

聚谷氨酸发酵过程中 ATR-FTIR 光谱信号的分数阶基线校正

何年, 单鹏*, 贺忠海, 王巧云, 李志刚, 吴缀

东北大学秦皇岛分校控制工程学院, 河北 秦皇岛 066000

摘要 采用衰减全反射傅里叶变换红外光谱法(ATR-FTIR), 结合多元校正模型对 γ -聚谷氨酸(γ -PGA)发酵过程中两种主要底物葡萄糖和谷氨酸钠的浓度进行间接测量, 为优化发酵系统控制提供重要的反馈信息。光谱测量中经常出现的基线漂移会严重影响后续多元校正模型的性能, 需要采用基线校正算法对光谱进行预处理。现有流行的基线校正算法多数是基于 Whittaker Smoother(WS)平滑算法, 这些算法均采用整数阶微分对拟合基线进行约束, 表达能力有限。针对现有基线校正算法中的整数阶微分自适应性差的问题, 利用更加灵活的分数阶微分对基线进行约束, 提出了一种基于分数阶的基线校正算法, 实现对整数阶基线校正的扩展。总共进行了5个批次的 γ -PGA发酵实验, 并对不同批次和全部批次的 ATR-FTIR 光谱数据分别进行了分数阶基线校正, 模型的预测精度均得到不同程度的提升。实验结果表明, 只有在批次2时, 基于整数阶的基线校正效果最好; 其他批次的基线校正效果最好时的阶次均为分数阶。这也表明了分数阶微分(包含整数阶微分)对基线的约束更加合理。同时发现全部批次的整体基线校正效果远远差于单一批次的效果, 原因可能是各批次发酵光谱的基线是不同的, 对不同的批次需要选用不同的阶次以获得最佳的基线校正。此外, γ -PGA 发酵样品的 ATR-FTIR 光谱测量是以蒸馏水为背景, 会在 $3\ 100\sim 3\ 600\ \text{cm}^{-1}$ 波数范围内出现负水峰, 形成有害的干扰信号; 分数阶基线校正后的光谱表明, 分数阶基线校正算法将负的水峰当作基线, 在一定程度上进行了消除。综合分析, 分数阶基线校正算法不仅扩展了传统整数阶基线校正算法的应用范围, 也为消除 ATR 光谱中负的水峰提供了新的解决思路。

关键词 分数阶微分; 光谱预处理; 基线校正

中图分类号: O657.33

文献标识码: A

DOI: 10.3964/j.issn.1000-0593(2022)06-1848-07

引言

发酵过程中对物料浓度的在线检测是保证产品质量的关键, 目前仍有大部分的工厂采用传统的分离分析方法, 费时繁琐, 无法实现发酵的实时控制。而衰减全反射傅里叶变换红外光谱(attenuation reflectance Fourier transformation infrared spectrometry, ATR-FTIR)技术具有快速、绿色、无损等优点, 应用广泛^[1-3], 十分适合发酵过程中参数的在线检测。本文通过对聚谷氨酸发酵过程中采集的 ATR-FTIR 光谱与建立的多元校正模型间接的测量发酵过程中主要底物葡萄糖和谷氨酸钠的浓度。因光谱中存在基线漂移, 影响定量分析精度, 所以在建立多元校正模型前需进行光谱预处理, 去除噪声和基线。

其中基于 WS 平滑算法的基线校正算法应用广泛, 已成

功应用于各种光谱分析中。Eilers 等首先将 WS 算法应用于信号的平滑和插值^[4], 随后通过添加非对称权重, 在峰值信号处实现较强的光滑, 将光滑曲线作为拟合基线, 提出非对称最小二乘基线校正算法^[5] (asymmetric least squares, AsLS); Zhang 等通过改进非对称权重, 实现自适应迭代, 提出自适应迭代加权惩罚最小二乘^[6] (adaptive iteratively reweighted penalized least squares, airpls); Baek^[7] 等将权重按 logistic 函数分配, 实现权重信息柔化。 (asymmetrically reweighted penalized least squares smoothing, arpls); 姜安^[8] 等通过引入拟合残差的一阶导数作为新的惩罚项, 用直方图估计背景设定阈值, 加快迭代速度 (jiang improved asymmetric least squares baseline correction algorithm, jasls); He^[9] 等通过改进 jasls 的阈值, 形成 (improved asymmetric least squares method, iasls); 近年 Xu^[10] 等给惩罚项加上新的权重, 形成重加权惩罚基线校正 (doubly reweighted penalized

