

文章编号:1673-1689(2008)04-0012-04

基于 OSC/PLS 的茶叶中 EGCG 含量的 近红外光谱法测定

赵杰文, 郭志明, 陈全胜

(江苏大学食品与生物工程学院, 江苏 镇江 212013)

摘要:采用正交信号校正法(OSC)对茶叶的近红外光谱进行预处理,并结合偏最小二乘法(PLS)建立表没食子儿茶素没食子酸酯(epigallocatechin gallate, EGCG)含量的预测模型。PLS 校正模型采纳的最佳因子数会随着 OSC 因子的增加而逐渐减少以达到简化模型的效果。试验以预测残差平方和(PRESS)来评价模型的整体性能,当滤除前 6 个 OSC 因子时为最佳,此时校正模型采纳的 PLS 因子数为 5,校正时的相关系数 r^2 和校正标准偏差 SEC 分别为 0.955 90 和 0.382 79,预测时的相关系数 r^2 和预测标准偏差 SEP 分别为 0.936 61 和 0.444 13。研究结果表明,应用近红外光谱法结合 OSC/PLS 可以快速准确地测定茶叶中 EGCG 的含量。

关键词:近红外光谱;偏最小二乘法;正交信号校正法;茶叶;表没食子儿茶素没食子酸酯

中图分类号:TP 242.62;O 657

文献标识码:A

Near Infrared Spectroscopy Determination of EGCG Content in Tea Based on OSC and PLS

ZHAO Jie-wen, GUO Zhi-ming, CHEN Quan-sheng

(School of Food & Biological Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: Orthogonal signal correction (OSC) was used as a method to preprocess the near infrared (NIR) spectra of tea, to establish the calibration model of EGCG content against tea spectra by partial least square (PLS). The number of optimal factors of PLS model used to predict EGCG content against tea spectra would be reduced in accordance with OSC factors reduction, to obtain terse model. The model has been verified by cross-validation, the predicted residual error sum of squares (PRESS) is the minimum. The optimum PLS calibration model was obtained when 6 OSC factors were filtered, to have obtained the coefficient r^2 0.955 90, with the standard error of calibration (SEC) of 0.382 79 and the standard error of prediction (SEP) of 0.444 13. The study demonstrated that NIR spectroscopy technology with OSC/PLS could be successful applied as a reliable and rapid method to determine the content of EGCG in tea.

Key words: near infrared spectra; PLS; OSC; tea; EGCG

收稿日期:2008-04-14.

基金项目:江苏省自然科学基金重点项目(BK2006707-1).

作者简介:赵杰文(1945-),男,江苏苏州人,工学博士,教授,博士生导师,主要从事食品与农产品无损检测的研究. Email:zhao@ujs.edu.cn

茶叶中的儿茶素属于黄烷醇类化合物,是多酚类物质的主体成分,占茶叶干重的 12%~24%。其中以表没食子儿茶素没食子酸酯(epigallocatechin gallate, EGCG)含量最高,其质量分数约占儿茶素的 80%^[1]。近年来大量科学研究表明,EGCG 具有明显的清除体内自由基、抗癌、抗炎、抑菌、抗病毒、抗突变、抗衰老及改善肝功能等生物活性,现已被中美等国列为潜在的抗癌药物进行研究^[2-3]。EGCG 也是决定茶叶饮料品质好坏的关键因子之一。由于 EGCG 不能人工合成,茶叶是 EGCG 的唯一来源,因此寻求一种快速简便的方法来测定茶叶中 EGCG 的含量,不仅有利于加快茶叶原料的选择速度,还有利于提高茶叶的品质检测水平。

