

文章编号: 0253-2239(2010)02-0585-05

基于直接正交信号校正的水稻冠层叶瘟光谱诊断

刘 飞¹ 冯 雷¹ 柴荣耀² 孙光明¹ 楼兵干³ 孙 测⁴ 何 勇¹

(¹浙江大学生物系统工程与食品科学学院, 浙江 杭州 310029; ²浙江省农业科学院植物保护研究所, 浙江 杭州 310021)
³浙江大学生物技术研究所, 浙江 杭州 310029; ⁴浙江省慈溪市农业机械化管理局, 浙江 慈溪 315300)

摘要 采用直接正交信号校正-连续投影算法(DOSC-SPA)联用的组合处理方法,实现了应用可见/近红外光谱技术对水稻冠层叶瘟病的快速准确诊断。采集水稻冠层样本共 120 个,其中健康和染病样本各 60 个。用 DOSC-SPA 方法,对水稻冠层光谱数据进行直接正交化处理,然后通过连续投影算法提取有效波长(EW),建立了有效波长(775 nm)与叶瘟诊断判别的直接线性方程: $Y = 5.283X$ 。应用该方程对预测集样本进行诊断判别,其判别准确率为 95.0%,获得了满意的判别精度。结果表明,DOSC-SPA 组合处理方法提取水稻冠层叶瘟诊断的特征波长是非常有效的,并且通过直接线性方程能获得满意的判别精度,为后续水稻大田叶瘟病情监测、喷药处理及相应病情监测仪器的开发提供了方法和依据。

关键词 光谱学;叶瘟病判别;直接正交信号校正;连续投影算法;偏最小二乘法;水稻

中图分类号 O657.33;S435.111.4 **文献标识码** A **doi**: 10.3788/AOS20103002.0585

Discrimination of Rice Canopy Leaf Blast Based on Spectroscopic Techniques and Direct Orthogonal Signal Correction

Liu Fei¹ Feng Lei¹ Chai Rongyao² Sun Guangming¹ Lou Binggan³
Sun Ce⁴ He Yong¹

¹ College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, Zhejiang 310029, China
² Institute of Plant Protection, Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang 310021, China
³ Institute of Biotechnology, Zhejiang University, Hangzhou, Zhejiang 310029, China
⁴ Bureau of Agriculture Mechanization of Cixi, Cixi, Zhejiang 315300, China

Abstract Visible and near infrared (Vis/NIR) spectroscopy was investigated for the fast and accurate discrimination of rice canopy leaf blast using the combining processing of direct orthogonal signal correction-successive projections algorithm (DOSC-SPA). 120 rice canopy samples were collected with 60 samples for healthy and blast rice canopy leaves, respectively. The calibration set consisted of 80 samples with 40 of healthy and 40 of blast. The remaining 40 samples were used for validation set. Using the new combination of direct orthogonal signal correction-successive projections algorithm, the spectral data of rice canopy was directly orthogonal processed, and then the successive projections algorithm (SPA) was applied for the selection of most important variables, named effective wavelengths (EWs). Finally, a direct linear function was developed between the effective wavelength (775 nm) and rice canopy leaf blast for the discrimination rice canopy leaf blast, which was shown as $Y=5.283X$. The validation set was used to evaluate and assess the prediction performance of the developed direct linear function, and the discrimination ratio was 95%. An acceptable discrimination result and a parsimonious discrimination model were achieved by this direct

收稿日期: 2009-03-09; 收到修改稿日期: 2009-04-16

基金项目: 国家 863 计划(2007AA10Z210), 国家自然科学基金(60605011), 公益性行业(农业)科研专项(200803037)和浙江省重大科技专项(2009C12002)资助课题。

作者简介: 刘 飞(1983—),男,博士研究生,主要从事精细农业技术和农业装备自动化等方面的研究。

E-mail: fliu@zju.edu.cn

导师简介: 何 勇(1963—),男,教授,博士生导师,主要从事精细农业技术和农业装备自动化等方面的研究。

E-mail: yhe@zju.edu.cn

linear function. The results indicated that the new combination of direct orthogonal signal correction and successive projections algorithm was a powerful way for the selection of effective wavelengths, and Vis/NIR spectroscopy was successfully applied for the discrimination of the rice canopy leaf blast, and the direct linear function achieved an acceptable discrimination ratio. This study provided a new approach for the further studies such as the on-field monitoring of rice blast, the treatment of spraying rice leaf blast, and the development of instrument for leaf blast monitoring.

