

背景和光程对茶多酚含量近红外预测模型的影响

王胜鹏¹ 龚自明¹ 何远军² 高士伟¹
郑鹏程¹ 滕靖¹ 王雪萍¹ 叶飞¹ 卢素芳¹

1. 湖北省农业科学院果树茶叶研究所, 武汉 430064; 2. 湖北省恩施市农业局, 恩施 445000

摘要 为比较空气和纯水背景的模型效果, 扫描5个(2、3、5、8和10 mm)不同光程的红茶汤近红外光谱, 应用排列组合原理和偏最小二乘法(PLS)建立2种背景下单光程和多光程组合茶多酚近红外预测模型, 并用 t -test检验模型预测效果。结果表明: 单光程模型中, 纯水背景与空气背景模型预测结果间无显著性差异; 多光程组合模型中, 空气背景模型要全部优于纯水背景模型, 并且最佳组合模型((5+8) mm)具有比单光程模型更好的预测精度。空气背景茶多酚模型要优于纯水背景茶多酚模型。

关键词 近红外光谱; 单光程模型; 多光程模型; 空气背景; 纯水背景

中图分类号 O 657.33 **文献标识码** A **文章编号** 1000-2421(2015)02-0120-05

现代近红外(NIR)光谱分析技术在农业、医药、石化、烟草和食品等行业应用广泛。近红外光谱技术在茶叶方面的应用主要是用来快速评定成品茶叶的等级^[1]和测定成分含量, 如用来测定咖啡碱含量、茶多酚总量等^[2-4]; 将近红外光谱技术应用于茶汤内含物含量的检测报道较少^[5], 原因是茶汤光谱背景和光程选择影响了内含物近红外模型预测效果。与茶汤中内含物含量相比, 溶剂水的含量是最多的, 而—OH在近红外区域有极强的吸收峰, 茶汤内含物的近红外吸收较水弱, 在空气背景的近红外光谱中, 水峰极可能掩盖了茶汤内含物的近红外信息; 而纯水背景的近红外光谱则会扣除纯水的近红外信息, 近似地认为是茶汤内含物的信息, 所以茶汤光谱建模时背景选择非常重要。陈斌等^[6]尝试了以纯水为背景来检测啤酒的乙醇含量, 得到了较好的结果; 刘宏欣等^[7]比较了空气和纯水背景测定啤酒主要品质参数的结果, 认为可以以空气为背景检测透射光谱, 但上述结论都没有经过统计学检验, 缺乏说服力。扫描茶汤光谱时光程的选择也很重要, 张林等^[8]、杜敏等^[9]分别研究了光程变化对分析模型的影响; 李刚等^[10]利用光谱的非线性特性建立了多光程组合建模, 提高了血液成分检测的精度, 但上述研究中

光程选择的代表性并不强, 多光程模型的可选择组合不多, 也没有对多光程模型效果进行深入地探讨。

红茶是世界流行的三大非酒精饮料之一, 茶多酚具有苦涩味, 对茶汤滋味和茶叶保健功能有非常重要的影响。药理学研究表明, 茶多酚对心血管系统疾病^[11]、癌症^[12]等具有一定的防治效果。鉴于茶多酚的天然非稳定性, 快速和准确的测定茶汤中茶多酚的含量很重要, 而传统的化学检测方法费时费力。近红外方法具有分析速度快、不污染环境等优点。因此, 为建立稳定的红茶茶汤茶多酚近红外预测模型, 笔者以空气和纯水为背景, 扫描2、3、5、8和10 mm光程的红茶汤近红外光谱, 应用排列组合原理对光程进行各种优化组合, 利用偏最小二乘法(PLS)建立2种背景下茶多酚含量的单光程和多光程近红外预测模型, 用 t -test检验模型预测结果间有无显著性差异, 探讨影响模型预测精度的2个因素(光谱背景和光程)对近红外模型的影响, 以便建立稳健的近红外光谱预测模型, 方便红茶品质检测, 达到省时省力的目的。

收稿日期: 2014-07-22

基金项目: 国家青年科学基金项目(31400586); 国家茶叶产业技术体系项目(CARS-23); 湖北省青年科学基金项目(2014CFB224); 湖北省农业科学院果树茶叶研究所青年基金项目(GCJJ201301)