近红外光谱检测技术具有快速、无损,以及可以同时检测多种品质的优点。近年来,该检测技术越来越广泛地被应用到茶叶品质的检测中。但在所采集的原始近红外光谱数据中,常常含有因系统不稳定造成的噪音信息以及有待测品质不相关的冗余信息等,这些不相关的信息会影响校正模型的质量和精度。常用的光谱数据预处理方法如多元散射校正(MSC)和二阶求导(SOD)等在滤除噪音的同时会损失部分与待测品质相关的光谱信息^[4-5]。作者在利用近红外漫反射光谱技术检测茶叶中 EGCG 含量的前期研究中发现,对原始光谱进行常规预处理后,再采用偏最小二乘法(PLS)建立的模型过于复杂。为此,研究中尝试采用一种全新的光谱数据预处理方法——正交信号校正法(Orthogonal Signal Correction——OSC)对茶叶的近红外原始光谱进行预处理,并结合 PLS 方法建立预测茶叶中 EGCG 含量的校正模型。

1 正交信号校正法

正交信号校正法主要用于近红外光谱矩阵的预处理,由著名的瑞典计量化学家 Svante Wold^[6]于 1998 年提出,随后 Tom Fearn^[7]、Robert N. Feudale^[8]及 Nigel G. Yee^[9]等先后对这种预处理方法作了改进与发展。该法主要用于滤除近红外光谱矩阵中的系统噪声(如基线漂移、光的散射等)。其基本思想是:利用数学上正交的办法,将原始光谱矩阵 X 中与待测品质 Y 不相关的部分信息滤除。换句话说就是 X 中被滤除的信息是与待测品质 Y 在数学上正交的。因此,正交信号校正法能确保被滤除掉的信息与待测品质无关。

本研究中作者利用 Tom Fearn 介绍的算法对

苹果的近红外光谱进行预处理,其算法^[4-5,7]如下:设茶叶训练集的近红外原始光谱数据矩阵为 $X(I \times J, I$ 为测试集样本数, J 为波数点数),该矩阵中的某一元素 x_{ij} ($i = 1, 2, \dots, I; j = 1, 2, \dots, J$) 的含义是第 i 个样本在第 j 个波数点处的反射率,茶叶中 EGCG 实测值向量为 $y(I \times 1)$ 。

1) 计算第一权重向量 w : 该问题可描述为 $\max(w^T X^T X w)$, 且 $w^T w = 1, w^T X^T Y = 1$, 式中上标符号“ T ”表示向量或矩阵的转置;

2) 计算得分向量 t : $t = Xw$;

3) 计算载荷向量 p : $p = X^T t / (t^T t)$;

4) 计算从原始光谱矩阵 X 中滤除第一 OSC 因子后的矩阵 $X_{0,1}$: $X_{0,1} = X - tp^T$;

以 $X_{0,1}$ 代替 X , 重复上述步骤, 计算出滤除第二个 OSC 因子后的矩阵 $X_{0,2}$, 如此循环, 可求出 $X_{0,3}, X_{0,4}, \dots$ 。

预测集光谱数据矩阵 X_{un} 的 OSC 校正则利用测试集 OSC 因子的权重及载荷向量进行:

5) $t_{un} = X_{un} w_i$, w_i 为测试集第 i 个 OSC 因子的权重向量;

6) $X_{un,0} = X_{un} - t_{un} p_i$, p_i 为测试集第 i 个 OSC 因子的载荷向量, $X_{un,0}$ 为经正交信号校正法预处理后的预测集茶叶光谱。

