第30卷，第7期
2010年7月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 30, No. 7, pp1934-1938
July, 2010

基于组合模拟波段的油菜菌核病早期诊断方法研究

刘 飞[1]，冯 雷[1]，楼兵干[2*]，孙光明[1]，王连平[3]，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. 浙江大学生物技术研究所，浙江 杭州 310029
3. 浙江省农业科学院植物保护与微生物研究所，浙江 杭州 310021

摘 要 应用组合模拟波段建立的线性和非线性判别模型实现了油菜菌核病的早期诊断。采集油菜健康叶片80个，菌核病染病叶片100个，采用预处理算法与连续投影算法(SPA)相结合提取组合模拟波段，分别建立偏最小二乘法(PLS)、多元线性回归(MLR)和最小二乘-支持向量机(LS-SVM)模型。通过比较，最优PLS判别的预处理分别为直接正交信号校正(DOSC)、De-trending和原始反射光谱(Raw)，准确率分别为100%，95.7%和95.7%。应用组合模拟波段的最优线性模型为SPA-MLR(DOSC)和SPA-PLS(DOSC)模型，准确率均为100%，基于DOSC、De-trending和Raw组合模拟波段的SPA-LS-SVM模型的判别准确率均为100%。结果表明，基于组合模拟波段进行油菜菌核病早期诊断是可行的，为油菜菌核病的早期诊断及病害监测仪器的开发提供了方法和依据。

关键词 可见/近红外光谱；油菜菌核病；直接正交信号校正；连续投影算法；最小二乘-支持向量机

中图分类号：O657.3；S435.6 **文献标识码**：A **DOI**：10.3964/j.issn.1000-0593(2010)07-1934-05

引 言

油菜是我国四大油料作物之一，具有适应性强、用途广、经济价值高等特点，占我国油料作物总面积的40%以上，占我国油料作物总产量的30%以上[1]。油菜的生长状况决定了油菜籽的产量和品质。油菜菌核病作为油菜生产中重要的病害之一，尤其在越夏越冬油菜菌核病菌核数量多，天气潮湿多雨，容易造成油菜菌核病的爆发。常年株发病率高达10%～30%，严重的达80%，病株一般减产70%以上，严重影响了油菜的产量和品质。目前，油菜菌核病的诊断多依靠人眼进行判别，容易造成发病早期的不及时预测，从而错过最佳的防治时期。同时，人眼判别的主观性强，需要的时间和精力较多，无法满足现代农业生产和管理的要求。因此，急需一种能够快速、准确进行油菜菌核病检测的方法和技术。

近红外光谱技术作为一种绿色分析技术，因具有快速、准确、无损、无污染等特点，已被广泛应用于农业、食品、化工、医药等行业[2, 3]。在油菜生长信息的检测中，近红外光谱技术已有应用，主要集中于油菜营养信息中氮素含量的检测[4-6]、油菜叶片氨基酸含量的检测[7]、除草剂胁迫下油菜叶片中乙酰乳酸合成酶和蛋白含量[8, 9]的检测等方面的研究。应用可见/近红外光谱技术进行油菜菌核病检测的研究还少有报道。本研究通过获取油菜菌核病叶片和健康叶片的光谱数据，在全面比较各种光谱数据预处理方法的基础上，采用连续投影算法提取有效波长作为油菜菌核病快速检测的组合模拟波段，结合最小二乘-支持向量机方法和线性判别方法，建立了油菜菌核病的早期诊断模型。

1 材料与方法

1.1 仪器设备

光谱采集使用美国ASD(analytical spectral device, Boulder, USA)公司的Handheld FieldSpec光谱仪。该光谱仪测定的光谱范围为325～1 075 nm，探头视场角为20°，光谱扫描次数设定为30次。试验采用漫反射模式，光源采用14.5 V卤素灯，光源入射角度为45°，光谱仪探头与油菜叶片所在平面垂直距离大约为150 mm。光源、叶片和光谱仪探头保持在同一条直线上。光谱数据分析软件采用ASD ViewSpecPro，Unscrambler® 9.8(CAMO AS, Oslo, Nor-

收稿日期：2009-11-29，**修订日期**：2010-02-26

基金项目：国家高技术研究发展计划(863计划)项目(2007AA10Z210)，国家自然科学基金项目(60605011，60802038)，浙江省重大科技专项重点农业项目(2009C12002)，浙江省研究生创新科研项目(YK2008014)和中央高校基本科研业务费专项资金项目资助

