤有机质的有效波长。关于这两种光谱预处理选择方法哪种更科学合理的研究很少。因此，本研究选用 9 组光谱数据集，采用已有预处理方法的 120 种排列组合结合 PLS 建模，旨在探究光谱预处理的必要性及科学的预处理选择方法。

1　实验部分

1.1　光谱数据集

采用 9 组光谱数据集进行实验，数据集 3 和 8 分别为实验数据和公司提供数据，其余数据从网上下载，详细信息见表 1。其中，数据集 3 包含 51 个四元调和油样品的近红外光谱及香油、玉米油、大豆油和稻米油四种组分的质量百分比含量。数据集 3 和 5 采用 KS 方法划分训练集和预测集，其他数据采用网上的分组方式。图 1 为 9 个数据训练集的光谱，其中数据集 5 的第 115 号样本和数据集 6 的第 680 和 681 号样本为奇异样本未被使用。

表 1　数据集统计信息

Table 1　Statistical information of datasets

数据集	样品	分析组分	波长范围/nm	波长点	训练集	预测集	仪器/测量	来源
1	血液	胆固醇	1 100～2 498	700	143	47	NIR Systems 6500/漫反射	http://www.idrc-chambersburg.org/shootout2010.html
2	橙汁	蔗糖	1 100～2 498	700	150	68	Model 6250 /反射	http://www.ucl.ac.be/mlg/index.php? page=databases
3	四元调和油	稻米油大豆油	833.3～2500	4148	34	17	Vertex70/反射	实验数据
4	燃油	密度	750～1 550	401	142	121	未知/反射	http://www.eigenvector.com/Data/SWRI
5	汽油	双环芳烃	200～400	572	71(70)	44	Cary 3 UV-Vis. /吸收	http://myweb.dal.ca/pdwentze/downloads.html
6	小麦	蛋白质	400～2498	1050	775(773)	107	Foss Model 6500/漫反射	http://www.idrc-chambersburg.org/shootout2008.html
7	玉米	蛋白质	1 100～2 498	700	53	27	mp6 NIR/反射	http://software.eigenvector.com/Data/Corn/index.html
8	烟草	还原糖	1 100～2 500	1296	180	90	Vector22/N FT-NIR/漫反射	烟草公司提供[8]
9	三元混合物	乙醇	850～1 049	200	63	32	Vectra XM2 PC/漫反射	文献[9]

1.2　光谱预处理方法

根据预处理的目的，光谱预处理方法可以分为基线校正、散射校正、平滑处理和尺度缩放四类。基线校正是为了扣除仪器背景或漂移对信号的影响，包括一阶导、二阶导及 CWT 等。一阶导和二阶导分别可以扣除斜线和曲线背景，提高光谱分辨率，基本公式如下

$$x_{i,1\mathrm{st}} = \frac{x_{i+g} - x_i}{g} \tag{1}$$

$$x_{i,2\mathrm{nd}} = \frac{x_{i+g} - 2x_i + x_{i-g}}{g^2} \tag{2}$$

式中，x_i 为第 i 个样品的光谱，g 为窗口宽度。CWT 将信号分解为一系列小波函数的叠加，能够同时起到背景校正和噪声去除的作用，基本公式为

$$C[f(t),\varphi_{a,b}(t)] = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{a}}\varphi\left(\frac{t-b}{a}\right)\mathrm{d}t \tag{3}$$

式中，$f(t)$是输入信号，在这里可看作光谱信号，t 是时域信号，可以看作是波数，$\varphi_{a,b}(t)$是小波基函数，a 为平移参数，C 为小波系数。

散射校正用来消除由于颗粒分布不均匀及颗粒大小不同产生的散射对光谱的影响，包括 MSC 和 SNV。MSC 首先计算校正集的平均光谱，然后将每条光谱与平均光谱作一元线性回归。SNV 从原始光谱中减去该条光谱的平均值后，再除以校正集光谱的标准偏差。

平滑处理为了消除光谱信号中的随机噪声，提高样本信号的信噪比。Savitzky-Golay(SG)平滑法是利用多项式对原始光谱的移动窗口内的数据进行多项式分解并用最小二乘进行数据拟合，其实质是一种加权平均法。