王胜鹏, 博士。研究方向: 茶叶加工和茶叶品质无损检测。E-mail: wsspp0426@163.com

通信作者: 龚自明, 研究员。研究方向: 茶叶加工。E-mail: zimingong@163.com

1 材料与方法

1.1 红茶样品

红茶茶样 90 个(红碎茶),购买时间为 2014 年 5 月。

1.2 茶多酚含量测定

按照国标方法(GB/T 8313 — 2008 茶叶中茶多酚和儿茶素类含量的检测方法)测定样品茶多酚含量。根据样品茶多酚含量高低,选取 60 个样品作为校正集,其余 30 个样品作为验证集,用来检验模型性能,统计结果见表 1。

表 1 红茶茶多酚含量分析结果

Table 1 The contents of black tea's total tea polyphenols

红茶茶样 Black tea samples	数量 Numbers	范围/% Range	平均值/% Mean	SD
校正集样品 Calibration set samples	60	6.08~20.52	11.87	2.15
验证集样品 Prediction set samples	30	6.41~13.36	10.81	1.82

1.3 茶汤近红外光谱采集

在 25 °C 下,分别以空气和纯水为背景采用不同光程的石英比色皿(2、5、10 mm)和低 OH 透射管(3、8 mm)在 Antaris 型傅里叶变换近红外光谱仪(American)上采集光谱。光谱扫描范围 4 000~12 500 cm^{-1} ;分辨率 4 cm^{-1} ;检测器 InGaAs;每个茶汤样品采集 3 次光谱,取 3 次采集光谱的平均值

作为该红茶汤样品的光谱。

1.4 数据分析

利用 TQ ANALYST 软件对光谱进行预处理,包括平滑、多元散射校正(multiplicative scatter correction, MSC)、一阶导数(1st derivative)和二阶导数(2nd derivative)等,经比较,采用 MSC 对原始光谱进行预处理效果最好。采用 Excel 2007 软件处理试验数据,用相关系数(R)、交互验证均方差(root mean square error of cross validation, RMSECV)、预测均方差(root mean square error of prediction, RMSEP)和概率 P 评价模型性能。

2 结果与分析

2.1 在空气和纯水背景下采集的不同光程茶汤原始光谱

由图 1A 可看出,茶汤中水—OH 伸缩振动的一级倍频和合频吸收区域在 6 800 cm^{-1} 和 5 100 cm^{-1} 附近,且随着光程的增加,液体光谱的吸光度也逐渐增加。图 1B 为扣除纯水吸收后的茶汤中茶多酚的近红外光谱,光谱吸收峰区间和图 1A 水的强吸收区域在 6 800 cm^{-1} 和 5 100 cm^{-1} 正好重合。图 1B 中茶多酚最大吸光度不超过 0.30,远远小于图 1A 中水的最大吸收峰(4.97),因此,水的强吸收峰可能掩盖和干扰了茶多酚的吸收信息,进而影响茶多酚的近红外模型效果。在对光谱进行 MSC 预处理后,利用偏最小二乘法(PLS)建立空气背景和纯水背景下的茶多酚近红外预测模型。

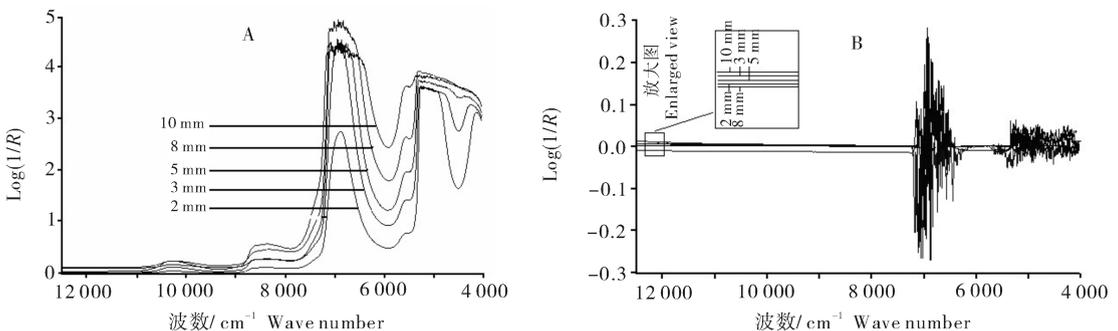


图 1 空气背景(A)和纯水背景(B)的红茶 1 号茶汤不同光程的近红外光谱

Fig. 1 NIR spectra of different optical path lengths of No. 1 black tea on air background(A) and pure water background(B)

