

## 基于高光谱技术的厌氧发酵液固形物含量检测的研究

叶 辉, 李晓丽, 余克强, 夏益华, 张 初, 何 勇\*

浙江大学生物系统工程与食品科学学院, 浙江 杭州 310058

**摘 要** 实时监测发酵液中固形物含量的变化, 对控制厌氧发酵过程的稳定性具有重要作用。研究中采用近红外高光谱技术结合化学计量学方法, 对水葫芦和稻草秸秆混合厌氧发酵过程中的固形物含量进行定量检测研究。与传统 2540G(APHA, 1990)标准方法相比, 近红外高光谱技术具有无损、快速的优点。实验过程中, 首先获取发酵液样本的高光谱信息, 应用移动平均平滑法(MAS)进行光谱预处理, 并采用竞争自适应重加权采样算法(CARS)、连续投影算法(SPA)和 Random frog 算法提取光谱特征信息, 然后基于全谱和所选特征波长下的光谱信息分别建立总固形物含量(TS)和挥发性固形物含量(VS)的校正模型, 建模方法包括偏最小二乘回归(PLSR)和最小二乘-支持向量机(LS-SVM)。研究表明, SPA-LS-SVM 模型的预测结果最好, 其中 TS 的预测均方根误差(RMSEP)及相关系数( $R_p$ )分别为 0.005 8 和 0.841; 而 VS 的 RMSEP 和  $R_p$  分别为 0.004 1 和 0.874。结果表明, 利用近红外高光谱结合化学计量学方法可以实现厌氧发酵液中的固形物含量的检测, 为布置光谱传感器以便定量检测厌氧发酵过程中的固形物含量奠定了理论依据。

**关键词** 近红外高光谱; 总固形物含量; 挥发性固形物含量; 厌氧发酵; 化学计量学方法

**中图分类号:** TS201

**文献标识码:** A

**DOI:** 10.3964/j.issn.1000-0593(2017)12-3833-06

### 引 言

厌氧发酵是一种变废为宝的环境友好型的方式, 可以将富含纤维素等有机物的稻草秸秆和水葫芦转化成甲烷等清洁能源。在厌氧发酵过程中, 对于一定容积的反应器, 发酵液固形物含量过高易导致某些中间产物的累积, 进而抑制发酵菌生长<sup>[1]</sup>。其中, 总固形物含量(total solid, TS)和挥发性固形物含量(volatile solid, VS)被认为是评估发酵液固形物含量的重要指标, 为保证厌氧发酵产沼气的过程稳定, 实时检测固形物含量具有重要意义<sup>[2]</sup>。本研究中 TS 是指稻草秸秆和水葫芦的干重, 主要是纤维素、半纤维素以及少量木质素、蛋白质、脂肪和少许污泥等。VS 是指经过高温(通常在 575 °C 左右)灼烧后挥发掉的固体有机物。传统检测发酵液中的 TS 和 VS 方法是根据 2540G(APHA, 1990)的标准方法<sup>[3]</sup>进行测定, 此方法操作复杂, 耗时长。

近红外光谱分析方法是基于光谱范围在 780~2 500 nm 的近红外光与待测有机物中 C—H, N—H, O—H 等化学键之间作用的一种方法。它可以替代传统实验室分析, 而且是一种易于操作, 低成本且快速的分析方法。具体分析步骤

是, 首先建立初始的训练模型, 然后利用训练样本建立光谱信息与化学参考值之间的关系, 最后应用建立的模型来反演预测集样本。这种方法不仅可以定量(感兴趣物质的浓度, 如挥发性脂肪酸等属性的程度)而且可以定性(如微生物分解、产酸等发酵阶段)分析<sup>[4]</sup>。Reed 等<sup>[5]</sup>利用傅里叶变换-近红外光谱(FT-NIR)对污水污泥发酵液中的 TS 和 VS 进行了建模预测, 其中相关系数  $R_{TS} = 0.75$ ,  $R_{VS} = 0.75$ 。Krapf 等<sup>[6]</sup>利用近红外光谱建立偏最小二乘法(PLS)预测了农业饲料的厌氧发酵过程参数, 其中挥发性固体含量(VS)的预测均方根误差(RMSEP)为  $5 \text{ g} \cdot \text{kg}^{-1}$ (干物质), 相关系数为 0.938。目前, 厌氧发酵过程在线监测的技术手段主要分为光谱、色谱、质谱和电化学方法等<sup>[1]</sup>, 其中近红外光谱被广泛应用到实验和实践中。

