

3 步混合变量选择策略在鸡肉近红外水分检测中的应用

Research on the application of three-step hybrid variable selection strategy in chicken moisture detection by near infrared

袁 凯 张志勇 席 前 伍莹芮 郭东升 何国康

YUAN Kai ZHANG Zhi-yong XI Qian WU Ying-rui GUO Dong-sheng HE Guo-kang

(山西农业大学工学院, 山西 太谷 030801)

(College of Engineering, Shanxi Agricultural University, Taigu, Shanxi 030801, China)

摘要:基于光谱多元校正中有效变量选择的 3 步混合策略(初筛、精挑、细选),提出了间隔偏最小二乘(iPLS)、区间变量迭代空间收缩法(iVISSA)和迭代保留信息变量(IRIV)联用的特征变量选择方法,对生鲜鸡胸肉的近红外光谱进行特征波长选择,建立了鸡肉水分检测模型。结果表明,建模波长数量经 iPLS-iVISSA-IRIV 3 步选择后减少为全光谱建模的 0.76%,但模型精确度和稳定性逐步提高。选定 8 个特征波长建模,其校正相关系数 $R_C = 0.907\ 7$,校正均方根误差 $RMSEC = 0.516\ 1$;预测相关系数 $R_P = 0.943\ 5$,预测均方根误差 $RMSEP = 0.612\ 3$ 。表明基于 3 步混合策略提出的 iPLS-iVISSA-IRIV 方法能有效选择鸡肉水分检测的特征波长。

关键词:近红外光谱;混合策略;变量选择;鸡肉;水分

Abstract: Based on the three-step hybrid strategy of effective variable selection (rough selection, fine selection and optimal selection) in multivariate calibration of spectra, a feature variable selection method combining interval partial least squares (iPLS), interval variable iterative space shrinkage approach (iVISSA) and iteratively retaining informative variables (IRIV) was proposed to select the feature wavelengths of the near-infrared spectra of fresh chicken breast, and established a chicken moisture detection model. The number of modeled wavelengths was reduced by 0.76% after iPLS-iVISSA-IRIV, but the accuracy and stability of the model are gradually improved. The modeling results using the selected 8 characteristic wavelengths were as follows: correlation

coefficient of calibration $R_C = 0.907\ 7$, root mean square error of calibration $RMSEC = 0.516\ 1$, correlation coefficient of prediction $R_P = 0.943\ 5$, root mean square error of prediction $RMSEP = 0.612\ 3$. The result shows that the iPLS-iVISSA-IRIV method based on the three-step hybrid strategy can effectively select the characteristic wavelengths of chicken moisture detection.

Keywords: NIR spectra; hybrid strategy; variable selection; chicken; moisture

水分是生鲜肉品质评价的一个重要参数,对肉质、口感以及肉的加工、运输、贮藏有直接影响^[1]。研究^[2-4]表明,近红外光谱能应用于生鲜肉品质的快速、无损检测。高分辨率的近红外光谱仪器能采集到丰富的物质成分信息,但数据中往往包含大量冗余信息,直接用其建模不仅增加了建模难度,且影响模型的精确度和泛化能力,因此特征变量选择成为解决这一问题的有效途径^[5-7]。当面对高维数据时,一些寻找最优变量子集的方法容易过度拟合且需要大量计算,有些方法则不适用,因此结合两种或两种以上的有效变量选择方法成为解决这一问题的新思路。孔庆明等^[8]利用 iPLS-SPA 对小麦秸秆近红外光谱进行特征波长选择,波长数量从 520 降至 10; Tang 等^[9]利用 CARS-SPA 选择的变量比单独使用 CARS 选择的变量更少,且用所选波长建模精度要优于全光谱建模; Xu 等^[10]利用 CARS-GA 选取特征波长,波长数量从 1 557 降至 53,建立了预测竹笋粗纤维含量的模型; Li 等^[11]用 MC-UVE-SPA 选取特征波长,建立了预测梨果实 SSC 和硬度的 LV-SVM 模型,这些研究主要使用了两种方法相结合进行特征波长选择,第 2 种方法对第 1 种方法选择的变量进一步组合、优化。