作者简介：刘 飞，1983年生，浙江大学生物系统工程与食品科学学院博士研究生 *通讯联系人 e-mail：bglou@zju.edu.cn

way)，Matlab® 7.0(Math Works，Natick，USA)。

1.2　样本准备

供试材料为双低品系甘蓝型油菜浙双 758(*Brassica napus* L. cv. ZS758)，油菜采用盆栽方式种植于人工气候室。供试菌为油菜菌核病菌 *Sclerotinia sclertiorum*(浙江省农科院植物保护与微生物研究所提供)。菌核病菌在培养基上培养，接种时，将长满菌丝的菌丝块放在油菜植株的叶片上，人工气候室温度设定为 20 ℃，相对湿度设定 85%。实验分为健康对照和接菌两组，在相同条件下同时进行培养。每隔 6 h 观察一次。接菌 18 h 小部分油菜叶片出现轻微的发病症状，为保证染病样本的准确性，待接菌 24 h 开始进行光谱数据采集。本实验中，健康样本为完全健康的油菜叶片，染病样本包括发病症状非常微弱以及稍微明显一点的叶片样本，且通过后续的观察，确定有微弱发病症状的样本后期的症状变得明显。本次实验共采集染病叶片样本 100 个，健康叶片样本 80 个。随机选择 60 个染病样本和 50 个健康样本组成建模集样本(共 110 个样本)，剩余的 40 个染病样本和 30 个健康样本组成预测集样本(共 70 个样本)。

1.3　数据预处理及波长选择

本文比较了常用的数据预处理算法对光谱检测性能的影响，包括 Savitzky-Golay 平滑(SG)、变量标准化(SNV)、多元散射校正(MSC)、一阶及二阶导数处理(1-Der and 2-Der)、去趋势处理(De-trending)及直接正交信号校正(DOSC)处理等[10]。上述预处理方法中，DOSC 是在正交信号校正(OSC)的基础上改进的预处理算法，DOSC 在对光谱数据矩阵进行处理的同时考虑了样本类别矩阵的信息[11]，而其他预处理方法只对光谱数据进行预处理。因此，通过上述预处理方法的比较，有利于获得提高油菜菌核病早期诊断准确率的最优预处理方法。

为减少模型的输入变量，提高模型计算速度，采用连续投影算法(SPA)进行了油菜菌核病早期诊断有效波长的选取。连续投影算法通过正交投影变换，比较投影的大小，选取含有最低冗余度和最小共线性的有效波长。选取的有效波长作为模型的输入变量，在不影响模型总体预测性能的情况下有效地简化了模型，提高模型的预测速度。因预处理后的波长并非原始光谱仪采集的反射率数据，所以结合连续投影算法选取的有效波长为一种组合模拟波段。本文将直接正交信号校正与连续投影算法联用，提出一种新的 DOSC-SPA 组合模拟波段选取方法，可大大减少光谱数据的冗余信息，有效解决光谱信息的共线性问题。通过选取的组合模拟波段可建立油菜菌核病早期诊断模型。

1.4　建模方法

多元线性回归(MLR)和偏最小二乘法(PLS)是应用最为广泛的光谱建模分析方法[12, 13]。MLR 直接利用输入的光谱波长建立模型，可以有效地反映输入变量与油菜菌核病诊断的相关性。PLS 通过提取输入光谱波长的特征变量(LV)，建立光谱与油菜菌核病诊断的相关关系，通过交互验证得到油菜菌核病检测的判别模型。建模过程中，将油菜的健康样本和染病样本分别赋予虚拟变量 0 和 1，设定判别误差绝对值的阈值为 0.5，即预测值小于 0.5 的样本为健康样本，大于 0.5 的样本为染病样本。同时，为比较模型预测性能的稳定性，进行了阈值为 0.2 时的准确率对比分析。通过光谱预处理结合组合模拟波段，可分别建立油菜菌核病早期诊断的 MLR 和 PLS 模型。

因为 MLR 和 PLS 均为线 性建模方法，为了充分提取光谱数据中潜在非线性有效信息，建立了油菜菌核病早期诊断的最小二乘-支持向量机(LS-SVM)模型。LS-SVM 是一种新型的统计学习方法[14-16]，它通过支持向量的线性组合，有效的提高了模型的性能，并能通过径向基(RBF)核函数充分利用输入变量的线性和非线性信息，实现油菜菌核病的早期诊断。评判模型性能的指标为预测集样本的判别准确率，准确率越高，说明模型性能越好。

