

可见-近红外光谱联合随机蛙跳算法检测生物柴油含水量

陈立旦^{1,2}, 赵艳茹²

(1. 浙江经济职业技术学院, 杭州 310018; 2. 浙江大学生物系统工程与食品科学学院, 杭州 310058)

摘要:生物柴油是一种优质清洁柴油, 可从各种生物质中提炼, 其特有的优势受到越来越广泛的关注。该文应用可见-近红外光谱技术原理对生物柴油的含水率进行了检测。配置含水率分别为 0、2.5%、5.0%、7.5% 和 10.0% 的试验样品并获取可见-近红外光谱, 进行主成分分析, 观察不同含水率生物柴油的聚类性, 并采用 Random Frog 算法进行特征波段的提取, 最后采用随机蛙跳算法(Random Frog)挑选出的特征波段作为偏最小二乘回归(partial least squares regression, PLSR)和最小二乘支持向量机(least squares-support vector machine, LS-SVM)模型的输入量, 建立生物柴油含水量的预测模型。结果发现: 采用 Random Frog 提取出的 8 条特征波段(563、560、642、565、562、493、559 和 779 nm)所建立非线性模型 LS-SVM 所得到的结果较好, 其中 Random Frog-LS-SVM 的结果中 R 均大于 0.95, 校正集均方根误差 RMSEC=0.722, 预测集均方根误差 RMSEP=0.520。结果表明采用 Random Frog-LS-SVM 模型可以准确的预测生物柴油的含水量, 为实际应用提供参考。

关键词:生物柴油; 含水量; 近红外光谱; 主成分分析; Random Frog; 偏最小二乘回归; 最小二乘支持向量机
doi: 10.3969/j.issn.1002-6819.2014.08.020

中图分类号: O657.33; TK6

文献标识码: A

文章编号: 1002-6819(2014)-08-0168-06

陈立旦, 赵艳茹. 可见-近红外光谱联合随机蛙跳算法检测生物柴油含水量[J]. 农业工程学报, 2014, 30(8): 168—173.

Chen Lidan, Zhao Yanru. Measurement of water content in biodiesel using visible and near infrared spectroscopy combined with Random-Frog algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014, 30(8): 168—173. (in Chinese with English abstract)

0 引言

生物柴油是指可供柴油机使用的清洁的、原料可再生的液体燃料。其主要成分是脂肪酸甲酯, 是以油菜籽油、大豆油、玉米油、麻风果、花生油、棕榈油、回收烹饪油及动物油等可再生资源制成的。它具备与石化柴油相近的性能, 可与柴油以任意比例混合或直接在柴油机上使用, 能减少温室气体排放、降低空气污染, 具有良好的安全性能。由植物油或脂通过酯交换反应得到的生物柴油, 其具有不含芳香烃、不含硫、十六烷值较高和可生物降解等特性, 生物柴油受到越来越广泛的关注^[1]。

然而生物柴油中的水分含量不仅对柴油的燃烧性能有很大的影响, 而且会对柴油机有腐蚀作用。肖华等得出柴油掺水比例对油耗有影响, 掺水比例过高时, 反而会费油, 而且柴油机全负荷运转时, 在各种转速下掺水均会使功率和扭矩下降, 且

掺水越多, 功率和扭矩下降得越厉害^[2]。并且油脂中的水分会提高生物柴油的化学活性, 使其容易变质。国内报道柴油掺水事件也对柴油市场造成了一定的影响。因此对应用到柴油机的生物柴油进行水分检测是非常有必要的。

传统的油品水分检测通常有蒸馏法、色谱法、卡尔-费休法、重量法等, 但操作较复杂, 检测用时长, 受环境影响大, 负反应多、试剂有毒、精度不高, 不能满足大多数用户要求在线快速检测油品水分的需求, 因此, 需要一种可以快速检测生物柴油中含水量的新方法^[3]。

可见-近红外光谱法具有分析简便快速、低成本、无污染及样品的非破坏性和多组分同时测定等优点^[4-10]。在石油化工领域, 近红外光谱技术已应用于汽油、柴油的测试及其成分的定量分析与研究^[11]。张瑜等研究了基于可见-近红外光谱技术的润滑油酸值无损检测方法, 获得了润滑油在 475~975 nm 范围内的可见-近红外光谱^[12]。张瑜等研究了基于可见-近红外光谱技术的发动机润滑油含水量的快速检测, 得到较好的结果^[13]。对生物柴油中含水量的检测研究较少。本文基于近红外技术的广泛使用对生物柴油含水量进行无损检测。