2 试验方法及数据

试验中选用市售茶叶 22 种, 均产于 2007 年 4 月到 5 月。从中随机地选取 15 种作为测试集, 余下的 7 种作为预测集。试验前先将茶叶粉碎并过 40 目筛, 每种茶叶分成 5 个样本, 每个样本称取 1 g, 这样测试集共 75 个样本, 预测集共 35 个样本, 然后将它们分别编号后置于 4℃ 冰柜中贮藏。光谱采集试验在温度稳定的试验室(温度控制为 25℃)内进行。试验前, 将冰柜中取出的茶叶置于试验室中 12 h, 以使茶叶整体温度达到与环境温度的一致。试验时, 将茶叶样本放于样品杯中, 由近红外光谱仪(Antaris II, 美国 Thermo Scientific 公司)进行光谱扫描(扫描波数范围为 10 000~4 000 cm^{-1} , 扫描次数 32 次, 分辨率 4 cm^{-1}), 然后将样本倒出样品杯再重新倒入, 重复采集 4 次, 并取 4 次的平均光谱作为整个样本的原始光谱。采集光谱后, 将每个茶叶样本经浸泡、超声、离心、过 0.45 μm 微孔滤膜等一系列预处理后, 所得样本用高效液相色谱测定 EGCG 含量, 色谱条件参考 ISO-14502-2006。表 1 列出了茶叶中 EGCG 的含量的测定结果。

表 1 HPLC 测定的茶叶中 EGCG 的质量分数
Tab. 1 Contents of EGCG in tea samples by HPLC

成分	含量性质	测试集 75 个样本			预测集 35 个样本		
		范围	均值	标准方差	范围	均值	标准方差
EGCG	质量分数/%	7.340~14.304	11.236	1.824	7.651~14.088	11.184	1.764

3 结果与讨论

3.1 光谱的 OSC 预处理

采用 OSC 法对茶叶近红外原始光谱进行预处理。图 1 显示了采用 OSC 法处理前后光谱图的变化,其中图 1(a)为原始光谱图,图 1(b)为从原始光谱中滤除 5 个 OSC 因子后的谱图。从图中可以看出,经 OSC 法预处理后,光谱形状总体上与处理前的光谱没有差别,但处理后的光谱曲线变得更为光滑,且光谱排列也比处理前更为整齐。这说明正交信号校正法滤除了原光谱中的部分噪声,但又保留了与测试样本相关的主要信息。

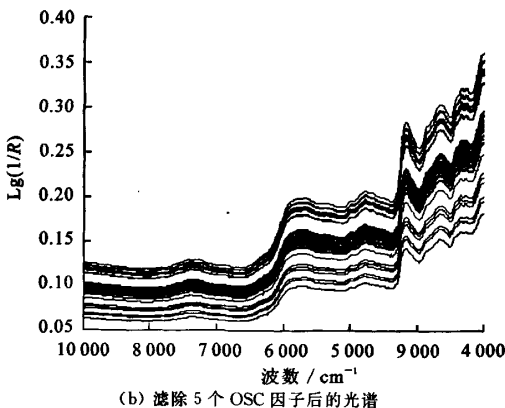
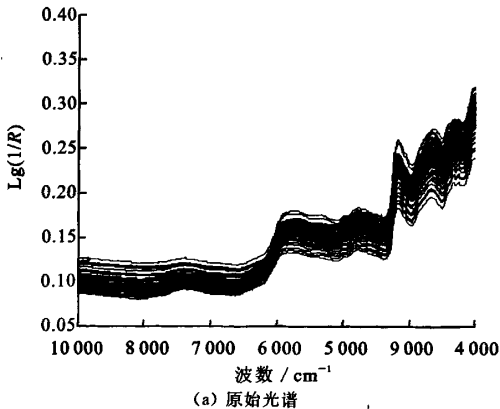


图 1 茶叶的原始光谱(a)和 OSC 校正后的光谱(b)

Fig. 1 Original spectra (a) and OSC corrected spectra (b) of tea sets

3.2 PLS 校正

3.2.1 PLS 校正结果 采用偏最小二乘法对茶叶的原始光谱和经 OSC 法处理后的茶叶光谱分别进

行多变量校正,在不同的 OSC 因子下,校正模型的最佳因子数(主成分数)通过交互验证法(Cross-Validation)确定,模型交互验证时最小的预测残差平方和(PRESS 值)所对应的主成分因子数最佳。图 2 中分别显示了滤除前 0~13 个 OSC 因子时的 $\lg(\text{PRESS 值})$ 与 PLS 因子的关系。利用图 2 可以确定不同 OSC 因子下的主成分因子数。表 2 为经 OSC 法处理前后,试验数据的偏最小二乘法校正结果。

