

基于近红外光谱与组合间隔偏最小二乘法的 稻米镉含量快速检测

The feasibility of rapid determination of the cadmium content in rice based on near infrared spectroscopy and synergy interval partial least squares

朱向荣^{1,2} 李高阳^{1,2} 苏东林^{1,2} 刘伟^{1,2} 单杨^{1,2}

ZHU Xiang-rong^{1,2} LI Gao-yang^{1,2} SU Dong-lin^{1,2} LIU Wei^{1,2} SHAN Yang^{1,2}

(1. 中南大学研究生院隆平分院, 湖南 长沙 410125; 2. 湖南省食品测试分析中心, 湖南 长沙 410125)

(1. Longping Branch, Graduated School of Central South University, Changsha, Hunan 410125, China;

2. Hunan Food Test and Analysis Centre, Changsha, Hunan 410125, China)

摘要:采用近红外光谱(near infrared spectroscopy, NIRS)结合组合间隔偏最小二乘法(synergy interval partial least squares, siPLS)建立稻米镉含量快速检测的方法。收集并分析 72 个稻米样品的 NIRS 谱图。对光谱前处理方法进行优化, 确定平滑、多元散射校正与均值中心化处理为最优方法。采用 siPLS 法筛选特征波数, 建立稻米镉含量的定量模型。稻米镉 siPLS 模型交叉验证均方根(RMSECV)与预测均方根(RMSEP)值分别为 0.247 与 0.261, 训练集相关系数(Rc)与预测集相关系数(Rp)值分别为 0.919 与 0.895。结果表明: 运用 siPLS 法选择特征波长后, 不但可以降低模型的复杂度, 同时还能够提高预测精度。NIRS 作为一种快速、无损与便捷的初筛方法, 可用于检测稻米中镉含量是否超标。

关键词: 镉污染; 稻米; 定量测定; 近红外光谱; 组合间隔偏最小二乘法

Abstract: Near infrared spectroscopy (NIRS) combined with synergy interval partial least squares (siPLS) were used to establish the rapid method for the determination of cadmium content in rice. The NIR spectra of 72 rice samples were collected and analyzed. After optimization, an approach (SG smoothing combined with MSC and mean-centering) was adopted as a pre-treatment method for raw spectrum.

基金项目: 科技部“十二五”国家科技支撑计划(编号: 2012BAK17B17); 国家科技惠民计划(编号: 2012GS430202); 农业部农业科研杰出人才培养计划(农产品加工与质量安全创新团队); 湖南省科技重大专项(编号: 2011FJ1002-4)

作者简介: 朱向荣(1978—), 男, 助理研究员, 中南大学研究生院隆平分院在读博士生。E-mail: xiangrongchu@163.com

通讯作者: 单杨

收稿日期: 2015-05-06

SiPLS method was used as extraction methods of characteristic variables, and the NIRS quantitative model was built and predicted. The RMSECV and RMSEP of the siPLS model were 0.247 and 0.261, respectively. The Rc and Rp were 0.919 and 0.895, respectively. The results showed that siPLS algorithm not only could decrease the complexity of the model, but also improved the predictive precision. NIRS can be useful as a rapid, non-destructive and convenient analytical method for primary screening and detecting cadmium-polluted rice.

Keywords: cadmium-polluted; rice; quantitative determination; near-infrared spectroscopy; synergy interval partial least squares.

镉(cadmium, Cd)是一种毒性大, 迁移性强的重金属元素, Cd 在土壤中有较高活性^[1], Cd 污染是土壤和食品中重金属污染关注的首要问题之一。水稻被认为是镉吸收最强的大宗谷类作物^[2], 很容易吸收土壤中的镉。水稻镉污染不仅导致水稻产量下降, 更为严重的是镉元素在水稻植株内大量累积, 并通过食物链进入人体, 对人类健康带来严重危害^[3]。中国 GB 2762—2012《食品中污染物限量》规定稻米中镉的最高限量为 0.2 mg/kg, 因此, 对稻米中镉含量进行质量控制, 保障稻米食用安全, 采用灵敏、准确的分析方法检测稻米镉含量尤为重要。目前, 国内外研究者^[4]常采用电感耦合等离子体原子发射光谱与石墨炉原子吸收光谱法等仪器分析方法测定农产品中 Cd 含量。上述方法虽然准确性好, 专属性强、灵敏度高, 但也存在着样品消化时间长、强酸试剂消耗量大、环境危害较大、仪器操作复杂等缺点。