尺度缩放是为了消除数据尺度差异过大带来的不良影响，包括：中心化、Pareto 尺度化、最大最小归一化和标准化等。中心化是将每个样品光谱减去训练集的平均光谱。Pareto 尺度化将光谱除以标准偏差的开方。最大最小归一化(max-min scaling)是将每个光谱点减去所在变量列的最小值后，再除以光谱所在列最大值和最小值的差值。标准化(normalization，又称作 pearson scaling 或 auto scaling)与 SNV 类似，不同之处在于前者是对光谱的列取平均，后者是对行取平均。

图 2(a—j)为血液数据近红外光谱采用上述 10 种预处理方法处理后的光谱。从(a—c)可以看出，经一阶导数和二阶导数处理后的光谱，明显消除了基线和背景的干扰；(d，e)分别经 SNV 和 MSC 的光谱重合度变高，减弱了散射对原始光谱的影响；(f)经 SG 平滑处理后的光谱的噪声明显减弱；(g—j)经归一化处理后的光谱都落入了一个特定的区间之内，去除了尺寸差异、信息结构不同的干扰。

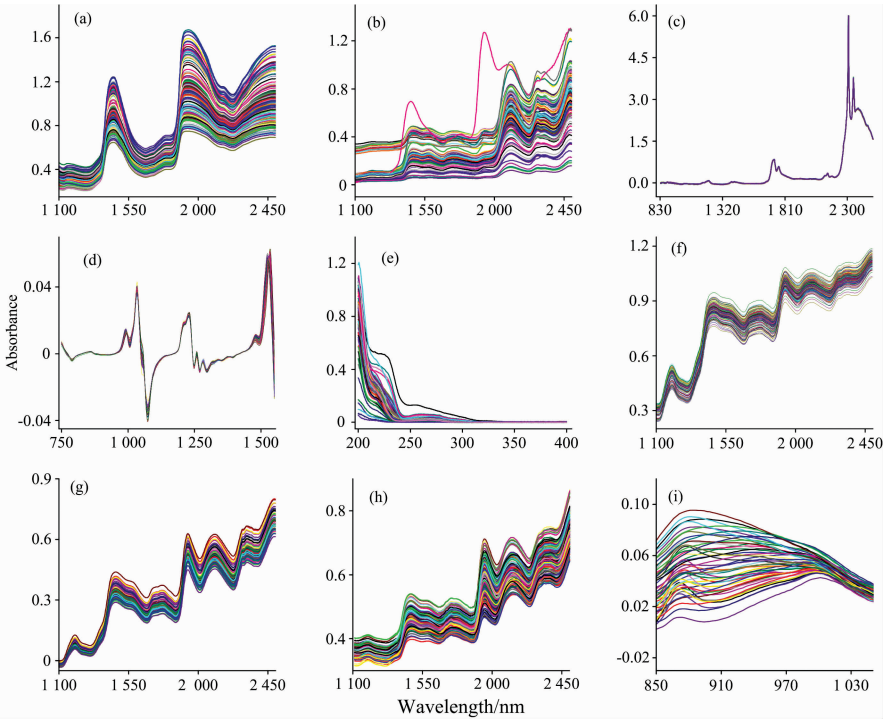

图 1 9 个数据训练集的光谱图

(a)：血液；(b)：橙汁；(c)：四元调和油；(d)：燃油；(e)：汽油；(f)：小麦；(g)：玉米；(h)：烟草；(i)：三元混合物

Fig. 1 The spectra of training set in nine datasets

(a)：Blood；(b)：Orange juice；(c)：Quarternary blended oil；(d)：Fuel oil；
(e)：Gasoline；(f)：Wheat；(g)：Corn；(h)：Tobacco；(i)：Ternary mixture

