

〈红外应用〉

基于 siPLS 的强化生物除磷污泥胞内糖原近红外分析

李卫华^{1,2}, 徐玲^{1,2}, 杨英^{1,2}, 严国兵^{1,2}, 赵庆¹, 帅磊¹

(1.安徽建筑大学 环境与能源工程学院, 安徽 合肥 230601;

2.安徽建筑大学 水污染控制与废水资源化安徽省重点实验室, 安徽 合肥 230601)

摘要: 为了快速、准确地测定强化生物除磷(EBPR)过程中污泥胞内糖原的含量,采用4种预处理方法分别对污泥近红外光谱进行预处理,并结合联合区间偏最小二乘(siPLS)进行变量优选,建立光谱吸光度数据与糖原含量的定量分析模型。结果表明,将一阶S-G(Savitzky-Golay)平滑处理后的光谱等分为20个子区间,联合子区间[10 13 16 19]建立的siPLS模型预测效果最优,预测集的均方根误差(RMSEP)和相关系数(r_p)分别达到0.0048、0.9105,且该模型的交互验证和外部验证相对分析误差(RPD)均大于3.0。一阶平滑处理后的光谱siPLS模型预测精度高、建模变量少,可实现糖原含量的快速测定。

关键词: 强化生物除磷; 糖原; 近红外光谱; Savitzky-Golay平滑法; 联合区间偏最小二乘

中图分类号: O657.33 **文献标识码:** A **文章编号:** 1001-8891(2016)12-1053-08

Near-infrared Spectroscopy Analysis of the Intracellular Glycogen Content During the Enhanced Biological Phosphorus Removal Process Based on siPLS

LI Weihua^{1,2}, XU Ling^{1,2}, YANG Ying^{1,2}, YAN Guobing^{1,2}, ZHAO Qing¹, SHUAI Lei¹

(1.School of Environment and Energy Engineering, Anhui Jianzhu University, Hefei, 230601, China;

2.Key Laboratory of Anhui Province of Water Pollution Control and Wastewater Reuse, Anhui Jianzhu University, Hefei, 230601, China)

Abstract: In order to develop a procedure to determine the intracellular glycogen content in the sludge from the enhanced biological phosphorus removal (EBPR) process rapidly and accurately, four Near-infrared spectral pre-processing methods were separately used to improve the performance of multivariate calibrations. The quantitative analysis models between the near-infrared (NIR) spectra and the glycogen contents were established based on synergy interval partial least squares (siPLS) method. Results showed that the optimal model was established by four sub-intervals [10 13 16 19] with siPLS method after the NIR spectra were treated with Savitzky-Golay smoothing and first derivative as pre-processing method, and the preprocessed spectra were divided into 20 sub-intervals. The root mean square errors (RMSEP) and the correlation coefficients (r_p) of the prediction sets were 0.0048 and 0.9105 respectively. The residual predictive deviation (RPD) values of cross-validation and prediction sets were all above 3.0. The NIR spectral model based on S-G smoothing and siPLS method can not only improve the predictive accuracy and reduce the variable numbers, but also realize the rapid and accurate determination of intracellular glycogen content in the sludge.

Key words: enhanced biological phosphorus removal (EBPR), glycogen, near infrared spectroscopy (NIR), Savitzky-Golay smoothing (S-G), interval partial least squares (siPLS)

收稿日期: 2016-02-29; 修订日期: 2016-05-11.

作者简介: 李卫华(1977-), 男, 江苏人, 博士, 副教授, 硕士生导师, 主要从事污水生物处理、红外光谱分析研究。E-mail: weihuali@mail.ustc.edu.cn.

通讯作者: 杨英, 博士, 教授, 硕士生导师。E-mail: yangying5918@163.com.

基金项目: 国家自然科学基金(51378017); 安徽省教育厅科学研究重大项目(KJ2015ZD12); 安徽省高校省级自然科学基金项目(KJ2013A067); 省级质量工程教学研究重点项目(2015jyxm242)。

0 引言

强化生物除磷 (enhanced biological phosphorus removal, EBPR) 是污水生物处理的重要环节, 糖原 (glycogen) 作为强化生物除磷过程中污泥胞内主要的聚合物, 在聚磷菌代谢过程中被不断地循环利用和储存^[1-4]。糖原含量的测定通常采用蒽酮比色法, 该操作过程复杂、耗时长、偶然误差大^[5-6]。近年来, 红外光谱分析技术以分析速度快、分辨率高、无损检验、成本低等优点被广泛地应用于污水生物处理的相关领域^[7-8]。中红外技术结合计量学方法定量分析污泥胞内聚合物已有相关报道^[9-10], 但目前国内外关于近红外 (near-infrared, NIR) 定量分析生物活性污泥胞内糖原含量的相关文献报道较少。近红外光谱吸收峰主要来源于含氢基团倍频和合频的贡献, 吸收强度弱、谱带较宽, 易受噪声和背景等非目标因素的影响^[11-12]。此外, 强化生物除磷污泥是多组分的混合体系, NIR 吸收峰来自污泥胞内多种聚合物的贡献, 对单组分信息的提取干扰性强。因此, 选择恰当的多变量数据分析 (MVA) 方法并对光谱进行有效地预处理是污泥胞内糖原 NIR 光谱优化模型建立的关键。

