

# 基于近红外光谱技术的油菜叶片丙二醛含量快速检测方法研究

孔汶汶, 刘飞, 邹强, 方慧\*, 何勇\*

浙江大学生物系统工程与食品科学学院, 浙江杭州 310029

**摘要** 应用近红外光谱技术实现了油菜叶片中丙二醛(MDA)含量的快速无损检测。对90个油菜叶片样本进行光谱扫描, 用60个样本建模, 30个样本验证。经过平滑、变量标准化、一阶及二阶求导、去趋势等预处理后, 建立了MDA预测的偏最小二乘法(PLS)模型。将PLS提取的有效特征变量(LV)和连续投影算法(SPA)提取的有效波长作为最小二乘-支持向量机(LS-SVM)的输入变量, 分别建立了LV-LS-SVM和SPA-LS-SVM模型。以预测集的预测相关系数( $r$ ), 预测标准偏差(RMSEP)作为模型评价指标。结果表明, 油菜叶片中MDA含量预测的最优模型为LV-LS-SVM模型, LV-LS-SVM在去趋势处理后的预测效果为 $r=0.9999$ , RMSEP=0.5302; 在二阶求导处理后的预测效果为 $r=0.9999$ , RMSEP=0.3957。说明应用光谱技术检测油菜叶片中MDA的含量是可行的, 并能获得满意的预测精度, 为油菜大田生长状况的动态连续监测提供了新的方法。

**关键词** 近红外光谱; 油菜; 丙二醛; 偏最小二乘法; 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(2011)04-0988-04

## 引言

油菜(*Brassica napus* L.)是我国主要的油料作物之一, 在我国农业生产中占有重要地位。油菜的生长状况对油菜籽的品质和产量有重要影响。叶片衰老是植物生长发育不可分割的一部分, 衰老引起的叶片同化功能减退及一些作物的早衰现象极大地影响并限制作物产量潜力的发挥, 因此研究油菜叶片衰老和抗逆性对油菜生产理论研究和应用等方面都具有非常重要的意义。植物体内丙二醛(malondialdehyde, MDA)是植物器官在衰老时或逆境条件下发生膜脂过氧化的产物之一, 它的含量能够代表膜脂过氧化的程度, 因此在植物衰老和抗性生理研究中, 丙二醛含量是一个非常重要的指标。传统的测定丙二醛含量的方法为硫代巴比妥酸(TBA)比色法, 该方法需要在实验室进行破坏性的实验才能得到结果, 实验过程费时、费力, 且无法实现油菜生长过程中丙二醛的连续监测。因此, 研究开发一种简便、无损的丙二醛含量快速检测方法是十分必要的, 对实现油菜生长状况的大田检测与连续监测, 油菜生产过程的精细化管理和作业, 提高油菜籽的产量与品质具有重要意义。

近红外光谱技术(NIRS)是一种快速、准确、无损的定性和定量分析技术, 它已被广泛应用于农业、食品、石油工业、医药等行业<sup>[1-3]</sup>。在油菜叶片生命信息检测方面, 光谱技术已应用于油菜叶片及冠层的氮含量、乙酰乳酸合成酶含量、蛋白质含量、氨基酸含量等快速检测<sup>[4-6]</sup>。但目前应用近红外光谱技术对油菜叶片MDA含量检测的研究还少有报道。本研究的目的在于通过近红外光谱分析, 比较不同光谱预处理方法, 并结合化学计量学方法, 建立油菜叶片MDA含量的快速无损检测模型。

## 1 材料与方法

### 1.1 仪器设备

采用美国ASD(Analytical Spectral Device)公司生产的ASD FieldSpec FR2500光谱仪, 光谱测定范围325~2500 nm, 光谱分辨率在1000~2500 nm区间为2 nm, 采集视场角为25°, 实验采用漫反射模式。分析软件为ASD View Spec Pro, Unscrambler V9.8(CAMO AS, Oslo, Norway), MATLAB V7.0(The Math Works, Natick, USA)。