2.2 单光程近红外模型

1) 2 种背景下不同单光程近红外模型结果。从图 2 可以看出,在光程一定的情况下(2、3、5、10 mm),纯水背景的茶多酚近红外模型校正集和预测 R 要大于空气背景模型, RMSECV 和 RMSEP 均小

于空气背景模型,说明空气背景下红茶汤茶多酚的近红外信息被纯水掩盖,纯水背景下茶多酚近红外模型效果得到了提高,但模型效果并没有提高很多,甚至出现了相反的情况,在 8 mm 光程下,纯水背景近红外模型效果要差于空气背景模型效果。

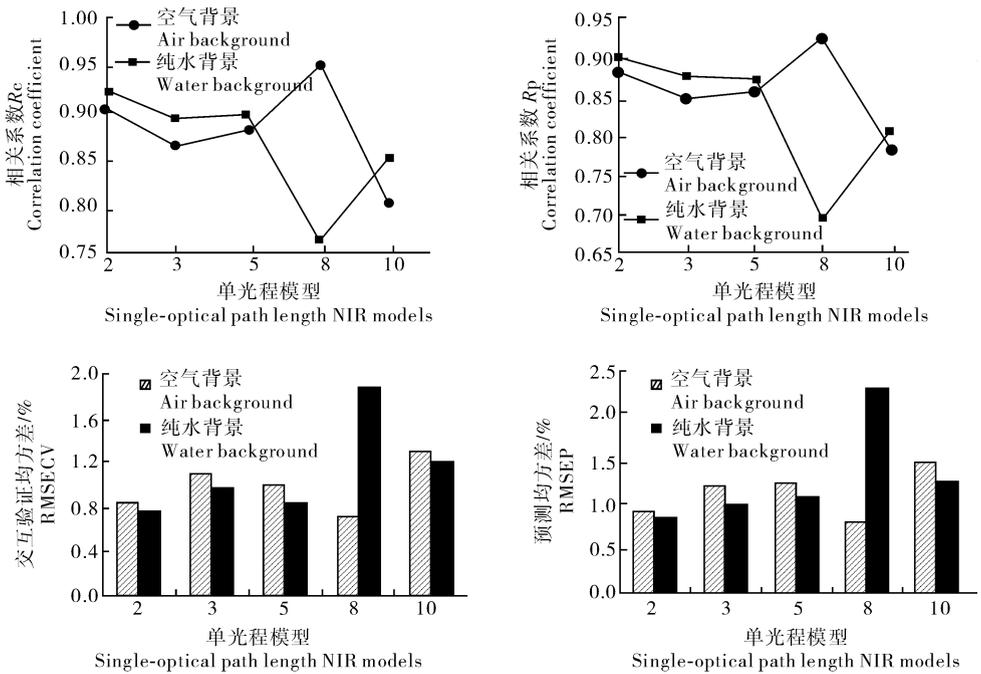


图 2 空气和纯水背景下茶多酚单光程近红外模型结果

Fig. 2 The results of different optical path length tea polyphenols NIR models on air and pure water background

2) 显著性检验。在空气和纯水背景下, 相同单光程模型的预测结果间无显著性差异 ($P > 0.05$), 表明茶汤中水对茶多酚近红外模型的影响有限。

3) 不同光程模型性能的显著性检验。从图 3 可以看出, 空气背景的不同光程茶多酚近红外预测模型结果间 P 值都大于 0.05, 都没有显著性差异; 纯水背景的不同光程茶多酚近红外预测模型结果间 $P(2 \text{ vs. } 3)$ 和 $P(2 \text{ vs. } 10)$ 小于 0.05, 有显著性差异, 表明纯水背景下的茶汤近红外光谱模型对光程的变