收稿日期: 2021-05-11, 修订日期: 2021-07-29

基金项目: 国家自然科学基金青年科学基金项目(61601104), 教育部基本科研业务费国家项目培育基金项目(N2023021)资助

作者简介: 单鹏, 1985年生, 东北大学秦皇岛分校控制工程学院讲师 e-mail: 2031816313@qq.com

* 通讯作者 e-mail: peng.shan@neuq.edu.cn

least squares, drpls); Ye^[11]等通过改进 arpls 的权重赋值函数, 实现进一步优化(improved asymmetricly reweighted penalized least squares, iarpls)。

这些算法多是以改变权重或自动迭代权重的方式进行改进, 但对基线的约束都采取相同策略, 选择阶次较低的整数阶微分; 此外还引入新的惩罚项来改进算法性能。这种改进可以通过引入不同的惩罚项, 实现对基线不同的约束, 原模型中的惩罚项是对基线的粗糙度进行约束, 在 jpls 算法中加入了真实光谱和拟合基线残差的一阶微分作为新的惩罚项, 他们认为拟合出的基线不仅与原始数据之间的误差很小, 而且还要求它们的一阶导数很接近。这使得对拟合基线的约束性更强, 能够适应不同的光谱。

上述提及的各种改进算法都未曾涉及对粗糙度描述方法进行改进, 都沿用整数阶微分或直接固定阶次, 使得对基线的约束不够灵活。低阶整数阶微分通常只取 1, 2 和 3 阶, 可选择性极差; 同时考虑到整数阶微分不能很好的描述基线的特点, 且在实际信号中整数阶阶次的信号很少见。故引入分数阶微分的概念, 提高算法灵活性, 扩展对粗糙度的描述, 从而进一步研究微分阶次对基线校正效果的影响。我们提出的分数阶基线校正算法涵盖了原来的整数阶算法, 理论上认为分数阶基线校正效果不会差于整数阶基线校正效果; 这一推断在 γ -PGA 发酵光谱分析中进行了仔细的检验。

1 原 理

1.1 AsLS 基线校正算法原理

AsLS 基线校正是在 WS 平滑算法的基础上改进而来, 首先对 WS 平滑算法进行简单介绍^[4]。WS 优化目标如式(1)

$$\arg \min_z \left[\sum_i (y_i - z_i)^2 + \lambda \sum_i (\Delta z_i)^2 \right] \quad (1)$$

式(1)中, y_i 为原始信号的第 i 个点, z_i 为平滑序列的第 i 个点, λ 为正则化系数, Δ 为微分算子, 一阶微分可表示 $\Delta z_i = z_i - z_{i-1}$ 。Eilers 在文中将最小化目标函数的第一项称为保真度, 第二项称为粗糙度。通过调节 λ 得到合适的平滑序列。当 λ 越大, 对粗糙度的惩罚越大, 就要求序列越平滑。对于一阶微分, 当 λ 越大时, 平滑序列 z 就越趋向于一条水平的直线。同理二阶微分, 在 λ 越大时, 滑序列 z 就越趋向于一条倾斜的直线。

为简化代数运算用矩阵表示目标函数

$$Q = \|y - z\|^2 + \lambda \|D_d z\|^2 \quad (2)$$

最小化式(2)可得

$$z = (E + \lambda D_d^T D_d)^{-1} y$$

式(2)中, y 为原始信号列向量, z 为平滑序列列向量, E 为单位矩阵, D_d 为微分矩阵, 只能表示整数阶。当信号长度为 5, 取整数一阶、二阶微分时, 表示如下

$$D_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

$$D_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{bmatrix}$$

AsLS 在 WS 算法基础上引入非对称权重作用于保真度^[5]

$$Q = W \|y - z\|^2 + \lambda \|D_d z\|^2 \quad (3)$$

最小化式(3)可得

$$z = (W + \lambda D_d^T D_d)^{-1} W y$$

式(3)中, W 为权重对角矩阵, 式中权重系数 W_i 根据非对称的方式选择

$$W_i = \begin{cases} p & y_i > z_i \\ 1-p & y_i \leq z_i \end{cases}$$

一般 p 取值范围为 0.001~0.1。 λ 取值范围为 $10^2 \sim 10^9$, 固定迭代次数, W 一般迭代 10 次。

AsLS 也可以看作是平滑算法, 通过对含有峰的光谱信号进行平滑, 得到一条光滑的不含有峰的曲线作为基线, 这种基线校正方法不需要任何的先验信息, 只需要通过调节反对称权重和正则化系数就能得到一条适合的基线。

1.2 基于分数阶的基线校正算法原理

AsLS 算法中的 D_d 微分算子只适用于整数阶, 在 D_d 的基础上扩展到分数阶, 较为简便的实现分数阶基线校正。分数阶微分定义有不同的形式, 为了更好的包含原有整数阶, 选用 Grunwald-Letnikov (GL) 分数阶微分定义, 表示如式(4)^[12-14]

$$d^a f(x) = \lim_{h \rightarrow 0} \frac{1}{h^a} \sum_{m=0}^{\frac{x-a}{h}} (-1)^m \frac{\Gamma(a+1)}{m! \Gamma(a-m+1)} f(x-mh) \quad (4)$$

