

# DOSC 在花椒挥发油含量近红外光谱分析中的应用<sup>\*</sup>

祝诗平 王刚 尹雄 兰雪冬 任德齐

**【摘要】** 应用近红外光谱分析技术,采用偏最小二乘法,对 141 份花椒粉末样品近红外光谱建立挥发油含量定量模型,交叉验证测定系数  $R^2$  为 0.936,交叉验证误差均方根  $R_{MSECV}$  为 0.421,经直接正交信号校正(DOSC)预处理后,相应的交叉验证测定系数  $R^2$  提高到 0.95,  $R_{MSECV}$  减小为 0.374。使用 105 份样品近红外光谱所建立的模型对 36 份样品的预测集进行预测,光谱采用 DOSC 预处理前后,预测测定系数  $R^2$  由 0.923 提高到 0.969,  $R_{MSEP}$  由 0.452 减小到 0.289,  $R_{SD}$  由 11.65% 减小到 7.44%,  $R_{PD}$  由 3.622 增加到 5.573。研究结果表明,使用 DOSC 预处理后的花椒挥发油含量近红外光谱定量模型的预测效果有较大的提高,且具有较好的稳定性和较强的预测能力,可用于实际的花椒挥发油检测。

**关键词:** 花椒 挥发油 含量 近红外光谱 直接正交信号校正

中图分类号: O657.33

文献标识码: A

## Application of DOSC in Detection of Volatile Oil Content from Zanthoxylum Bungeagum Maxim by Near-infrared Spectroscopy Technique

Zhu Shiping<sup>1,2</sup> Wang Gang<sup>2</sup> Yin Xiong<sup>2,3</sup> Lan Xuedong<sup>2,3</sup> Ren Deqi<sup>4</sup>

(1. Chongqing University 2. Southwest University 3. The Second Agricultural School of Chongqing  
4. Chongqing College of Electronic Engineering)

### Abstract

Based on near infrared spectroscopy technique and partial least squares (PLS), calibration model of volatile oil content of 141 powder samples with particle size of 40 meshes was established to classify Zanthoxylum bungeagum Maxim more rapidly. After spectra were preprocessed by direct orthogonal signal correction (DOSC), the determination coefficient ( $R^2$ ) increased from 0.936 to 0.95 and the root mean square error of cross validation ( $R_{MSECV}$ ) decreased from 0.421 to 0.374. Applying the model established by 105 samples to the test set with 36 samples,  $R^2$  increased from 0.923 to 0.969, the root mean square error of prediction ( $R_{MSEP}$ ) decreased from 0.452 to 0.289,  $R_{PD}$  increased from 3.622 to 5.573 and  $R_{SD}$  decreased from 11.65% to 7.44% after spectra were preprocessed by DOSC. The research indicated that the model preprocessed by DOSC is better, and the precision of model got higher. The result showed that rapid detection of volatile oil content in Zanthoxylum bungeagum Maxim by NIR and DOSC is feasible and efficient.

**Key words** Zanthoxylum bungeagum Maxim, Volatile oil, Content, Near-infrared spectroscopy, DOSC

### 引言

挥发油含量是评定花椒质量等级的重要理化指

标<sup>[1]</sup>。传统的花椒挥发油含量的测定方法目前有蒸馏法等<sup>[1~3]</sup>,这些提取分离法耗时、费工,很难做到快速测定,给花椒分级带来极大困难。

收稿日期: 2007-09-17

\* 国家自然科学基金资助项目(项目编号:30671198)、重庆市科委自然科学基金资助项目(项目编号:CSTC2005BB2211)和西南大学发展基金资助项目(项目编号:SWUF2007026)

祝诗平 重庆大学光电技术及系统教育部重点实验室 博士后 副教授(西南大学), 400030 重庆市

王刚 西南大学工程技术学院 硕士生, 400715 重庆市

尹雄 重庆第二农业学校 高级讲师 硕士生(西南大学), 402160 重庆市

兰雪冬 重庆第二农业学校 高级讲师 硕士生(西南大学)

(任德齐) 重庆电子工程职业学院电子信息系 副教授, 401331 重庆市

Copyright © 2008 China Agricultural Machinery Publishing House. All rights reserved. <http://www.cnki.net>