Key words spectroscopy; leaf blast discrimination; direct orthogonal signal correction; successive projections algorithm; partial least squares analysis; rice

1 引 言

水稻是我国极为重要的粮食作物之一,其产量接近全国粮食总产量的 1/2,在我国粮食生产及粮食安全保障中占有重要地位。水稻在生长过程中,容易受到多种病虫害的影响,最终影响到水稻的产量和质量。稻瘟病是水稻病害中最为常见的一种,严重发病情况下可使水稻减产 40%~50%^[1]。目前,水稻叶瘟的诊断多依靠人眼进行分级测报,由于在发病早期,病斑不显著,难以实现快速准确判别,从而错过防治的有利时期。同时,人眼识别的主观性强,需要的时间和精力较多,无法满足现代农业的发展要求。因此,研究一种能够快速、准确对水稻叶瘟进行判别的方法是非常必要的。

近红外光谱技术作为一种快速、简便和准确的分析技术,已被广泛应用于农业、食品、石油、化工和医药等行业^[2-6]。在水稻相关的研究中,应用光谱技术进行了水稻品种快速鉴别^[7]、水稻不同生育期光谱有效波段选择^[8]、水稻叶面积指数的估算^[9]、稻谷年份的鉴别^[10]、水稻辐照剂量的判别^[11]、水稻白背飞虱虫害与冠层光谱特性的关系^[12]等。吴曙雯等^[13]研究了稻叶瘟对水稻光谱特性的影响,但只分析了水稻光谱反射率的变化,未建立定量的判别分析模型。目前,应用可见/近红外光谱技术对水稻冠层叶瘟的定量化诊断判别报道甚少。本研究通过获取水稻冠层叶瘟的光谱数据,采用直接正交信号校正-连续投影算法(DOSC-SPA)联用的方法,提取有效波长(EW),建立了光谱有效波长与水稻冠层叶瘟诊断的定量化判别模型,从而实现了对水稻冠层叶瘟的快速准确判别。

2 材料与方法

2.1 仪器设备

试验使用美国 ASD 公司的 Handheld FieldSpec 光谱仪,其光谱测定范围 325~1075 nm,采样间隔为 1.5 nm,探头视场角为 20°,光谱扫描次数设定为 30 次。光源采用 14.5 V 卤素灯。试验采

用漫反射模式,光谱仪探头距离水稻冠层 150 mm,光谱采集的仪器位置设定如图 1 所示。分析软件采用 ASD View Spec Pro, Unscrambler 9.6 (CAMO AS, Oslo, Norway), Matlab 7.0 (The Math Works, Natick, USA)。

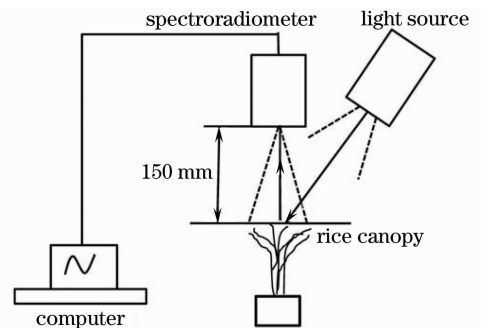


图 1 光谱采集仪器设定

Fig. 1 Set up of spectra acquisition instrument

2.2 水稻染病试验设计

供试水稻品种为籼稻原丰早品种,供试菌为稻瘟病菌原菌稻灰梨孢 *Pyricularia grisea* 菌株 ZB₁ (浙江省农业科学院植物保护研究所提供)。稻瘟病菌的分生孢子在燕麦片培养基上培养,接种时,配置孢子悬浮液浓度为 $2 \times 10^5 \sim 3 \times 10^5$ cfu/mL (cfu/mL 为每毫升中菌群数)。试验分为对照(健康)和染病两组,在相同条件下同时进行。稻苗在人工气候箱培养,欲染病的稻苗在 3~4 片叶时进行喷雾接种,环境条件是: 25 °C, $n_{RH} > 86\%$ (n_{RH} 为空气相对湿度),然后在 25 °C~28 °C 恒温下黑暗保湿 24 h,再移出室外进行隔离光照、喷水保湿培养。在第 3, 4, 7, 10, 15, 20, 25, 30, 35 天进行健康和染病水稻冠层的可见/近红外光谱数据采集。每个时间段均采集了 60 个健康样本和 60 个染病样本的光谱数据。为了研究应用光谱技术进行水稻冠层叶瘟的早期诊断的可行性,本试验采用第 4 天采集的 120 个水稻样本的光谱数据,其中健康和染病样本各 60 个。这是因为第 3 天时,染病样本还未显现症状,第 4 天染病样本开始出现细微症状,但需要非常仔细的辨认,而第 7 天及以后,水稻叶片叶瘟症状已比较明显,肉眼