式(4)中: α 为阶数; h 为微分步长; t 与 a 分别为微分的上、下限; $\Gamma(x)$ 为 Gamma 函数。当函数 $f(x)$ 定义域为 $x \in [a, t]$ 且 $h=1$ 时, 由式(4)可得出 $f(x)$ 的分数阶微分表达式

$$\frac{d^a f(x)}{dx^a} \approx f(x) + (-a)f(x-1) + \frac{(-a)(-a+1)}{2} f(x-2) + \dots + \frac{\Gamma(-a+n)}{n! \Gamma(-a)} f(x-n) \quad (5)$$

式(5)中: α 为阶数, $f(x)$ 的 0 阶微分为 $f(x)$ 本身。同样的将微分差值运算构造成矩阵的形式, 记为 D_a 。

$$D_a = \begin{bmatrix} d_1 & 0 & \dots & 0 & \dots & 0 \\ d_2 & d_1 & 0 & & & \ddots \\ d_3 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & d_2 & d_1 & 0 & 0 \\ d_k & & \ddots & \ddots & \ddots & \ddots \\ \vdots & \ddots & & \ddots & d_2 & d_1 & 0 \\ d_{n+1} & \dots & d_k & \dots & d_3 & d_2 & d_1 \end{bmatrix}$$

其中 $d_{n+1} = \frac{\Gamma(-\alpha+n)}{n! \Gamma(-\alpha)}$, $n = t - a$; 通常真实光谱波数数目在 1 000 以上, 但当 n 较大时, d_{n+1} 很小, 基本可以忽略不计, 同时为了加快计算速度, 节省存储空间, 设置一个长度 k , 将 k 以后的系数全部置为零。矩阵结构如下

$$D_{\alpha} = \begin{bmatrix} d_1 & 0 & \cdots & 0 & \cdots & 0 \\ d_2 & d_1 & 0 & & & \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ d_k & \cdots & d_2 & d_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & d_k & \cdots & d_2 & d_1 & 0 \\ 0 & \cdots & 0 & d_k & \cdots & d_2 & d_1 \end{bmatrix}$$

实验表明,当 k 较小时,基线校正效果较差,一般取 20 以上的值。但计算速度会随 k 的增大而变慢。用 D_{α} 替换 D_d 即可实现分数阶基线校正(fractional differential asymmetric least squares, FdAsLS)。当 α 取整数阶时,比原来的整数阶微分矩阵多了几项,这对于一些信号的校正是不利的,因为起始部分点的微分变化较大,容易造成基线的突变,但对于原始信号起始部分基本为零的光谱信号,并不会产生影响。在取分数阶时,该现象尤为明显,因为前 k 行的微分表达式都是不相同的,从矩阵的前 k 行,可以明显的观察到。GL 定义下的整数阶矩阵相比于原来整数阶矩阵,略有不同。当信号长度为 5, GL 定义下的整数一阶、二阶微分时,表示如下

$$D_1^1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

$$D_2^1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{bmatrix}$$

2 实验部分

2.1 数据采集

对于 γ -PGA 发酵实验,选用的菌种为枯草芽孢杆菌亚种,从中国工业微生物菌种保藏管理中心(China Center of Industrial Culture Collection, CICC)购买,菌种编号为 20643。将以冻干粉的形式存储的菌种先进行活化培养,然后在培养好的固体菌落中,用接种环挑选一株生长状态良好的菌体,接种于种子培养基(500 mL 三角瓶装液量 100 mL),然后在 37 °C 和 180 r·min⁻¹ 的恒温振荡培养箱中(THZ-92A, 跃进医疗器械有限公司,中国上海)中培养 10~16 h。所用种子培养基为:葡萄糖(10 g·L⁻¹),蛋白胨(10 g·L⁻¹),牛肉膏(5 g·L⁻¹),氯化钠(5 g·L⁻¹)。发酵培养基由葡萄糖(40 g·L⁻¹),酵母膏(5 g·L⁻¹),谷氨酸钠(35 g·L⁻¹),氯化铵(2 g·L⁻¹),磷酸氢二钾(5 g·L⁻¹)和硫酸镁组成(0.5 g·L⁻¹)组成。种子培养基和发酵培养基均在 121 °C 下灭菌 20 min。将经过种子培养的菌株接种到接种量为 2% 的发酵培养基中,并将 3 L 的发酵培养液放入工作容积为 5 L 的发酵罐(GRJB-5D, 绿色生物工程有限公司,中国镇江)中,在 37 °C 恒温和 300 r·min⁻¹ 搅拌速度的条件下进行发酵。