近红外光谱法(near-infrared spectroscopy, NIRS)是目前发展较为迅速的分子光谱快速检测技术, 具有快速、无损、绿色、环境友好等优点。NIRS 在稻米中淀粉、蛋白质、氨基酸、脂肪、矿质元素、水分、脂肪酸值等营养指标的含量检测

等领域均有应用。从原理上讲,无机元素在近红外光谱区并没有吸收,但是有机物质通过螯合或络合的方式与无机元素形成螯合物或络合物,而这些物质在近红外光谱区有相应的响应与吸收^[5]。稻米的主要有机成分为淀粉和蛋白质,其含量分别约为 80%和 8%。镉元素与淀粉及蛋白质中的 C—H、N—H 与 O—H 等基团结合,而使得采用近红外光谱检测镉成为可能。近年来,国外学者采用 NIRS 结合化学计量学方法对动物样品^[6, 7]、植物样品^[8-11] 以及水样^[12, 13] 中重金属元素定性、定量分析,Kumagai 等^[14] 采用 NIRS 对镉污染稻米进行定性研究,但尚未见到采用 NIRS 对稻米镉进行定量研究的报道。本研究拟通过 NIRS 与组合间隔偏最小二乘法(synergy interval partial least squares, siPLS)组合,建立稻米镉含量是否超标的 NIRS 定量模型,并对光谱预处理方法与光谱变量筛选进行系统的考察,以达到预期的效果。

1 材料与方 法

1.1 仪器与试剂

石墨炉原子吸收光谱仪:AA-6800 型,日本岛津公司;
傅里叶变换近红外光谱仪:Nicolet Antaris II 型,美国 Thermo 公司;
硝酸、高氯酸:优级纯,国药集团化学试剂有限公司。

1.2 试验方法

1.2.1 样品收集 在湖南省长沙市周边收集了 72 个稻米样品(40 个样品采自湘阴县浩河 口 镇,10 个样品采自长沙县大同桥镇,12 个样品采自长沙县江背镇,10 个样品采自长沙县干杉镇)。样品经烘干(85 ℃,2 h)、去壳、粉碎、过 60 目筛 4 个步骤。

1.2.2 镉含量的测定 按 GB/T 5009.15—2003《食品中镉的测定》执行。

1.2.3 NIRS 方法 利用积分球漫反射检测系统完成样品近红外光谱的采集,NIRS 扫描波数范围 4 000~10 000 cm⁻¹。NIRS 方法经优化后,样品光谱采集的工作流(workflow)确 定 为:分 辨 率 8 cm⁻¹,扫 描 次 数 32 次,增 益 为 2,以 内 置 背 景 为 参 考。每个样品重复 3 次扫描,取其平均光谱,以消除样品不均匀性带来的干扰。

1.2.4 组合间隔偏最小二乘法 偏最小二乘回归法(partial least squares regression,PLSR)是多元校正中经典的定量建模方法之一。在建模的过程中,为了不丢失光谱信息,PLSR 常采用全光谱(full spectrum)数据进行建模,但是全光谱中无信息变量(uninformative variables)过多,将会导致模型过拟合,预测结果变差。因此,需采用变量筛选方法提取特征变量,提高 NIRS 模型的精度。间隔偏最小二乘法^[15](interval partial least square,iPLS)是将全光谱分为 *n* 个等宽子区间,然后对每个子区间进行局部建模,得到误差最小,相关系数最大的子区间。组合间隔偏最小二乘法(siPLS)是 iPLS 的一个扩展,该方法的主要原理是以交叉验证均方根误差为指标,对全光谱不同子区间(比如 2、3 以及 4 个区间)进行任意组合,采用多元回归方法建立定量模型,最终得到最优的光谱变量区间组合以及相应的模型^[16]。