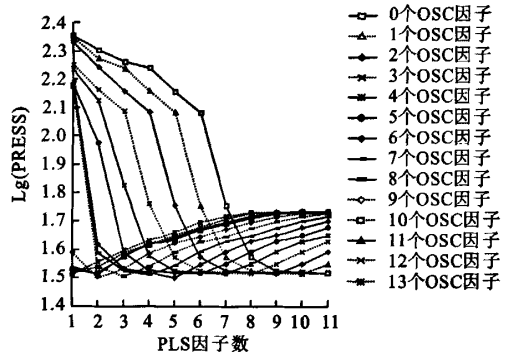


图 2 PLS 因子数与 $\lg(\text{PRESS})$ 的关系

Fig. 2 The number of PLS factors vs. diagram of $\lg(\text{PRESS})$

3.2.2 分析 从表 2 可以看出,原始光谱的 PLS 模型已能较好地预测茶叶中 EGCG 的含量,其校正时的相关系数 r^2 和标准偏差 SEC 分别为 0.956 26 和 0.381 50,预测时模型的相关系数 r^2 和标准偏差 SEP 分别为 0.936 93 和 0.443 01。但该模型采纳的最佳因子数为 11,这使得模型显得过于复杂。对原始光谱进行 OSC 预处理后,PLS 模型所采纳的最佳因子数随着 OSC 因子数的增加而逐渐减少,甚至可减少至 1。最初,被滤除的 OSC 因子每增加 1 个,模型所采纳的最佳因子数就减少 1,滤去前 9 个 OSC 因子后,模型的最佳因子数减到 2。滤除前 12 个 OSC 因子后,模型采纳的因子数减为 1,继续滤除下去,模型的最佳因子个数都保持为 1。从表 2 中还可以看出,滤除 OSC 因子前后,PLS 模型对茶叶中 EGCG 的含量的校正和预测精度仅有很小的波动,而没有发生太大的变化。研究中以 PRESS 来评价模型的整体性能。由图 2 可以看出,当滤除前 6 个 OSC 因子时, $\lg(\text{PRESS 值})$ 最小,此时校正模型采纳的最佳因子数为 5,校正时的相关系数 r^2

和标准偏差 SEC 分别为 0.955 90 和 0.382 79, 用 0.936 61 和 0.444 13。与 OSC 法处理前的 PLS 模型相比,其精度虽有微小降低,但模型更简洁。

表 2 正交信号校正(OSC)处理前后的偏最小二乘法(PLS)校正结果

Tab. 2 Summary of PLS results before and after the spectral of teas being preprocessed by orthogonal signal correction (OSC)

OSC 因子数	PLS 因子数	测试集 相关系数 r^2	校正标准 偏差 SEC	预测集 相关系数 r^2	预测标准 偏差 SEP
0	11	0.956 26	0.381 50	0.936 93	0.443 01
1	10	0.956 26	0.381 50	0.936 93	0.443 01
2	9	0.956 26	0.381 50	0.936 93	0.443 01
3	8	0.956 25	0.381 55	0.936 92	0.443 04
4	7	0.956 23	0.381 65	0.936 90	0.443 10
5	6	0.955 98	0.382 72	0.936 70	0.443 82
6	5	0.955 90	0.382 79	0.936 61	0.444 13
7	4	0.954 42	0.389 46	0.935 34	0.448 57
8	3	0.950 73	0.404 89	0.931 77	0.460 76
9	2	0.944 62	0.429 27	0.924 97	0.483 24
10	2	0.935 07	0.464 81	0.913 68	0.518 27
11	2	0.960 84	0.360 98	0.931 55	0.461 03
12	1	0.909 40	0.549 05	0.884 54	0.599 40
13	1	0.915 13	0.531 40	0.891 37	0.581 41