用配备有水平铂金刚石 ATR 采样附件(ZnSe, 单反射)的布鲁克 Alpha 型傅里叶变换红外光谱仪(德国,埃特林根)上收集光谱数据。在 35 °C 下,以 8 cm⁻¹ 的分辨率在 4 000~600 cm⁻¹ 的波数范围内进行 64 次扫描。每个样品测量之前,用蒸馏水作为参考获取背景光谱。对于每个样品,重复测量两次。所得的平均光谱用于进一步分析。作为主要底物,葡萄糖(g·L⁻¹)和谷氨酸钠(g·L⁻¹)是用于监测 γ -PGA 发酵的参数。总共进行了 48 h 的 5 次发酵实验,获得 151 个样品的光谱及其发酵参数的标准值。各批次样本数分别为 14, 27, 40, 40 和 30, 它们的光谱如图 1 所示。

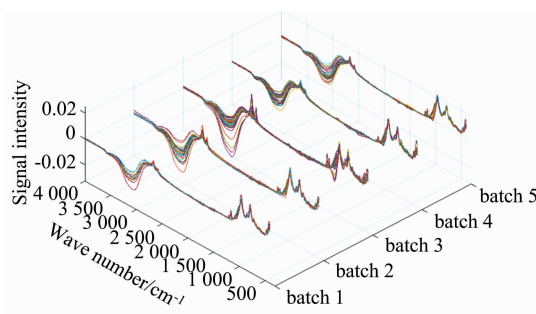


图 1 各批次光谱

Fig. 1 Spectra of each batch

2.2 模型建立

各批次样本根据 Kennard-Stone(KS)算法,按 3:1 比例划分样本,3 份作为校正集,建立模型。其余 1 份构成测试集,验证模型。多元校正模型选择偏最小二乘回归(partial least squares regression, PLS)^[15], PLS 采用 5 折交叉验证,从 1~15 中选出最佳潜变量个数。评价指标选择 PLS 模型的校正集均方根误差(root mean square error of calibration, RMSEC)和测试集均方根误差(root mean square error of prediction, RMSEP)。对 5 个批次数据分别建立原始光谱、原始光谱+FdAsLS 预处理、原始光谱+AsLS 预处理的 PLS 模型。最后合并所有批次样本,重新划分校正集和测试集,建模过程不变。

3 结果与讨论

为验证分数阶基线校正效果,进行六组实验。每组实验以 PLS 模型的 RMSEP 为评价指标。固定反对称权重 $p=0.001$ 和分数阶微分长度 $k=20$ 。微分阶次从 0.5 到 4.5,间隔 0.1,共 41 个阶次; λ 从 10^0 到 10^9 ,幂指数间隔 0.5,共 19 个数,通过网格搜索法筛选出最佳参数组合。为了比较分数阶和整数阶基线校正算法性能的优劣,我们将所有整数阶参数组合(其中 0 阶是用未经过基线校正的原始数据建模)和最佳分数阶参数组合(对应最小的 RMSEP)在六组实验上的结果进行了汇总(见表 1)。考虑到两种整数阶微分表示不同,所有实验,在取整数阶时,选择 AsLS 中的原始定义。其中只有批次 2 的预测模型误差减小最多所对应的基线校正微分阶次为整数阶,其余模型最佳阶次都为分数阶。可以反映出分数阶微分基线校正有着不低于整数阶的校正效果,大多数

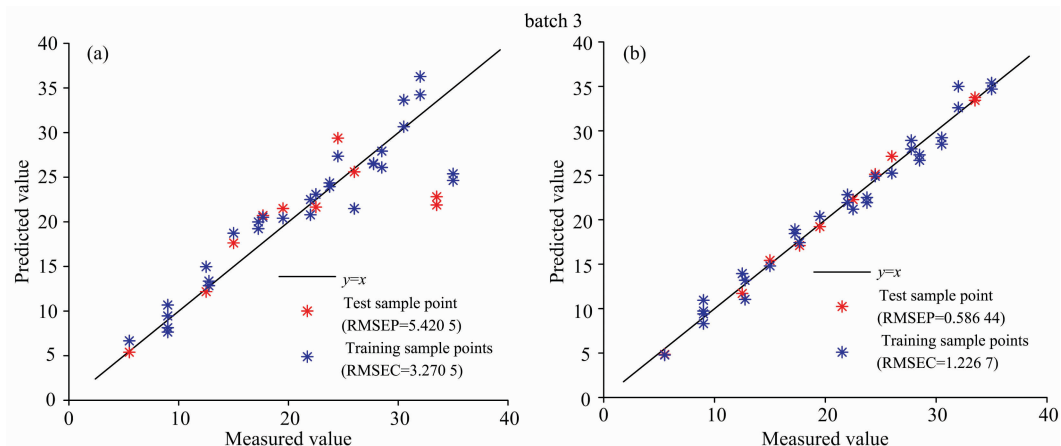
情况下都超过了整数阶。同时分数阶微分阶次取值任意，运用灵活。通过网格搜索法得到的最佳阶次中，取值基本没有重复，足以说明分数阶微分具有自适应性，能够灵活地提取真实基线，因此相比于从前单一的整数阶微分，分数阶微分能够充分发挥基线校正的优势，能最大限度的提高多元校正模型的预测精度。在批次 1 中，底物的预测误差大幅减小，葡萄糖预测模型的 RMSEP 从 2.098 降到 0.857，预测精度提高较多。同时其他批次预测精度都大幅提升，表明了基线校正有助于后续光谱定量分析。将 5 个批次的样本合并后建立模型(未进行预处理的原始数据)，可以明显地观察到，全