### 1.2 样本光谱扫描及MDA测定

收稿日期: 2010-06-28, 修订日期: 2010-10-02

基金项目: 浙江省自然科学基金重点项目(Z3090295), 国家自然科学基金项目(31071332), 浙江省科技厅项目(2009R50001), 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资助

作者简介: 孔汶汶, 1987年生, 浙江大学生物系统工程与食品科学学院博士生 e-mail: kongwenwen1987@163.com

\* 通讯联系人 e-mail: yhe@zju.edu.cn; newxfl@yahoo.com.cn

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

油菜叶片样本来自浙江大学试验农场,油菜生长过程采用一般田间管理方式。在油菜生长苗期,共采集样本90个,从中随机选取60个样本作为建模集样本( $n=60$ ),剩余30个样本作为预测集样本( $n=30$ )。用ASD光谱仪对油菜叶片所有样本进行光谱扫描,每个样本保留三条光谱曲线,将其取平均得到该样本的光谱曲线,之后利用获取的光谱数据结合化学计量学方法,实现油菜叶片中MDA含量的快速检测。油菜叶片的MDA含量采用硫代巴比妥酸比色法测定<sup>[7]</sup>。MDA含量单位为 $\text{nmol} \cdot \text{g}^{-1}$ 。

### 1.3 光谱模型

#### 1.3.1 光谱数据预处理

为了消除来自高频随机噪声、基线漂移、样本不均匀、光散射等不利因素影响,优化建模效果,需要对光谱进行预处理。本文比较了Savitzky-Golay平滑(SG)、变量标准化(SNV)、多元散射校正(MSC)、一阶及二阶(1-Der and 2-Der)求导和去趋势(Detrending)六种预处理方法的优化效果。

为了减少模型的输入变量,提高模型的计算速度,本文进行了特征向量和有效波长的提取。采用偏最小二乘法(PLS)提取光谱数据的特征变量(LV),通过累计贡献率确定特征向量的个数,实现光谱数据的压缩和降维。为了比较分析,本文还采用连续投影算法(SPA)提取光谱信息中的有效波长(EW),该方法能在光谱信息矩阵中寻找含有最低限度冗余信息的变量组,使变量之间的共线性达到最小,减少建模变量个数,从而提高了建模的效率<sup>[8]</sup>。

#### 1.3.2 建模方法

偏最小二乘法(PLS)是一种近红外光谱技术中广泛采用的线性建模方法<sup>[9]</sup>,该方法通过对光谱数据进行特征向量的提取,实现光谱数据的压缩和降维,通过交互验证方法保证建模效果的准确性和适应性。该模型采用相关系数( $r$ )和预测集均方根误差(RM SEP)作为评价指标,相关系数越接近1, RM SEP越小,说明模型的预测性能越好。

最小二乘-支持向量机(LS-SVM)是一种新型的统计学习方法<sup>[10-12]</sup>,它能够很好地解决小样本、非线性、高维数等实际问题,充分挖掘光谱数据的线性和非线性有效信息,已经在光谱分析领域得到很好的应用。本文应用PLS提取的特征向量和SPA提取的特征波长作为LS-SVM模型的输入变量。采用径向基函数(RBF)作为核函数,RBF核函数作为非线性函数能够简化训练过程中的计算。LS-SVM模型的两个模型参数 $\gamma$ 和 $\sigma^2$ 通过带有交互验证的网格搜索(Grid search)法

来确定,最终建立油菜叶片丙二醛含量预测的非线性模型。

## 2 结果与分析

### 2.1 油菜叶片反射光谱图

图1为90个油菜叶片样本的原始近红外反射光谱图。油菜叶片的光谱曲线趋势相似,但在不同的波段上升和下降的趋势不同。油菜叶片反射率的不同,说明其内部化学成分等存在差异。油菜叶片中MDA的统计值如表1所示。建模和预测样本的MDA均包含了较大的范围,这有益于建立的模型的稳定性和代表性。