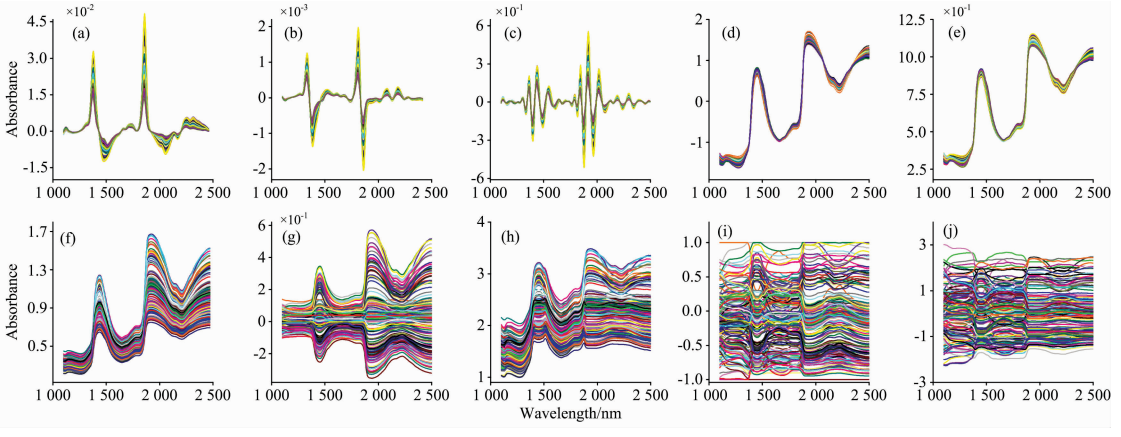

图 2 血液数据进行单一预处理方法的近红外光谱图

(a)：一阶导；(b)：二阶导；(c)：CWT；(d)：SNV；(e)：MSC；(f)：SG 平滑；
(g)：中心化；(h)：Pareto 尺度化；(i)：最大最小归一化；(j)：标准化

Fig. 2 The NIR spectra of single preprocessing methods for blood data

(a)：1st derivative；(b)：2nd derivative；(c)：CWT；(d)：SNV；(e)：MSC；(f)：SG smoothing；
(g)：Mean centering；(h)：Pareto scaling；(i)：Max-min scaling；(j)：Normalization

1.3 预处理组合方法

把无预处理及 10 种预处理方法按照基线校正、散射校正、平滑处理和尺度缩放的顺序进行 4 重 for 循环[4, 10]，各个预处理步骤包含的预处理方法见表 2。按照表中从上到下

的顺序依次从每个预处理步骤中选择一种预处理方法(包含0代表的无预处理)，全排列组合可以得到 4×3×2×5=120 种预处理及其组合方法，用数字 1～120 对这些预处理方法进行编号。

表 2　四个预处理步骤包含的预处理方法

Table 2　Specific preprocessing methods used in the four preprocessing steps

基线校正	散射校正	平滑处理	尺度缩放
0	0	0	0
一阶导	SNV	SG 平滑	中心化
二阶导	MSC		Pareto 尺度化
CWT			最大最小归一化
			标准化

1.4　模型的建立及评价

使用 PLS 建立模型，利用交叉验证均方根误差(root mean squared error of cross validation，RESECV)来优化模型的相关参数。采用预测均方根误差(root mean squared error of prediction，RMSEP)来评价模型优劣。

2　结果与讨论

2.1　数据集的光谱特征

图 1(a—i)为 9 个数据集的光谱，直接观察其光谱信号特点可以看出，(a，b，f，g，h，i)的光谱信号存在漂移和背景；(a，b，e，f，g，h，i)散射较大；(b，h)光谱信号存在噪声；(c，d)光谱重合度非常好，也不存在信号漂移和噪声。因此，数据集 1，6，7 和 9 需要扣除背景及散射校正；数据集 2 和 8 需要扣除背景，散射校正及平滑处理；数据集 3 和 4 无需进行预处理；数据集 5 需要散射校正。

2.2　参数优化

在建模之前 PLS 的因子数及 10 种预处理方法中一阶导数、二阶导数、SG 平滑的窗口参数，CWT 的小波函数和分解尺度参数都需要进行优化，优化结果如表 3 所示。在利用 PLS 方法建立模型时，因子数从 1～25 进行变化，RMSECV 值最小时所对应的因子数，就是 PLS 建模所需要的最佳因子数[11]。SG 平滑、一阶导数、二阶导数窗口的优化方法是窗口数从 3～59 进行变化，间隔为 2，分别计算各自的 RMSEP 值，RMSEP 最小值对应的窗口即为最佳窗口。CWT 的小波函数和分解尺度的优化方法是小波函数采用 Haar、Daubechies(db2，db3，…，db20)，Coiflets(coif1，coif2，…，coif5)，Symmlets(sym2，sym3，…，sym8)32 个函数[12]，分解尺度从 1～40 进行改变，分别计算不同小波函数和分解尺度对应的 RMSEP 值。RMSEP 最小值对应的小波函数和分解尺度分别为 CWT 最佳的小波函数和分解尺度。