常见的 NIR 光谱预处理方法有数据增强算法、多元散射校正、标准正态变换、微分求导、平滑处理等。不同预处理方法的优化效果存在一定差异, 需要进行对比分析^[13]。目前已有大量报道探讨了不同的预处理方法或者预处理方法的不同组合形式对某一种物质 NIR 光谱定量分析模型的影响。梁家杰等人探讨了波段优选对白砂糖 NIR 定量分析的影响, 先后对各波段进行 MSC 和 S-G 平滑求导预处理。发现 780~1100 nm 波段一阶导数谱的预测精度最高, 表明平滑求导参数的合理设置和波段的恰当选择, 可有效改善 NIR 光谱的分析效果^[14]。Ravikanth 等尝试利用 NIR 超光谱成像技术区分小麦和其中掺杂物, 对比分析了 5 种光谱预处理技术和 3 种分类算法的各种组合形式对分类模型预测效果的影响, 研究表明, 当采用 k-NN 分类算法并对光谱进行 SNV 预处理时分类精度最高^[15]。联合区间偏最小二乘法 (synergy interval partial least squares, siPLS) 是一种基于偏最小二乘法 (partial least squares, PLS) 改进的 NIR 光谱定标建模方法。该方法有效地压缩光谱建模变量数, 可同时实现最佳建模区间的选择和待测物质含量的预测, 有效过滤噪声和去除无关信息变量, 预测精度和稳健性较 PLS 优越^[16]。Jiang 等利用近红外光谱技术结合 siPLS 测定小麦秸秆固态发酵过程中的 PH 值和含水率, PH 和含水率

NIR 光谱建模预测集的 RMSEP 分别为 0.0963、0.0149; 相关系数 r_p 分别达到为 0.9686、0.8684。表明采用近红外技术快速测定上述工艺参数具有可行性^[17]。Wang 等利用 siPLS 对 53 个不同酒精含量待测样品的 NIR 光谱进行定量分析, 所建模型的预测精度较高, 预测效果较好^[18]。本研究拟采用不同的预处理方法对强化生物除磷污泥 NIR 光谱数据进行预处理, 结合 siPLS 选择最佳建模区间, 建立污泥胞内糖原含量与光谱吸光度数据的定量分析模型。对比分析不同预处理方法对建模预测能力和精度的影响, 以期得到污泥近红外光谱的最佳预测模型, 实现污泥胞内糖原含量的快速、准确测定。

1 材料与方法

1.1 实验装置、设备和检测方法

1.1.1 实验装置、设备

采用序批式活性污泥法 (sequencing batch reactor, SBR) 富集好氧聚磷菌。反应器采用有效体积为 4 L 的圆柱体有机玻璃容器, 底部装有搅拌叶片、微孔曝气头, 配备蠕动泵和电磁阀等设备。整个操作系统利用微电脑定时控制器实现进水、搅拌、曝气、沉淀、排水等过程的自动转换。污泥样品取自反应器的典型周期各个阶段, 每次取样沉淀得到的污泥需先经真空泵抽滤, 再冷冻干燥 22 h, 最后取出放在干燥皿中以备用。

1.1.2 糖原的化学测定

糖原的化学测定采用蒽酮比色法, 包括硫酸消解、溶液稀释、蒽酮比色反应 3 个过程。具体步骤如下: 称取 10 mg 冷冻干燥后的污泥样品置于干燥的玻璃试管中, 加入 10 mL 75% 的硫酸溶液, 摇匀后置于 100℃ 的沸水浴中消解 10 min, 取 1 mL 稀释两次, 每次稀释 10 倍。然后加入 5 mL 现配的蒽酮试剂, 置于冰水浴冷却再转至沸水浴中加热 10 min。最后, 利用冰水浴冷却至室温, 采用紫外分光光度计测定吸光度 (波长 620 nm) ^[7-8]。

1.1.3 污泥样品的近红外光谱采集

选用德国布鲁克公司生产的 MPA 型傅里叶近红外光谱仪, 光谱的扫描范围为 12000~4000 cm^{-1} , 扫描次数为 64 次, 分辨率为 2 cm^{-1} 。实验在较干燥的环境中进行, 室温约 26℃ 左右。每次测量前需将颗粒污泥研磨均匀并覆盖测量专用瓶底部, 测量时应先测定背景值, 再测污泥样品光谱数据。每个待测污泥样品平行测量 3 次, 最终以 3 次测量的平均值作为该污泥样品的理想光谱^[19]。

1.2 NIR 光谱数据的分析方法

1.2.1 NIR 光谱预处理方法

综合考虑 NIR 光谱吸收峰的特点和污泥样品组成的构成,拟采用 0~1 归一化(normalization)、多元散射校正(multiplicative scatter correction, MSC)、标准正态变换(standard normal variate, SNV)、卷积平滑法(Savitzky-Golay smoothing, S-G) 4 种预处理方法对污泥胞内糖原的近红外光谱进行预处理。

1) 归一化(0~1)

不同近红外光谱仪的量纲单位各异,数据归一化可以使数据标准化,剔除冗余信息、减小光程对 NIR 光谱定量建模的影响^[13],是常用的光谱数据增强方法,具体表达式如下:

$$x_{ij} = (x_{ij} - x_{i,\min}) / (x_{i,\max} - x_{i,\min}) \quad (1)$$

式中: x_{ij} 为第 i 条光谱第 j 个波长点对应的吸光度数据; $x_{i,\max}$ 为第 i 条光谱所有波长点对应的吸光度数据最大值; $x_{i,\min}$ 为第 i 条光谱所有波长点对应的吸光度数据最小值。