2　实验结果与分析

2.1　油菜样本的光谱特性

油菜叶片样本的原始可见/近红外反射光谱如图 1 所示。图中横坐标为波长，范围为 500～1 000 nm，纵坐标为反射率值。从图 1 可知，健康和染病的油菜叶片样本在波长 550 nm 附近和 750～1 000 nm 两个波段范围存在比较明显的反射率差异，这是因为染病油菜样本的发病状况不同，病斑的大小不同，造成反射率的差异。但直接从光谱图难以准确区分油菜样本是否发病，需要进一步结合化学计量学方法，实现油菜菌核病的早期诊断。

2.2　PLS 判别模型的建立

首先，将全部油菜样本的原始光谱数据作为输入变量，通过留一交互验证，建立油菜菌核病早期诊断的 PLS 判别模型。提取前三个特征变量，得到油菜健康和染病样本的散点分布图，如图 2 所示。观察可知，油菜健康样本和染病样本大体分为两类，中间有少部分样本分布空间相重叠。健康样本的分布相对更加集中，而菌核病染病样本分布相对分散，这是因为染病样本因发病状况不同，病斑的大小差别较大，而且部分轻微发病样本用肉眼难以进行判别，这类样本的光谱特性与健康样本区分不明显。为准确进行油菜菌核病早期诊断的研究，需要进一步结合化学计量学方法建立判别模型。

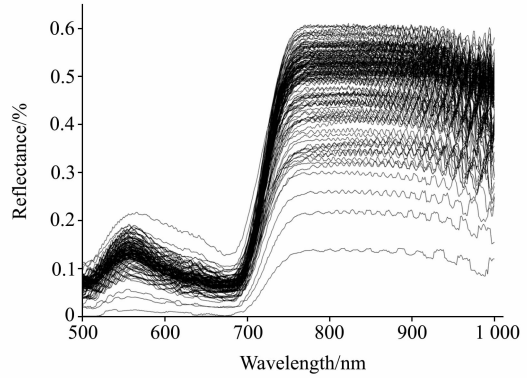

Fig. 1　Original visible/near infrared reflectance spectra of oilseed rape

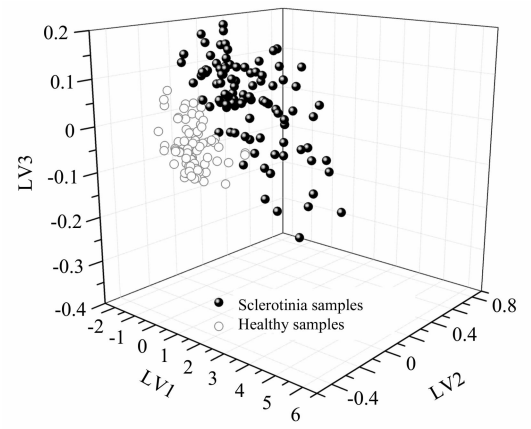

Fig. 2 Scatter plots by LV1×LV2×LV3 of oilseed rape samples

将上述预处理后的光谱数据和原始光谱数据作为 PLS 模型的输入变量，建立油菜菌核病早期诊断的判别模型。建模中，采用留一交互验证法保证模型的稳定性和可靠性，用预测集样本对模型的性能进行验证。通过计算，得到上述各种预处理下 PLS 判别模型的判别准确率，结果如表 1 所示。当判别阈值设定为 0.5 时，只有二阶导数预处理下出现 1 个误判样本，其预测集判别准确率为 98.6%，其余预处理下 PLS 判别模型的准确率均为 100%。在实际的判别分析中，为区别模型预测性能的稳定性，确保模型的预测精度，需要设定一个更为严格的阈值条件，本文设定判别误差绝对值的阈值为 0.2。由表 1 可知，经过 DOSC 预处理后的 PLS 判别模型的判别准确率最高为 100%，全部样本判别准确。其次为原始光谱数据(Raw)和 De-trending 预处理后所建模型，判别准确率均为 95.7%。

Table 1 Prediction results of oilseed rape *sclerotinia* by PLS with different preprocessing

Preprocessing	Validation(n=70)	
	0.5/%	0.2/%
Raw	100	95.7
Savitzky-Golay	100	87.1
SNV	100	94.3
MSC	100	94.3
1-Der	100	85.7
2-Der	98.6	95.7
De-trending	100	95.7
DOSC	100	100