2.3　预处理后 PLS 建模的效果

对 9 组数据集分别进行 120 种预处理及其组合方法后，建立 PLS 模型对预测集预测的 RMSEP 值显示在图 3 中。不同预处理方法用编号 1—120 来表示，红线表示 PLS 的预测结果，蓝色散点表示每种预处理后 PLS 的预测结果。

表 3　预处理方法及 PLS 参数优化结果

Table 3　Results of preprocessing methods and PLS parameter optimization

数据集	样品	组分	PLS 因子数	平滑窗口	一阶导窗口	二阶导窗口	小波参数优化	
							分解尺度	小波函数
1	血液	胆固醇	18	15	21	43	26	db10
2	橙汁	蔗糖	8	13	27	49	40	sym5
3	四元调和油	稻米油	14	59	59	41	32	sym8
		大豆油	14	11	47	51	34	db13
4	燃油	燃料密度	18	3	3	11	5	Haar
5	汽油	双环芳烃	4	5	59	47	40	Haar
6	小麦	蛋白质	25	23	17	37	19	sym7
7	玉米	蛋白质	10	3	3	5	1	db11
8	烟草	还原糖	16	15	49	43	26	Haar
9	三元混合物	乙醇	12	9	13	27	15	Haar

从图 3(a—c)可以看出，对血液、橙汁、四元调和油数据，大部分预处理都有效果。观察血液数据的光谱信号可以看出该数据需要多元散射校正及背景扣除，但是从图 3(a)中可以看出，加入基线校正和多元散射校正的方法后预处理效果都比无预处理效果好。血液数据的最佳预处理方法是编号 69 对应的二阶导-SG 平滑-最大最小归一化。观察橙汁数据的光谱信号可以看出该数据需要背景扣除，但是从图 3(b)中可以看出，虽然加入二阶导数(60—90 区间)预处理效果有所上升，但当加入 CWT 时(90—120 区间)，预处理效果反而有所下降。橙汁数据的最佳预处理方法是编号 78 对应的二阶导-SNV-SG 平滑-Pareto 尺度化。观察四元调和油数据的光谱信号可以看出该数据无需进行预处理，但是从图 3(c)中可以看出，大部分预处理方法对于该数据集都是有效的。四元调和油数据的最佳预处理方法是编号 19 对应的 SNV-SG 平滑-最大最小归一化。

从图 3(d—f)可以看出，对燃油、汽油、小麦数据，大部分预处理方法几乎无效果。观察燃油数据的光谱信号可以看出该数据需要选择散射校正的方法进行预处理，但是从图 3

(d)中可以看出，编号 2，7，12，17，42，47 的预处理方法对数据的处理效果优于无预处理。燃油数据的最佳预处理方法是编号为 2 对应的中心化。观察汽油数据的光谱信号可以看出该数据需要散射校正和背景去除，但是从图 3(e)中可以看出，该汽油数据并不需要进行预处理。观察小麦数据的光谱信号可以看出该数据需要散射校正及背景扣除，但是从图 3(f)中可以看出，加入基线校正和散射校正方法后(除 1—10 区间外)，预处理方法对于建模效果都没有提高，反而不如无预处理。编号 6 和 8 的预处理方法对小麦数据的建模能力有所提高，SG 平滑处理效果最好。