2) MSC

在散射系数不变的前提下, NIR 光谱漫反射遵守 Kubelka - Munk 定律,然而污泥颗粒大小、形状、分布等因素都会对近红外光的散射产生一定的影响,致使散射系数发生改变^[20]。MSC、SNV 是常见的校正散射效应的预处理方法。MSC 具体表达式如下:

①最小二乘线性拟合

以平均光谱矩阵 \bar{X} 作为理想光谱矩阵,利用最小二乘法建立 \bar{X} 和每个样品光谱矩阵 X_i 的回归曲线,表达式如下:

$$X_i = a_i \bar{X} + b_i + e_i \quad (2)$$

式中: a_i 、 b_i 为拟合系数; e_i 为残差光谱。

②光谱多元散射校正

经 MSC 校正后的样品光谱矩阵 $X_{i, MSC}$ 为:

$$X_{i, MSC} = (X_i - b_i) / a_i \quad (3)$$

3) SNV

假设每条光谱对应的吸光度数满足正态分布,通过对每条光谱吸光度数据标准化,即光谱矩阵按行标准正态化实现光谱校正^[21]。表达式如下:

$$x_{ij, SNV} = \frac{(x_{ij} - \bar{x}_i)}{\sqrt{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 / (n-1)}} \quad (4)$$

式中: \bar{x}_i 为第 i 条光谱吸光度数据的平均值; $x_{ij, SNV}$ 为经标准正态变换校正后的第 j 个波长点对应的吸光度; n 为每条光谱吸光度变量总数。

4) S-G 平滑法

常用的光谱求导方法有 S-G 平滑和直接差分求导两种, S-G 平滑处理又被称作“数字滤波器”,具有平滑处理和微分求导的双重作用。可有效滤噪、提高信噪比、消除基线和背景等因素对光谱的干扰,是近红外光谱定标建模常用的预处理方法^[14]。表达式为:

①对窗口内的 $2m+1$ 个吸光度数据点 (x_i, y_i) 进行最小二乘线性拟合:

假设拟合曲线 n 次多项式:

$$f(x) = \sum_{k=0}^n a_k x_i^k$$

令:

$$r_i = f(x_i) - y_i;$$

$$I = \sum_{i=0}^m r_i^2 = \sum_{i=0}^m \left[\sum_{k=0}^n a_k x_i^k - y_i \right]^2,$$

当 $I = I_{\min}$, $\partial I / \partial a_k = 0$ ($k=0, 1, 2, \dots, n$) 时,即可求得所有的多项式系数 a_k ,且唯一。

②由 $f(x)$ 和窗口内中心点的 y_i (波长值) 得到拟合 \tilde{x}_i , 即该窗口的平滑值。

③对 $f(x)$ 求 m 阶导,计算窗口内中心点的 m 阶导的导数值。

④每次窗口移动一个数据点,在全光谱范围内移动,得到平滑光谱和 m 阶导数光谱。

1.2.2 NIR 光谱建模方法

近红外光谱中包含大量与待测物质无关的信息变量,适当的光谱信息筛选,将有助于进一步提高优化模型的灵敏度和精度。拟采用 Nørgaard 等开发的联合区间偏最小二乘 (siPLS) 建立 NIR 光谱定标分析模型,实现变量优选,压缩变量数。siPLS 的基本原理:将 NIR 光谱建模区间 n 等份,利用自由组合 C_n^m 的原理 (一般 $m \leq 5$),从 n 个建模子区间中随机的选择 m 个子区间建立糖原含量的 PLS 定量模型^[22]。具体包括校正模型的建立和模型的验证两部分。首先对已建立的 C_n^m 个 PLS 定量分析模型进行自测,以交互验证均方根误差 (root mean square errors of cross-validation, RMSECV)、相关系数 r 和交互验证相对分析误差 (relative percent deviation, RPD——SD(standard deviation)/RMSECV) 作为自测效果优劣的评价指标,确定最佳校正模型;再利用该校正模型对待测污泥样品胞内糖原含量进行预测,验证模型的最终评价指标为预测集均方根误差 (root mean square errors of prediction, RMSEP)、相关系数 r_p 和外部验证相对分析误差 (relative percent deviation, RPD——SD/RMSEP)。

2 实验结果与讨论

2.1 污泥 NIR 原始光谱解

36个待测污泥样品的 NIR 光谱图如图 1 所示, 由于污泥样品的取样时间点和取样时反应器运行状态都不完全相同, 待测污泥样品光谱存在较大差异。波数为 $9000\sim 12000\text{ cm}^{-1}$ 的光谱区域不存在明显吸收峰, 推测该波段基本不具有与待测污泥样品胞内糖原分子结构中化学键相关的信息。光谱曲线平滑度较低, 基线漂移现象明显。