简单易操作、低成本及高准确性使得近红外光谱在包括厌氧发酵在内的过程控制领域得以迅速发展。高光谱能够获得样本感兴趣区域的光谱值, 提高了数据的可靠性和稳定性<sup>[7]</sup>。本工作采用了近红外高光谱来对水葫芦和稻草秸秆混合厌氧发酵过程中发酵液的性质进行无损快速检测研究。

收稿日期: 2015-10-17, 修订日期: 2016-04-20

基金项目: 国家自然科学基金项目(61201073)资助

作者简介: 叶 辉, 1989 年生, 浙江大学生物系统工程与食品科学学院硕士研究生

e-mail: yehui03@126.com

\* 通讯联系人 e-mail: yhe@zju.edu.cn

## 1 实验部分

### 1.1 样品

稻草秸秆来自浙江省上虞市的农田, 风干之后将其粉碎成小于 5 mm 粒径大小的颗粒, 贮存备用。水葫芦从浙江省杭州市西溪湿地取样, 清洗干净后去除坏死部分, 挤去根部水分, 然后风干粉碎成小于 5 mm 粒径大小的颗粒, 封存备用。厌氧发酵接种污泥来源于浙江省杭州正兴牧业有限公司正在运行中的沼气厂。经过混合厌氧发酵后, 每天定时取 200 g 发酵液样本, 存储于  $-20\text{ }^{\circ}\text{C}$  冰箱内以备采集光谱信息。

### 1.2 TS 和 VS 的测定方法

TS 和 VS 根据 2540G 标准方法 (APHA, 1999) 测定。首先, 将烘干的干锅称重为  $W_1$ , 将样品置于干锅内称重为  $W_2$ , 然后置于恒温干燥箱 ( $105\text{ }^{\circ}\text{C}$ ) 至恒重称量为  $W_3$ , 则  $\text{TS} = \frac{W_3 - W_1}{W_2 - W_1} \times 100\%$ ; 然后将样品同干锅置于马弗炉内 ( $575\text{ }^{\circ}\text{C}$ ) 至恒重, 在干燥皿内冷却至室温后称重为  $W_4$ , 则  $\text{VS} = \frac{W_3 - W_4}{W_2 - W_1} \times 100\%$ 。具体见参考文献 [3]。

### 1.3 高光谱数据采集

将存储于  $-20\text{ }^{\circ}\text{C}$  冰箱的发酵液样品取出, 在  $4\text{ }^{\circ}\text{C}$  恒温冰箱中解冻, 然后按照采样的时间顺序依次编号, 置于底面积为 5 cm 的培养皿中, 采集样本高光谱数据。采集系统的硬件主体构造如图 1 所示, 主要包括: 镜头 (OLES23; Specim, spectral Imaging Ltd., Oulu, Finland); ImSpector 光谱仪 (ImSpectorV10E, Spectral Imaging Ltd., Oulu, Finland), 其光谱分辨率为 2.8 nm; CCD 相机 (C8484-05, Hamamatsu, Hamamatsu city, Japan); 线光源由两个卤素灯 (Fiber-Lite DC950 Illuminator; Dolan Jenner Industries Inc., Boxborough, MA, USA) 组成, 暗箱和计算机等组成。整个系统为线扫描高光谱图像采集系统。

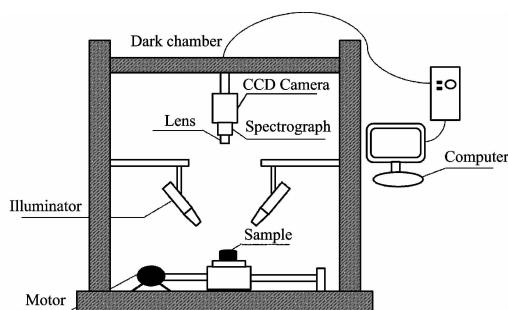


图 1 高光谱成像系统的结构

Fig. 1 Structure of hyperspectral imaging system

### 1.4 数据处理

为使提取的光谱信息更有效, 需要对原始光谱数据预处理, 以达到去除噪音的目的, 同时尽量消除或降低环境因素和随机误差造成的影响, 采用了移动平均平滑法对原始光谱数据进行预处理。