试验拟基于近红外光谱多元校正中有效变量选择的 3 步混合策略(初筛、精挑、细选)^[12],提出一种 iPLS-

基金项目:山西省自然科学基金(编号:201701D121103);国家重点研发计划项目(编号:2016YFD0701801)

作者简介:袁凯,男,山西农业大学在读硕士研究生。

通信作者:张志勇(1979—),男,山西农业大学副教授,博士。

E-mail: zzyzyy1979@163.com

收稿日期:2020-07-03

iVISSA-IRIV 3 种方法相结合的特征变量选择方法,旨在为基于分立波长元件的便携式鸡肉水分检测仪的设计提供依据。

1 试验与方法

1.1 主要仪器与设备

近红外光谱仪:Field Spec3 型,波长范围 350 ~ 2 500 nm,美国 ASD 公司;

电热鼓风箱:101-2AB 型,北京心雨仪器仪表有限公司;

分析天平:BSM120.4 型,精度 0.1 mg,上海卓精电子科技有限公司;

海尔电冰箱:BC/BD-101HBZ 型,青岛海尔特种电冰柜有限公司;

恒温水浴锅:HH-W600 型,济南欧莱博科学仪器有限公司;

手动绞肉机:NS-001 型,4 mm 孔径,浙江省荣博厨具有限公司。

1.2 样品及样品处理

冷鲜鸡胸肉:60 块,购买后立即运往无菌实验室,在超净工作平台上将鸡胸肉修整成 4 cm×4 cm×1.5 cm 的肉块,随后装入保鲜袋并放入 0 °C 的冰箱中保存,5 d 内完成鸡胸肉反射光谱的采集和水分测定,市售。

1.3 反射光谱采集

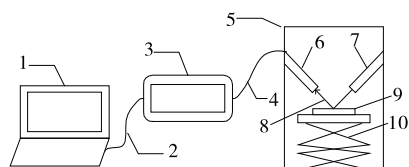
自制的光谱采集装置如图 1 所示,每次采集样本光谱前,将待采集的鸡胸肉从冰箱取出,室温下解冻 30 min,且光谱仪开机预热 30 min 后开始采集。采集时将肉块置于样品台上,为了减小每个样本距离光源和探头不一致的误差,调节平台使所有肉块上表面高度保持一致。考虑到鸡肉内部组织差异性较大,光谱仪扫描次数设为 3 次,取平均值作为样本原始光谱。

1.4 水分测定

每个鸡肉样本在采集完光谱数据后,根据国家食品安全标准《食品中水分的测定》中的直接干燥法进行水分测定。

1.5 数据处理及方法

采用 Matlab 2015b 软件进行数据处理。由于 350~



1. 计算机 2. 数据线 3. 光谱仪 4. 光纤 5. 暗箱 6. 采集探头
7. 光源 8. 入射与反射光线 9. 样品 10. 样品台(高度可调节)

图 1 光谱采集系统

Figure 1 Spectral acquisition system

400 nm 的光谱数据存在较大噪声信号,且 1 451 ~ 2 500 nm 的光谱数据几乎没有起伏变化,可视为“无信息变量”,故选取 401~1 450 nm 作为原始光谱,如图 2 所示。多元散射校正(Multiplicative scatter correction, MSC)能有效改善样本颗粒大小不一、表面散射不均以及光程变化对漫反射光谱采集的影响,而中心化法(Mean centering, MC)能增强样品光谱之间的特异性。因此,后续模型构建中,采用两种方法联合对光谱数据进行预处理,图 3 为 60 个样本经预处理后的光谱图。

1.5.1 3 步混合策略 有效波长变量选择的 3 步混合策略分为初筛、精挑和细选。初筛的目的是选出几个含信息量丰富的波长区间,通过波长间隔选择算法来实现。精挑是从第 1 步所选的波长间隔中进一步选择,滤除波长间隔中的噪声和非信息变量,再次缩小变量空间。细选是对第 2 步所保留的波长变量进一步组合、优化,达到使所选变量最优、最少的目的。此策略能避免变量选择方法的随机组合,克服单种方法对大量变量难以达到最优选择和组合的缺点^[12]。

1.5.2 特征波长选择方法

(1) 区间偏最小二乘(Interval partial least squares, iPLS):iPLS 是一种最具代表性的波长区间选择算法,具有对变量解释性强的优点^[13]。首先按设定宽度将整个光谱区域划分为 n 个子区间,在每个子区间上建立待测品