1.2.5 模型评价 采用模型相关系数(*R*)、交叉验证均方根(RMSECV)与预测集验证均方根(RMSEP)进行评价,分别按式(1)~(3)计算:

$$R = \sqrt{1 - \frac{\sum_{i=1}^m (\hat{c}_i - c_i)^2}{\sum_{i=1}^m (c_i - \bar{c}_i)^2}}$$
 (1)

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (\hat{c}_i - c_i)^2}{n}}$$
 (2)

$$RMSEP = \sqrt{\frac{\sum_{i=1}^m (\hat{c}_i - c_i)^2}{m}}$$
 (3)

式中:
 $\bar{c}_i$ ——包括样本集的所有样本试验值的均值,mg/kg;
 c_i ——试验值,mg/kg;
 $\hat{c}_i$ ——预测值,mg/kg;
n——校正集样本数;
m——预测集样本数。

1.2.6 统计分析 利用 MATLAB 软件进行。

2 结果与讨论

2.1 校正集的选择

采用 NIRS 数据进行建模,选择有代表性的校正集样品进入到校正模型的数据库,不但可以降低过拟合风险、提高模型的预测精度,还可以减少建模的工作量。本研究共收集 72 个稻米样本,采用基于马氏距离的 Kennard-Stone(KS)法^[17]依次从整个样本集(samples set)挑选 54 个样品作为校正集(calibration set),18 个样品作为预测集(prediction set)。

2.2 光谱分析

图 1 为代表性样品在 4 000~10 000 cm⁻¹ 的 NIRS 图,主要吸收峰光谱解析为^[18]:8 423.5 cm⁻¹附近的吸收峰为 C—H 键振动的二级倍频,6 907 cm⁻¹附近的吸收峰为 O—H 键伸缩振动的一级倍频,6 480 cm⁻¹附近的吸收峰为 N—H 键伸缩振动的一级倍频,5 720 cm⁻¹与 5 900 cm⁻¹两处吸收峰对应的是 C—H 键的一级倍频,5 200 cm⁻¹附近的吸收峰为 O—H 键振动组合频,4 830 cm⁻¹附近的吸收峰为 N—H 键伸缩振动与弯曲振动的组合频,4 390 cm⁻¹附近的吸收峰为 C—H 振动的组合频。

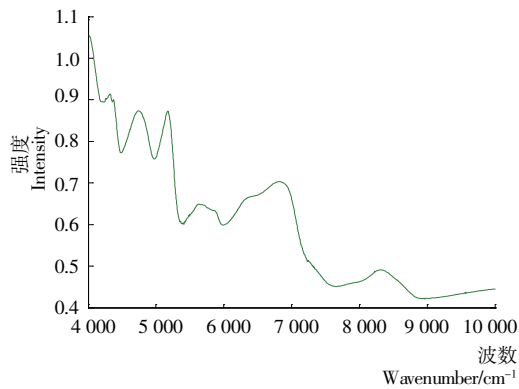


图 1 稻米样品代表性的近红外光谱图
Figure 1 NIR spectrum of rice sample

2.3 光谱预处理比较

采集的 NIRS 中,光谱的背景噪声、物理因素导致的光散射与量纲不一致等将对建模的准确性产生重要影响^[19]。因此,需要对光谱进行预处理,以减少上述影响。在全光谱范围内,以 RMSECV 值及 R_c (训练集相关系数)进行评价指标,考察了 Savitzky-Golay(SG)平滑、中心化(meaning)、自归一化(autoscaling)及多元散射校正(multiple scattering correction, MSC)等 5 种数据预处理方法对稻米样品的 NIR 光谱模型的影响。最终确定了平滑、中心化以及多元散射校正作为光谱预处理最优方法。