近红外光谱反映了待测有机物在近红外波段合频和倍频的吸收, 而一般谱带信息冗余严重, 影响模型的稳定性和预测能力。特征变量提取能够在很大程度上消除变量冗余信息和仪器噪音干扰。连续投影算法 (successive projection algorithm, SPA) 是一种降低变量共线性的有效方法。相比于遗传算法, 蒙特卡罗无信息变量消除法等, 连续投影算法提取出的信息变量更少且能够提高建模速度 [8]。Random frog (RF) 算法通过输出变量选择概率的方式, 进行特征变量的选取。分为以下三个步骤: (1) 随机将包含  $Q$  个变量的变量子集  $V_0$  初始化; (2) 基于初始化的子集  $V_0$ , 提出一个包含  $Q^*$  个变量的候选变量子集  $V^*$ , 在某一概率下使  $V_1 = V^*$ , 并将  $V_1$  替换  $V_0$ , 这个步骤循环至  $N$  次迭代完成; (3) 计算每一个变量的选择概率, 并将此概率用于评价变量的权重 [9], RF 算法的具体步骤如图 2 所示。竞争自适应重加权算法 (CARS) 是一种基于达尔文进化论中“适者生存”的原则而形成的提取关键变量的有效策略。CARS 方法能够同时有效去除无信息变量和有效压缩共线性信息, 最终选出预测目标最为关键的信息变量, 具体算法步骤见文献 [10]。

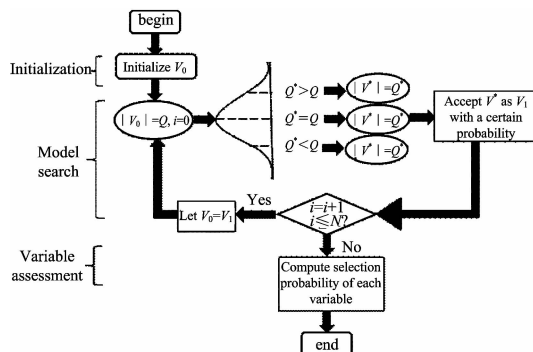


图 2 随机蛙跳算法的流程图

Fig. 2 Flowchart of the random frog algorithm

偏最小二乘 (PLS) 是一种多元线性回归算法 [11], 最小二乘-支持向量机 (LS-SVM) 是基于 SVM 的机器学习算法, 是一种非线性回归算法 [8]。这两种化学计量学方法, 经常被用于多元变量的光谱分析。模型评价参数为相关系数  $r$ , 建模集均方根误差 (RMSEC) 和预测集均方根误差 (RMSEP), 其中 RMSEC 和 RMSEP 值越小,  $r$  越大, 这说明模型预测能力越好 [12]。

## 2 结果与讨论

### 2.1 TS 和 VS 的统计分析

根据 Kennard-Stone (K-S) 算法 [13] 将发酵液样品分成校正集和预测集, 结果如表 1 所示, 其中校正集的 TS 和 VS 的数值范围包含了预测集的范围, 表明集的划分是合理的 [14]。

### 2.2 光谱曲线分析

本研究获取的 870~1700 nm 近红外波段范围内有 256 个波段, 由于波段首尾部分存在很大噪声, 故只保留 975~1645 nm 范围的 200 个波段的光谱信息进行分析和后续处理。经移动平均平滑法对光谱进行预处理, 得到如图 3 所示

表 1 建模集和预测集的 TS 和 VS 统计描述

Table 1 Statistical description of TS and VS of calibration and prediction

	Calibration set			Prediction set		
	N.	Range	Mean±SD	N.	Range	Mean±SD
TS	61	0.045~0.097	0.068±0.011	31	0.046~0.079	0.063±0.007
VS	61	0.023~0.066	0.042±0.009	31	0.029~0.054	0.039±0.007