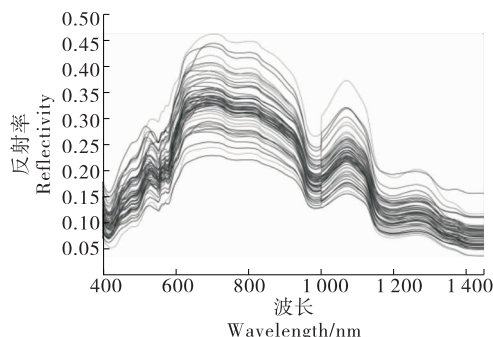


图 2 原始光谱

Figure 2 Original spectrum

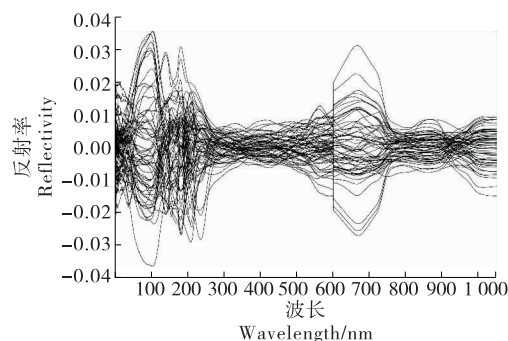


图 3 预处理光谱

Figure 3 Pretreatment spectrum

质的局部 PLS 回归模型,得到 n 个子模型,取前 k 个交叉验证均方根误差(RMSECV)最低的子模型所在区间为特征波长区间。

(2) 区间变量迭代空间收缩法(Interval variable iterative space shrinkage approach, iVISSA):该算法也是一种波长间隔选择算法,基于模型总体分析(MPA),利用权重二进制矩阵采样(WBMS)法更改各变量权重,若所有子模型的均方根误差与预测误差不再改变,取权重为 1 的变量为特征变量,且将选中变量的相邻变量放入子模型中进行回归分析,全局分析和局部分析交替进行,能同时对光谱间隔、位置和组合进行优化^[14]。

(3) 迭代保留信息变量算法(Iteratively retaining informative variables, IRIV):该算法将波长变量按重要性分成 4 个等级,分别为强信息变量、弱信息变量、无信息变量和干扰信息变量^[15]。将采样矩阵采集到的变量随机组合,并分别建立各组合的 PLS 回归模型,用 RMSECV 衡量各模型精度。基于 MPA,每次循环迭代都计算包含和不包含此变量时的 RMSECV 平均值之差,并根据 Mann-Whitney U 检验的 P 值来判定该变量的等级,最后经反向消除保留强信息变量和弱信息变量,从而获得最优特征波长变量。

1.5.3 模型构建方法 偏最小二乘回归(Partial least squares regression, PLSR)是一种多元统计分析方法。首先对光谱矩阵 $\mathbf{X}$ 和水分参照值矩阵 $\mathbf{Y}$ 进行特征分解,分别得到 $\mathbf{X}$ 和 $\mathbf{Y}$ 的特征因子矩阵(得分矩阵),然后建立特征因子矩阵的多元线性回归模型,最优潜在变量数由交叉验证法得到的 RMSECV 值确定,潜在变量数确定回归系数矩阵 $\mathbf{B}$ 也随之确定,最后经数据重构得到 $\mathbf{X}$ 和 $\mathbf{Y}$ 的回归模型。

2 结果与分析

2.1 水分含量标准值测量

表 1 为 60 个鸡肉样本水分含量统计表,用 Kennard-Stone^[16]将样本按 4:1 比例划分为 48 个校正集和 12 个预测集。

2.2 全波段建模

偏最小二乘回归(PLSR)是近红外分析中最典型、最常用的建模方法。校正集用来构建模型,预测集用来检验模型的稳定性,并根据 RMSECV 值确定最佳主因子

数。用全波段光谱数据建模,当 RMSECV 值最小为 0.782 9 时,主因子数为 7,此时校正模型 $R_c=0.886 6$, $RMSEC=0.568 8$;预测模型 $R_p=0.877 9$, $RMSEP=0.905 4$,结果如图 4 所示,表明模型精度有待提高。