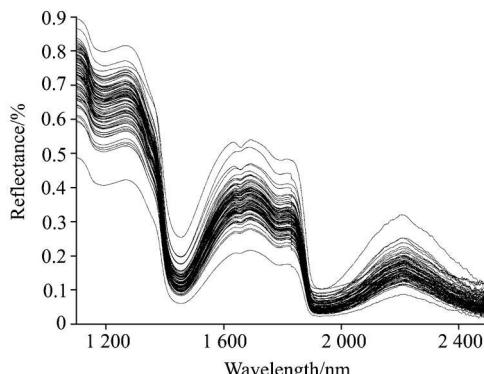


Fig 1 Original near infrared reflectance spectra in oilseed rape leaves

Table 1 Statistical values of MDA in oilseed rape leaves( $\text{nmol} \cdot \text{g}^{-1}$ )

Dataset	No	Range	Mean	Standard deviation
Calibration	60	6.97~116.52	40.45	30.4060
Prediction	30	8.00~125.42	43.10	33.3035
All	90	6.97~125.42	41.33	31.2389

### 2.2 PLS模型

油菜叶片的光谱数据经过不同预处理后作为PLS模型的输入变量,通过特征向量的提取和留一交互验证,建立油菜叶片MDA含量的预测PLS模型。基于不同预处理方法的PLS模型对油菜叶片中MDA含量的预测结果如表2所示。由表可知,最优PLS模型所对应的预处理方法为二阶(2-Der)求导,其次为去趋势(Detrending)处理。二阶求导对应的预测结果为 $r=0.9103$ 和 $\text{RM SEP}=13.5941$ 。

Table 2 Prediction results of MDA by PLS with different pretreatment

Pretreatment	LV	r	RM SEP	Bias	Slope	Offset
Raw	17	0.7825	22.5607	-2.7379	0.8337	4.4317
SG	18	0.7407	25.3053	-3.3989	0.8208	4.3242
SNV	8	0.4848	29.6515	-3.2671	0.3381	25.2613
MSC	8	0.4831	29.5525	-3.0341	0.3292	25.8785
1-Der	15	0.8308	19.8227	2.3505	0.8789	7.5695
2-Der	10	0.9103	13.5941	-0.7171	0.8479	5.8383
Detrending	18	0.8537	17.5733	-1.6640	0.8307	5.6314

### 2.3 SPA 提取有效波长

应用连续投影算法(SPA)对上述两种最优预处理(二阶求导和去趋势处理)后的光谱数据进行有效波长提取。为便于比较分析, 对原始反射光谱数据也采用 SPA 提取有效波。

SPA 提取的有效波长的最大数设定为 30, 计算过程采用留一交叉验证方法, 不同预处理下提取的有效波长如表 3 所示, 表中提取的有效波长按其重要性排列, 排序越靠前, 说明该预处理下该波长对油菜叶片 MDA 含量预测越重要。

Table 3 Selected EWS by SPA with different pretreatments

Pretreatment	No	Selected EWS/ nm
Raw	5	2 463, 2 482, 2 344, 2 439, 1 837
2 Der	17	1 917, 1 877, 2 222, 2 082, 1 272, 1 932, 2 161, 1 919, 1 769, 1 805, 1 605, 2 039, 2 207, 1 592, 1 231, 1 310, 2 197
Detrending	15	1 494, 2 433, 2 097, 2 496, 2 377, 2 437, 2 393, 2 419, 2 406, 2 483, 2 500, 2 400, 2 385, 1 873, 2 429

### 2.4 SPA-PLS, SPA-LS SVM 和 LV-LS SVM 模型

采用不同预处理方法下 SPA 提取的有效波长作为输入变量, 建立油菜叶片 MDA 含量预测的 SPA-PLS 和 SPA-LS-SVM 模型。为了比较分析, 将 PLS 提取的特征向量作为 LS-SVM 模型的输入变量, 建立油菜叶片 MDA 含量预测的 LV-LS-SVM 模型。

