

# Optimization of Identification Model of Adulterated Dairy Products Based on Near Infrared Spectroscopy

Zhongyue Bian, Ruoyu Jia, Qianqian Guo, Rui Chen, Tianjiao Li

Dezhou College of Food Quality and Safety, Liaocheng Shandong  
Email: 17862667572@163.com

Received: Mar. 29<sup>th</sup>, 2019; accepted: Apr. 12<sup>th</sup>, 2019; published: Apr. 19<sup>th</sup>, 2019

## Abstract

Near infrared spectroscopy (NIR) combined with partial least squares (PLS) was used to identify rapidly adulterated dairy products. Eliminating abnormal points and other methods were employed to optimize the model. The results showed that for the identification of dairy products adulterated with water, the ideal model established by PLS was under transmission mode, using Spectrum, S-G smoothing method and additional scattering correction light path method, and the factor was 5, in which the correlation coefficient of prediction set (RP) was 0.9975 and the coefficient of determination ( $R^2$ ) was 0.9943. For the identification of dairy products adulterated with starch, the best model was under transmission condition, first-order derivative spectrum, Norris smoothing method and no correction of optical path, and the factor was 4, in which the RP was 0.9913 and the  $R^2$  was 0.9827.

## Keywords

Adulterated Dairy Products, Near Infrared Spectroscopy, Rapid Identification, Model Optimization

# 基于近红外光谱的掺伪乳制品鉴别模型优化

卞中悦, 贾若玉, 郭倩倩, 陈蕊, 李天骄

德州学院食品质量与安全, 山东 聊城  
Email: 17862667572@163.com

收稿日期: 2019年3月29日; 录用日期: 2019年4月12日; 发布日期: 2019年4月19日

## 摘要

为实现掺伪乳制品的快速鉴别, 基于近红外光谱技术, 采用偏最小二乘法(PLS法)建立掺伪乳制品的鉴别

模型, 并利用剔除异常点等方法对模型进行优化。结果表明: 掺水乳制品在透射方式下, 因子数为5, 采用原始光谱、S-G平滑、附加散射矫正方式时定量模型最为理想, 其预测集相关系数为0.9975, 决定系数为0.9943; 掺淀粉乳制品在透射方式下, 因子数为4, 采用一阶导数谱、Norris平滑、不矫正时定量模型最为理想, 其预测集相关系数为0.9913, 决定系数为0.9827。

## 关键词

掺伪乳制品, 近红外光谱, 快速鉴别, 模型优化

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近年来, 乳制品的销售量逐渐增加, 但由于经济利益的驱使, 乳制品掺伪事件频发。例如: “三聚氰胺毒奶粉事件”、“阜阳奶粉事件”, 这些事件严重影响了消费者的身体健康和生命安全。因此, 快速准确地对乳制品品质进行检测具有十分重要的意义。NY/T 802-2004《乳与乳制品中淀粉的测定酶-比色法》等国标文件中规定了有关乳制品掺伪检测的传统方法。传统的检测方法操作方法虽然简单, 但是过程较繁琐[1] [2]、耗时长、使用的化学试剂易导致环境污染[3], 不利于大规模检测[4]。为了实现掺伪乳制品的高通量快速检测, 本文基于近红外光谱技术对乳制品的品质进行全面分析, 解决了传统方法的不足。近红外光谱技术是光谱测量技术和化学计量学学科的有机结合, 被誉为“分析巨人”[5], 在20世纪90年代后期逐渐应用到石油[6] [7]、农业[8]、制药[9]和食品[10] [11]等领域。该技术具有分析速度快、分析效率高、使用范围广、低消耗、可实现多组分指标同时检测及远程在线分析等突出优点, 适用于对各种乳制品的质量检测, 例如: 李凯歌等[12]考察了近红外结合马氏距离判别分析和仿生模式识别在还原奶检测中的适用性。程文字等[13]利用近红外结合支持向量机法对液态奶中是否含有三聚氰胺进行了定性研究。韩东海等[14]利用近红外对还原奶掺假进行了鉴别分析, 同时快速测定了原料奶的新鲜度。