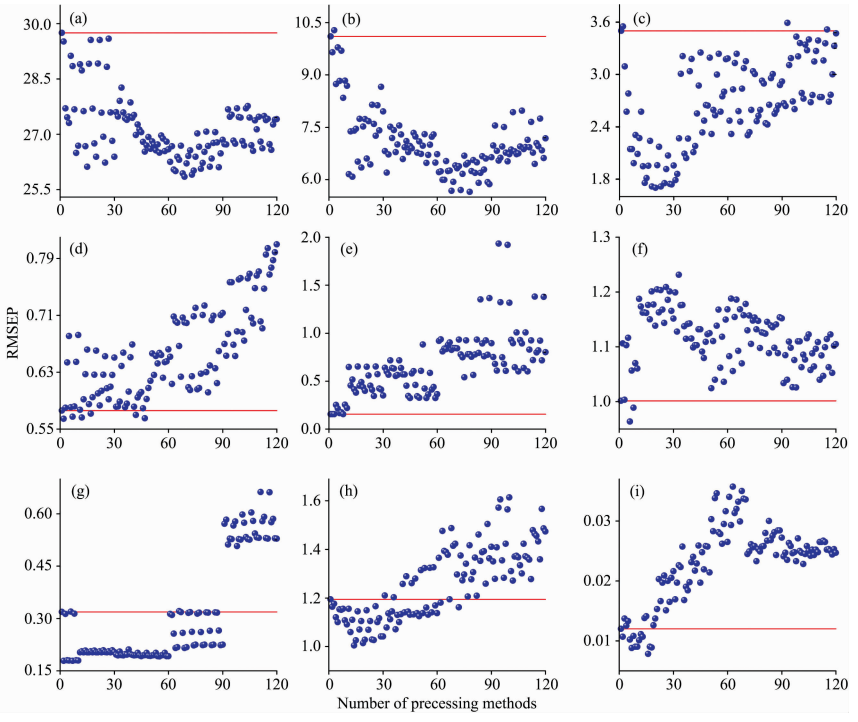

图 3 采用 PLS(红线)和 120 种预处理及其组合结合 PLS(蓝点)预测的 RMSEP 图

(a)：血液；(b)：橙汁；(c)：四元调和油；(d)：燃油；(e)：汽油；(f)：小麦；(g)：玉米；(h)：烟草；(i)：三元混合物数据集

Fig. 3 RMSEP values of PLS (red line) and 120 preprocessing and their combinations coupled with PLS (blue dots)

(a): Blood; (b): Orange juice; (c): Quarternary blended oil; (d): Fuel oil; (e): Gasoline; (f): Wheat; (g): Corn; (h): Tobacco; (i): Ternary mixture datasets

从图 3(g—i)可以看出，对玉米、烟草、三元混合物数据，部分预处理方法有效果。观察玉米数据的光谱信号可以看出该数据需要多元散射校正及背景扣除，但是从图 3(g)中可以看出，虽然加入多元散射校正和导数处理(11—90 区间)可以提高预处理效果，但是加入 CWT(90—120)预处理效果反而有所下降。玉米数据的最佳预处理方法是编号 4 对应的标准化。观察烟草数据的光谱信号可以看出该数据需要背景扣除，但是从图 3(h)中可以看出，加入一阶导时(30—60 区间)预处理效果有所加强，加入二阶导和 CWT 时(60—120 区间)，预处理效果反而有所下降。烟草数据的最佳预处理方法是编号 14 对应的 SNV-最大最小归一化的组合方式。观察三元混合物数据的光谱信号可以看出该数据需要多元散射校正及背景扣除，但是从图 3(i)中可以看出，加入基线校正的方法(30—120 区间)预处理效果不如无预处理，加入散射校正的方法，只是部分的提高了预处理效果(11—18 区间)。三元混合物数据的最佳预处理方法是编号 16 对应的 SNV-SG 平滑的组合方式。综合上述数据的实验结果，光谱信号特点选择的预处理方法大多都不是建模效果最优的预处理方法，因此，光谱预处理方法的选择应结合建模效果而不是根据光谱信号特点直接选择。

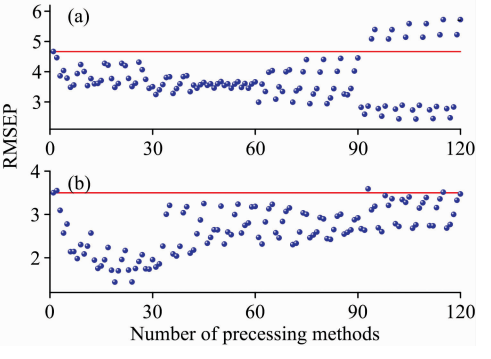

图 4 四元调和油数据 PLS(红线)和 120 种预处理及其组合结合 PLS(蓝点)预测的 RMSEP 图

(a)：大豆油组分；(b)：稻米油组分

Fig. 4 RMSEP values of PLS (red line) and 120 preprocessing and their combinations coupled with PLS (blue dots) for quarternary blended oildataset