收稿日期: 2013-11-19 修订日期: 2014-03-24

基金项目: 国家自然科学基金面上项目(50975258); 2013年浙江省科技厅公益项目(2013C37078)资助

作者简介: 陈立旦(1968—), 男, 浙江宁波人, 教授, 主要从事汽车油液的检测技术研究。杭州 浙江经济职业技术学院, 310018。

Email: cld1121@163.com

1 材料与方法

1.1 试验样品

本试验中生物柴油样品由浙江东江能源科技有限公司提供, B100 产品(以植物、动物油脂等可再生生物资源为原料生产的)。按照体积分数配置含水率分别为 0、2.5%、5.0%、7.5%、10.0% 的试验样品, 将水分掺入生物柴油中时以玻璃棒迅速搅拌, 使水滴以尽量小的颗粒与生物柴油混合。每种类型样品选择 35 个样本, 共计 175 个样本。剔除异常样本 1 个, 将剩余的 174 个样本按照 1:2 的比例采用 Kennard-Stone (K-S)^[14]方法进行分类, 其中 116 个样本被选作校正集, 58 个样本被选为预测集样本。

1.2 仪器设备

本试验采用美国 ASD (Analytical Spectral Device) 公司生产的 Handheld FieldSpec 光谱仪, 其光谱采样间隔 1.5 nm, 测定范围为 325~1 075 nm。光源采用与仪器配套的 14.5 V 的卤素灯^[15]。光谱扫描采用透射方式, 每个样本采样 45 mL, 油样本的厚度大约为 2 cm 左右。每个样本扫描 30 次, 保存 3 条平均光谱, 在处理光谱数据时, 选择 3 条光谱曲线的平均光谱作为相应样本的光谱曲线。

1.3 生物柴油光谱特征的提取

为了避免光谱中噪声所带来的干扰, 将存在较大噪声的 325~410 nm 以及 1 000~1 075 nm 之间的光谱去除。保留 410~1 000 nm 之间光谱进行生物柴油含水率的检测。

1.3.1 主成分分析

主成分分析是将多指标化为少数几个综合指标的一种统计分析方法。本文获得的生物柴油光谱变量波长为 591 条, 由于分析元素较多, 且彼此存在一定的相关。主成分分析则可以在保证尽可能多的反映原始信息的基础上, 采用较少的互不相关的综合变量来代替较多的分析元素^[16]。

1.3.2 随机蛙跳 (Random Frog) 算法提取特征波段

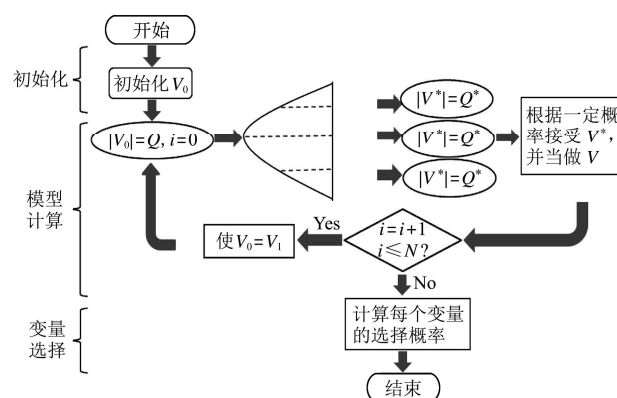
Random Frog 是一种新型特征波段算法。其能够利用少量的变量迭代进行建模, 是一种非常有效的高维数据变量选择方法。它能够输出每个变量选择可能性, 从而进行变量的选择^[17]。

图 1 表示 Random Frog 算法的计算步骤, 其主要的运算步骤包括以下 3 步:

1) 一个变量子集 V_0 , 初始化时包含 Q 个变量;

2) 基于原始的变量子集, 提出一个候选变量子集 V^* , 包含 Q^* 个变量; 选择 V^* 作为 V_1 来代替原始的变量子集 V_0 。直到 N 次迭代终止, 这个过程完成。

3) 最后计算每个变量的选择概率并以此作为选择变量的标准。Random Frog 选择变量的具体步骤参见文献[18]。



注: 来源于文献[17]。 V_0 : 变量子集; Q : 变量个数; V^* : 候选变量子集; Q^* : 变量个数; i : 迭代次数。

Note: From reference [17]. V_0 : Variable subset; Q : Number of variable; V^* : Candidate variable subset; Q^* : Number of variable; i : Iterations.