注: N. 代表建模所用样本数, SD 为 TS 和 VS 值的标准偏差。

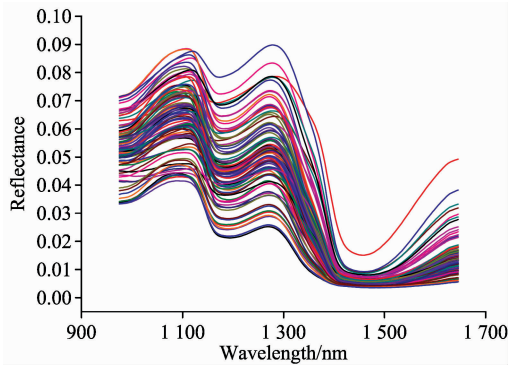


图 3 移动平均平滑处理后的光谱  
Fig. 3 Preprocessed spectra after moving average smoothing (MAS)

的光谱曲线图。

从图 3 可以看出: 在 1 450 nm 处的强吸收波谷是由 O—H 一级倍频振动引起的<sup>[7]</sup>, 说明了 TS 和 VS 中水分等化学信息。通常由于 N—H 键一级倍频伸缩振动在 1 500 nm 处形成波谷, 但被 O—H 键一级倍频吸收峰覆盖<sup>[15]</sup>。在 1 210 nm 附近的波谷吸收峰为 C—H 键的二级倍频伸缩振动所引起的<sup>[16]</sup>。光谱曲线图与被测样本中含有大量水分和样本固形物中包含的纤维素、半纤维素和木质素及少量蛋白质等相吻合。

2.3 特征波段提取

SPA 是一种有效的消除变量共线性的算法, 能够很好的从光谱信息中提取出最低冗余信息。根据 TS 和 VS 的光谱和浓度信息关系, 应用 SPA 算法分别得到 TS 的 5 个特征波长和 VS 的 14 个特征波长, 波长数据见表 2。

表 2 基于 CARS, SPA 和 RF 算法所选择的特征波段

Table 2 Selected characteristic wavelengths based on CARS, SPA and RF algorithm

Methods	Number	Wavelengths/nm
TS	CARS	11 1 066, 1 079, 1 093, 1 099, 1 180, 1 187, 1 197, 1 335, 1 345, 1 355
	SPA	5 975, 1 056, 1 321, 1 450, 1 534
	RF	8 1 079, 1 089, 1 096, 1 187, 1 200, 1 210, 1 345, 1 359
VS	CARS	17 975, 982, 988, 1 082, 1 089, 1 096, 1 103, 1 109, 1 177, 1 183, 1 190, 1 197, 1 214, 1 325, 1 348, 1 433, 1 450
	SPA	14 1 012, 1 164, 1 281, 1 332, 1 389, 1 416, 1 429, 1 446, 1 460, 1 477, 1 500, 1 544, 1 612, 1 646
	RF	7 1 072, 1 079, 1 086, 1 338, 1 345, 1 352, 1 359

基于 CARS 算法对发酵液总固形物含量(TS)的光谱变量选择过程如图 4 所示。图 4 中(a)–(c)分别表示样本采样个数、交叉验证均方根误差(RMSECV)及回归系数随采样次数变化的情况。图 4(a)说明变量的选择随采样次数的增大而减小。图 4(b)中 RMSECV 随采样次数增大呈下降和微弱波动趋势, 当运行 32 次时值最小; 运行次数大于 32 次时, RMSECV 显示轻微波动; 图 4(c)显示回归系数与采样运行次数的变化, 在前面的 32 次运行中与 TS 有关的大量冗余信息被删除, 之后的有用变量“存活”下来, 其中星号线与横坐标垂直的点即为最佳采样次数 32 次, 此时 RMSECV 最小, 选取最优变量个数为 11 个, 所选的特征波长见表 2。基于同样的方法, 对挥发性固形物含量(VS)进行计算, 在运行 31 次时, RMSECV 最小, 选取最优的 17 个变量, 特征波长见表 2。