## 2. 材料和方法

### 2.1. 仪器和材料

Antaris II型傅里叶近红外光谱分析仪、RESULT-Integration光谱采集软件、TQ Analyst建模软件(美国Thermo Fisher Scientific公司), 纯牛奶(购自德州市大型超市, 主要成分如下: 蛋白质3.0 g/100 mL、脂肪3.8 g/100 mL), 纯净水, 淀粉(天津市致远化学试剂有限公司-太阳牌, ar级-分析纯)。

### 2.2. 试验方法

#### 2.2.1. 样品制备

向品牌纯牛奶样品中不断掺入水或淀粉乳, 通过外观观察和气味鉴别的感官试验得出所掺入物质质量的上限值。按照此方法, 得出纯牛奶中掺入水的质量百分比为55.00%、掺入淀粉的上限含量为84.21 g/L。按照掺伪质量比为0.10%~70.00%制得73个掺水样品, 掺伪含量为0.16~84.21 g/L制得53个掺淀粉样品, 见表1。

**Table 1.** Adulterated dairy products system**表 1.** 掺伪乳制品体系

| 掺伪物质 | 掺入水的质量百分比/% | 掺入淀粉的含量/(g/L) | 样品数量/个 |
|------|-------------|---------------|--------|
| 水    | 0.10~70.00  | 0             | 73     |
| 淀粉   | 0.09~47.37  | 0.16~84.21    | 53     |

### 2.2.2. 光谱的采集和平均光谱的建立

采集光谱前将仪器开机预热 1 小时, 掺伪样品在实验室提前放置 1~2 小时。在室温(20°C~25°C)、室内湿度为 50%~60%下进行透射和漫反射方式下的光谱采集。光程设置为 2 mm, 扫描范围为 10,000~4000  $\text{cm}^{-1}$ , 分辨率 8  $\text{cm}^{-1}$ , 采集次数 32 次, 每个样品均扫描 3 次, 将所得光谱数据导入 TQ-Analyst 软件中建立其平均光谱。

### 2.2.3. 模型的建立、优化及验证

基于 TQ Analyst 软件, 采用 PLS 法, 从每组样品中随机选取三分之二个样品用于建模, 三分之一个样品用于预测, 其中预测集的掺伪质量分数在建模集的掺伪质量分数范围内。剔除异常点, 选取合适的光谱波段建立乳制品掺伪鉴别模型, 将原始光谱先后采用不处理、一阶导数谱、二阶导数谱加上无平滑、S-G 平滑、Norris 平滑三种方法的顺序进行预处理, 确立模型因子数, 并通过不矫正、附加散射矫正、标准正则变换三种矫正方式对模型进行优化。基于 Excel 软件对单个样品的预测值与真实值进行比较分析得出相关性以验证建立模型的准确性。

### 2.2.4. 定量模型参数

在定量模型中相关系数(Corr. Coeff.)和误差均方根(RMSE)是反映模型预测性和适用性的重要参数, 相关系数越接近 1, 相关性就越高; 误差均方根结果越接近 0, 误差性就越小; 此时定量模型的预测性和适用性最为理想。