图 1 随机蛙跳算法 (Random Frog) 的计算步骤

Fig.1 Key steps of Random Frog

1.4 建模方法

1.4.1 偏最小二乘回归 (partial least squares regression, PLSR)

偏最小二乘是基于成分提取思想, 可解决变量之间多重相关性的问题^[19-22]。PLSR 是对光谱矩阵 X 和浓度矩阵 Y 同时进行分解, 并在分解时考虑两者之间的相互关系, 加强对应计算关系, 从而保证获得最佳的校正模型。

1.4.2 最小二乘支持向量机 (least squares-support vector machine, LS-SVM)

LS-SVM 能够通过核函数将低维的非线性问题转化成高维的线性问题, 同时利用结构风险最小化原理来提高泛化能力, 从而有效地处理线性和非线性多元变量的建模问题, 也能较快地解决多元变量的复杂性问题, 最终降低了计算的复杂性并提高计算效率^[23]。本研究中采用径向基函数 (radial basis function, RBF) 进行模型的构建。

1.5 模型评价标准

本文的光谱模型采用以下参数进行评估: 建模和预测集的相关系数 (R), 校正集均方根误差 (root mean square error for calibration, RMSEC) 预测集均方根误差 (RMSEP)。好的模型需要有较低的均方根误差以及较高的 R ^[24-26]。

1.6 分析软件

本试验分析软件为 ASD View Spec Pro、Unscramble10.1、MatlabR2009a 以及 LS-SVM 工具包和 OriginPro 8 绘图软件。

2 结果与分析

2.1 光谱曲线分析

5 种不同含水率的生物柴油的平均可见近红外吸收光谱如图 2 所示。可以发现, 在可见光区域, 含有水分的生物柴油与原始生物柴油的光谱曲线波形不同, 含有水分的生物柴油的光谱曲线在 550 和 610 nm 左右处, 有强烈反射峰, 可能是由于当将水与生物柴油混合, 水以微小的悬浮颗粒存在于生物柴油中, 因而所造成的这种反射峰^[27], 除此外, 在 480 以及 700 nm 等处的吸收曲线也是不同的。然而, 所有的光谱曲线在 930 nm 左右处有较强的吸收峰, 推测是由于 C-H 振动的三级倍频在 900~950 nm。所有的含有水分的生物柴油光谱曲线波形是相似的, 随着含水率的增加, 生物柴油的吸光度增大。

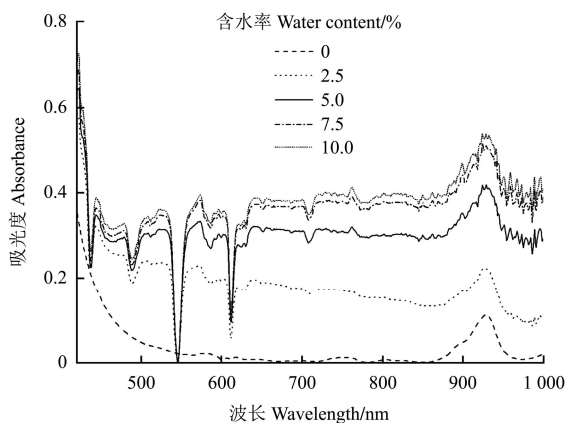


图 2 不同含水率生物柴油的可见/近红外吸收光谱
Fig.2 Vis-near infrared absorbance spectra of biodiesel with different water contents

2.2 主成分分析

选取主成分的个数取决于主成分的累计方差贡献率, 一般使累计方差贡献率大于 85%~95%所需的主成分数能够代表原始变量所能提供的绝大部分信息。本文中仅采用 PC-1 与 PC-2 就可以使累计方差贡献率达到 99.3%。图 3 中发现, 含水率 5.0% 的生物柴油中存在异常样本, 观察其光谱数据, 发现吸收度偏小, 可能当测量该样本时, 没有及时搅拌, 导致了悬浮颗粒的沉淀。在后续的处理中, 将该样本剔除以提高模型精度。