在农业、食品等行业中,近红外光谱分析(near infrared spectroscopy, 简称 NIR)已得到了广泛应用<sup>[4~5]</sup>。在建立近红外光谱模型前,一般需要对光谱进行预处理,常用的谱图预处理方法有数据增强变换、平滑、导数、标准正态变量变换、多元散射校正、傅里叶变换、小波变换等。但是这些光谱预处理方法,只是对谱图本身的数据进行处理,并未考虑浓度矩阵的影响。所以在进行预处理时,极有可能损失部分对建立校正模型有用的信息,又可能对噪声消除不完全,而影响所建模型的质量<sup>[6]</sup>。1998年 Wold 提出了一种光谱预处理算法——正交信号校正方法(orthogonal signal correction, 简称 OSC)<sup>[7]</sup>,文献<sup>[8~15]</sup>先后提出了 DO、DOSC、O-PLS、Piece-wise-OSC、OWAVEC 等几种 OSC 改进算法。

本文应用近红外光谱分析技术,建立花椒挥发油含量的预测模型,分析直接正交信号校正(DOSC)算法对近红外光谱模型预测效果的改善情况。

## 1 直接正交信号校正基本原理

2001年, Westerhuis J A 提出了一种直接正交信号校正(DOSC)<sup>[12]</sup>。DOSC 很好地解决了其他 OSC 方法中遇到的难题:得分矢量  $t$  不能同时满足最大地表达光谱矩阵  $A$  中与浓度矩阵  $C$  无关的信息和保持在  $C$  的正交补空间上。

与 Wold 算法不同, DOSC 算法首先将  $A$  和  $C$  正交,再对正交后  $A$  进行主成分分析,具体算法如下:①将校正集原始光谱矩阵和原始浓度矩阵分别进行均值中心化(mean-centering)或标准化(autoscaling)处理,得矩阵  $A(n \times m)$  和  $C(n \times 1)$ ,这里  $n$  为校正集样品数, $m$  为波长点数。②计算  $C$  在  $A$  张成的线性空间上的投影: $M=A((A')^{-1})'C$ 。③计算  $A$  在  $M$  的正交补空间上的投影: $Z=A-MM^{-1}A$ 。④对  $ZZ'$  进行主成分分析,取前  $f$  个需正交处理的主成分得分矩阵  $S_f$ 。⑤计算权重矩阵  $W_f=A^{-1}S_f$ ,广义逆  $A^{-1}$  通过 PLS 得到。⑥计算新的  $S_f$ ,  $S_f=AW_f$ 。⑦计算载荷矩阵  $P_f$ ,  $P_f=A'S_f/(S_f'S_f)$ 。⑧对  $A$  进行校正,过滤掉  $A$  中与  $C$  正交的噪声信息,  $A_{DOSC}=A-S_fP_f'$ 。⑨对于预测集原始光谱矩阵经过与校正集相同的均值中心化或标准化处理,得光谱矩阵  $A_{new}(n' \times m)$ ,由权重  $W$  和载荷  $P$  求出校正后的光谱为: $S=A_{new}W$ ,  $A_{DOSC}=A_{new}-SP'$ 。

## 2 实验材料与方法

### 2.1 样品采集与制备

分别采购来自陕西韩城、四川茂纹、汉源、西昌

甘肃兰州、重庆巴南、江津、云南等地散装或袋装去籽干花椒。其中:红花椒样品 74 份、青花椒样品 67 份,共计 141 份,取出每份花椒样品 150 克使用粉碎机进行粉碎,过 40 目筛,得花椒粉末样品。

### 2.2 样品光谱采集

对每份花椒粉末样品采用 MATRIX-I 型傅里叶变换近红外光谱仪进行近红外光谱扫描,扫描条件为:扫描范围  $12\ 798\sim\ 3\ 599\ \text{cm}^{-1}$ ,扫描次数 64 次,分辨率  $4\ \text{cm}^{-1}$ ,光谱点数 4 771,带漫反射积分球附件,PbS 检测器。