和计算, 得到 7 个特征波长见表 2。

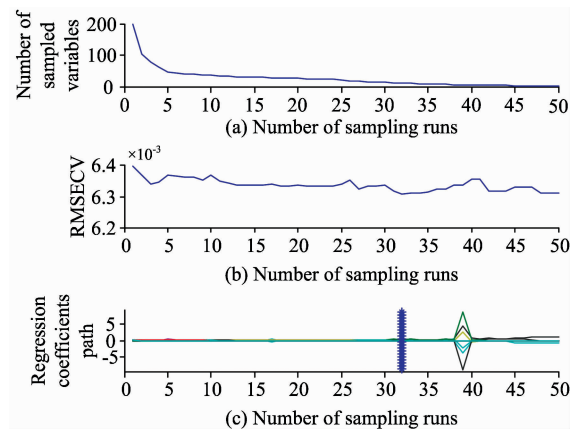


图 4 CARS 算法应用于 TS 的结果  
Fig. 4 The results of applying CARS algorithm for TS

RF 是基于对每一个变量的选择概率选取变量的方法。根据计算过程中每一个波长出现的频率大小选取 8 个 TS 的光谱特征波长见图 5 和表 2; 对 VS 的光谱信息作同样处理

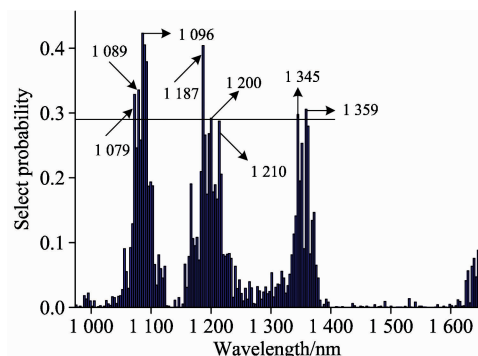


图 5 通过 RF 算法选择的波长

Fig. 5 The selected wavelengths by applying the RF algorithm (TS)

所选波长中, 1 450 nm 附近及 980~1 000 nm 分别是水分子(O—H)一级和二级倍频振动峰, 1 550 nm 附近及 1 050~

1 100 nm 是由于液氨(N—H)一级和二级倍频伸缩所引起, 1 210和 1 550 nm 附近是 C—H 的一级和二级倍频伸缩振动引起的特征吸收峰<sup>[16]</sup>。以上所述与发酵液样本含有大量水分以及少量有机物的事实可相互印证。

#### 2.4 模型预测结果

通过以上三种方法得到的特征波长以及相应的全谱波段信息, 分别作为 PLS 和 LS-SVM 模型的输入变量, 其建模和预测结果如表 3 所示。

表 3 中 Full Spectral-PLS 代表着以选取的 200 个波段建立 PLS 模型, CARS-PLS, SPA-PLS, RF-PLS 分别是以 CARS, SPA 和 RF 选取的特征波长为输入变量建立的 PLS 线性回归模型。同样, Full Spectral-LS-SVM 是以选取的 200 个波段建立 LS-SVM 非线性回归模型, CARS-LS-SVM, SPA-LS-SVM, RF-LS-SVM 分别是以 CARS, SPA 和 RF 算法选取的特征波长为输入变量建立的 LS-SVM 非线性回归模型。

表 3 基于不同波长选择方法建立的 PLS 和 LS-SVM 模型的结果

Table 3 The results of PLS and LS-SVM models based on different wavelength selection methods

	Models	Number	Calibration		Prediction	
			$R_c$	RMSEC	$R_p$	RMSEP
TS	Full Spectral-PLS	200	0.845	0.005 6	0.841	0.005 8
	CARS-PLS	11	0.848	0.005 5	0.839	0.005 8
	SPA-PLS	5	0.846	0.005 6	0.837	0.005 8
	RF-PLS	8	0.848	0.005 5	0.840	0.005 8
	Full Spectral-LS-SVM	200	0.846	0.005 6	0.840	0.005 8
	CARS-LS-SVM	11	0.848	0.005 5	0.839	0.005 8
	SPA-LS-SVM	5	0.846	0.005 6	0.841	0.005 8
	RF-LS-SVM	8	0.848	0.005 5	0.839	0.005 8
VS	Full Spectral-PLS	200	0.836	0.004 8	0.826	0.004 1
	CARS-PLS	17	0.841	0.004 8	0.832	0.004 6
	SPA-PLS	14	0.832	0.004 9	0.808	0.004 5
	RF-PLS	7	0.836	0.004 8	0.828	0.004 4
	Full Spectral-LS-SVM	200	0.955	0.002 6	0.865	0.004 7
	CARS-LS-SVM	17	0.927	0.003 3	0.849	0.004 5
	SPA-LS-SVM	14	0.937	0.003 1	0.874	0.004 4
	RF-LS-SVM	7	0.897	0.003 9	0.816	0.004 5