## 3. 结果与分析

### 3.1. 光谱结果分析

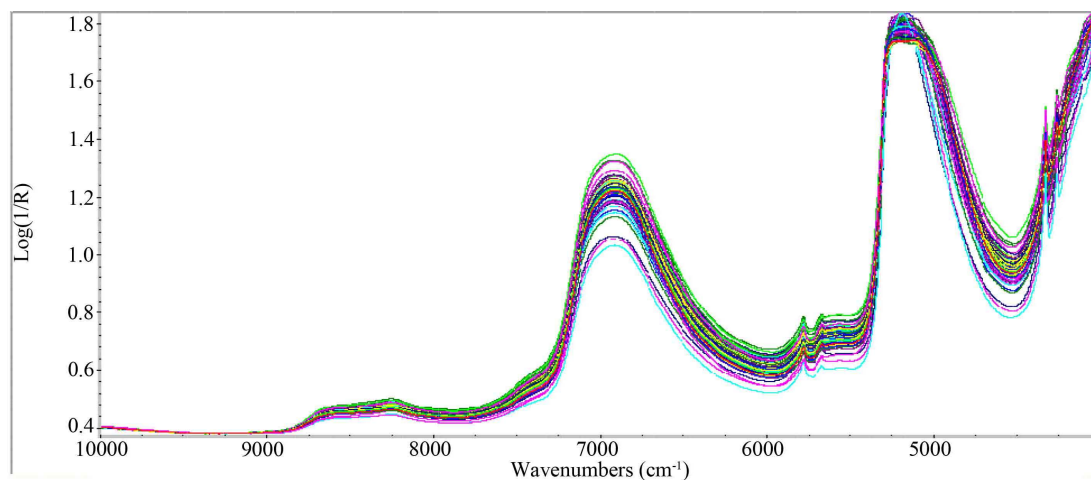
近红外光谱区与有机分子中含氢基团 X-H (O, N, C)振动的合频和分子中单个化学键的基频振动倍频的吸收区一致, 通过扫描样品的近红外光谱, 可以得到样品中有机分子含氢基团的特征信息, 因为乳制品中的有机成分的组成、化学键等比较相近[15], 所以不同的掺伪乳制品的光谱图比较相近, 无法从原始光谱图上得到掺伪程度, 因此需要结合化学计量学进一步分析。通过分析可知, 样品之间的差异在全谱的范围中均较明显, 所以试验建立模型时选择的光谱波长范围均为 10,000~4000  $\text{cm}^{-1}$ 。漫反射、透射方式下掺水乳制品体系的光谱图见图 1、图 2 所示, 漫反射、透射方式下掺淀粉乳制品体系的光谱图见图 3、图 4 所示。

### 3.2. 异常样本的剔除

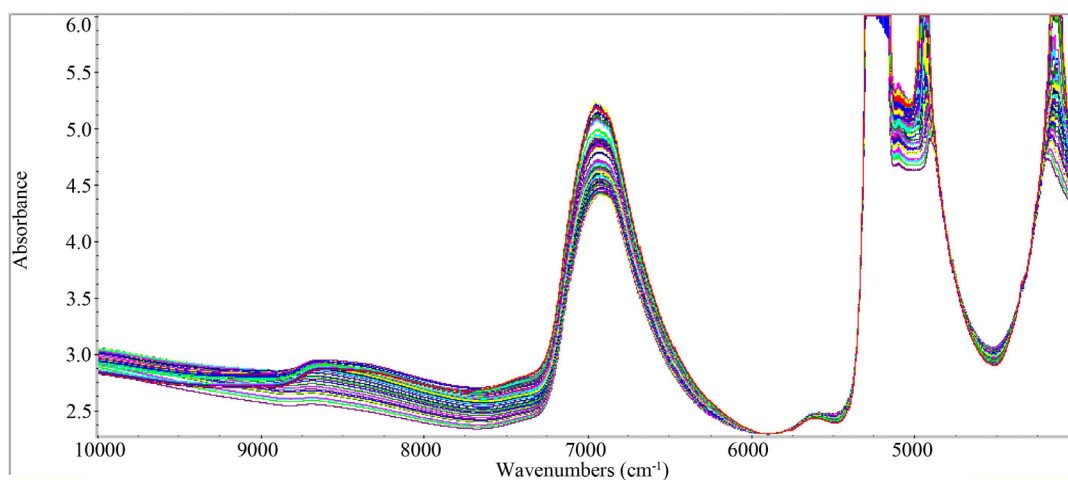
本试验采用马氏距离法进行异常点的剔除, 见图 5~8 所示。在漫反射方式下, 掺水乳制品存在 3 个异常样品, 为样品 34、64、65; 掺淀粉乳制品存在 4 个异常样品, 为样品 4、8、23、34。在透射方式下, 掺水乳制品不存在异常点; 掺淀粉乳制品存在 2 个异常样品, 为样品 24 和 31。