在 SPA-PLS 模型中, 计算过程采用留一交叉验证方法, 得到 MDA 含量预测的 SPA-PLS 模型。该模型的预测结果如表 4 所示。在 SPA-LS-SVM 和 LV-LS-SVM 中, 核函数为径向基函数(RBF), 模型参数( $\gamma$ ,  $\sigma^2$ )的取值范围设定为  $10^{-3} \sim 10^8$ , 预测结果如表 4 所示。通过比较分析可知, 去趋势(Detrending)和二阶(2 Der)求导两种预处理方法结合 LV-LS-SVM 建模方法均获得了满意的预测精度, 且其预测效果远优于 SPA-PLS 和 SPA-LS-SVM 模型。LV-LS-SVM 在 Detrending 处理后预测效果为  $r = 0.9999$ ,  $RMSEP = 0.5302$ , 此时模型参数的最优组合为  $\gamma = 4.2 \times 10^7$ ,  $\sigma^2 = 2.4 \times 10^6$ 。LV-LS-SVM 在 2-Der 求导处理后预测效果为  $r = 0.9999$ ,  $RMSEP = 0.3957$ , 此时模型参数的最优组合为  $\gamma = 5.8 \times 10^6$ ,  $\sigma^2 = 5.9 \times 10^3$ , 图 2 为其预测样本集的真实值与预测值的散点分布图。LV-LS-SVM 模型的预测效果优于 SPA-PLS 和 SPA-LS-SVM 模型, 其原因在于 LV-LS-SVM 模型的输入变量为对所有光谱数据进行压缩降维提取的特征向量, 提取

的 LV 能代表光谱数据 94% 以上的有效信息, 而 SPA 提取的特征波长仅为一些特征波长点的信息, SPA 在提取过程中, 可能造成光谱中部分重要信息的丢失。与表 2 中全波段的 PLS 模型相比, LV-LS-SVM 模型的预测效果更优, 说明 LS-SVM 模型能充分利用光谱数据中的线性和非线性信息。结果表明, 应用近红外光谱技术结合 LV-LS-SVM 模型可以实现油菜叶片中 MDA 含量的快速检测, 并获得了满意的预测精度。

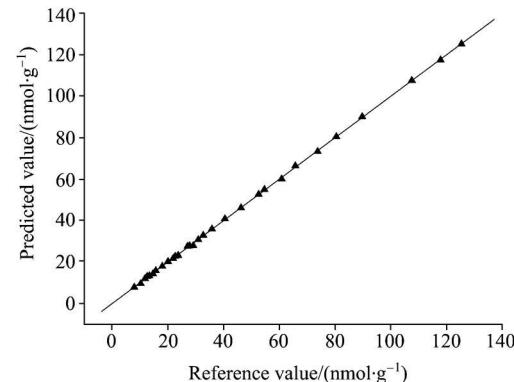


Fig 2 Predicted vs reference values of MDA in oilseed rape by LS-SVM model

Table 4 Prediction results for different models

Model	Pretreatment	LV/EW/(\gamma, σ²)	r	RMSEP	Bias	Slope	Offset
SPA-PLS	Raw	3/5/-	0.1832	37.7405	-3.3038	0.1422	33.6677
	2-Der	10/15/-	0.6189	29.7016	-6.3482	0.6368	9.3044
	Detrending	7/17/-	0.5987	34.6627	-2.9340	0.7694	7.0054
SPA-LS-SVM	Raw	-/5/(7.9×10³, 4.2×10³)	0.1817	35.6501	-0.1811	0.1180	37.8300
	2-Der	-/15/(4.6×10⁵, 2.7×10⁶)	0.6185	28.9720	-6.2490	0.6047	10.7891
	Detrending	-/17/(8.7×10⁶, 3.8×10⁶)	0.5865	33.9668	-2.5969	0.7220	9.3480
LV-LS-SVM	Raw	20/-/(2.2×10⁶, 2.6×10⁴)	0.9995	0.9896	-0.1783	1.0049	-0.0321
	2-Der	17/-/(5.8×10⁶, 5.9×10³)	0.9999	0.3957	-0.0340	1.0002	0.0233
	Detrending	21/-/(4.2×10⁷, 2.4×10⁶)	0.9999	0.5302	-0.0713	1.0083	-0.2846