化较为敏感, 光程影响了茶多酚近红外模型预测效果, 选择合适的光程可以提高茶多酚近红外模型效果。

2.3 多光程组合近红外建模

图 4 为应用排列组合原理得到的 26 组不同光程组合的空气和纯水背景茶多酚近红外模型结果。从图 4A, B 可以看出, 空气背景模型的 R 均大于纯水背景的 R , 而图 4C, D 中, 空气背景模型的 RMSECV 和 RMSEP 均小于纯水背景模型, 说明空气背景的多光程组合近红外模型全部优于纯水背景近红外模型, 进一步说明了茶汤中水对茶多酚近红外模型的影响有限。将图 4 与图 2 结果进行比较, 可以得出: 多光程模型结果并不一定会优于单光程模型, 多光程近红外模型效果也不是单光程模型效果的简单叠加, 这是因为虽然多光程模型可以提供更多样本成分含量信息, 但同时也引入了更多的噪声, 这会对模型效果产生负面作用。所以, 要得到最佳光程模型结果, 需通过排列组合原理, 逐渐筛选出最佳组合, 所得模型效果才会优于单光程模型, 本文最佳组合为以空气为背景的 (5+8) mm 光程的近红外模型 ($R_p = 0.984, RMSEP = 0.245$, 图 5)。