结果表明, 对于总固形物含量(TS)的光谱变量, 应用竞争性自适应重加权算法(CARS)、连续投影算法(SPA)和 RF 三种算法提取的特征波长都是有效的, 预测结果相近, 模型稳定性较好。而对于挥发性固形物含量(VS)不同建模预测结果, LS-SVM 建模除了 RF-LS-SVM 相比较差外, 其他 LS-SVM 预测结果都优于 PLS, 表明, 对于 VS 的预测模型, 非线性的 LS-SVM 的建模方法优于线性的 PLS 模型, 并得到以下结果:

(1)总固形物含量(TS)全波长和特征波长的八种建模集和预测集结果非常相近, 表明上述三种特征波长提取方法都是有效的, 所建立的预测模型稳定性较好。其中, SPA-LS-SVM 预测结果稍优于基于 CARS 和 RF 的预测模型结果。

(2)挥发性固形物含量(VS)的模型中, 最小二乘-支持向

量机(LS-SVM)建模和预测结果都明显优于偏最小二乘回归(PLS)建模方法, 其中 SPA-LS-SVM 建模方法和预测结果最佳。

(3)基于 SPA-LS-SVM 建模预测厌氧发酵液中 TS 和 VS 的方法是可行的, 其中 VS 的预测相关系数为 0.874 优于 TS 的预测相关系数 0.841。

基于 SPA-LS-SVM 模型的 TS 和 VS 回归模型如图 6。

从图 6 可以看出, 模型预测值与实际参考值相近, 都分布在拟合直线附近, 其中相关系数  $R_{TS}=0.874$ ,  $R_{VS}=0.841$  都高于文献[5]所述的  $R_{TS}=0.75$ ,  $R_{VS}=0.75$ ; VS 的预测均方根误差(RMSEP)为 0.004 1 也小于文献[6]所述的 0.005, 验证了近红外高光谱技术结合 SPA-LS-SVM 算法检测发酵液中固形物含量的可行性。

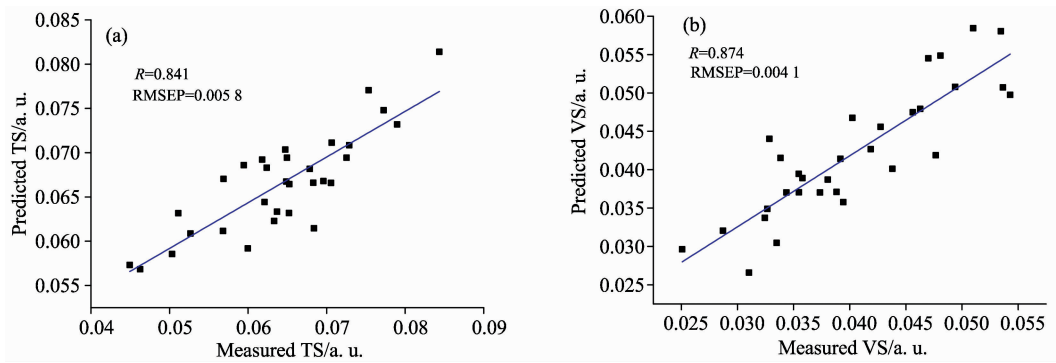


图 6 基于 SPA-LS-SVM 模型的 TS (a) 和 VS (b) 回归模型结果

Fig. 6 Regression results of TS (a) and VS (b) based on SPA-LS-SVM model

### 3 结 论

应用近红外高光谱技术结合化学计量学方法可以用于检测厌氧发酵液中总固形物含量(TS)和挥发性固形物含量(VS),便于控制发酵过程中固形物含量不至于过高,进而有利于发酵过程的稳定,并得到如下结论:采用 PLS 和 LS-SVM 作为发酵液中 TS 和 VS 含量的预测模型都是可行的,

相比 SPA-LS-SVM 取得了更好的结果,其中 TS 模型的  $R_c = 0.846$ ,  $R_p = 0.841$ ; VS 模型的  $R_c = 0.937$ ,  $R_p = 0.874$ , 相比于文献所述结果,取得了较好地结果。