局模型的预测结果不如各批次单独建模(批次 3 除外)的局部模型；且各批次合并后的样本即使经过基线校正后，模型精度提升也远差于各批次单独进行基线校正后的预测精度，即校正效果不明显。这一现象的产生可以归因于不同发酵批次的基线是不相同的，通过固定一组参数对所有批次光谱进行相同的基线校正是不合理的。从实际情况来考虑，在不同的发酵批次过程中，很多因素(如：谷草芽孢杆菌的活性、测量条件、仪器性能的差异等)都会造成基线的变化，因此应该对不同批次光谱单独进行基线校正，才能充分发挥基线校正的效果。

表 1 不同阶次的基线校正效果比较

Table 1 Comparison of baseline correction effects of different orders

| ingredient | | glucose | | Sodium Glutamate | | ingredient | | glucose | | Sodium Glutamate | |
|------------|--------------|---------|--------------|------------------|------------|--------------|-------|--------------|-------|------------------|--|
| Order | RMSEP | RMSEC | RMSEP | RMSEC | Order | RMSEP | RMSEC | RMSEP | RMSEC | | |
| All batch | | | | | | batch1 | | | | | |
| 0 | 2.703 | 2.059 | 0.937 | 0.386 | 0 | 2.098 | 0.551 | 0.502 | 1.119 | | |
| 0.6 | 2.436 | 1.650 | 0.873 | 0.418 | 1 | 1.280 | 1.271 | 0.306 | 0.897 | | |
| 1 | 2.243 | 1.886 | 0.911 | 0.381 | 2 | 1.124 | 1.037 | 0.317 | 0.474 | | |
| 1.3 | 2.200 | 1.477 | 0.899 | 0.493 | 2.4 | 1.165 | 0.964 | 0.093 | 1.173 | | |
| 2 | 2.322 | 1.667 | 0.882 | 0.451 | 3 | 0.898 | 1.098 | 0.279 | 0.822 | | |
| 3 | 2.353 | 1.777 | 0.888 | 0.292 | 3.3 | 0.857 | 0.905 | 0.401 | 0.425 | | |
| 4 | 2.414 | 1.342 | 0.942 | 0.350 | 4 | 1.110 | 1.464 | 0.401 | 0.434 | | |
| batch2 | | | | | | batch3 | | | | | |
| 0 | 1.493 | 0.361 | 1.052 | 0.201 | 0 | 5.421 | 3.271 | 0.861 | 0.499 | | |
| 1 | 1.183 | 0.789 | 0.752 | 0.602 | 1 | 0.935 | 1.420 | 0.932 | 0.720 | | |
| 2 | 1.564 | 1.091 | 0.637 | 0.424 | 2 | 1.081 | 1.758 | 0.941 | 0.273 | | |
| 2.1 | 1.265 | 0.174 | 0.697 | 0.258 | 3 | 0.647 | 1.267 | 0.915 | 0.421 | | |
| 3 | 1.514 | 0.520 | 0.663 | 0.432 | 3.2 | 0.586 | 1.227 | 0.572 | 0.388 | | |
| 4 | 1.537 | 0.532 | 0.670 | 0.150 | 4 | 0.606 | 1.447 | 0.898 | 0.518 | | |
| batch4 | | | | | | batch5 | | | | | |
| 0 | 1.471 | 0.599 | 0.685 | 0.628 | 0 | 0.786 | 0.607 | 0.732 | 0.530 | | |
| 1 | 1.459 | 0.659 | 0.674 | 0.722 | 1 | 0.953 | 1.003 | 0.453 | 0.279 | | |
| 2 | 1.177 | 0.566 | 0.571 | 0.733 | 2 | 0.949 | 0.633 | 0.317 | 0.295 | | |
| 2.3 | 1.411 | 0.430 | 0.464 | 0.655 | 3 | 0.881 | 0.399 | 0.264 | 0.300 | | |
| 3 | 1.285 | 0.810 | 0.637 | 0.742 | 4 | 1.047 | 0.239 | 0.253 | 0.302 | | |
| 3.8 | 0.652 | 0.524 | 0.620 | 0.618 | 4.1 | 0.402 | 0.920 | 0.271 | 0.299 | | |
| 4 | 1.137 | 0.960 | 0.621 | 0.435 | 4.3 | 0.474 | 0.888 | 0.208 | 0.315 | | |