2.3 特征波段的提取

采用 Random Frog 算法进行敏感波段的提取,

Random Frog 选择变量的依据是不同的变量具有不同的选择可能性, 图 4 中大部分敏感波段集中在 550、650、780 以及 900 nm 附近。本文设定选择可能性的阈值为 0.5, 最终选出 563、560、642、565、562、493、559 和 779 nm 共 8 条敏感波段 (8 条敏感波段按照被选择概率降序排列), 将这些敏感波段与图 2 对照, 可以发现在这些敏感波段处, 含水柴油光谱曲线均有一定的波动。另外可以发现大部分的敏感波段集中于 500~600 nm 之间, 这与含水的生物柴油的光谱曲线特征是相吻合的。采用 8 条敏感波段的光谱信息作为建立模型的输入变量进行预测模型的建立。

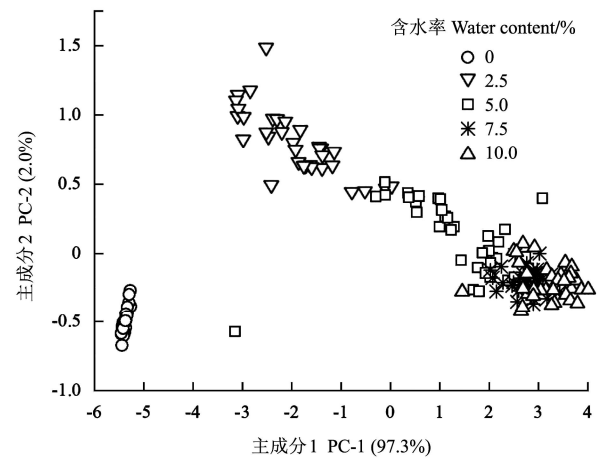


图 3 主成分得分图

Fig.3 Scores plot obtained from principal component analysis

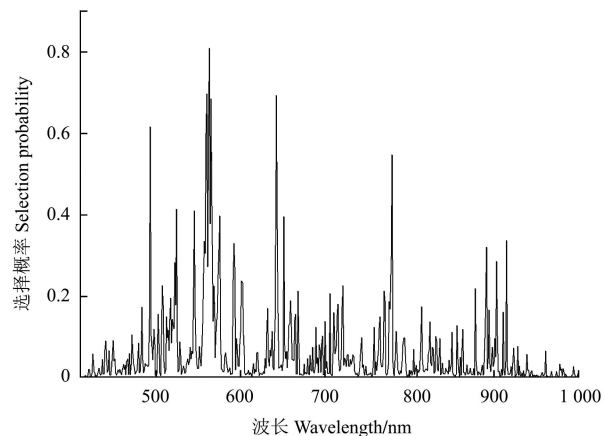


图 4 采用 Random Frog 进行特征波段提取

Fig.4 Extracting feature spectra using Random Frog

2.4 预测模型的建立

将经过 Random Frog 选择得到的 8 条特征波段的光谱信息作为建立预测模型 PLSR 和 LS-SVM 的输入变量 X_c , 生物柴油不同含水率, 0、2.5%、5.0%、7.5% 和 10.0% 作为 Y_c 值进行预测模型的建立。本文中, 采用 RBF 作为核函数的结果如表 1 所示。

表 1 不同含水率生物柴油的预测计算结果

Table 1 Predicting results of spectral model for measurement of water content in biodiesel

模型 Models	方法 Methods	变量 Variables	校正集 Calibration set		预测集 Prediction set	
			均方根误差 RMSEC Root mean square error	相关系数 R_c Correlation coefficient	均方根误差 RMSEP Root mean square error	相关系数 R_p Correlation coefficient
偏最小二乘回归 PLSR Partial least squares regression	全波段 Full	591	0.642	0.944	0.524	0.926
	随机蛙跳算法 Random-Frog	8	0.913	0.942	1.611	0.729
最小二乘支持向量机 LS-SVM Least squares-support vector machine	全波段 Full	591	0.365	0.990	0.383	0.981
	随机蛙跳算法 Random-Frog	8	0.722	0.965	0.520	0.966

注：校正集样本 $n=116$ ，预测集样本 $n=58$ 。

Note: Calibration samples $n=116$, prediction samples $n=58$.