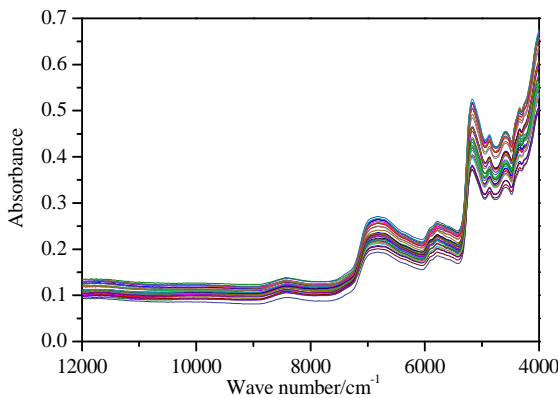


图 1 36个污泥样品原始近红外光谱图

Fig.1 Raw NIR spectra of the 36 activated sludge samples

2.2 污泥 NIR 光谱预处理

污泥样品原始光谱预处理后的效果如图 2 所示, 经过归一化处理, 光谱吸光度数据最终映射值在 $0\sim 1$ 范围内, 较好地去除了无关信息变量对光谱的干扰。MSC 和 SNV 预处理光谱图相似, 有效地削弱光谱的散射效应, 光谱束变窄、平滑度增强。S-G 预处理效果随着平滑参数的改变存在一定差异, 零阶平滑处理后的光谱图没有明显变化, 仅是平滑度略有增加。一阶、二阶平滑处理明显的改善了基线漂移现象; 重叠峰得到有效的分离, 但随着求导阶数的增加, 噪声信息也被扩大。相比之下, 一阶平滑处理可同时实现 NIR 光谱滤噪和重叠峰分离, 与有关文献中的描述相吻合^[14-15,19]。

2.3 siPLS 建模

2.3.1 污泥样品校正集和预测集的划分

有效地划分校正样品集和预测样品集是定量模型建立的前提, 本研究共获得取自 7 个典型周期的 63 个待测污泥样品, 随机选择其中 4 个典型周期的 36 个污泥样品作为校正集, 剩下 3 个典型周期的 27 个污泥样品作为预测集。校正集和预测集中胞内糖原含量的分布情况如表 1。

2.3.2 siPLS 校正模型的建立

采用 siPLS 结合预处理后的污泥 NIR 光谱建立污泥胞内糖原含量的定量分析模型, 光谱等份数 n 和联合子区间数 m 是建模的重要参数, 当 m 值一定时, n 较大则计算量较大且有用信息大量丢失; n 小则难以有效地变量数压缩^[22]。本研究将整个光谱 ($4000\sim 12000\text{ cm}^{-1}$) 划分为 20 个子区间, 联合 4 个子区间建立局部 PLS 回归模型。以交互验证均方根误差 (RMSECV) 和相关系数 r 作为评价指标确定最佳建模子区间, 确定最佳 siPLS 校正模型。并采用相对分析误差 RPD 对校正模型进行深入分析, 当 $1.5\leq\text{RPD}<2.0$, 表明模型具有一定的可行性; 当 $2.0\leq\text{RPD}<2.5$, 表明模型可以定量分析; $2.5\leq\text{RPD}<3.0$, 表明模型预测效果较好; $3\leq\text{RPD}$, 表明模型预测效果很好^[23-24]。

2.3.3 siPLS 校正模型的分析

采用 4 个典型周期的 36 个污泥样品建立校正模型, 不同预处理后的 NIR 光谱建模的预测效果存在一定的差异, 污泥样品的近红外光谱 siPLS 校正模型预测结果如表 2 所示。

由表 2 可知污泥胞内糖原 NIR 光谱预测模型校正集的相关系数均大于 0.8940, 交互验证均方根误差均小于 0.0058, 交互验证相对分析误差 RPD 均大于 2.5, 说明校正模型相关性较好, 利用 NIR 光谱结合 siPLS 建模对糖原含量进行定量分析可行性较高。其中一阶平滑处理后的 siPLS 校正模型对定标集的拟合效果最好, 图 3 为该校正集中测量值和预测值相关性图, 校正集的均方根误差 RMSECV 和相关系数 r 分别达到 0.0043、0.9292 且 $\text{RPD}>3.3$ 。图 4 为该校正模型定标建模的最佳联合区间, 建模区间 $[10\ 13\ 16\ 19]$ 对应的光谱波数为 $4389\sim 4775\text{ cm}^{-1}$ 、 $5577\sim 5970\text{ cm}^{-1}$ 、 $6780\sim 7174\text{ cm}^{-1}$ 、 $7984\sim 8377\text{ cm}^{-1}$ 。NIR 光谱的吸收峰来自待测物质分子结构中化学键的贡献, $4389\sim 4775\text{ cm}^{-1}$ 建模波段可能与糖原分子结构中聚合体 ($\cdot\text{O}-\text{H}$ 和 $\cdot\text{O}-\text{C}$) 的 $\text{O}-\text{H}$ 键、 $\text{C}-\text{O}$ 键伸缩组合频密切相关。 $5577\sim 5970\text{ cm}^{-1}$ 波段的吸收峰可能来源于糖原聚合体分子结构中亚甲基 ($\cdot\text{CH}_2$) 的 $\text{C}-\text{H}$ 键的一级倍频 (2ν) 贡献。 $6780\sim 7174\text{ cm}^{-1}$ 可能归属于糖原分子结构中聚合体 ($\cdot\text{O}-\text{H}$) 的 $\text{O}-\text{H}$ 键一级倍频区域 (2ν)^[25]。建模的变量数由 1037 减少为 203, 压缩 80% 光谱数据。有效去除无关信息变量, 加快模型预测的速度。