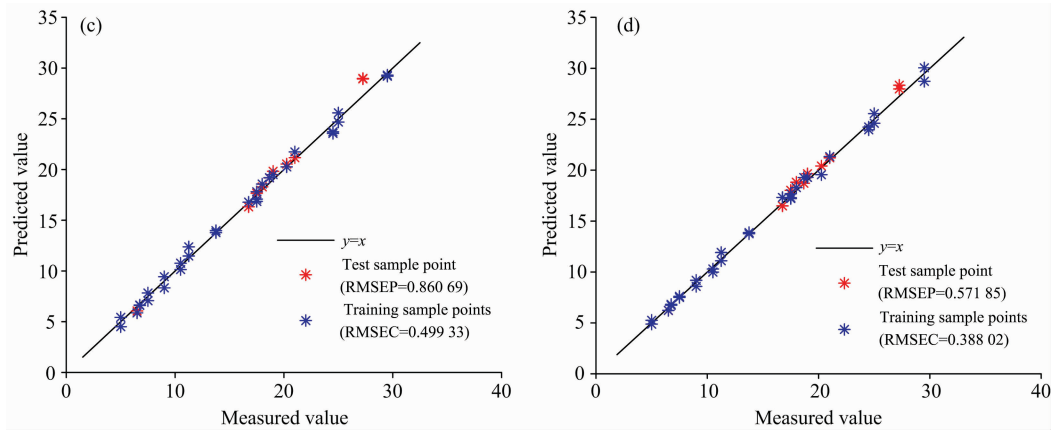


图 2 偏最小二乘回归分析

(a): 原始光谱葡萄糖预测结果; (b): 基线校正后葡萄糖预测结果;
(c): 原始光谱谷氨酸钠预测结果; (d): 基线校正后谷氨酸钠预测结果

Fig. 2 Partial least squares regression analysis

(a): Predicted result of glucose concentration by the original spectrum; (b): Predicted result of glucose concentration after baseline correction;
(c): Predicted result of sodium glutamate concentration by the original spectrum; (d): Predicted result of sodium glutamate concentration after baseline correction

以批次 3 为例, 对原始光谱建立 PLS 模型, 发现预测结果很差, 但经过基线校正以后, 预测均方根误差减小, 图 2 可以直观的反映出这种变化, 其余各批次经过基线校正以后, 结果如图 3 和图 4 所示, 其中图 3 是预测葡萄糖浓度 RMSEP 最小时所对应的基线校正处理后的结果, 图 4 对应于谷氨酸钠。

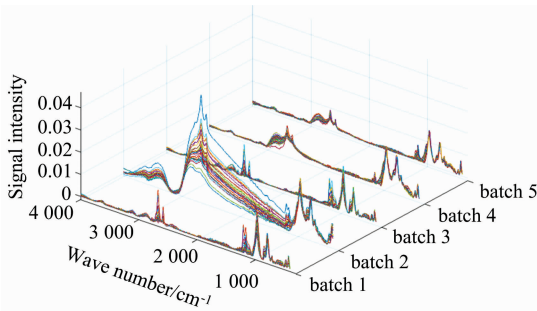


图 3 葡萄糖 RMSEP 最小时对应各批次基线校正后的光谱

Fig. 3 Baseline corrected spectra of glucose for each batch with the minimum RMSEP

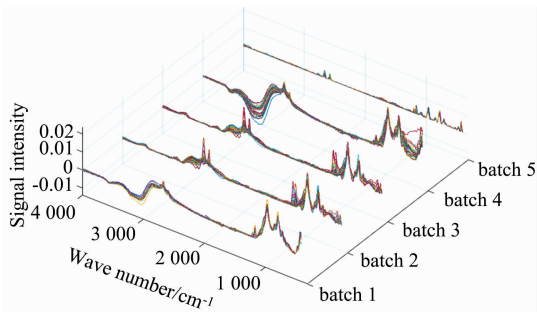


图 4 谷氨酸钠 RMSEP 最小时对应各批次基线校正后的光谱

Fig. 4 Baseline corrected spectra of sodium glutamate for each batch with the minimum RMSEP