### 3.3. 漫反射方式下的建模结果分析

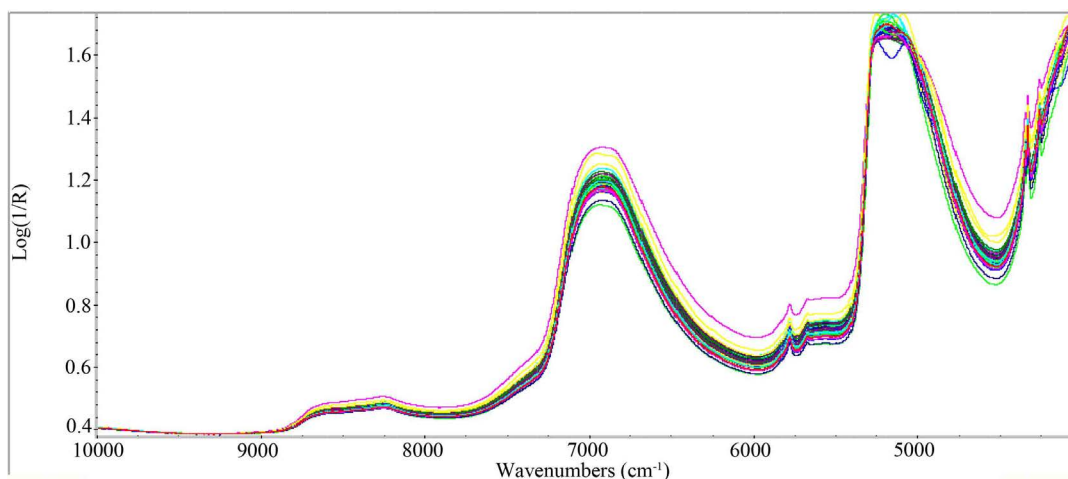
见表 2 分析得:



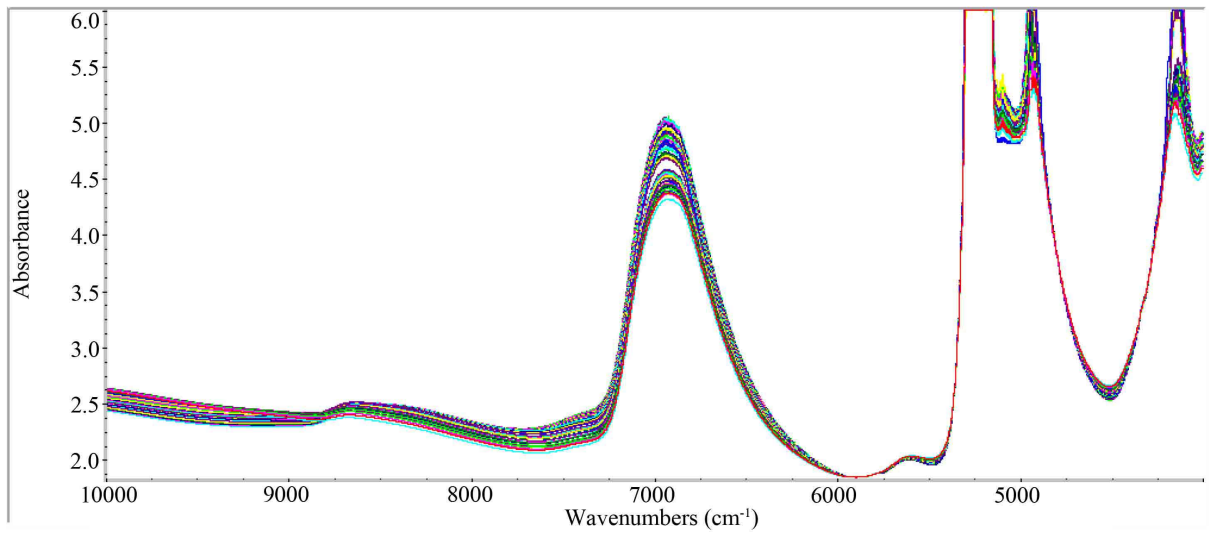
**Figure 1.** Spectrogram region of 10,000 - 4000  $\text{cm}^{-1}$  of dairy product adulterated with water under diffuse reflection mode  
**图 1.** 漫反射方式掺水乳制品体系 10,000~4000  $\text{cm}^{-1}$  的光谱图