可以非常明确的辨认,因此不再需要做光谱判别。试验中完全健康的水稻作为健康样本,染病的样本不论是轻微的还是症状稍微明显一点的,均作为染病样本对待。且非常轻微染病的样本后续均表现出了明显的染病症状,也确保核对了第 4 天染病样本均已染病成功。随机选择建模集样本 80 个(健康和染病样本各 40 个),其余 40 个样本(健康和染病各 20 个)为预测集样本。

2.3 DOSC-SPA 组合处理

为消除光谱数据的噪声、温度等外部不利因素的影响,建立较优化的判别模型,通常需要对光谱数据进行预处理,常用的光谱预处理方法包括 Savitzky-Golay 平滑、变量标准化、一阶及二阶求导处理等^[14]。上述预处理方法只针对光谱数据矩阵进行处理,并未考虑待测物浓度矩阵的信息,因此数据处理缺乏针对性。直接正交信号校正(DOSC)是在正交信号校正(OSC)的基础上,改进得到的一种新的预处理算法^[15,16]。直接信号校正不仅对光谱数据进行处理,同时考虑被测物浓度矩阵,再通过正交化投影,提取最优的特征变量,从而改善处理后模型的预测性能。

连续投影算法(SPA)是一种新的波长选择方法^[17~19]。它可以通过投影变换,提取含有最低冗余度和最小共线性的有效波长,大大减少模型输入变量的数量,在不影响预测性能的情况下简化模型,提高模型的预测速度。将直接信号校正与连续投影算法联用,提出一种新的 DOSC-SPA 组合处理方法,可以将光谱信息与待测物浓度矩阵相关联,提取最为有效的特征波长,进而可以建立更加简便、快速的分析模型。采用 DOSC-SPA 组合处理,提取水稻冠层叶瘟光谱的有效波长,进而建立水稻冠层叶瘟光谱诊断判别模型。

2.4 偏最小二乘法

偏最小二乘法(PLS)是一种应用广泛的光谱建模分析方法^[20]。PLS 通过提取全波段光谱的特征变量(LV),建立光谱与水稻叶瘟的相关关系,通过交互验证得到水稻叶瘟检测的 PLS 判别模型。建模中,将健康样本和染病水稻样本分别赋予虚拟变量 1 和 0,其判别阈值为 0.5,预测值大于 0.5 的样本为健康样本,小于 0.5 的样本为染病样本。通过对光谱数据进行预处理,建立水稻叶瘟诊断的 PLS 判别模型。模型性能的优劣以对预测集样本的准确判别率为标准,正确判别率越高,说明模型性能越好。

3 试验结果与分析

3.1 水稻冠层光谱特性

第 4 天采集的 120 个健康和染病水稻冠层样本的原始可见/近红外反射光谱图如图 2 所示。图中横坐标为波长,范围为 400~1000 nm,纵坐标为反射率值。从图 2 可知,健康和染病的水稻样本在波长 500~650 nm 和 750~1000 nm 两个波段存在比较明显的差异,这是因为染病水稻样本的病斑随时间推移越来越大,引起光谱反射率的变化。但直接从光谱图难以准确区分各个时期健康和染病的水稻样本,因此需要进一步的建模分析,实现健康和染病水稻样本的判别区分。

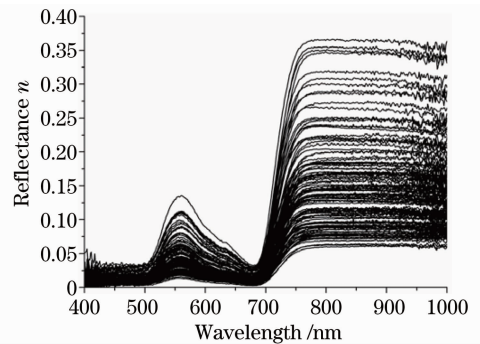


图 2 水稻冠层原始可见/近红外反射光谱图
Fig. 2 Original visible/near infrared reflectance spectra of rice canopy