经过基线校正后的光谱相较于原始光谱, 除批次 4 用以预测谷氨酸钠浓度的校正光谱还保留负的水峰, 其余校正光谱基本不含或含有少量负的水峰。因此认为峰信号全部是有用的, 通常 AsLS 被用于只含有全部为正峰或全部为负峰的信号的校正, 但是对于用水溶液测得的 ATR 光谱, 负的水峰对于后续的分析是无用的甚至是有害的, 所以可以通过 AsLS 基线校正将水峰扣除。改进后的 FdAsLS 同样具有该效果, 且对负峰的扣除效果更好。

各批次校正后的光谱各不相同, 间接反映出发酵过程的特殊性。且校正后的光谱各批次存在较大的差异, 如批次 4 预测谷氨酸钠时, 基线校正效果作用不大, 校正后的光谱基本不变, 且经过基线校正后, RMSEC 反而变差了。理论上 RMSEC 应该小于 RMSEP, 但是在该数据中, 各批次经过基线校正后部分批次出现 RMSEP 小于 RMSEC 的情况。这可能是样本太小所造成的, 同时测试集样本中个别样本预测很差, 但经过基线校正后, 预测精度立刻提升, 图 2(a, b) 中红色标记的测试集样本就属于这种情况。图 2(a) 中存在明显偏离斜线 $y=x$ (x, y 分别为测量值和 PLS 模型预测值。样本点越偏离斜线, 说明模型预测精度越差, 反之预测精度越高) 的点, 但经过基线校正 [图 2(b)], 红色点十分贴近斜线且基本均匀分布在斜线的两侧。一方面样本的均匀分布, 说明 KS 算法样本划分合理; 另一方面样本点经过基线校正后靠近斜线, 说明分数阶基线校正算法的有效性。当样本量变大时, 即所有批次合并后, 重新划分校正集测试集, 所得实验结果与理论相符。

4 结 论

基于 WS 平滑算法的各种基线校正算法 (包括 WS 平滑算法) 都可以通过微分算子扩展至任意阶次, 具有更好的灵

活性和可选择性, 从而进一步提高多元校正模型的预测精度。 γ -PGA 发酵实验的光谱数据分析结果表明, 分数阶微分基线校正效果优于整数阶。同时发现 AsLS 和 FdAsLS 基线校正算法在去除基线的同时, 实现了对 ATR 光谱水峰的扣

除, 消除水峰对后续光谱定量的影响, 扩展了该基线校正算法应用范围。同时为用水溶液测得的 ATR 光谱消除水峰提供了新的思路。

References

- [1] Schorn-García D, Cavaglia J, Giussani B, et al. *Microchemical Journal*, 2021, 166: 106215.
- [2] Pu Y Y, O'Donnell C, Tobin J T, et al. *International Dairy Journal*, 2020, 103: 104623.
- [3] Cavaglia J, Schorn-García D, Giussani B, et al. *Food Control*, 2020, 109: 106947.
- [4] Eilers P H C. *Analytical Chemistry*, 2003, 75: 3631.
- [5] Eilers P, Boelens H. *Baseline Correction with Asymmetric Least Squares Smoothing*, Leiden University Medical Center Report 1, 2005.
- [6] Zhang Zhimin, Chen Shan, Liang Yizeng. *Analyst*, 2010, 135(5): 1138.
- [7] Baek S J, Park A, Ahn Y J, et al. *Analyst*, 2015, 140(1): 250.
- [8] JIANG An, PENG Jiang-tao, XIE Qi-wei, et al(姜 安, 彭江涛, 谢启伟, 等). *Computers and Applied Chemistry(计算机与应用化学)*, 2012, 29(5): 537.
- [9] He Shixuan, Zhang Wei, Liu Lijuan, et al. *Anal. Methods*, 2014.
- [10] Xu D G, Liu S, Cai Y Y, et al. *Applied Optics*, 2019, 58(14): 3913.
- [11] Ye Jianfeng, Tian Ziyang, Wei Haoyun, et al. *Applied Optics*, 2020, 59(34): 10933.
- [12] Loverro A. *Fractional Calculus: History, Definitions and Applications for the Engineer*. Report, 2004, Corpus ID: 18520731.
- [13] Hong Yongsheng, Liu Yaolin, Chen Yiyun, et al. *Geoderma*, 2019, 337: 758.
- [14] Xia Zhenzhen, Yang Jie, Wang Jing, et al. *Applied Spectroscopy*, 2020, 74(4): 417.
- [15] Fourie E, Alexandre-Tudo J L, Mihnea M, et al. *Food Control*, 2020, 115: 107303.