### 2.3 样品挥发油含量化学测定

采用文献<sup>[1]</sup>的标准测定每份花椒样品的挥发油含量。

### 2.4 模型校正方法与评价

光谱经均值中心化处理,采用 PLS 法建立校正模型,以全部样品进行留一法内部交叉验证(PLS-LOOCV)测定系数  $R^2$  和交叉验证误差均方根  $R_{MSECV}$  来检验近红外光谱模型的预测效果。然后将样品划分校正集和预测集,以校正集模型对预测集样品的预测值与测定值之间的测定系数  $R^2$ 、预测标准差  $R_{MSEP}$ 、预测相对标准差  $R_{SD}$ <sup>[16]</sup>和预测相对分析误差  $R_{PD}$ <sup>[17]</sup>作为评价校正模型预测能力的指标。其中, $R_{SD}$ 为  $R_{MSEP}$ 与  $y_{mp}$ (预测集样品测定值的平均值)的比值, $R_{PD}$ 为  $S_D$ (预测集样品测定值的标准偏差)与  $S_{EP}$ 的比值。一般认为当  $R_{SD}<10\%$ ,模型可用于实际的检测, $R_{PD}$ 在 10 以上为极好,7~10 之间为非常好,5~7 之间为好,3~5 之间为普通,3 以下为差<sup>[18~19]</sup>。文中涉及算法均基于 Matlab 2006b 平台实现。

## 3 实验结果与分析

### 3.1 花椒样品近红外光谱

141 份花椒粉末样品近红外光谱如图 1 所示。

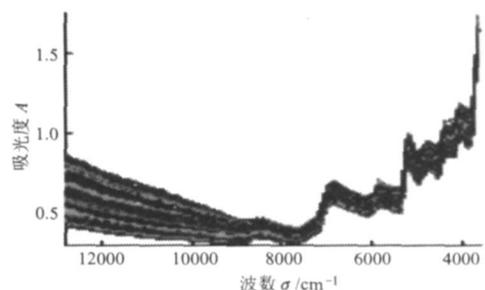


图 1 141 份花椒粉末样品的近红外光谱

Fig. 1 NIR spectra of 141 Zanthoxylum bungeagum Maxim samples

### 3.2 花椒样品挥发油含量测定结果

141 份样品挥发油含量范围为  $11.9\sim\ 75.4\ \text{mL/kg}$

根据 141 份花椒粉末样品挥发油含量分布范围, 选取 105 份样品作为校正集, 余下的 36 份样品作为预测集, 样品数目比  $105/36 \approx 3$ 。花椒挥发油测定值的变化范围、平均值及标准偏差如表 1 所示。

表 1 花椒挥发油测定值的变化范围、平均值及标准偏差

Tab.1 Summary of the variation ranges, means and standard deviations of volatile oil content of *Zanthoxylum bungeagum* Maxim mL/kg

样品集	样品数	最小值	最大值	平均值	标准偏差
总样品	141	11.9	75.4	36.90	16.8
校正集	105	11.9	75.4	36.30	16.9
预测集	36	13.7	66.9	38.78	16.5

### 3.3 141 份花椒粉末样品 PLS-LOO-CV 结果

使用 141 份花椒粉末样品近红外光谱, 对花椒挥发油含量进行 PLS-LOO-CV, 取主成份数为 16,  $R^2=0.936$ ,  $R_{MSECV}=0.421$ 。

141 份花椒粉末样品 PLS-LOO-CV 挥发油含量预测值与测定值之间的相关关系如图 2 所示。

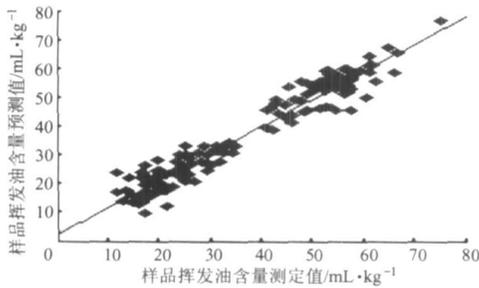


图 2 挥发油含量 PLS-LOO-CV 预测值与测定值之间的相关关系

Fig.2 Relationship between predicted and chemical value of volatile oil content by PLS-LOO-CV