(a): Soybean oil; (b): Rice oil

2.4　同一光谱不同组分的预处理效果

对于相同样品的不同组分进行预测，由于扫描的原始光谱是相同的，因此，根据光谱信号特点选择的预处理方法肯定是相同的，但是根据建模效果选择的预处理方法未必相同。我们对四元调和油数据中大豆油和稻米油组分分别进行120种预处理组合后PLS建模，相应的RMSEP显示在图4(a，b)中。图4(a)大部分预处理方法都是有效的，但是当CWT与最大最小归一化或标准化同时存在时，预处理效果明显降低。四元调和油数据中的大豆油组分的最佳预处理方法是编号107对应CWT-SNV-SG平滑-中心化。图4(b)加入散射校正的方法(11—30区间)，预处理效果增强。四元调和油数据中稻米油组分的最佳预处理方法是编号24对应的MSC-最大最小归一化。通过两个组分的比较表明了对于相同样品集，即使光谱相同，但不同组分的预处理方法也不相同。所以，预测结果除了与光谱有关，还有预测组分有关。

3　结　论

对不同数据及相同数据的不同组分分别进行120种预处理及其组合，分析光谱信号特点及预处理后PLS建模的RMSEP值。结果表明：(1)相比直接观察光谱信号特点，根据光谱与预测组分的建模效果可以更为准确地选择最佳预处理方法；(2)对于不同的数据集，因为其数据集信息和复杂性不同，所以其最佳预处理方法也不同；对于相同样品集，即使光谱相同，但不同组分的预处理方法也不相同。因此，不存在普适性的最佳预处理方法，预测结果除了与光谱有关，还有预测组分有关；(3)通过对已有预处理方法按照预处理目的进行分类再排列组合是选择最佳预处理方法的一种有效途径。

References

[1] Li P, Du G R, Cai W S, et al. Journal of Pharmaceutical and Biomedical Analysis, 2012, 70: 288.
[2] Devos O, Downey G, Duponchel L. Food Chemistry, 2014, 148: 124.
[3] Liu Y J, Yu Y D, Zhou X G, et al. Chemometrics and Intelligent Laboratory Systems, 2017, 161: 8.
[4] Gerretzen J, Szymanska E, Jansen J J, et al. Analytical Chemistry, 2015, 87(24): 12096.
[5] Engel J, Gerretzen J, Szymanska E, et al. Trends in Analytical Chemistry, 2013, 50: 96.
[6] Zhu Z Q, Yuan H F, Song C F, et al. Sensors and Actuators B: Chemical, 2018, 268: 299.
[7] Qiao X X, Wang C, Feng M C, et al. Spectroscopy Letters, 2017, 50(3): 156.
[8] Li Y K, Shao X G, Cai W S. Talanta, 2007, 72(1): 217.
[9] Wulfert F, Kok W T, Smilde A K. Analytical Chemistry, 1998, 70: 1761.
[10] Gerretzen J, Szymanska E, Bart J, et al. Analytica Chimica Acta, 2016, 938: 44.
[11] Bian X H, Li S J, Lin L G, et al. Analytica Chimica Acta, 2016, 925: 16.
[12] Liu P, Wang J, Li Q, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, doi:10.1016.