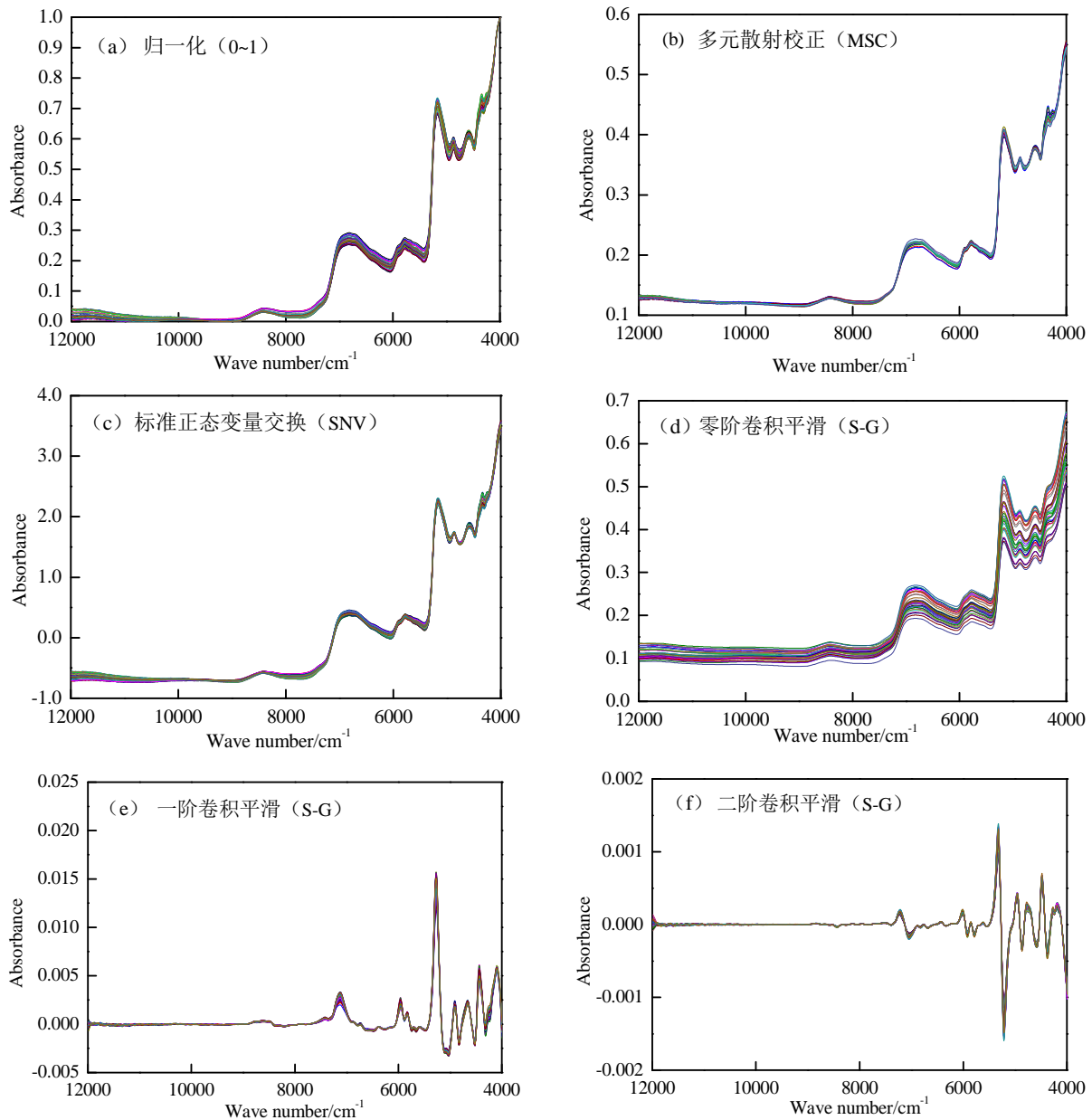


图2 不同预处理后的污泥近红外光谱图

Fig.2 NIR spectra of the activated sludge pre-processed using different spectral pre-processing technique

表1 校正集和预测集中污泥胞内糖原含量的分布范围

Table 1 Content distribution of the intracellular glycogen for calibration set and prediction set

样品集	样品参数/%			
	最大值	最小值	均值	标准差
校正集	12.9651	7.4031	9.8567	1.4531
预测集	12.3062	7.3908	9.8330	1.4644

2.3.4 siPLS 模型验证

每一种预处理方法对糖原含量的定量分析模型的预测能力都会产生一定影响,不同预处理的 NIR 光

谱 siPLS 建模预测效果对比分析见表 3。

对比分析不同预处理对光谱 siPLS 建模预测结果的影响,仅有二阶 S-G 平滑和归一化处理后的 NIR 光谱 siPLS 模型的外部验证 RPD 值小于 2.5,其他预处理后的模型外部验证 RPD 值均大于 2.5,表明污泥样品的 siPLS 模型预测结果较为理想。归一化预处理后的建模预测集相关系数 $r_p < 0.8600$, RMSEP 达到 0.0062,预测效果一般。SNV、MSC 预处理后的 siPLS 建模预测效果较好。而 S-G 平滑处理建模效果与平滑参数密切相关,当移动窗口内平滑点数、多项式次数不变时,求导阶数的增加可能会导致参与建模的噪声信息也相应增多,从而影响模型预测能力。二阶 S-G