### 3.4 直接正交信号校正对光谱进行预处理

对 141 份花椒样品原始光谱, 结合挥发油含量数据, 进行 PLS-LOO-CV, 在每次循环中, 对参与建模的 140 份样品原始光谱和待预测的 1 份样品的原始光谱均使用直接正交信号校正进行预处理, 去掉 2 个 OSC 因子。当模型主成份数为 1 时, 测定系数  $R^2=0.95$ ,  $R_{MSECV}=0.374$ , 挥发油含量预测值与测定值之间的相关关系如图 3 所示。

经过直接正交信号校正预处理后, PLS-LOO-CV 效果得到改善, 测定系数  $R^2$  由 0.936 提高到 0.95,  $R_{MSECV}$  由 0.421 减小到 0.374。而且, 模型主成份数由 16 降到 1, 模型得到简化。

### 3.5 校正集模型对预测集预测结果

使用 105 份花椒粉末样品原始的和经 DOSC 预处理后的近红外光谱, 分别建立校正集挥发油含量近红外光谱定量模型(主成份数分别为 13 和 1), 对

预测集(36 份花椒粉末样品原始的和经 DOSC 预处理后的近红外光谱)进行预测, 其结果如表 2 所示。

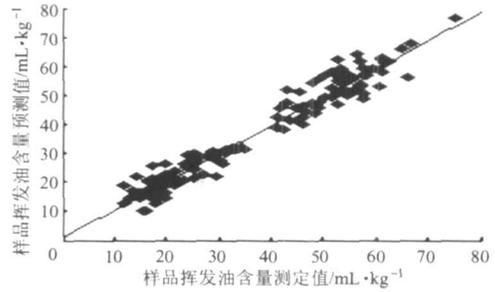


图 3 经 DOSC 预处理后挥发油含量 PLS-LOO-CV 预测值与测定值之间的相关关系

Fig.3 Relationship between predicted and chemical value of volatile oil content by PLS-LOO-CV after DOSC

表 2 DOSC 预处理前后模型各项评价指标对比表

Tab.2 Evaluation indices of models before and after DOSC

光谱类型	$R^2$	$R_{MSEP}$	$R_{SD}/\%$	$S_{EP}$	$R_{PD}$
原始光谱	0.923	0.452	11.650	0.457	3.622
DOSC 校正光谱	0.969	0.289	7.440	0.291	5.573

从表 2 可以看出, 经 DOSC 预处理后的花椒挥发油近红外光谱定量模型的预测效果, 比未经 DOSC 预处理时有较大提高, 同时, 主成份数由 13 降到 1, 模型得到简化。

原始光谱预测集挥发油含量预测值与测定值相关关系如图 4 所示, 经 DOSC 预处理后的预测集挥发油含量预测值与测定值相关关系如图 5 所示。

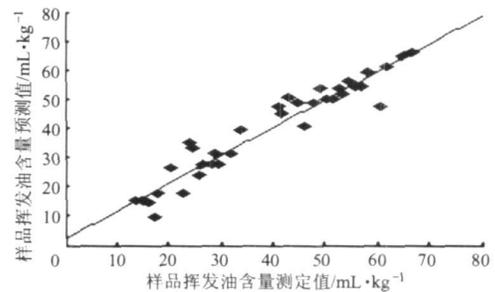


图 4 原始光谱预测集挥发油预测值与测定值的相关关系

Fig.4 Relationship between predicted and chemical value of volatile oil content for raw spectra

## 4 结论

(1) 应用偏最小二乘法, 建立 40 目花椒粉末样品中挥发油含量预测模型, 经 DOSC 预处理后, 模型预测能力有较大提高。同时, 主成份数减少到 1, 模型得到简化。因此, 在建立近红外光谱定量模型前, 通过 DOSC 将与浓度矩阵无关的光谱信号滤除, 可以有效地提高模型预测能力和简化模型。