表 1 中可以看出，Random-Frog-PLSR 模型与全波段建立的 PLSR 模型相比， R_p 较差（ $0.729 < 0.926$ ），RMSEP 较大（ $1.611 > 0.524$ ），整体结果较差；然而从变量个数来说，Random Frog-PLSR 仅采用了 8 条波段，相比 591 条全波段的数据矩阵，Random Frog-PLSR 模型已经是非常简化。LS-SVM 非线性模型所得到的结果明显优于 PLSR 模型，尤其是 Random Frog-LS-SVM 模型中 R 均大于 0.95，此方法能够准确预测生物柴油的含水量。

3 结 论

本文应用可见-近红外光谱技术原理对生物柴油的含水率进行了检测。并采用随机蛙跳算法（Random Frog）算法进行特征波段的提取，经过 Random Frog 提取出的 8 条特征波段（563、560、642、565、562、493、559 和 779 nm）大部分集中于 500~600 nm 之间，与光谱曲线的特性符合。并且采用这 8 条特征波段所建立的线性模型 PLSR 以及非线性模型 LS-SVM 所得到的结果与全波段得到的结果差不多，其中 Random Frog-LS-SVM 的结果中 R 均大于 0.95，校正集均方根误差 RMSEC=0.722，预测集均方根误差 RMSEP=0.520。结果表明采用 Random Frog-LS-SVM 模型可以准确的预测生物柴油的含水量，为实际应用提供参考。

【参 考 文 献】

- [1] 袁银南, 江清阳, 孙平, 等. 柴油机燃用生物柴油的排放特性研究[J]. 内燃机学报, 2003, 21(6): 423—426.
Yuan Yinnan, Jiang Qingyang, Sun Ping, et al. Study of emission characteristic of diesel engine fueled with biodiesel[J]. Transactions of Csice, 2003, 21(6): 423—426. (in Chinese with English abstract)
- [2] 肖华, 黄荣华. 重柴油掺水燃烧对柴油机性能影响的试

验研究[J]. 柴油机设计与制造, 2007, 15(1): 31—33.
Xiao Hua, Huang Ronghua. An experimental research on performance effect with water-diesel blends combustion in a diesel engine[J]. Design and Manufacture of Diesel Engine, 2007, 15(1): 31—33. (in Chinese with English abstract)

- [3] 焦昭杰, 张贤明, 梁新元, 等. 近红外光谱法快速测定油品中的水分[J]. 光谱实验室, 2011, 28(6): 3120—3124.
Jiao Zhaojie, Zhang Xianming, Liang Xinyuan, et al. Fast determination of water content in oil by near infrared spectroscopy[J]. Chinese Journal of Spectroscopy Laboratory, 2011, 28(6): 3120—3124. (in Chinese with English abstract)
- [4] 李晓丽, 何勇, 裘正军. 一种基于可见-近红外光谱快速鉴别茶叶品种的新方法[J]. 光谱学与光谱分析, 2007, 27(2): 279—282.
Li Xiaoli, He Yong, Qiu Zhengjun. Application PCA-ANN method to fast discrimination of tea varieties using visible/Near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2007, 27(2): 279—282. (in Chinese with English abstract)
- [5] He Cheng, Yang Zengling, Huang Guangqun. Qualitative and quantitative analysis of straw content in straw-coal blends using Vis/NIR spectroscopy[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2013, 29(17): 188—195.
贺城, 杨增玲, 黄光群, 等. 可见/近红外光谱分析秸秆—煤混燃物的秸秆含量[J]. 农业工程学报, 2013, 29(17): 188—195. (in English with Chinese abstract)
- [6] 胡永光, 陈培培, 赵梦龙. 绿茶杀青叶料含水率可见-近红外光谱检测[J]. 农业机械学报, 2013, 44(8): 174—179.
Hu Yongguang, Chen Peipei, Zhao Menglong. Determination of water content in de-enzyming green tea leaves based on visible-near infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(8): 174—179. (in Chinese with English abstract)

English abstract)