2.4 基于 siPLS 法的特征变量选择

本研究采集的近红外原始光谱包含 1 557 个变量,在多元变量校正时包含着不相关或冗余信息,利用全光谱建立校正模型存在着过拟合风险。因此,采用特征变量选择方法对全光谱变量进行筛选,能减少模型的变量数,降低模型的复杂性,提高模型的预测性能。本试验分析了 siPLS 特征变量提取方法,并与全光谱建模结果进行比较。SiPLS 法具体步骤:将全光谱划分为 60 个区间,每个区间 26 个变量,然后对不同区间进行组合,得到 RMSECV 值最小的子区间组合。本试验考察了 siPLS 法不同子区间个数的组合对模型的影响,确定 3 个区间组合次数时,RMSECV 值最低, R_c 值最大。

图 2 为训练集建模时前 10 个主因子数与所对应的 RMSECV 值的相关图。由图 2 可知,当主因子数为 9 时, RMSECV 值最小。

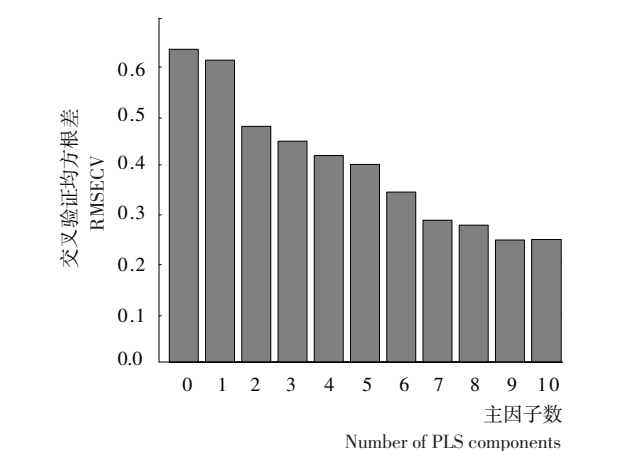


图 2 主因子数与 RMSECV 的相关图

Figure 2 The correlation diagram of PLS components and RMSECV

表 1 列出了前 10 个区间组合所对应的波数位置、主因子数以及 RMSECV 值,由表 1 可知,组合区间为[21 30 38]时,对应的主因子数为 9,此时 RMSECV 值最小,为 0.247。

采用 siPLS 法得到的最优区间组合为[21 30 38],21,30 与 38 三区间在全光谱对应的波数位置见图 3,分别为 6 005.2~6 101.6,6 907.8~7 004.2 与 7 710.0~7 806.4 cm^{-1} 。具体 NIRS 解析为^[18]: 6 005.2~6 101.6 cm^{-1} 为 C—H 伸缩振动

表 1 RMSECV 最低的前 10 个组合区间及其对应的主因子数

Table 1 The top 10 intervals corresponding to PLS component and RMSECV using siPLS

区间	主因子数	RMSECV
[21 30 38]	9	0.247
[21 30 39]	10	0.248
[6 21 30]	10	0.258
[21 23 30]	10	0.260
[21 27 30]	10	0.261
[9 21 30]	8	0.262
[14 21 30]	10	0.264
[21 28 30]	9	0.267
[21 30 31]	8	0.269
[20 21 42]	6	0.272

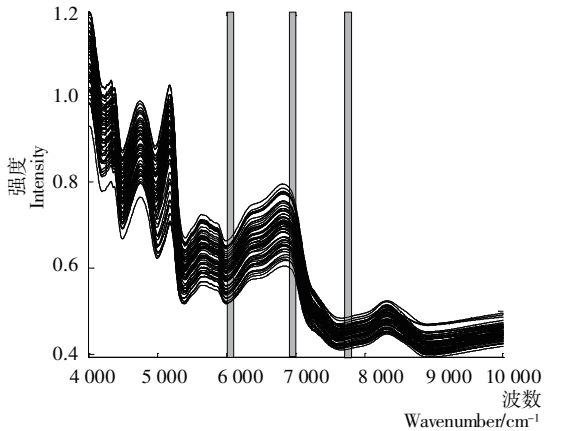


图 3 采用 siPLS 算法选择波段在全谱中的位置

Figure 3 Spectral subintervals were selected by siPLS algorithm

的一级倍频,6 907.8~7 004.2 cm^{-1} 为蛋白质 N—H 反对称伸缩振动一级倍频以及淀粉 O—H 伸缩振动的一级倍频。7 710.0~7 806.4 cm^{-1} 为 C—H 伸缩振动的一级倍频伸缩振动的第二组合频。