2.3 三步混合策略优选特征波段(波长)建模

2.3.1 初筛 将 401~1 450 nm 整个光谱区以 40 nm 的波长间隔分成 26 个区间,选出前 10 个 RMSECV 值最低的波长区间,结果如图 5 所示。用所选波段建模,结果如图 6 所示,模型的 RMSECV 值最小为 0.699 4 时的主因子数为 7,此时校正模型 $R_c=0.900 0$, $RMSEC=0.536 1$;预测模型 $R_p=0.893 7$, $RMSEP=0.820 3$,相比全波段建模,模型精度有所提高。由图 5 可知,选中波段大多是波峰和波谷段,例如 760,980 nm 附近有明显的吸收峰,

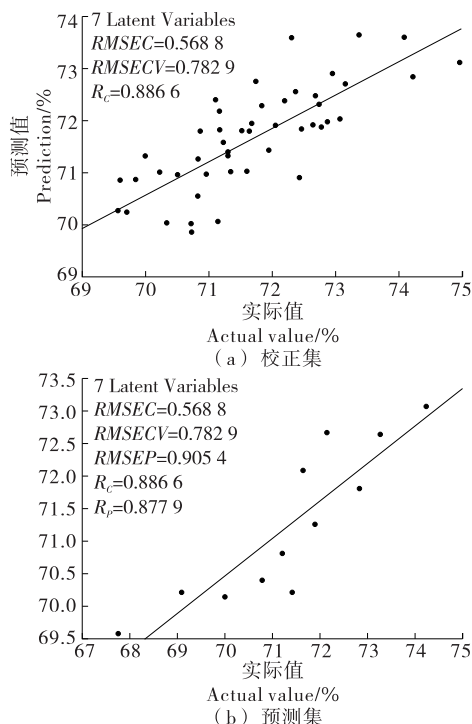


图 4 全波段建模

Figure 4 Full spectrum modeling

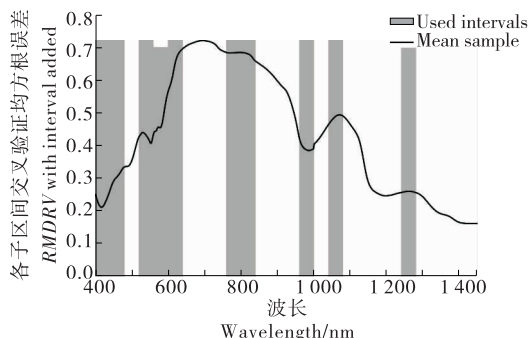


图 5 iPLS 所选波段

Figure 5 iPLS selected band

表 1 样本水分统计表

Table 1 Sample moisture statistics table

样本集	数量	水分范围/%	平均值/%
校正集	48	69.58~74.96	71.67
预测集	12	67.78~74.23	71.36
总样本集	60	67.78~74.96	71.61