- [7] Liao Y T, Fan Y X, Cheng F. On-line prediction of pH values in fresh pork using visible/near-infrared spectroscopy with wavelet de-noising and variable selection methods [J]. *Journal of Food Engineering*, 2012, 109(4): 668—675.
- [8] Kavdir I, Lu R, Ariana D, et al. Visible and near-infrared spectroscopy for nondestructive quality assessment of pickling cucumbers[J]. *Postharvest Biology and Technology*, 2007, 44(2): 165—174.
- [9] Wang J, Nakano. K, Ohashi S, et al. Comparison of different modes of visible and near-infrared spectroscopy for detecting internal insect infestation in jujubes[J]. *Journal of Food Engineering*, 2010, 10(1): 78—84.
- [10] Xie L J, Ying Y B, Ying T J. Rapid determination of ethylene content in tomatoes using visible and short-wave near-infrared spectroscopy and wavelength selection[J]. *Chemometrics and Intelligent Laboratory Systems*, 2009, 97(2): 141—145.
- [11] 马骊军. 近红外光谱技术在油品检测中的应用[D]. 北京: 中国石油大学, 2008.
- [12] 张瑜, 吴迪, 何勇, 等. 基于可见-近红外光谱技术的润滑油酸值无损检测方法研究[J]. *红外*, 2011, 32(12): 39—43.
Zhang Yu, Wu Di, He Yong, et al. Non-invasive measurement of acid value of lubricant using visible and near infrared spectroscopy[J]. *Infrared*, 2011, 32(12): 39—43. (in Chinese with English abstract)
- [13] 张瑜, 蒋璐璐, 吴迪, 等. 基于可见-近红外光谱技术的润滑油含水量无损检测方法研究[J]. *光谱学与光谱分析*, 2010, 30(8): 2111—2114.
Zhang Yu, Jiang Lulu, Wu Di, et al. Non-invasive measurement of water content in engine lubricant using visible and near infrared spectroscopy[J]. *Spectroscopy and Spectral Analysis*, 2010, 30(8): 2111—2114.
- [14] Macho S, Rius A, Callao M P, et al. Monitoring ethylene content in heterophasic copolymers by near-infrared spectroscopy standardization of the calibration model[J]. *Analytica Chimica Acta*, 2001, 445(2): 213—220.
- [15] 王莉, 何勇, 刘飞, 等. 应用光谱技术和支持向量机分析方法快速检测啤酒糖度和 pH 值[J]. *红外与毫米波学报*, 2008, 27(1): 51—55.
Wang Li, He Yong, Liu Fei, et al. Rapid detection of sugar content and pH in beer by using spectroscopy technique combined with support vector machines[J]. *Journal of Infrared and Millimeter Waves*, 2008, 27(1): 51—55. (in Chinese with English abstract)
- [16] 刘韬, 田洪祥, 郭文勇. 主成分分析在某型柴油机光谱数据分析中的应用[J]. *光谱学与光谱分析*, 2010, 30(3): 779—782.
Liu Tao, Tian Hongxiang, Guo Wenyong. Application of PCA to diesel engine oil spectrometric analysis[J]. *Spectroscopy and Spectral Analysis*, 2010, 30(3): 779—782. (in Chinese with English abstract)
- [17] Li H D, Xu Q S, Liang Y Z. Random frog: An effective reversible jump Markov Chain Monte Carlo-like approach for variable selection with applications to gene selection and disease classification[J]. *Analytical Chimica Acta*, 2012, 740: 20—26.
- [18] Yun Y H, Li H D, Wood L R. E, et al. An efficient method of wavelength interval selection based on random frog for multivariate spectral calibration[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscop*, 2013, 111: 31—36.
- [19] Shetty N, Gislum R. Quantification of fructan concentration in grasses using NIR spectroscopy and PLS[J]. *Field Crops Research*, 2011, 120(1): 31—37.
- [20] B Braga J W, Junior A A S, Martins I S. Determination of viscosity index in lubricant oils by infrared spectroscopy and PLSR[J]. *Fuel*, 2014, 120: 171—178.
- [21] Chen B S, Wu H N, Li S F Y. Development of variable pathlength UV-vis spectroscopy combined with partial-least-squares regression for wastewater chemical oxygen demand (COD) monitoring[J]. *Talanta*, 2014, 120: 325—330.
- [22] Liu Y D, Sun X D, Ouyang A G. Nondestructive measurement of soluble solid content of navel orange fruit by visible-NIR spectrometric technique with PLSR and PCA-BPNN[J]. *LWT - Food Science and Technology*, 2010, 43(4): 602—607.
- [23] 谢传奇, 王佳悦, 冯雷, 等. 应用高光谱图像光谱和纹理特征的番茄早疫病早期检测研究[J]. *光谱学与光谱分析*, 2013, 33(6): 1603—1607.
Xie Chuanqi, Wang Jiayue, Feng Lei, et al. Study on the early detection of early blight on tomato leaves using hyperspectral imaging technique based on spectroscopy and texture[J]. *Spectroscopy and Spectral Analysis*, 2013, 33(6): 1603—1607. (in Chinese with English abstract)
- [24] Zhang X L, He Y. Rapid estimation of seed yield using hyperspectral images of oilseed rape leaves[J]. *Industrial Crops and Products*, 2013, 42: 416—420.
- [25] Yu K Q, Zhao Y R, Li X L, et al. Identification of crack features in fresh jujube using Vis/NIR hyperspectral imaging combined with image processing[J]. *Computers and Electronics in Agriculture*, 2014, 103: 1—10.
- [26] Huang H, Liu L, Ngadi M O, et al. Rapid and non-invasive quantification of intramuscular fat content of intact pork cuts[J]. *Talanta*, 2014, 119: 385—395.
- [27] 乐成峰, 李云梅, 查勇, 等. 太湖悬浮物对水体生态环境的影响及其高光谱反演[J]. *环境科学学报*, 2008, 28(10): 2148—2155.
Le Chengfeng, Li Yunmei, Zha Yong, et al. Inversion of hyperspectral reflectance for estimating suspended matter concentration[J]. *Acta Scientiae Circumstantiae*, 2008, 28(10): 2148—2155. (in Chinese with English abstract)