采用 siPLS 法选择 NIRS 特征吸收波数,建立稻米镉 PLS 定量模型并进行预测,训练集中 54 个样品 ICP—MS 测定的试验值与 NIRS 预测值的相关图见图 4。由图 4 可知,样本号为 8 井、30 井与 31 井的样品预测相对较差。此时,模型校正集 RMSECV 为 0.247,训练集相关系数 (R_c) 为 0.919。

图 5 为预测集中 18 个样品 ICP—MS 测定的试验值与 NIRS 预测值的相关图。预测集 RMSECV 为 0.261,预测集相关系数(R_p)为 0.895。

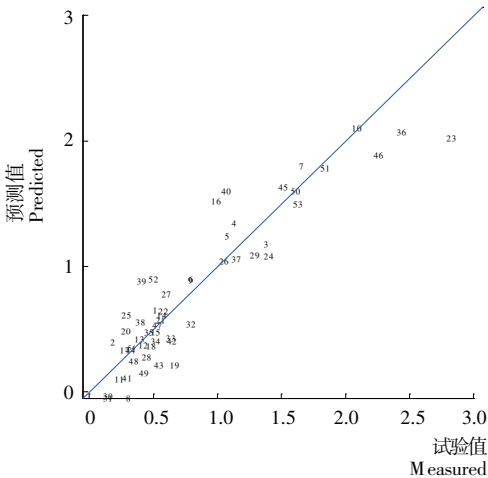


图 4 训练集中试验值与预测值的相关图

Figure 4 Prediction value vs. ICP—MS measurement for the determination of actual content in the calibration set ($n=54$)

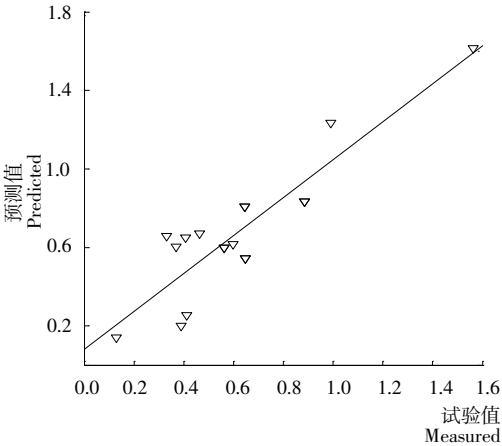


图 5 预测集中试验值与预测值的相关图

Figure 5 Prediction value vs. ICP—MS measurement for the determination of actual content in the calibration set ($n=18$)

2.5 全谱和特征波长选择模型的比较

为了评价 siPLS 法在稻米镉定量建模变量选择中的重要性,本研究采用 2.3 节相同的光谱预处理手段,建立了全光谱的 PLS 模型。由表 2 可知,采用全光谱建立的 PLS 模型的预测结果比 siPLS 法都要差,这说明采用 siPLS 法对原始光谱进行特征波长选择,主要成分信息得到最大保留,冗余信息得到部分去除,预测能力得到提高。

表 2 siPLS 法与 PLS 建模结果比较

Table 2 Comparison of the results of siPLS and PLS

方法	RMSECV	R_c	RMSEP	R_p
PLS	0.289	0.883	0.302	0.858
SiPLS	0.247	0.919	0.261	0.895

3 结论

本研究采用近红外光谱与组合间隔偏最小二乘法相结合,初步实现了镉污染稻米中镉含量的定量检测。结果表明,采用 siPLS 法选择特征变量再进行建模,所得结果优于全谱建模结果。本研究将进一步扩大样品量,提高模型的稳健性与准确性。本研究结果可为近红外光谱技术在稻米镉快速识别上提供初步依据。