3.2 PLS 判别模型的建立

应用 DOSC 处理后的光谱数据建立水稻冠层健康和染病样本的 PLS 判别模型,建模中采用留一交互验证保证模型的稳定性。同时与常用的 Savitzky-Golay 平滑、变量标准化、一阶及二阶求导处理等预处理方法进行比较分析,用预测集样本对模型进行检验,以 0.5 作为健康(标准值为 1)和染病(标准值为 0)样本的识别阈值,通过比较,得到最优 PLS 判别模型为 DOSC-PLS 模型,应用的特征变量(L)为 2 个。应用 $L_1 \times L_2$ 得到水稻样本的散点分布图如图 3 所示。水稻健康样本和染病样本能够很好的区分,并且染病样本的分布相对更加集中,而健康样本的分布相对比较分散。该模型对建模集样本的判别准确率为 100%,能准确地将健康和染病的样本区分。对预测集样本的判别准确率为 95.0%,将其中 2 个健康样本误判为染病样本。具体的判别结果如表 1 所示。考虑到采用全波段 600 个变量建模,所需时间较长,并且 600 个变量中含有大量的共线性和冗余变量,因此,本文采用了连续投影算法提取有效波长,实现水稻冠层叶瘟的快速诊断判别。

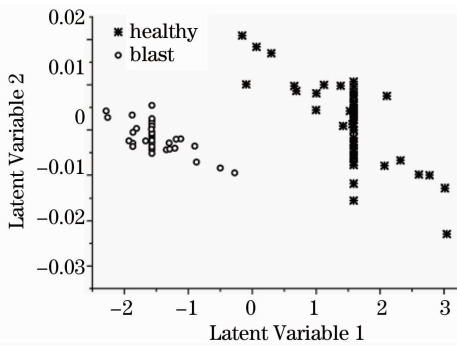


图3 应用 $L_1 \times L_2$ 的水稻样本散点分布图

Fig. 3 Scatter plots of rice samples using $L_1 \times L_2$

3.3 DOSC-SPA 提取有效波长及直接线性模型的建立

在 DOSC 处理过的光谱数据基础上,应用连续投影算法,提取水稻冠层叶瘟诊断的有效波长。将 SPA 选取波长的最大值设定为 30,波长选择中采用留一交叉验证法,当交叉验证均方根误差 (RMSECV) 最小时所对应的变量个数即为选定的波长个数。通过计算,得到基于 DOSC-SPA 方法的 RMSECV 最小为 0.0014,此时选取的有效波长个数为 6 个,分别为 775,408,406,

表 1 不同模型对水稻冠层叶瘟的判别结果

Table 1 Discrimination results of rice canopy leaf blast by different models

Models	Calibration ($n=80$)		Validation ($n=40$)		
	Healthy ($n=40$)	Blast ($n=40$)	Healthy ($n=20$)	Blast ($n=20$)	Total ($n=40$)
DOSC-PLS	100%, (0)	100%, (0)	90%, (2)	100%, (0)	95%, (2)
DOSC-SPA-PLS	100%, (0)	100%, (0)	90%, (2)	100%, (0)	95%, (2)
$Y=5.283X$	100%, (0)	100%, (0)	90%, (2)	100%, (0)	95%, (2)

通过图 4 可知,波长 408,406,412,417 和 422 nm 均分布在光谱曲线的起始部分,而由图 4 观察得出,在 400~450 nm 之间,光谱曲线含有较大的噪音。因此,将 DOSC-SPA 提取的 6 个波长中最为重要的 775 nm 单独取出,建立直接线性判别方程。将 DOSC 处理后的 775 nm 处的数值作为输入变量 X ,将样本健康或染病的种类作为输出变量 Y ,应用建模集 80 个样本得到水稻冠层叶瘟诊断的直接线性方程。其结果如下

$$Y = 5.283X, \quad (1)$$

其中 X 为 DOSC 处理后 775 nm 处的光谱值, Y 为水稻冠层叶瘟的判别值。当 $Y > 0.5$ 时,判定该样本为健康样本;当 $Y < 0.5$ 时,判定该样本为染病样本。用建模集样本对该直接线性方程的判别性能进行验证,对建模集样本的判别准确率为 100%,对预测集样本的判别准确率为 95.0%,同样是将其中 2 个健康样本误判为染病样本,具体结果如表 1 所示。对原始光谱数据进行直接正交信号校正处理后,如

412,417 和 422 nm,如图 4 所示。所选的波长按照其重要性大小排列,越靠前,说明该波长点越重要。在所选的 6 个波长点中,775 nm 最为重要。

将所提取的 6 个有效波长作为 PLS 判别模型的输入,采用留一交叉验证,得到 DOSC-SPA-PLS 模型,用预测集样本进行验证,判别结果如表 1 所示。对所有建模集样本判别准确率为 100%,对预测集样本判别准确率为 95.0%,将其中 2 个健康样本误判为染病样本。