4 结 语

研究中利用 OSC 法对茶叶近红外光谱进行了预处理,并结合 PLS 法建立了茶叶中 EGCG 含量的预测模型。研究发现,OSC 法预处理后,茶叶光谱的形状变得更光滑、简洁,排列更整齐,又保留了其中的主要信息。同时,随着所用 OSC 因子的逐个增加,PLS 模型的最佳因子数也逐渐减少,模型的精度变化很小,这表明 OSC 法能在保证精度的前提

下简化茶叶中 EGCG 含量的预测模型。试验通过交互验证来优化 OSC 因子数及 PLS 最佳主成分因子数,当滤除 6 个 OSC 因子,校正模型采纳的最佳因子数为 5 时,所对应的 PRESS 最小,模型交互验证时的相关系数 r^2 和标准偏差 SEC 分别为 0.955 90 和 0.382 79,预测时的相关系数 r^2 和标准偏差 SEP 分别为 0.936 61 和 0.444 13。研究结果表明,应用近红外光谱法结合 OSC/PLS 既可以准确地测定茶叶中 EGCG 的含量,又可以简化 PLS 模型,以提高模型的鲁棒性。

参考文献(References):

- [1] 杨贤强,王岳飞,陈留记. 茶多酚化学[M]. 上海:上海科学技术出版社,2003.
- [2] 赵丽萍,邵宛芳. 茶叶中 EGCG 功效研究进展[J]. 中国农学通报,2007,23(7):143-147.
- [3] ZHAO Li-ping, SHAO Wan-fang. Research effect of epigallocatechin gallate in the tea[J]. *Chinese Agricultural Science Bulletin*, 2007, 23(7):143-147. (in Chinese)
- [4] 仇燕嵘,李楠,韩国柱,等. 表没食子儿茶素没食子酸酯的研究进展[J]. 中草药,2006,37(2):303-306.
- [5] ZHANG Yan-lai, LI Nan, HAN Guo-zhu, et al. Advance of researches on epigallocatechin gallate[J]. *Chinese Traditional and Herbal Drugs*, 2006, 37(2):303-306. (in Chinese)
- [6] 赵杰文,张海东,刘木华. 简化苹果糖度预测模型的近红外光谱预处理方法[J]. 光学学报,2006,26(1):136-140.
- [7] ZHAO Jie-wen, ZHANG Hai-dong, LIU Mu-hua. Preprocessing methods of near-infrared spectra for simplifying prediction model of sugar content of apples[J]. *Acta Optica Sinica*, 2006, 26(1):136-140. (in Chinese)
- [8] 张海东,赵杰文,刘木华. 基于正交信号校正和偏最小二乘(OSC/PLS)的苹果糖度近红外检测[J]. 食品科学,2005,26(6):189-192.
- [9] ZHANG Hai-dong, ZHAO Jie-wen, LIU Mu-hua. Near infrared determination of sugar content in apples based on orthogonal signal correction and partial least square(OSC/PLS) method[J]. *Food Science*, 2005, 26(6):189-192. (in Chinese)
- [10] Wold S, Antti H, Lindgren F. Orthogonal signal correction of near-infrared spectra[J]. *Chemometrics and Intelligent Laboratory Systems*, 1998, 44(1):175-185.
- [11] Fearn T. On orthogonal signal correction[J]. *Chemometrics and Intelligent Laboratory Systems*, 2000, 50(1):47-52.
- [12] Feudale R, Tan H, Brown S. Piecewise orthogonal signal correction[J]. *Chemometrics and Intelligent Laboratory Systems*, 2002, 63(2):129-138.
- [13] Yee N, Coghill G. Factor selection strategies for orthogonal signal correction applied to calibration of near-infrared spectra[J]. *Chemometrics and Intelligent Laboratory Systems*, 2003, 67(2):145-156.

(责任编辑:秦和平)