参考文献

- 1 Chaney R L, Reeves P G, Ryan J A, et al. Improved understanding of hyperaccumulation yields commercial phytoextraction and phytomining technologies [J]. *Biometals*, 2004, 17 (5): 549~553.
- 2 Lindén A, Olsson I M, Bensryd I, et al. Monitoring of cadmium in the chain from soil via crops and feed to pig blood and kidney [J]. *Ecotoxicology and Environmental Safety*, 2003, 55 (2): 213~222.
- 3 Chen Gui-ping, Mei Yuan, Tao Wei, et al. Micro near infrared spectroscopy (MicroNIRS) based on on-line enrichment; determination of trace copper in water using glycidyl methacrylate-based monolithic material [J]. *Analytica Chimica Acta*, 2011, 670(1~2): 39~43.
- 4 Akamatsu S, Yoshioka N, Mitsuhashi T. Sensitive determination of cadmium in brown rice and spinach by flame atomic absorption spectrometry with solid-phase extraction[J]. *Food Additives & Contaminants: Part A*, 2012, 29(11): 1 696~1 700.
- 5 Kumagai M, Ohisa N, Amano T, et al. Canonical discriminant analysis of cadmium content levels in unpolished rice using a portable near-infrared spectrometer[J]. *Analytical Science*, 2003, 19 (11): 1 553~1 555.
- 6 González-Martín I, González-Pérez C, Hernández-Méndez J, et al. Mineral analysis (Fe,Zn,Ca,Na,K) of fresh Iberian pork loin by near infrared reflectance spectrometry; Determination of Fe, Na and K with a remote fibre-optic reflectance probe[J]. *Analytica Chimica Acta*, 2002, 468(2): 293~301.
- 7 Font R, Del Río-Celestino M, Vélez D, et al. Visible and near-infrared spectroscopy as a technique for screening the inorganic arsenic content in the red crayfish (*procambarus clarkii* Girard) [J]. *Analytical Chemistry*,2004,76(14): 3 893~3 898.
- 8 Font R, Del Río-Celestino M, Vélez D, et al. Use of near-infrared spectroscopy for determining the total arsenic content in prostrate amaranth[J]. *Science of the Total Environment*,2004,327 (1~3): 93~104.
- 9 Font R, Vélez D, Del Río-Celestino M, et al. Screening inorganic arsenic in rice by visible and near-infrared spectroscopy[J]. *Microchimica Acta*,2005,151(3~4): 231~239.
- 10 张龙, 潘家荣, 朱诚. 基于近红外光谱的重金属汞、镉和铅污染水稻叶片鉴别[J]. *浙江大学学报(农业与生命科学版)*, 2013, 39(1): 50~55.
- 11 刘燕德, 施宇. 香根草叶片铅含量近红外光谱快速检测[J]. *农业机械学报*, 2014, 45(3): 232~236.

表 3 广东省各地居民食用大米铅镉暴露风险评估结果

Table 3 Exposure risk assessment of lead and cadmium in edible rice in Guangdong Province

地区	Pb		Cd		
	均值	MOE	均值	EMI	占 PTMI 百分比/%
梅州	0.010	21.2	0.097	16.4	66
潮州	0.005	42.5	0.088	14.9	60
汕头	0.020	10.6	0.076	12.9	52
揭阳	0.005	42.5	0.059	10.0	40
汕尾	0.010	21.2	0.088	14.9	60
河源	0.014	15.2	0.090	15.3	61
惠州	0.010	21.2	0.131	22.2	89
韶关	0.010	21.2	0.161	27.3	109
清远	0.010	21.2	0.120	20.3	81
广州	0.020	10.6	0.100	17.0	68
东莞	0.014	15.2	0.137	23.2	93
深圳	0.020	10.6	0.064	10.8	43
珠海	0.010	21.2	0.077	13.1	52
中山	0.010	21.2	0.086	14.6	58
佛山	0.020	10.6	0.109	18.5	74
顺德	0.005	42.5	0.101	17.1	68
江门	0.010	21.2	0.095	16.1	64
肇庆	0.010	21.2	0.092	15.6	62
云浮	0.005	42.5	0.092	15.6	62
阳江	0.010	21.2	0.059	10.0	40
茂名	0.010	21.2	0.053	9.0	36
湛江	0.010	21.2	0.073	12.4	49

† 以人群的标准体重(60 kg)进行计算;大米摄入量按 339 g/d 计^[12]。

3 结论

根据 2012~2014 年广东省食品安全风险监测所采集的数据显示,广东省所辖各地大米中铅的含量尚处于安全水平,而大米镉的污染则不容乐观但具有很强地域性差异。根据本研究结果表明,广东省人群通过大米而摄入铅的量总体

是安全的,广东省内大部分地区人群通过大米而摄入镉的量也是安全的,但部分地区大米镉的暴露潜在一定的风险,主要集中于粤中北部地区,应在今后加大此部分地区大米食用安全的监控力度。

参考文献

1 赖利. 食品的金属污染[M]. 第 1 版. 北京:轻工业出版社,1986: 90~130.