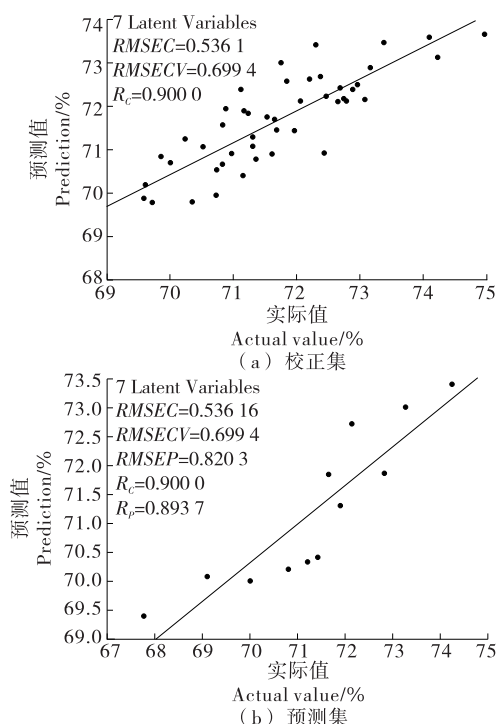


图6 iPLS所选波段建模

Figure 6 iPLS selected band modeling

特别是 980 nm 附近。而 760, 980 nm 处为水的特征吸收波长, 说明 iPLS 能有效筛选出与水分相关的变量, 供下一步选择。

2.3.2 精选 从第 1 步选择的 400 个波长中选中 191 个波长, 如图 7 所示, 760, 980 nm 附近的波长再次被选中, 说明经 iVISSA 精选, 能实现对波长间隔、位置和组合的同时优化, 在保留重要变量的同时, 滤除了 iPLS 所选变量中的噪声和非信息变量。使用所选波长建立 PLS 模型, 结果如图 8 所示, $RMSECV$ 最小值为 0.690 3 时的主因子数为 8, 此时校正模型 $R_c=0.915 6$, $RMSEC=0.494 3$; 预测模型 $R_p=0.923 4$, $RMSEP=0.732 3$, 模型精度和稳定性再次提高。

2.3.3 细选 由第 2 步所选变量结果可知, 变量大都集

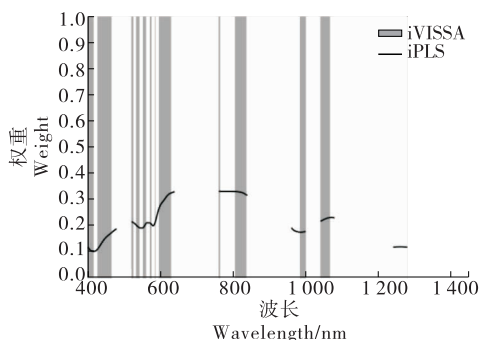


图7 iVISSA所选波段

Figure 7 iVISSA selected band band

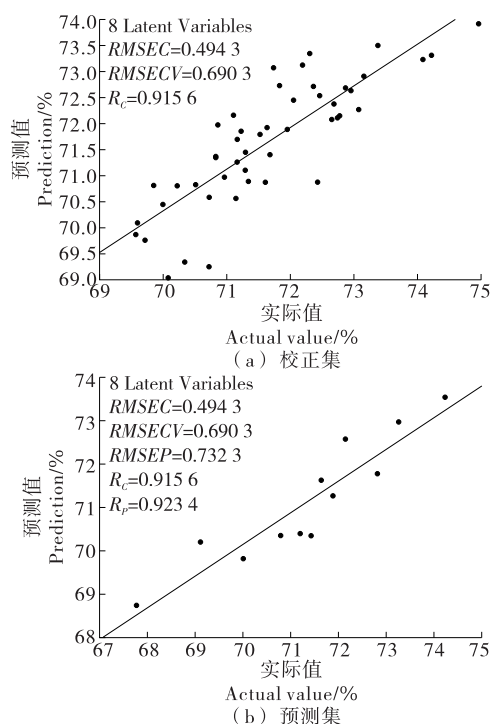


图8 iVISSA所选波段建模

Figure 8 iVISSA selected band modeling

中于 401~450 nm 和 500~600 nm, 而连续波长之间存在较强的共线性关系, 说明此结果仍存在继续优化、精简的可能性。经 IRIV 细选, 从第 2 步所选的 191 个波长中选出 8 个波长, 如图 9 所示, 分别为 402, 442, 452, 553, 555, 586, 999, 1 042 nm, 用所选波长建立 PLS 模型, 结果如图 10 所示, $RMSECV$ 值最小为 0.652 6, 校正模型 $R_c=0.907 7$, $RMSEC=0.516 1$; 预测模型 $R_p=0.943 5$, $RMSEP=0.612 3$ 。与第 2 步建模结果相比, 校正相关系数基本保持不变, 预测相关系数再次提高, 且 $RMSECV$ 值持续降低。说明经 IRIV 细选, 变量空间大幅缩减, 模型精度和稳定性再次有所提高。

2.4 结果统计

由图 11 可知, 经 3 步混合选择, 变量空间从 1 050 维

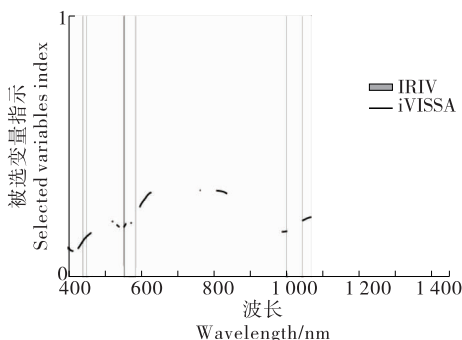


图9 IRIV所选波长

Figure 9 IRIV selected variables

缩减至 8 维, $RMSECV$ 值逐步降低。虽然经过第 3 步 IRIV, 模型的校正相关系数有略微下降, 但预测相关系数再次提高, 整体而言, 模型在逐步简化、稳定。最终选定的 8 个特征波长包含 999 nm, 此波长位于水的特征吸收