### 3 结语

应用近红外光谱技术结合最小二乘-支持向量机方法实现了油菜叶片中 MDA 含量的快速检测。通过对不同预处理方法、不同化学计量学方法(PLS, LS-SVM)的比较, 并结合特征向量和特征波长提取方法, 得到最优预处理方法为二阶求导, 最优的预测模型为 LV-LS-SVM 模型, 预测结果为  $r$

= 0.999 9 和 RMSEP= 0.395 7, 获得了满意的预测精度。本文为研究油菜生长过程中叶片衰老及逆境胁迫表征提供了快

速、准确的检测方法, 也为实现油菜其他营养和生理指标的检测, 以及油菜大田生长状况活体监测提供了新的方法。

## References

- [1] Kulmyrzaev A A, Dufour E. Food Bioprocess and Technology, 2010, 3(2): 247.
- [2] XIAO Yan-chun, DOU Sen(肖彦春, 邓森). Chinese Journal of Analytical Chemistry(分析化学), 2007, 35(11): 1596.
- [3] LIU Xian, HAN Lu-jia, YANG Zeng-ling, et al(刘贤, 韩鲁佳, 杨增玲, 等). Chinese Journal of Analytical Chemistry(分析化学), 2007, 35(9): 1285.
- [4] Chen Xiaojing, Wu Di, He Yong, et al. Analytica Chimica Acta, 2009, 638(1): 16.
- [5] Liu F, Zhang F, Jin Z L, et al. Analytica Chimica Acta, 2008, 629: 56.
- [6] 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.
- [7] ZOU Qi(邹琦). Plant Physiology Lab Guide(植物生理学实验指导). Beijing: China Agriculture Press(北京: 中国农业出版社), 2000.
- [8] Galvao R K H, Araújo M C U, Fragoso W D, et al. Chemometrics Intell. Lab. Syst., 2008, 92: 83.
- [9] Geladi P, Kowalski B R. Anal. Chim. Acta, 1986, 185: 1.
- [10] Liu F, He Y, Wang L. Anal. Chim. Acta, 2008, 610: 196.
- [11] Suykens J A K, Vanderwalle J. Neural Process. Lett., 1999, 9: 293.
- [12] Suykens J A K, Van Gestel T, De Brabanter J, et al. Least Squares Support Vector Machines. Singapore: World Scientific, 2002.

## Fast Determination of Malondialdehyde in Oilseed Rape Leaves Using Near Infrared Spectroscopy

KONG Wen-wen, LIU Fei, ZOU Qiang, FANG Hui\*, HE Yong\*

College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China

**Abstract** Near infrared (NIR) spectroscopy was applied for the fast and nondestructive determination of malondialdehyde (MDA) content in oilseed rape leaves. A total of 90 leaf samples were collected, the calibration set was composed of 60 samples, and the prediction set was composed of 30 samples. Different preprocessing methods were used before the calibration stage, including smoothing, standard normal variate, first and second derivative, and detrending. Then partial least squares (PLS) models were developed for the prediction of MDA content in oilseed rape leaves. The latent variables selected by PLS and effective wavelengths selected by successive projections algorithm (SPA) were used as the inputs of least square-support vector machine (LS-SVM) to develop LV-LS-SVM and SPA-LS-SVM models. The correlation coefficients ( $r$ ) and root mean square error of prediction (RMSEP) were used as the model evaluation indices. Excellent results were achieved by LV-LS-SVM model, and the prediction results by LS-SVM model using detrending spectra were  $r = 0.999 9$  and RMSEP = 0.530 2, and those by LS-SVM model using 2-Der spectra were  $r = 0.999 9$  and RMSEP = 0.395 7. The results showed that NIR spectroscopy could be used for determination of MDA content in oilseed rape leaves, and an excellent prediction precision was achieved. This study supplied a new approach to the dynamic and continuous field monitoring of growing status of oilseed rape.

**Keywords** Near infrared spectroscopy; Oilseed rape; Malondialdehyde; Partial least squares analysis; Least squares-support vector machine

(Received Jun. 28, 2010; accepted Oct. 2, 2010)

\* Corresponding author