2 谭见安. 地球环境与健康[M]. 北京:化学工业出版社,环境科学与工业出版社,2004:154~155.

3 Hartwig A, Schwerdtle T. Interactions by carcinogenic metal compounds with DNA repair processes: toxicological implications [J]. Toxicol Lett,2002,127(1~3):47~54.

4 JECFA. WHO Technical Report Series 884. Evaluation of Certain Food Additives and Contaminants. Forty-ninth report of the Joint FAO/WHO Expert Committee on Food Additives[R]. Geneva:WHO,1999.

5 Joint FAO/WHO Expert Committee on Food Additives (JECFA). Summary and conclusions of the seventy-third meeting of joint FAO/WHO expert committee on food additives[R]. Geneva: FAO/WHO, 2010.

6 杨梦昕,付湘晋,李忠海,等. 湘江流域重金属污染情况及其对食物链的影响[J]. 食品与机械,2014,30(5):103~106.

7 刘雪源. 中式烹饪过程食品安全研究进展[J]. 食品与机械,2013, 29(6):262~267,272.

8 徐娇,邵兵. 试论食品安全风险评估制度[J]. 中国卫生监督杂志, 2011(4):342~350.

9 Joint FAO/WHO Expert Committee. Food Additives Summary and Conclusion, JECFA /73 / SC[R]. Geneva: WHO, 2010.

10 Reffstrup TK, Larsen JC, Meyer O. Risk assessment of mixtures of pesticides. Current approaches and future strategies[J]. Regul Toxicol Pharmacol, 2010, 56(2):174~192.

11 Agusa T, Kunito T, Sudaryanto A, et al. Exposure assessment for trace elements from consumption of marine fish in Southeast Asia[J]. Environmental Pollution,2007,145(3):766~777.

12 唐洪磊,郭英,孟祥周,等. 广东省沿海城市居民膳食结构及食物污染状况的调研——对持久性卤代烃和重金属的人体暴露水平评价[J]. 农业环境科学学报 2009,28(2):329~336.

2003, 19(11): 1 553~1 555.

15 Norgaard L, Saudland A, Wagner J, et al. Interval partial least squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy[J]. Applied Spectroscopy, 2000, 54(3): 413~419.

16 丁姣,蔡建荣,张海东,等. 近红外结合 Si-ELM 检测食醋品质指标[J]. 食品与机械, 2012, 28(1): 93~96.

17 Kennard R W, Stone L A. Computer aided design of experiments[J]. Technometrics, 1969(11): 137~148.

18 陆婉珍. 现代近红外光谱分析技术[M]. 第二版. 北京:中国石化出版社,2006.

19 张菊华,朱向荣,苏东林,等. 茶油品质鉴别的透射和透反射模式分析比较[J]. 食品与机械, 2012, 28(1): 101~104.

(上接第 46 页)

12 Ning Yu, Li Ji-hui, Cai Wen-sheng, et al. Simultaneous determination of heavy metal ions in water using near-infrared spectroscopy with preconcentration by nano-hydroxyapatite [J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2012(96): 289~294.

13 Huang Zi-xia, Tao Wei, Fang Juan-juan, et al. Multivariate calibration of on-line enrichment near-infrared (NIR) spectra and determination of trace lead in water[J]. Chemometrics and Intelligent Laboratory Systems, 2009, 98(2): 195~200.

14 Kumagai M, Ohisa N, Amano T, et al. Canonical discriminant analysis of cadmium content levels in unpolished rice using a portable near-infrared spectrometer [J]. Analytical Sciences,