(2) 经直接正交信号校正预处理后的花椒挥发

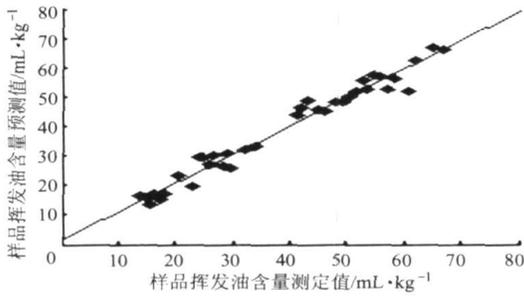


图5 经 DOSC 预处理后预测集样品挥发油含量  
预测值与测定值的相关关系

Fig. 5 Relationship between predicted and chemical  
value of volatile oil content for test set  
samples spectra preprocessed by DOSC

油近红外光谱模型,  $R_{SD}$  小于 10%, 因此该模型可用于实际的花椒挥发油含量检测, 而且模型预测  $R_{PD}$  大于 5, 因此该模型具有较好的稳定性和较高的预测能力。经 DOSC 预处理前后, 预测决定系数  $R^2$  均在 0.92 以上, 线性关系十分明显, 说明基于近红外光谱技术快速检测花椒挥发油含量是可行的。

### 参 考 文 献

- 1 LY/T 1652—2005 花椒质量等级[S].
- 2 秦军, 陈桐, 吕晴. 同时蒸馏萃取气质联用法测定花椒挥发油成分[J]. 贵州工业大学学报: 自然科学版, 2001, 30(6): 4~6.
- 3 陈振德, 许重远, 谢立. 超临界 CO<sub>2</sub> 流体萃取花椒挥发油化学成分的研究[J]. 中国中药杂志, 2006, 26(10): 687~688.
- 4 牛智有, 韩鲁佳. 鱼粉中氨基酸近红外光谱定量分析[J]. 农业机械学报, 2007, 38(5): 114~117.
- 5 田海清, 应义斌, 徐惠荣, 等. 西瓜可溶性固形物含量近红外透射检测技术[J]. 农业机械学报, 2007, 38(5): 111~113.
- 6 褚小立, 袁洪福, 陆婉珍, 等. 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.
- 7 Wold S, Antti H, Lindgren F, et al. Orthogonal signal correction of near-infrared spectra[J]. Chemometrics and Intelligent Laboratory Systems, 1998, 44(1~2): 175~185.
- 8 余浩. 基于正交信号校正算法的近红外光谱预处理[D]. 杭州: 浙江大学, 2004.
- 9 Sjöblom J, Svensson O, Josefson M, et al. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra[J]. Chemometrics and Intelligent Laboratory Systems, 1998, 44(1~2): 229~244.
- 10 Andersson C A. Direct orthogonalization[J]. Chemometrics and Intelligent Laboratory Systems, 1999, 47(1): 51~63.
- 11 Fearn T. On orthogonal signal correction[J]. Chemometrics and Intelligent Laboratory Systems, 2000, 50(1): 47~52.
- 12 Westerhuis J A, Jong S D, Smilde A K. Direct orthogonal signal correction[J]. Chemometrics and Intelligent Laboratory Systems, 2001, 56(1): 13~25.
- 13 Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS)[J]. Journal of Chemometrics, 2002, 16(3): 119~128.
- 14 Feudale R N, Tan H, Brown S D. Piecewise orthogonal signal correction[J]. Chemometrics and Intelligent Laboratory Systems, 2002, 63(2): 129~138.
- 15 Esteban-Díez I, Gonzalez-Saiz J, Pizarro C. OWAVEC: a combination of wavelet analysis and an orthogonalization algorithm as a pre-processing step in multivariate calibration[J]. Analytica Chimica Acta, 2004, 515(1): 31~41.
- 16 〇, ひ, . . . (七) vi 19 [M]. : バ ) ~, 2005.
- 17 〇, Γ ~ ⑤, ≡, . (9) 〇 〇 [J]. > , 2008, 39(3): 79~81, 85.
- 18 〇 ビば, x や, デ, . 〇 = vi 19 ② 〇 ギ [J]. > ..., 2004, 13(4): 13~25.
- 19 〇 Williams P C, Sobering D C. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds[J]. Journal of Near Infrared Spectroscopy, 1993, 1(1): 25~32.