波长(980 nm)附近, 表明基于 3 步混合策略提出的 iPLS-iVISSA-IRIV 方法能有效筛选出鸡肉光谱数据中与水分相关的信息。

3 结论

基于 3 步混合策略提出了 iPLS-iVISSA-IRIV 特征波长选择方法, 结合偏最小二乘回归建立了生鲜鸡肉水分检测模型, 取得了较为满意的结果。结果表明, iPLS、iVISSA 和 IRIV 3 步混合能有效筛选出检测鸡肉水分的特征波长, 在稳步缩减变量空间的同时, 减小了有效变量误剔除的可能性以及建模的复杂度, 并逐步提高模型的精确度和稳定性。最终选定 8 个特征波长所建模型的校正相关系数为 0.907 7, 预测相关系数为 0.943 5。

参考文献

- [1] 闵连吉. 肉的科学加工技术[M]. 北京: 中国食品出版社, 1988: 9.
- [2] 赵文英, 花锦, 张梨花, 等. 近红外光谱测定不同鲜肉肉糜中蛋白质含量[J]. 食品与机械, 2017, 33(1): 48-50, 118.
- [3] 谢安国, 康怀彬, 王飞翔, 等. 高光谱成像检测煎制中调理牛肉品质的变化[J]. 食品与机械, 2018, 34(11): 20-23, 54.
- [4] 田卫新, 何丹丹, 杨东, 等. 一种基于高光谱图像的熟牛肉 TVB-N 含量预测方法[J]. 食品与机械, 2016, 32(12): 70-74.
- [5] 魏文松, 彭彦昆, 郑晓春, 等. 基于优选波长的多光谱检测系统快速检测猪肉中挥发性盐基氮的含量[J]. 光学学报, 2017, 37(11): 382-393.
- [6] JIA Bei-bei, YOON S C, ZHUANG Hong, et al. Prediction of pH of fresh chicken breast fillets by VNIR hyperspectral imaging[J]. Journal of Food Engineering, 2017, 208(SEP): 57-65.
- [7] FENG Yao-ze, SUN Da-wen. Near-infrared hyperspectral imaging in tandem with partial least squares regression and genetic algorithm for non-destructive determination and visualization of Pseudomonas loads in chicken fillets[J]. Talanta, 2013, 109: 74-83.
- [8] 孔庆明, 苏中滨, 沈维政, 等. IPLS-SPA 波长选择方法在近红外秸秆生物量中的应用研究[J]. 光谱学与光谱分析, 2015, 35(5): 1 233-1 238.
- [9] TANG Guo, HUANG Yue, TIAN Kuang-da, et al. A new spectral variable selection pattern using competitive adaptive reweighted sampling combined with successive projections algorithm[J]. The Analyst, 2014, 139(19): 4 894-4 902.
- [10] XU Fu-bin, HUANG Xing-yi, DAI Huang, et al. Nondestructive determination of bamboo shoots lignification using FT-NIR with efficient variables selection algorithms[J]. Analytical Methods, 2014, 6(4): 1 090.
- [11] LI Jiang-bo, ZHAO Chun-jiang, HUANG Wen-qian, et al. A combination algorithm for variable selection to determine soluble solid content and firmness of pears[J]. Analytical Methods, 2014, 6(7): 2 170.