Measurement of water content in biodiesel using visible and near infrared spectroscopy combined with Random-Frog algorithm

Chen Lidan^{1,2}, Zhao Yanru²

(1. Zhejiang Technology Institute of Economy, Hangzhou 310018, China;

2. School of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China)

Abstract: Biodiesel (fatty acid methyl or ethyl esters) is made from vegetable oil or animal fat (triglycerides) reacting with methanol or ethanol using a catalyst (lye). It is safe, biodegradable, and produces less air pollutants than petroleum-based diesel or recycled restaurant greases. With the increasing demand of green energy source and the decreasing of fossil fuel, biodiesel has gained increasing attention as one of the alternative fuels. 100% biodiesel (B100) was used in this study. Experimental samples with water content of 0, 2.50%, 5.00%, 7.50% and 10.0% were set. There were 35 samples for every treatment with different water contents, and total 175 samples. 116 samples were selected for calibration set, and 58 samples for prediction set based with Kennard-Stone (K-S) method. Visible and near infrared spectra (Vis-NIR) technique which was a nondestructive and rapid method, was used to measure the water content in biodiesel. Samples were scanned using the ADS Handheld FieldSpec spectrometer and spectra of samples were acquired. Principal component analysis (PCA) was used to compress spectral data and observe the cluster's situation of biodiesel with different water contents. The scores plot showed a good cluster distribution and the total accumulated variance of PC-1 and PC-2 was up to 99.3%. Random Frog algorithm was applied to extract spectral feature. Then, 8 sensitive wavelengths (563, 560, 642, 565, 562, 493, 559 and 779 nm) were selected respectively. Spectral feature and different water contents were set as input values of partial least squares regression (PLSR) and least squares-support vector machine (LS-SVM) models. It was showed that LS-SVM and PLSR with full spectra had good results, while the variables were too much (116×591) compared with the regression models (116×8). Results of the Random Frog-LS-SVM were better than the Random Frog-PLSR. *R* of the non-linear LS-SVM models with spectral feature extracted by Random Frog was higher than 0.965, RMSEC of 0.722, RMSEP of 0.520. Sensitive wavelengths extracted were good for eliminating the interfering spectral and improving the accuracy of the model. Results indicated that the Random Frog-LS-SVM as a satisfactory model can measure the water content in biodiesel accurately, which could provide a reference for practical application.

Key words: biodiesel; water content; near infrared spectroscopy; principal component analysis; random frog; partial least squares regression; least squares-support vector machine

(责任编辑：刘丽英)