Study on the Selection of Spectral Preprocessing Methods

DIWU Peng-yao, BIAN Xi-hui*, WANG Zi-fang, LIU Wei

State Key Laboratory of Separation Membranes and Membrane Processes, School of Environmental and Chemical Engineering, Tianjin Polytechnic University, Tianjin 300387, China

Abstract Spectral signals of complex samples are usually disturbed by stray light, noise, baseline drift and other undesirable factors, which can affect the final qualitative and quantitative analysis results. Therefore, it is necessary to pretreat the raw spectra before modeling. How to find a proper preprocessing method from the existing spectral preprocessing methods is a difficult problem. One strategy is to choose the optimal preprocessing by observing the characteristics of the spectral signal directly, which does not require modeling and is more explanatory. However, it may be difficult and subjective for subtle or multiple interferences and lead to misleading results. Another strategy is based on the modeling performance, which does not need observe the spectral characteristics, but numerous processing methods need to investigate which is time-consuming for large datasets. In summary, it is necessary to explore which selection method is more scientific and reasonable. In this study, nine datasets were used to investigate the necessity of preprocessing and the choice of preprocessing methods by arranging and combining of 10 preprocessing methods. Firstly, the latent variables of partial least squares (PLS), the window size of first derivative (1st Der), second derivative (2nd Der) and SG smoothing, the wavelet function and decomposition scale of continuous wavelet transform (CWT) were optimized, respectively. Then, non-preprocessing and 10 preprocessing methods including 1st Der, 2nd Der, CWT, multiplicative scatter correction (MSC), standard normal variate (SNV), SG smoothing, mean centering, normalization,

Pareto scaling, auto scaling were combined in order of baseline correction, scattering correction, smoothing and scaling. A total of 120 preprocessing and their combinations were obtained. Finally, the characteristics of spectral signals and the root mean squared error of prediction (RMSEP) with PLS for 120 preprocessing methods were analyzed for the nine datasets and the same dataset with different components. Results show that compared with observing the characteristics of spectral signals, the optimal preprocessing method can be selected more accurately according to the modeling performance of the spectra and predictive components. For most datasets, appropriate preprocessing method can improve the modeling performance. For different datasets, the optimal preprocessing method is different because of the different information and complexity of the datasets. For the same dataset, the optimal preprocessing methods for different components are also different even if the spectra are the same. Thus, it can be concluded that no universal preprocessing method exists. The optimal preprocessing method is related to the spectra and the predictive components. Furthermore, it is an effective way to select the optimal pretreatment method by sorting and combining the existing preprocessing methods according to the preprocessing purpose.

Keywords Preprocessing method; Complex sample; Partial least squares; Parameter optimization; Method selection

(Received Aug. 2, 2018; accepted Dec. 18, 2018)

* Corresponding author

《光谱学与光谱分析》期刊社决定采用 ScholarOne Manuscripts 在线投稿审稿系统

《光谱学与光谱分析》期刊社与汤森路透集团签约，自 2010 年 12 月 1 日起《光谱学与光谱分析》决定采用 Thomson Reuters旗下的 ScholarOne Manuscripts 在线投稿审稿系统。

• ScholarOne Manuscripts，该系统不仅能轻松处理稿件，而且能提速科技交流。

• 全球已有 360 多家学会和出版社的 3 800 多种期刊选用了 ScholarOne Manuscripts 系统作为在线投稿、审稿平台，全球拥有超过 1 350 万的注册用户，代表着全球学术期刊在线投审稿的一流水平。

• ScholarOne Manuscripts 与 EndNote，Web of Science 无缝链接和整合；使科研探索、论文评阅和信息传播效率大为提高。

• ScholarOne Manuscripts 是汤森路透科技集团的一个业务部门，拥有丰富的学术期刊业务经验，为学术期刊提供综合管理工作流程系统，使期刊更有效管理投稿、同行评审、加工和发表过程，提高作者心中的专业形象，缩短论文发表时间，削减管理成本，帮助期刊提高科研绩效和实现学术创新。

《光谱学与光谱分析》采用“全球学术期刊首选的在线投稿审稿系统—ScholarOne Manuscripts”，势必对 2010 年 11 月 30 日以前向本刊投稿的作者在查阅稿件信息时，会带来某些不便，在此深表歉意！为了推进本刊的网络化、数字化、国际化进程，以实现与国际先进出版系统对接；为了不断提高期刊质量，加快网络化、数字化建设，加快与国际接轨的进程，希望能得到广大作者、读者们的支持与理解，对您的理解和配合深表感激。这是一件新事物，肯定有不周全、不完善的地方，让我们共同努力，不断改进和完善起来。

《光谱学与光谱分析》期刊社

2010 年 12 月 1 日