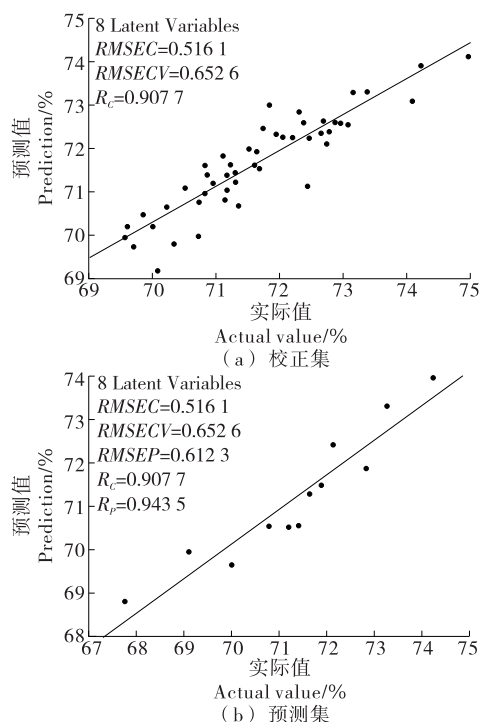


图 10 IRIV 所选波长建模

Figure 10 IRIV selected variables modeling

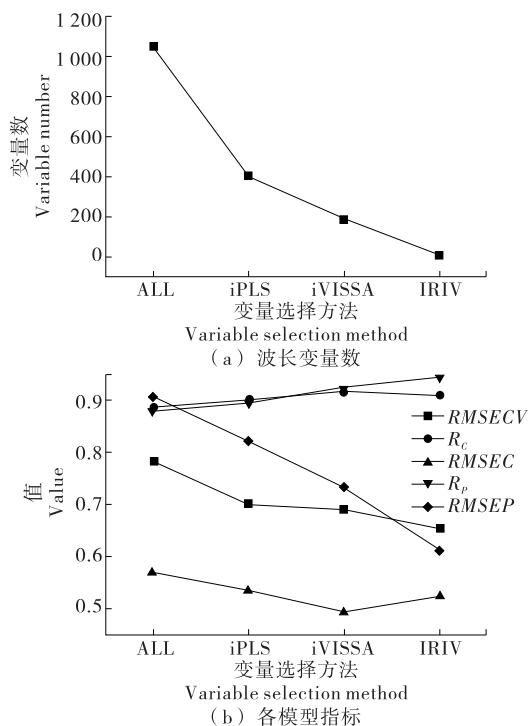


图 11 试验结果统计

Figure 11 Statistics of test results

(下转第 81 页)

3 结论

试验通过优化色谱条件、质谱条件和前处理条件,建立了同位素稀释—气相色谱—质谱法测定婴幼儿奶粉中8种己二酸酯类塑化剂含量的方法。该方法采用正己烷作为提取溶剂,超声提取法萃取,内标法定量。而采用超声提取可以实现快速、高效、低耗能,具有操作简单、方便、易于实施等特点。经添加回收的方式考察了方法的准确度和精确度,结果表明,试验建立的检测方法回收率高、重现性好、灵敏度高、能够快速实现,适用于婴幼儿配方奶粉中己二酸酯类塑化剂含量的分析测定。