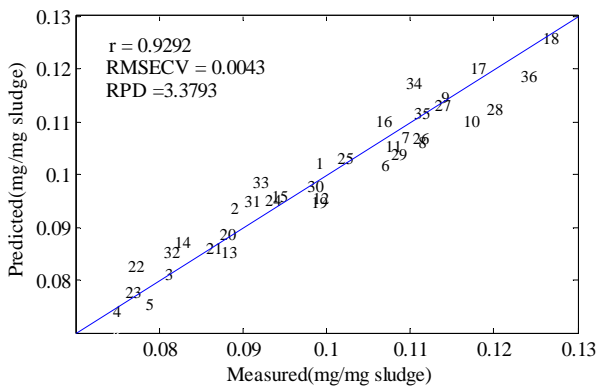


图3 一阶平滑处理后的光谱 siPLS 建模校正集中糖原测量值和预测值的相关性

Fig.3 Correlation between the determined and predicted values of glycogen in calibration set established by siPLS method with S-G smoothing and first derivative pretreatment

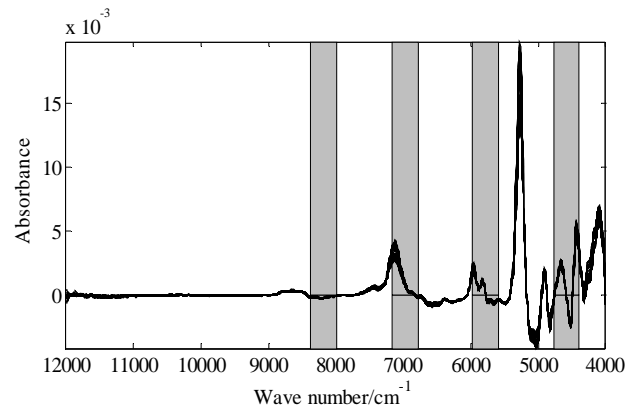


图4 一阶平滑处理后的 NIR 光谱 siPLS 建模的最佳子区间

Fig.4 Optimal intervals for modeling selected by siPLS method with S-G smoothing and first derivative pretreatment

表2 不同预处理后的光谱 siPLS 校正模型的预测结果

Table 2 Prediction results of calibration set with siPLS spectral models for different pretreatment methods

预处理方法	最佳联合区间	主因子	交互验证均方根误差 (RMSECV)	交互验证相关系数 (r)	交互验证相对分析误差 (RPD)
二阶 S-G	[5 10 18 19]	7	0.0057	0.8941	2.5493
RAW	[4 10 16 18]	5	0.0051	0.9162	2.8492
0~1	[8 11 18 20]	6	0.0049	0.9167	2.9655
零阶 S-G	[7 10 18 19]	5	0.0044	0.9272	3.3025
SNV	[9 13 16 20]	4	0.0045	0.9252	3.2291
MSC	[11 14 19 20]	4	0.0048	0.9196	3.0273
一阶 S-G	[10 13 16 19]	5	0.0043	0.9292	3.3793

表3 不同预处理后的光谱 siPLS 建模预测效果的比较

Table 3 Comparison of the prediction effect by siPLS model with different spectral pre-processing methods

	近红外光谱分析方法						
	二阶 S-G + siPLS	RAW + siPLS	归一化 + siPLS	零阶 S-G + siPLS	SNV + siPLS	MSC + siPLS	一阶 S-G + siPLS
预测均方根误差 (RMSEP)	0.0075	0.0070	0.0062	0.0057	0.0053	0.0058	0.0048
预测集相关系数 (r_p)	0.8104	0.8310	0.8539	0.8677	0.8744	0.8776	0.9105
外部验证相对分析误差 (RPD)	1.9525	2.0920	2.3619	2.5691	2.7630	2.5248	3.0508

平滑处理后的光谱建模效果较原始光谱差，零阶 S-G 平滑处理的建模效果较好。一阶 S-G 平滑处理后的光谱建模预测效果最佳，预测集的均方根误差 RMSEP 和相关系数 r_p 分别达到 0.0048、0.9105 且外部验证 RPD > 3.0。其相关性曲线如图 5 所示，除了个别样品

点存在偏差，其他所有数据点预测精度较高，样品点的分散度较低，均紧密分布在相关性曲线两侧。该模型预测值与实测值的对比分析和残差分析分别见图 6 和图 7。

由图 6 可知，预测值的变化趋势同污泥样品胞内

糖原含量的变化趋势基本保持一致。表明该模型作为 EBPR 机理研究的便捷手段具有可行性。由图 7 可得实测值和预测值的残差主要集中在 ± 0.008 , 实测值和预测值的绝对偏差较小, 模型预测准确性高。此外, 为了进一步论证 siPLS 预测模型的优越性, 建立一阶 S-G 平滑处理后的全光谱 PLS 预测模型。得到全光谱 PLS 模型预测集均方根误差 RMSEP 和相关系数 r_p 分别为 0.0082、0.8005。由此可见一阶 S-G 平滑处理后的 siPLS 建模不仅可以实现变量优化而且大幅度地提高模型的预测能力。