2.4 预测精度比较

将空气背景的 (5+8) mm 组合模型与空气背景的最佳 8 mm 近红外模型分别预测验证集样品的含量, 并用 RMSEP 来比较 2 个模型的预测精度。

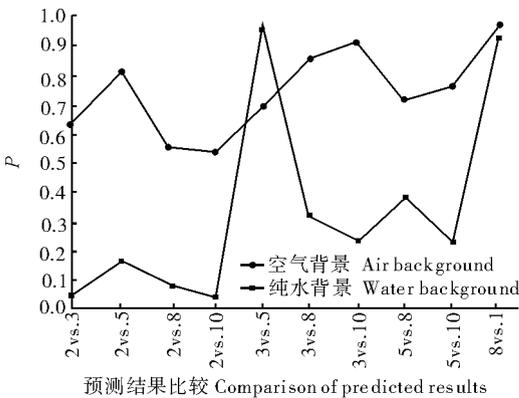


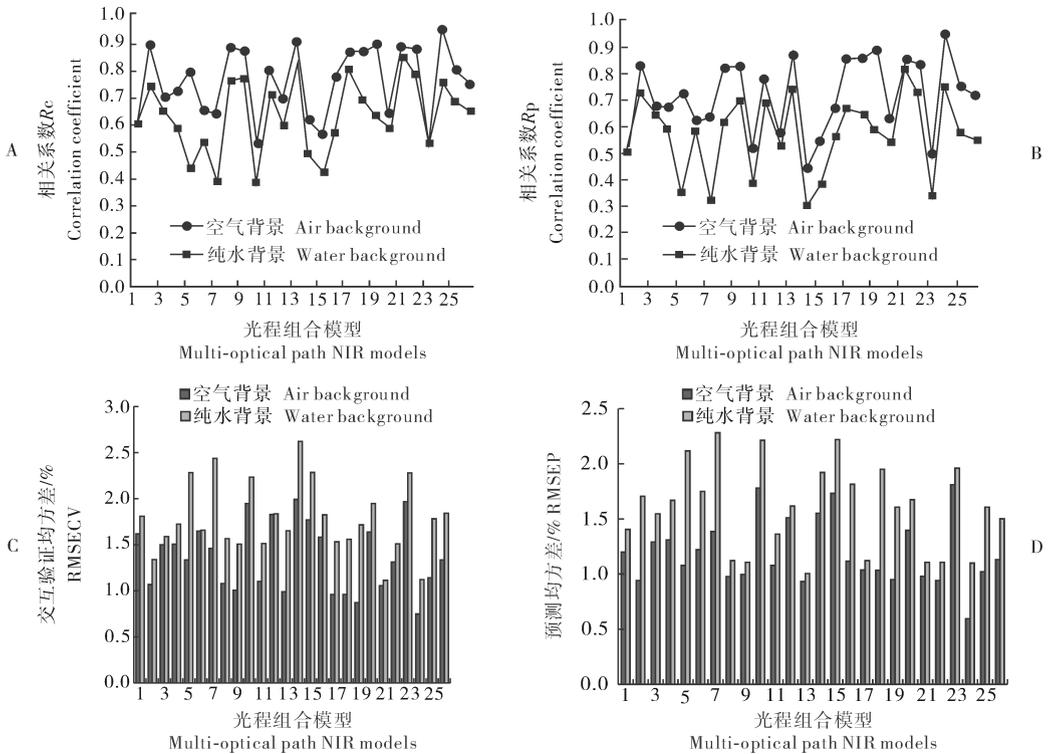
图 3 不同光程茶多酚近红

外预测模型结果间显著性检验 ($n = 30$)

Fig. 3 The significant testing between different optical path length NIR models on air and pure water background ($n = 30$)

其中(5+8) mm 光程组合近红外模型预测验证集样品 RMSEP 为 0.245, 8 mm 模型 RMSEP 为 0.361, (5+8) mm 组合模型的预测精度要优于单

光程 8 mm 的预测效果, 表明应用排列组合原理得到的茶多酚多光程组合近红外模型具有比单光程模型更好的预测精度, 模型的稳定性更好。



光程组合 Optical path combinations; 1: (2+3+5+8+10) mm; 2: (2+3+5+8) mm; 3: (2+3+5+10) mm; 4: (2+3+8+10) mm; 5: (2+5+8+10) mm; 6: (3+5+8+10) mm; 7: (2+8+10) mm; 8: (2+3+5) mm; 9: (2+3+8) mm; 10: (2+3+10) mm; 11: (2+5+8) mm; 12: (2+5+10) mm; 13: (3+5+8) mm; 14: (3+5+10) mm; 15: (3+8+10) mm; 16: (5+8+10) mm; 17: (2+3) mm; 18: (2+5) mm; 19: (2+8) mm; 20: (2+10) mm; 21: (3+5) mm; 22: (3+8) mm; 23: (3+10) mm; 24: (5+8) mm; 25: (5+10) mm; 26: (8+10) mm.

图 4 不同光程组合模型的近红外建模集和预测集结果

Fig. 4 The results of different optical path length combined NIR models

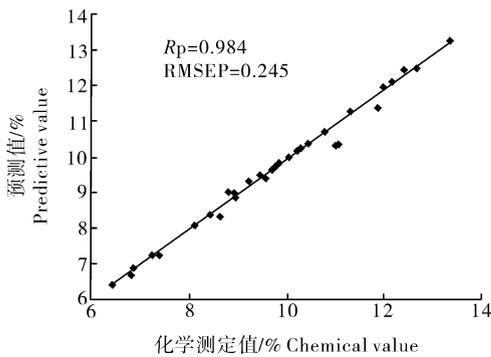


图 5 (5+8) mm 光程近红外模型

Fig. 5 The results of (5+8) mm optical path length NIR model

3 讨论

本文以红茶汤茶多酚为例, 研究了光谱的背景

和光程对近红外模型的影响, 而没有对光谱区段的选择做出论述。研究表明, 建立茶汤茶多酚近红外预测模型时, 以空气为背景扫描光谱, 因为以纯水为背景, 考虑到水的来源和水质因素, 每次扣除的水背景信息很可能存在不一致的情况, 影响试验的可重复性; 再加上纯水背景对茶多酚近红外预测模型的效果影响作用有限; 而以空气为背景, 空气要稳定得多, 光谱每次扣除的信息较一致, 试验的重复性会更好。在光程的选择上, (5+8) mm 的组合模型为最佳近红外模型, 模型的预测精度要好于单光程 8 mm 模型。与前人研究结果相比, 本文通过应用排列组合方法, 系统地研究了单光程和多光程对茶多酚近红外预测模型的影响, 得出的结果也更为全面和有说服力。本试验结果对红茶汤其他成分(如游