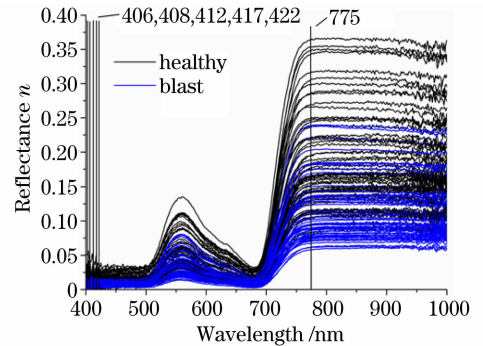


图 4 应用 DOSC-SPA 所提取的有效波长

Fig. 4 Effective wavelengths selected by DOSC-SPA

果只选用某一个特定的波长建模,在其他条件不变的情况下,得到的线性方程都应得到相同的效果。连续投影算法用来选取 DOSC 处理后水稻冠层叶瘟诊断判别的有效波长,实际选出了 6 个有效波长。通过逐步回归分析等方法是无法快速选出这 6 个有效波长的。但因为波长 408,406,412,417 和 422 nm 均分布在光谱曲线的起始部分,而由图 4 观察得出,在 400~450 nm 之间,光谱曲线含有较大的噪音,因此只选取了 775 nm 这一特定的波长建立线性方程进行预测。综合上述三个模型,采用直接线性方程最为简便,并且对水稻冠层叶瘟的判别准确率为 95.0%,获得了满意的判别精度。

4 结 论

应用可见/近红外光谱技术结合 DOSC-SPA 组合处理对水稻冠层叶瘟快速准确诊断判别是可行的。通过 DOSC-SPA 组合处理,选取有效波长为

775 nm, 建立水稻冠层叶瘟诊断的直接线性方程为 $Y=5.283X$, 对预测集样本判别准确率为 95%, 获得了满意的判别精度。结果表明 DOSC-SPA 组合处理是一种非常有效的波长选择方法, 所建直接线性方程对水稻冠层叶瘟诊断非常有效, 为后续进行水稻大田叶瘟病情监测、喷药处理及相应病情监测仪器的开发提供了方法和依据。