2.3 基于组合模拟波段的判别模型

考虑到采用全波段 501 个(500～1 000 nm)变量建模，所需时间较长，并且这些变量中含有大量的共线性和冗余信息，因此，本文采用 SPA 提取有效波长进行建模分析。将 3 种最优预处理(DOSC，De-trending 和 Raw)的光谱数据进行 SPA 运算，选取对油菜菌核病诊断的有效波长。SPA 计算中，采用留一交互验证法，设定最大选定波长数为 30，当交互验证均方根误差(RMSECV)最小或趋于稳定时所对应的变量个数即为选定的波长个数。3 种预处理选取的组合模拟波段如表 2 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 DOSC，De-trending 和 Raw 预处理下，最为重要的有效波长分别为 835，911 和 951 nm。

将上述组合模拟波段作为 MLR，PLS 和 LS-SVM 模型的输入变量，分别建立油菜菌核病早期诊断的判别模型。在 MLR 和 PLS 两种线性模型中，MLR 能够直接反映所选波长的预测性能和有效性，PLS 通过对所选波段进行特征变量提取，经过留一交互验证进一步建立油菜菌核病诊断模型。用预测集样本对所建模型进行验证，判别结果如表 3 所示。在阈值为 0.2 时，应用 Raw 和 De-trending 所选组合模拟波段的预测效果均不如全波段建模(Raw 和 De-trending)的预测效果好。考虑到所用波长的数量，以及应用 SPA-PLS(De-trending)模型 90.0%的判别准确率，应用组合模拟波段进行油菜菌核病早期诊断是可行的。应用 DOSC 处理后的组合模

Table 2 Selected combinational-stimulated bands by SPA

Preprocessing	No.	Selected bands/nm
DOSC	14	835，532，943，643，529，624，690，660，672，635，692，695，521，503
De-trending	8	911，854，910，913，915，919，899，957
Raw	12	951，989，946，979，944，992，997，1000，506，995，976，973

Table 3 Prediction results of oilseed rape *sclerotinia* by combinational-stimulated bands

Models	Preprocessing	No. of wavelengths/LVs	Validation(n=70)	
			0.5/%	0.2/%
SPA-MLR	Raw	12/—	100	74.3
SPA-PLS	Raw	12/4	98.6	68.6
SPA-MLR	De-trending	8/—	100	88.6
SPA-PLS	De-trending	8/4	100	90.0
SPA-MLR	DOSC	14/—	100	100
SPA-PLS	DOSC	14/1	100	100
SPA-LS-SVM	Raw	12/—	100	100
SPA-LS-SVM	De-trending	8/—	100	100
SPA-LS-SVM	DOSC	14/—	100	100

拟波段所建的 SPA-MLR 和 SPA-PLS 模型的判别准确率均为 100%，获得了很好的判别准确率。

为充分利用所选组合模拟波段的线性和非线性信息，应用上述组合模拟波段作为输入变量分别建立 LS-SVM 模型。建模中，应用径向基(RBF)函数作为核函数，采用网格搜索法(Grid-search)寻找模型的两个参数(γ 和 σ^2)，设定模型参数的取值范围为 $10^{-3}\sim10^{6}$。通过计算，获得 Raw，De-trending 和 DOSC 处理下 SPA-LS-SVM 模型的参数(γ，σ^2)的最优组合分别为(3 286，1.44)，(3.87，11.98)和(0.20，0.03)。所建 SPA-LS-SVM 模型的判别结果如表 3 所示。三个 SPA-LS-SVM 模型的判别准确率均达到了 100%，结果优于所建 SPA-MLR 和 SPA-PLS 线性模型，说明应用组合模拟波段建立 LS-SVM 进行油菜菌核病早期诊断是可行的，能获得满意的预测精度。DOSC-SPA 联用所提取的组合模拟波段，为油菜菌核病早期诊断提供了新的方法，有利于油菜菌核病的早期防治和监测，也为油菜病害监测仪器的开发奠定了基础。

3　结　论

应用可见/近红外光谱技术结合组合模拟波段进行油菜菌核病的早期诊断是可行的。将不同预处理算法与连续投影算法联用选取组合模拟波段，建立了油菜菌核病早期诊断的线性和非线性判别模型。应用直接正交信号校正-连续投影算法(DOSC-SPA)所得组合模拟波段所建的 MLR，PLS 和 LS-SVM 模型的判别准确率均达到了 100%，说明 DOSC-SPA 联用是一种非常有效的组合模拟波段提取方法，为后续油菜菌核病的防治以及油菜病害监测仪器的开发提供了方法和依据。

参　考　文　献

[1] ZHANG Guo-ping, ZHOU Wei-jun(张国平，周伟军). Cultivation of Crops(作物栽培学). Hangzhou: Zhejiang University Press(杭州：浙江大学出版社), 2001.