参考文献

- [1] 崔海波. 塑化剂己二酸二丁酯的合成现状和发展[J]. 西部皮革, 2016, 38(6): 24.
- [2] 李沂濛, 史建公, 张毅, 等. 聚氯乙稀耐寒增塑剂(DOA)的市场需求分析[J]. 工业技术创新, 2015, 2(2): 155-159.
- [3] 刘俊, 朱然, 田延河, 等. 气相色谱—质谱法对食品包装材料中邻苯二甲酸酯类与己二酸酯类增塑剂的同时测定[J]. 分析测试学报, 2010, 29(9): 943-947.
- [4] 吴景武, 张伟亚, 刘丽, 等. 气相色谱—质谱法测定 PVC 食品保鲜膜中 DEHA 等己二酸酯类增塑剂[J]. 中国卫生检验杂志, 2006, 16(7): 817-818.
- [5] 陈伟, 马满英, 周鹏, 等. 塑料瓶装食醋中塑化剂的检测及其毒性分析[J]. 绿色科技, 2018(8): 130-132.
- [6] 马腾洲, 周宇艳, 张琳. 己二酸酯类增塑剂向橄榄油中迁移量的测定方法[J]. 沈阳航空航天大学学报, 2011, 28(4): 92-94.
- [7] SILVA M J, SAMANDAR E, YE Xiao-yun, et al. In vitro metabolites of di-2-ethylhexyl adipate (DEHA) as biomarkers of exposure in human biomonitoring applications[J]. Chemical Research in Toxicology, 2013, 26(10): 1 498-1 502.
- [8] 金银龙, 鄂学礼, 陈昌杰, 等. GB 5749—2006 生活饮用水卫生标准[S]. 北京: 中国标准出版社, 2006: 5.
- [9] 陈志锋, 孙利, 雍炜, 等. 溶解沉淀—气相色谱法测定聚氯乙烯食品保鲜膜中增塑剂己二酸二(2-乙基)己酯(DEHA)含量[J]. 中国卫生检验杂志, 2006, 16(7): 772-774.
- [10] FIERENS T, HOLDERBEKE M V, WILLEMS H, et al. Phthalates in Belgian cow's milk and the role of feed and other contamination pathways at farm level[J]. Food and Chemical Toxicology An International Journal Published for the British Industrial Biological Research Association, 2012, 50(8): 2 945.
- [11] JIA Wei, CHU Xiao-gang, LING Yun, et al. Analysis of phthalates in milk and milk products by liquid chromatography coupled to quadrupole Orbitrap high-resolution mass spectrometry[J]. Journal of Chromatography A, 2014 (1 362): 110-118.
- [12] 黄永辉. SPE-GC-MS 法同时测定奶粉中 18 种邻苯二甲酸酯迁移量[J]. 食品研究与开发, 2010, 31(11): 167-171.
- [13] 康莉, 廖仕成, 陈裕华, 等. 婴幼儿奶粉中邻苯二甲酸酯的迁移残留分析[J]. 中国乳品工业, 2014, 42(6): 41-43.
- [14] 单晓梅, 沈登辉, 陆蓓蓓, 等. 固相萃取—气质联用法测定水中己二酸酯类增塑剂[J]. 环境卫生学杂志, 2013, 3(1): 53-59.
- [15] 蔡艳, 赵家青, 贺小雨, 等. 食用油中邻苯二甲酸酯和己二酸酯的 GC-MS 定量分析与迁移特性研究[J]. 食品科学, 2014, 35(12): 191-195.
- [16] 伍竞成, 方从容, 荫硕焱, 等. 同位素稀释—气相色谱—质谱法测定方便面调味料中 DEHA 和 16 种 PAEs 类增塑剂[J]. 分析实验室, 2015, 34(2): 146-150.
- [17] ANTONIOS E G, ELISAVET S, MICHAEL G K. Di(2-ethylhexyl) adipate migration from PVC-cling film into packaged sea bream (*Sparus aurata*) and rainbow trout (*Oncorhynchus mykiss*) fillets: Kinetic study and control of compliance with EU specifications[J]. European Food Research and Technology, 2008, 226(4): 915-923.
- [18] 焦逊, 赵鹏, 徐龙华, 等. PVC 保鲜膜中 DEHP 和 DEHA 向猪肉中迁移规律研究[J]. 食品科学技术学报, 2018, 36(4): 55-60.
- [19] FASANO E, CIRILLO T, ESPOSITO F, et al. Migration of monomers and plasticizers from packed foods and heated microwave foods using QuEChERS sample preparation and gas chromatography/mass spectrometry[J]. LWT-Food Science and Technology, 2015(64): 1 015-1 021.
- [20] ANTONIOS E G, PANAGIOTA Z, ANDREAS K, et al. Migration of di(2-ethylhexyl) adipate and acetyltributyl citrate plasticizers from food-grade PVC film into sweetened sesame paste (halawa tehneh): Kinetic and penetration study[J]. Food and Chemical Toxicology, 2007(45): 585-591.
- [12] YU Hai-dong, YUN Yong-huan, ZHANG Wei-min, et al. Three-step hybrid strategy towards efficiently selecting variables in multivariate calibration of near-infrared spectra[J]. Elsevier B V, 2020, 224: 1 386-1 425.
- [13] NØRGAARD 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]. SAGE Publications, 2000, 54(3): 413-419.
- [14] DENG Bai-chuan, YUN Yong-huan, MA Pan, et al. A new method for wavelength interval selection that intelligently optimizes the locations, widths and combinations of the intervals[J]. The Analyst, 2015, 140(6): 1 876-1 885.
- [15] YUN Yong-huan, WANG Wei-ting, TAN Min-li, et al. A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration[J]. Analytica Chimica Acta, 2014, 807: 36-43.
- [16] KENNARD R W, STONE L A. Computer aided design of experiments[J]. Technometrics, 1969(11): 137-148.

(上接第 76 页)