参 考 文 献

- 1 Sun Shuyuan, Jin Minzhong, Zhang Zhiming. Rice Blast and Its Prevention[M]. Shanghai: Shanghai Scientific and Technical Publishers, 1986
孙漱沅, 金敏忠, 张志明. 水稻稻瘟病及其防治[M]. 上海: 上海科学技术出版社, 1986
- 2 Yan Yanlu, Zhao Longlian, Han Donghai *et al.*. The Foundation and Application of Near Infrared Spectroscopy Analysis[M]. Beijing: China Light Industry Press, 2005
严衍禄, 赵龙莲, 韩东海等. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005
- 3 F. Liu, F. Zhang, Z. L. Jin *et al.*. Determination of acetolactate synthase activity and protein content of oilseed rape (*Brassica napus* L.) leaves using visible/near infrared spectroscopy[J]. *Analytica Chimica Acta*, 2008, **629**: 56~65
- 4 Liu Fei, He Yong, Wang Li. Methods for the prediction of sugar content of rice wine using visible-near infrared spectroscopy[J]. *Acta Optica Sinica*, 2007, **27**(11): 2054~2058
刘 飞, 何 勇, 王 莉. 黄酒糖度预测的可见-近红外光谱方法研究[J]. 光学学报, 2007, **27**(11): 2054~2058
- 5 Feng Fengqin, Zhang Hui, Wang Li *et al.*. Rapid detection of the quality indicators in glycerol monolaurate products using Fourier transform infrared spectroscopy[J]. *Acta Optica Sinica*, 2008, **28**(2): 326~330
冯凤琴, 张 辉, 王 莉等. 应用红外光谱技术快速检测月桂酸单甘油酯的品质指标[J]. 光学学报, 2008, **28**(2): 326~330
- 6 Shi Youming, Liu Gang, Liu Jianhong *et al.*. Identification of auricularia auricula from different regions by Fourier transform infrared spectroscopy[J]. *Acta Optica Sinica*, 2007, **27**(1): 129~132
时有明, 刘 刚, 刘剑虹等. 不同产地黑木耳的傅里叶变换红外光谱鉴别[J]. 光学学报, 2007, **27**(1): 129~132
- 7 Li Xiaoli, Tang Yueming, He Yong *et al.*. Discrimination of varieties of paddy based on Vis/NIR spectroscopy combined with chemometrics[J]. *Spectroscopy and Spectral Analysis*, 2008, **28**(3): 578~581
李晓丽, 唐月明, 何 勇等. 基于可见/近红外光谱的水稻品种快速鉴别研究[J]. 光谱学与光谱分析, 2008, **28**(3): 578~581
- 8 Wang Fumin, Huang Jingfeng, Xu Junfeng *et al.*. Wavebands selection for rice information extraction based on spectral bands inter-correlation[J]. *Spectroscopy and Spectral Analysis*, 2008, **28**(5): 1098~1101
王福民, 黄敬峰, 徐俊锋等. 基于光谱波段自相关的水稻信息提取波段选择[J]. 光谱学与光谱分析, 2008, **28**(5): 1098~1101
- 9 Yang Xiaohua, Huang Jingfeng, Wang Xiuzhen *et al.*. The estimation model of rice leaf area index using hyperspectral data based on support vector machine[J]. *Spectroscopy and Spectral Analysis*, 2008, **28**(8): 1837~1841
杨晓华, 黄敬峰, 王秀珍等. 基于支持向量机的水稻叶面积指数高光谱估算模型研究[J]. 光谱学与光谱分析, 2008, **28**(8): 1837~1841
- 10 Shao Yongni, Cao Fang, He Yong. Discrimination years of rough rice by using visible/near infrared spectroscopy based on independent component analysis and BP neural network[J]. *J. Infrared Millim. Waves*, 2007, **26**(6): 433~436
邵咏妮, 曹 芳, 何 勇. 基于独立组分析法和 BP 神经网络的可见/近红外光谱稻谷年份的鉴别[J]. 红外与毫米波学报, 2007, **26**(6): 433~436
- 11 Y. N. Shao, Y. He, C. Q. Wu. Dose detection of radiated rice by infrared spectroscopy and chemometrics[J]. *J. Agricultural and Food Chemistry*, 2008, **56**: 3960~3965
- 12 Qiu Baijing, Chen Guoping, Cheng Qiwen. Canopy spectral reflectance feature of rice infected with *sogatella furcifera* and insect number inversion[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2008, **39**(9): 92~95
邱白晶, 陈国平, 程麒文. 水稻白背飞虱虫害的冠层光谱特性与虫量反演[J]. 农业机械学报, 2008, **39**(9): 92~95
- 13 Wu Shuwen, Wang Renchao, Chen Xiaobin *et al.*. Effects of rice leaf blast on spectrum reflectance of rice [J]. *J. Shanghai Jiaotong University: Agricultural Science*, 2002, **20**(1): 73~76
吴曙雯, 王人潮, 陈晓斌等. 稻叶瘟对水稻光谱特性的影响研究[J]. 上海交通大学学报: 农业科学版, 2002, **20**(1): 73~76
- 14 Chu Xiaoli, Yuan Hongfu, Lu Wanzhen. Progress and application of spectral data pretreatment and wavelength selection methods in NIR analytical technique[J]. *Progress in Chemistry*, 2004, **16**(4): 528~542
褚小立, 袁洪福, 陆婉珍. 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
- 15 S. Wold, H. Antti, F. Lindgren *et al.*. Orthogonal signal correction of near-infrared spectra [J]. *Chemometrics and Intelligent Laboratory Systems*, 1998, **44**: 175~185
- 16 Westerhuis J A, De Jong S, Smilde A K. Direct orthogonal signal correction[J]. *Chemometrics and Intelligent Laboratory Systems*, 2001, **56**: 13~25
- 17 M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvão *et al.*. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis [J]. *Chemometrics and Intelligent Laboratory Systems*, 2001, **57**: 65~73
- 18 R. K. H. Galvão, M. C. U. Araújo, W. D. Fragoso *et al.*. A variable elimination method to improve the parsimony of MLR models using the successive projections algorithm [J]. *Chemometrics and Intelligent Laboratory Systems*, 2008, **92**: 83~91
- 19 F. Liu, Y. H. Jiang, Y. He. Variable selection in visible/near infrared spectra for linear and nonlinear calibrations: A case study to determine soluble solids content of beer[J]. *Analytica Chimica Acta*, 2009, **635**: 45~52
- 20 P. Geladi, B. R. Kowalski. Partial least-squares regression: a tutorial[J]. *Analytica Chimica Acta*, 1986, **185**: 1~17