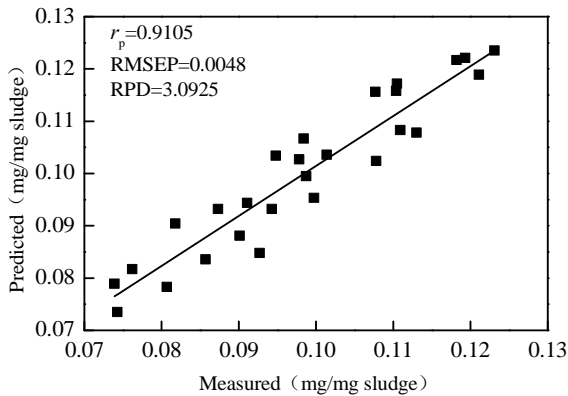


图5 一阶平滑处理后光谱 siPLS 建模预测集中糖原测量值和预测值的相关性

Fig.5 Correlation between the determined and predicted values of glycogen in prediction set by siPLS method with S-G smoothing and first derivative pretreatment

3 结论

利用联合区间偏最小二乘 (siPLS) 建立 EBPR 污泥 NIR 光谱吸光度数据与蒽酮比色法测定糖原含量的定量分析模型。对比 4 种不同预处理方法对 NIR 光谱建模预测效果的影响。结果显示, 经过一阶 S-G 平滑处理后的光谱建模效果最佳, 预测集的均方根误差 RMSEP 和相关系数 r_p 分别达到 0.0048、0.9105 且优于全光谱 PLS 建模的预测精度。siPLS 建模的最佳子区间为 [10 13 16 19], 变量数压缩 80%。有效去除无关信息变量, 加快模型预测速度。由此表明, 以 siPLS 为建模方法并对光谱进行一阶 S-G 平滑处理, 可获得较理想的污泥胞内糖原 NIR 光谱优化模型。虽然仅涉及污水的强化生物除磷过程, 但该方法构建的化学计量学平台可被推广应用于污水生物处理的其他相关领域, 具有一定的参考价值。

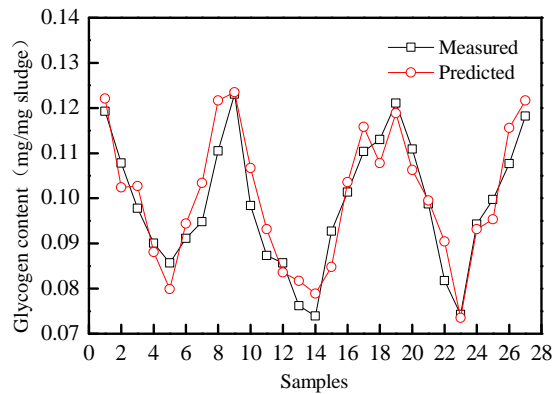


图6 糖原化学测量值与一阶平滑处理后的光谱 siPLS 建模预测值的比较

Fig.6 Comparison of the glycogen content detected by anthrone colorimetric method and that predicted by siPLS method with S-G smoothing and first derivative pretreatment

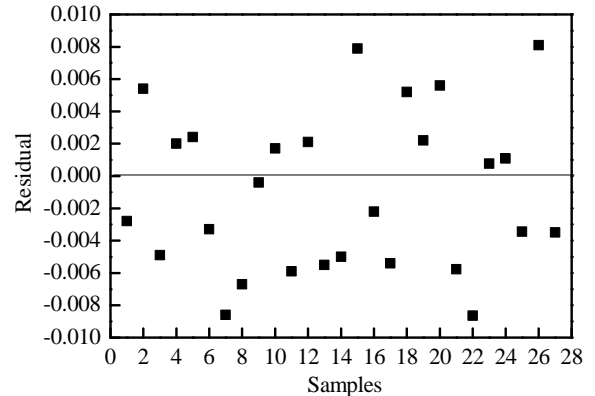


图7 一阶平滑处理后光谱 siPLS 建模预测集中糖原测量值和预测值的残差分布

Fig.7 Residual plot of the glycogen content detected by anthrone colorimetric method and that predicted by siPLS method with S-G smoothing and first derivative pretreatment

参考文献:

- [1] Mino T, Van loosdrecht M C M, Heijnen J J, et al. Microbiology and biochemistry of the enhanced biological phosphate removal process[J]. *Water Research*, 1998, **32**(11): 3193-3207.
- [2] ZHANG H L, SHENG G P, FANG W, et al. Calcium effect on the metabolic pathway of phosphorus accumulating organisms in enhanced biological phosphorus removal systems[J]. *Water Research*, 2015, **84**(1): 171-180.
- [3] Welles L, TIAN W D, Saad S, et al. Accumulibacter clades type I and II performing kinetically different glycogen-accumulating organisms metabolisms for anaerobic substrate uptake[J]. *Water Research*, 2015, **83**(15): 354-366.
- [4] Semerci N and Hasilci N B. Fate of carbon, nitrogen and phosphorus