离氨基酸总量、咖啡碱)近红外模型的建立有很好的参考作用,也对茶饮料等其他产品的成分在线检测提供一定的指导意义。

参 考 文 献

- [1] CHEN Q S, ZHAO J W, VITTAYAPADUNG S. Identification of the green tea grade level using electronic tongue and pattern recognition[J]. *Food Research International*, 2008, 41: 500-504.
- [2] CHEN Q S, ZHAO J W, HUANG X Y, et al. Simultaneous determination of total polyphenols and caffeine contents of green tea by near-infrared reflectance spectroscopy[J]. *Journal of Microchemical*, 2006, 83: 42-47.
- [3] CHEN Q S, ZHAO J W, LIU M H, et al. Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms[J]. *Journal of Pharmaceutical and Biomedical Analysis*, 2008, 46: 568-573.
- [4] CHEN Q S, ZHAO J W, ZHANG H D, et al. Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration[J]. *Analytica Chimica Acta*, 2006, 572: 77-84.
- [5] 姜礼义, 刘福莉, 陈华才, 等. 绿茶汤中茶多酚近红外定量分析的光程选择[J]. *中国计量学院学报*, 2009, 20(2): 135-138.
- [6] 陈斌, 王豪, 林松, 等. 基于相关系数法与遗传算法的啤酒酒精度近红外光谱分析[J]. *农业工程学报*, 2005, 21(7): 99-102.
- [7] 刘宏欣, 张军, 黄富荣, 等. 近红外光谱法快速测定啤酒的主要品质参数[J]. *光谱学与光谱分析*, 2008, 28(2): 313-316.
- [8] 张林, 陆辉山, 高强, 等. 光程对液态调味品定性定量分析模型的影响研究[J]. *中国调味品*, 2014, 39(1): 28-31.
- [9] 杜敏, 吴志生, 林兆洲, 等. 光程对清开灵注射液中黄芩苷近红外定量模型的影响[J]. *药物分析杂志*, 2012, 32(10): 1796-1800.
- [10] 李刚, 赵喆, 刘蕊, 等. 利用多光程光谱法检测血液多种成分含量的研究[J]. *光谱学与光谱分析*, 2010, 30(9): 2381-2384.
- [11] THEESHAN B, AMITABYE L R, TEELUCK K G, et al. Functional nutraceuticals in chronic and acute diseases: molecular approach to mechanisms of action[J]. *Toxicology*, 2010, 278(1): 68-74.
- [12] LING T, JOSHUA D L. Antioxidant and pro-oxidant activities of green tea polyphenols in cancer prevention [J]. *Polyphenols in Human Health and Disease*, 2014(2): 1191-1198.

Effects of background and optical path length on predicting the contents of tea polyphenol with NIR models

WANG Sheng-peng¹ GONG Zi-ming¹ HE Yuan-jun² GAO Shi-wei¹
ZHENG Peng-cheng¹ TENG Jing¹ WANG Xue-ping¹ YE Fei¹ LU Su-fang¹

1. *Institute of Fruit & Tea, Hubei Academy of Agricultural Sciences, Wuhan 430064, China;*
2. *Enshi City Agricultural Bureau, Enshi 445000, China*

Abstract In order to establish the stable models, the air and pure water background NIRS models were compared. 2 mm, 3 mm, 5 mm, 8 mm and 10 mm optical path length NIR spectra of black tea's solution on air and pure water background were attained to set up single-optical and multi-optical path length NIR models through using partial least squares (PLS) and principle of permutation and combination. The results of using NIR models to predict unknown samples of black tea were examined by *t*-test statistical method. The results showed that there was no difference between the NIR models with the same optical path length and different background. There was significant difference ($P < 0.05$) between models with different optical path lengths on the pure water background. In the multi-optical path length models, the performances of NIR models on air background were better than that of the models on water background. The best NIR model combined 5 mm and 8 mm optical path length, whose predictable accuracy was better than that of single-optical path length models. The NIRS models of air background was better than that of pure water background.

Key words near infrared spectroscopy; single-optical path length NIR modeling; multi-optical path NIR modeling; air background; pure water background

(责任编辑:陆文昌)