结果表明,利用近红外高光谱技术可以较好地检测厌氧发酵过程发酵液中固形物含量,为下一步研究工作中布置高光谱成像系统,以便在线监测厌氧发酵的构思提供理论依据。

### References

- [ 1 ] Madsen M, Holm-Nielsen J B, Esbensen K H. *Renewable and Sustainable Energy Reviews*, 2011, 15: 3141.
- [ 2 ] Reed J P, Devlin D, Esteves S R R. *Water Research*, 2011, 45: 2463.
- [ 3 ] ANPA. *Standard Methods for the Examination of Water and Wastewater*. Washington DC, 1990.
- [ 4 ] Serranti S, Fabbri A, Bonifazi G. *Optical Engineering*, 2012, 51(11): 111708.
- [ 5 ] Reed J P, Devlin D, Esteves S R R, et al. *Bioresource Technology*, 2013, 133: 398.
- [ 6 ] Krapf L C, Gronauer A, Schmidhalter U. *Near Infrared Spectrosc*, 2012, 20: 465.
- [ 7 ] Everard C D, Kim M S, Lee H. *Proc. SPIE, Sensing for Agriculture and Food Quality and Safety (VI)*, 2014, 9108: 91080A.
- [ 8 ] Liu F, He Y. *Food Chemistry*, 2009, 115: 1430.
- [ 9 ] Li H D, Xu Q S, Liang Y Z. *Analytica Chimica Acta*, 2012, 740: 20.
- [ 10 ] Wei F, Yang S, Gao Y. *Food Analysis Methods*, 2012, 5: 585.
- [ 11 ] Kennard R W, Stone L A. *Technometrics*, 1969, 11(1): 137.
- [ 12 ] Wu D, Yang H Q, Chen X J, et al. *Journal of Food Engineering*, 2008, 88: 474.
- [ 13 ] Jacobi H F, Moschner C R, Hartung E. *Bioresource Technology*, 2011, 102: 4688.
- [ 14 ] Madsen M, Ihungebo F N, Holm-Nielsen J B. *Journal of Near Infrared Spectroscopy*, 2012, 20: 635.
- [ 15 ] Cozzolino D, Parker M, Damberg R G. *Biotechnology and Bioengineering*, 2006, 95(6): 1101.
- [ 16 ] Liu D, Sun D W, Zeng X A. *Food and Bioprocess Technology*, 2014, 7: 307.

# Application of Hyperspectral Technology for the Determination of the Solid Concentration of the Anaerobic Digestion

YE Hui, LI Xiao-li, YU Ke-qiang, XIA Yi-hua, ZHANG Chu, HE Yong\*

College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

**Abstract** Timely monitoring variations of the solid concentration plays a significant role in the stability control of the anaerobic digestion process. In this study, infrared hyperspectral technology coupled with chemometrics methods is applied to detect the amount of solid concentration in which process water hyacinth and rice straw are co-digested. Compared to the traditional way (2540G APHA, 1990), it is faster and non-destructive. Firstly, the hyperspectral information of fermentation broth is obtained by application of infrared hyperspectral and the spectroscopy data is pretreated by utilizing moving average smoothing (MAS), and then adaptive weighted sampling competition (CARS), random frog (RF) and successive projections algorithm (SPA) are applied to extract characteristic wavelengths. Finally the calibration models of total solid (TS) and volatile solid (VS) are established based on the extracted characteristic wavelengths, partial least square (PLS) and least square-support vector machine (LS-SVM), which are taken to predict the solid concentration of fermentation broth. The study indicates that SPA-LS-SVM model achieves optimal result, among which the root mean square error prediction (RMSEP) and correlation efficient ( $R$ ) of the total solid concentration are respectively 0.005 8 and 0.841; the root mean square error prediction and correlation efficient of the volatile solid concentration are respectively 0.004 1 and 0.874. The study shows that it is feasible to utilize infrared hyperspectral combined with chemometrics methods for prediction of the solid concentration of the fermentation broth, and it can provide a theoretic and practical basis for setting up a spectral sensor to detect the solid concentration of anaerobic digestion process.

**Keywords** Infrared hyperspectral; Total solid; Volatile solid; Anaerobic digestion; Chemometrics

(Received Oct. 17, 2015; accepted Apr. 20, 2016)

\* Corresponding author