- removal in a post-anoxic system treating low strength wastewater[J]. *International Biodeterioration & Biodegradation*, 2016, **108**: 166-174.
- [5] CHEN Y G, CHEN Y S, XU Q, et al. Comparison between acclimated and unacclimated biomass affecting anaerobic-aerobic transformations in the biological removal of phosphorus[J]. *Process Biochemistry*, 2005, **40**(2): 723-732.
- [6] Raunkjær K, Jacobsen T H and Nielsen P H. Measurement of pools of protein, carbohydrate and lipid in domestic wastewater[J]. *Water Research*, 1994, **28**(2): 251-262.
- [7] 刘宏欣, 张军, 王伯光, 等. 水质监测中总磷无损的近红外光谱分析研究[J]. *分析科学学报*, 2008, **24**(6): 664-666.
- LIU H X, ZHANG J, WANG B W, et al. Investigation on the non-chemical method for monitoring total phosphate in water[J]. *Journal of Analytical Science*, 2008, **24**(6):664-666.
- [8] Herbert Jr R B, Winbjörk W, Hellman M, et al. Nitrogen removal and spatial distribution of denitrifier and anammox communities in a bioreactor for mine drainage treatment[J]. *Water Research*, 2014, **66**(1): 350-360.
- [9] 张华, 朱菁, 宋箭, 等. 反硝化除磷系统中 PHB 红外光谱解析及其与磷去除率的相关性[J]. *环境科学研究*, 2015, **28**(8): 1274-1280.
- ZHANG H, ZHU J, SONG J, et al. Infrared spectrum analysis of PHB and the correlation between PHB reduction and orthophosphate removal rate during denitrifying phosphorus removal[J]. *Research of Environmental Sciences*, 2015, **28**(8):1274-1280.
- [10] 李卫华, 严国兵, 申慧彦, 等. 强化生物除磷过程中污泥胞内糖原的红外光谱分析[J]. *环境科学学报*, 2015, **35**(3): 705-712.
- LI W H, YAN G B, SHEN H Y, et al. Analysis of intracellular glycogen in the activated sludge during the enhanced biological phosphorus removal process using FTIR spectra[J]. *Acta Scientiae Circumstantiae*, 2015, **35**(3): 705-712.
- [11] 李晋华, 杨志良, 王召巴, 等. 近红外漫透射技术检测玉米成分[J]. *红外技术*, 2013, **35**(11): 732-736.
- LI J H, YANG Z L, WANG Z B, et al. The corn content measurement with near infrared diffuse transmission[J]. *Infrared Technology*, 2013, **35**(11): 732-736.
- [12] Dardenne P, Sinnaeve G, Baeten V, et al. Multivariate calibration and chemometrics for near infrared spectroscopy: which method?[J]. *Journal of Near Infrared Spectroscopy*, 2000, **8**(1): 229-237.
- [13] 褚小立, 袁洪福, 陆婉珍, 等. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. *化学进展*, 2004, **16**(4): 528-542.
- CHU X L, YUAN H F, LU W Z, et al. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique[J]. *Progress in Chemistry*, 2004, **16**(4): 528-542.
- [14] 梁家杰, 潘涛, 陈星旦, 等. 白砂糖色值近红外光谱分析的波段选择[J]. *红外技术*, 2009, **31**(2): 90-94.
- LIANG J J, PAN T, CHEN X D, et al. Choice of wave band in near infrared spectroscopy analysis of color value of white granulated sugar[J]. *Infrared Technology*, 2009, **31**(2): 90-94.
- [15] Ravikanth L, Singh C B, Jayas D S, et al. Classification of contaminants from wheat using near-infrared hyperspectral imaging[J]. *Biosystems Engineering*, 2015, **135**: 73-86.
- [16] Kang Q, Ru Q G, Liu Y, et al. On-line monitoring the extract process of Fu-fang Shuanghua oral solution using near infrared spectroscopy and different PLS algorithms[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2016, **152**(5): 431-437.
- [17] JIANG H, LIU G H, MEI C L, et al. Measurement of process variables in solid-state fermentation of wheat straw using FT-NIR spectroscopy and synergy interval PLS algorithm[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2012, **97**: 277-283.
- [18] WANG X F, BAO Y F, LIU G L, et al. Study on the best analysis spectral section of NIR to detect alcohol concentration based on SiPLS[J]. *Procedia Engineering*, 2012, **29**: 2285-2290.
- [19] 江泽慧, 费本华, 杨忠, 等. 光谱预处理对近红外光谱预测木材纤维素结晶度的影响[J]. *光谱学与光谱分析*, 2007, **27**(3): 435-438.
- JIANG Z H, FEI B H, YANG Z, et al. Effects of spectral pretreatment on the prediction of crystallinity of wood cellulose using near infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2007, **27**(3): 435-438.
- [20] Rinnan A, Berg F V D, Engelsen S B, et al. Review of the most common pre-processing techniques for near-infrared spectra[J]. *Trac Trends in Analytical Chemistry*, 2009, **28**(10): 1201-1222.
- [21] CHEN H, LIN Z, WANG L, et al. Diagnosis of colorectal cancer by near-infrared optical fiber spectroscopy and random forest[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2015, **135**(25): 185-191.
- [22] Nørgaard L, Saudlaud A, Wagner J, et al. Interval partial least-squares regression(iPLS):a comparative chemometric study with an Example from near-infrared spectroscopy[J]. *Applied Spectroscopy*, 2000, **54**(3): 413-419.
- [23] Saeys W, Darius P, Ramon H. Potential for on-site and analysis of hog manure using a visual and near infrared diode array reflectance spectrometer[J]. *Journal of Near Infrared Spectroscopy*, 2004, **12**(1): 299-309.
- [24] Davies A, Baker R D, Grant S A, et al. Murray I. *Forage Analysis by Near Infrared Spectroscopy*[M]. The British Grassl and Society, 1993.
- [25] Wokeman J, Weyer L, 褚小立, 等. 近红外光谱解析使用指南[M]. 北京: 化学工业出版社, 2009.
- Wokeman J, Weyer L, ZHU X L, et al. *Practical Guide to Interpretive Near-infrared Spectroscopy*[M]. Beijing: Chemical industry press, 2009.