[2] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al(严衍禄，赵龙莲，韩东海，等). The Foundation and Application of Near Infrared Spectroscopy Analysis(近红外光谱分析基础与应用). Beijing: China Light Industry Press(北京：中国轻工业出版社), 2005.

[3] Liu F, He Y, Sun G M. Journal of Agricultural and Food Chemistry, 2009, 57: 4520.

[4] QIU Zheng-jun, SONG Hai-yan, HE Yong, et al(裘正军，宋海燕，何　勇，等). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2007, 23(7): 150.

[5] Müller K, Bøttcher U, Meyer-Schatz F, et al. Biosystems Engineering, 2008, 101: 172.

[6] ZHANG Xiao-dong, MAO Han-ping(张晓东，毛罕平). Transactions of the Chinese Society for Agricultural Machinary(农业机械学报), 2009, 40(2): 164.

[7] LIU Fei, ZHANG Fan, FANG Hui, et al(刘　飞，张　帆，方　慧，等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2009, 29(11): 3079.

[8] Liu F, Zhang F, Jin Z L, et al. Analytica Chimica Acta, 2008, 629: 56.

[9] LIU Fei, FANG Hui, ZHANG Fan, et al(刘　飞，方　慧，张　帆，等). Chinese Journal of Analytical Chemistry(分析化学), 2009, 37(1): 67.

[10] CHU Xiao-li, YUAN Hong-fu, LU Wan-zhen (褚小立，袁洪福，陆婉珍). Progress in Chemistry(化学进展), 2004, 16(4): 528.

[11] Westerhuis J A, De Jong S, Smilde A K. Chemometrics and Intelligent Laboratory Systems, 2001, 56: 13.

[12] Naes T, Mevik B H. Journal of Chemometrics, 2001, 15: 413.

[13] Cen H Y, He Y, Huang M. Journal of Agricultural and Food Chemistry, 2006, 54: 7437.

[14] Vapnik V N. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995.

[15] Suykens J A K, Vanderwalle J. Neural Processing Letters, 1999, 9: 293.

[16] Liu F, He Y, Wang L. Analytica Chimica Acta, 2008, 610: 196.

Study on the Early Detection of *Sclerotinia* of *Brassica Napus* Based on Combinational-Stimulated Bands

LIU Fei[1], FENG Lei[1], LOU Bing-gan[2*], SUN Guang-ming[1], WANG Lian-ping[3], HE Yong[1]

1. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China
2. Institute of Biotechnology, Zhejiang University, Hangzhou 310029, China
3. Institute of Plant Protection and Micrology, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, China

Abstract The combinational-stimulated bands were used to develop linear and nonlinear calibrations for the early detection of *sclerotinia* of oilseed rape (*Brassica napus* L.). Eighty healthy and 100 Sclerotinia leaf samples were scanned, and different pre-processing methods combined with successive projections algorithm (SPA) were applied to develop partial least squares (PLS) discriminant models, multiple linear regression (MLR) and least squares-support vector machine (LS-SVM) models. The results indicated that the optimal full-spectrum PLS model was achieved by direct orthogonal signal correction (DOSC), then De-trending and Raw spectra with correct recognition ratio of 100%, 95.7% and 95.7%, respectively. When using combinational-stimulated bands, the optimal linear models were SPA-MLR (DOSC) and SPA-PLS (DOSC) with correct recognition ratio of 100%. All SPA-LS-SVM models using DOSC, De-trending and Raw spectra achieved perfect results with recognition of 100%. The overall results demonstrated that it was feasible to use combinational-stimulated bands for the early detection of *Sclerotinia* of oilseed rape, and DOSC-SPA was a powerful way for informative wavelength selection. This method supplied a new approach to the early detection and portable monitoring instrument of *sclerotinia*.

Keywords Visible/near infrared spectroscopy; *Sclerotinia* of oilseed rape; Direct orthogonal signal correction; Successive projections algorithm; Least squares-support vector machine

(Received Nov. 29, 2009; accepted Feb. 26, 2010)

* Corresponding author