Study on the Fractional Baseline Correction Method of ATR-FTIR Spectral Signal in the Fermentation Process of Sodium Glutamate

HE Nian, SHAN Peng*, HE Zhong-hai, WANG Qiao-yun, LI Zhi-gang, WU Zhui

School of Control Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066000, China

Abstract In this paper, Attenuated Total Reflection Fourier Transformed Infrared Spectroscopy (ATR-FTIR) combined with the multivariate calibration model was used to realize the indirect measurement of the concentration of two main substrates (glucose and sodium glutamate) during the fermentation process of γ -polyglutamic acid (γ -PGA), which could provide feedback information for the fermentation process. The frequent baseline drift phenomenon in the spectrum measurement will seriously affect the performance of the subsequent multivariate calibration model, and it is necessary to use the baseline calibration algorithm to preprocess the spectrum. Most of the popular baseline correction algorithms are based on the Whittaker Smoother (WS) smoothing algorithm. And use integer-order differentials with limited expressive power to constrain the fitted baseline. Because of the poor adaptability of integer-order differential in the existing baseline correction algorithms, we use more flexible fractional-order differentials to constrain the baseline and then propose a baseline correction algorithm based on fractional-order, which realizes the extension of the integral order baseline correction. 5 batches of γ -PGA fermentation experiments were carried out, and the ATR-FTIR spectra of different batches and all batches were subjected to fractional baseline correction respectively; subsequently, the prediction accuracy of each model was improved to some extent. The experimental results show that only in batch 2 the baseline correction effect based on the integer-order is the best; the orders to obtain the best baseline correction effect for other batches were all fractional-order. It also reflects that the constraint of the fractional-order derivative (including the integer-order derivative) on the baseline is reasonable. At the same time, it is found that the overall baseline correction effect of all batches is far worse than that of a single batch. The reason may be that the baseline of the spectra for each fermentation batch is different. Different orders need to be selected for different batches to achieve the best effect of baseline correction. In addition, the background spectrum was acquired with distilled water as the reference before measuring each γ -PGA fermentation sample. A negative water peak thus inevitably appears in the wavenumber range of $3\ 100\sim 3\ 600\text{ cm}^{-1}$ and forms harmful interference signals; the fractional baseline-corrected spectra show that the fractional-order baseline correction algorithm regards the negative

water peak as the baseline and eliminates it to a certain extent. In summary, the fractional-order baseline correction algorithm expands the application range of the traditional integer-order baseline correction algorithm and provides a new solution to eliminate negative water peaks in the ATR spectra with water as the background spectrum.

Keywords Fractional differentiation; Spectral preprocessing; Baseline correction

(Received May 11, 2021; accepted Jul. 29, 2021)

* Corresponding author

欢迎投稿

欢迎订阅

欢迎刊登广告

《光谱学与光谱分析》2022年征订启事

国内邮发代码：82-68

国外发行代码：M905

《光谱学与光谱分析》1981年创刊，国内统一刊号：CN 11-2200/O4，国际标准刊号：ISSN 1000-0593，CODEN 码：GYGFED，国内外公开发行人，大 16 开本，332 页，月刊；是中国科协主管，中国光学学会主办，钢铁研究总院、中国科学院物理研究所、北京大学、清华大学共同承办的学术性刊物。北京大学出版社出版，每期售价 95 元，全年 1140 元。刊登主要内容：激光光谱测量、红外、拉曼、紫外、可见光谱、发射光谱、吸收光谱、X 射线荧光光谱、激光显微光谱、光谱化学分析、国内外光谱化学分析领域内的最新研究成果、开创性研究论文、学科发展前沿和最新进展、综合评述、研究简报、问题讨论、书刊评述。

《光谱学与光谱分析》适用于冶金、地质、机械、环境保护、国防、天文、医药、农林、化学化工、商检等各领域的科学研究单位、高等院校、制造厂家、从事光谱学与光谱分析的研究人员、高校有关专业的师生、管理干部。

《光谱学与光谱分析》为我国首批自然科学核心期刊，中国科协优秀科技期刊，中国科协择优支持基础性、高科技学术期刊，中国科技论文统计源刊，“中国科学引文数据库”，“中国物理文摘”，“中国学术期刊文摘”，同时被国内外的 CJCR, CNKI, CSCD, SCI, AA, CA, Ei, AJ, PJK, MEDLINE, Scopus 等文献机构收录。根据中国科学技术信息研究所发布信息，中国科技期刊物理类影响因子、引文量及综合评价总分《光谱学与光谱分析》都居前几位。欢迎国内外厂商在《光谱学与光谱分析》发布广告（广告经营许可证：京海市监广登字 20170260 号）。

《光谱学与光谱分析》的主编为高松院士。

欢迎新老客户到全国各地邮局订阅，若有漏订者可直接与《光谱学与光谱分析》期刊社联系。

联系地址：北京市海淀区学院南路 76 号（南院），

《光谱学与光谱分析》期刊社

邮政编码：100081

联系电话：010-62181070, 62182998

电子信箱：chnghxygpfx@vip.sina.com

修改稿专用邮箱：gp2008@vip.sina.com

网 址：http://www.gpxygpfx.com