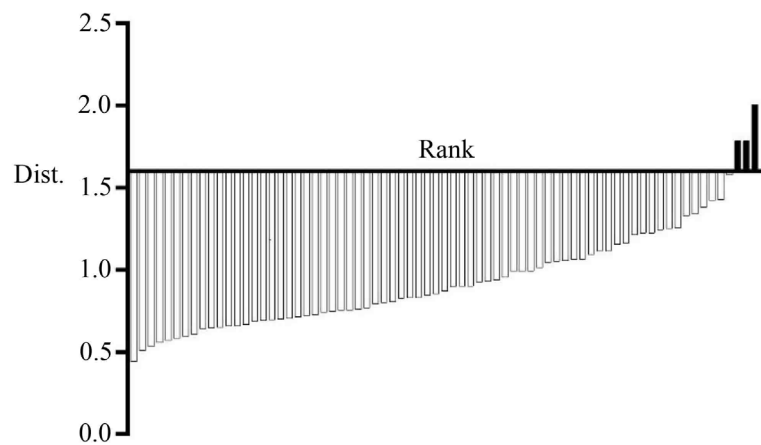
**Figure 2.** Spectrogram region of 10,000 - 4000  $\text{cm}^{-1}$  of dairy product adulterated with water under transmission mode  
**图 2.** 透射方式掺水乳制品体系 10,000~4000  $\text{cm}^{-1}$  光谱图



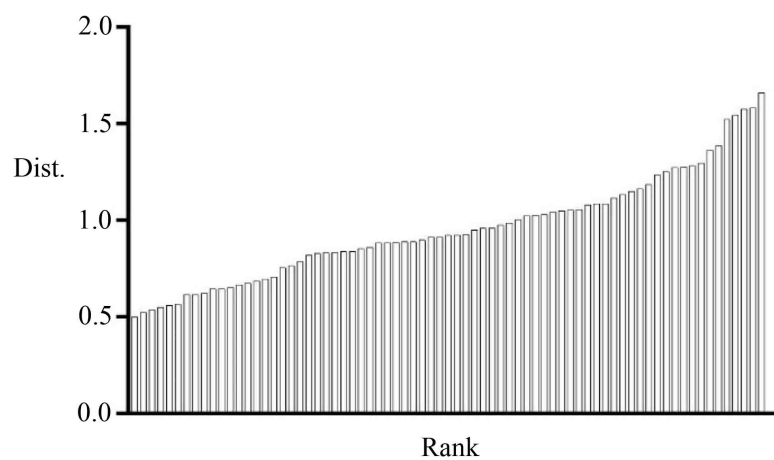
**Figure 3.** Spectrogram region of 10,000 - 4000  $\text{cm}^{-1}$  of dairy product adulterated with starch under diffuse reflection mode  
**图 3.** 漫反射方式掺淀粉乳制品体系 10,000~4000  $\text{cm}^{-1}$  光谱图



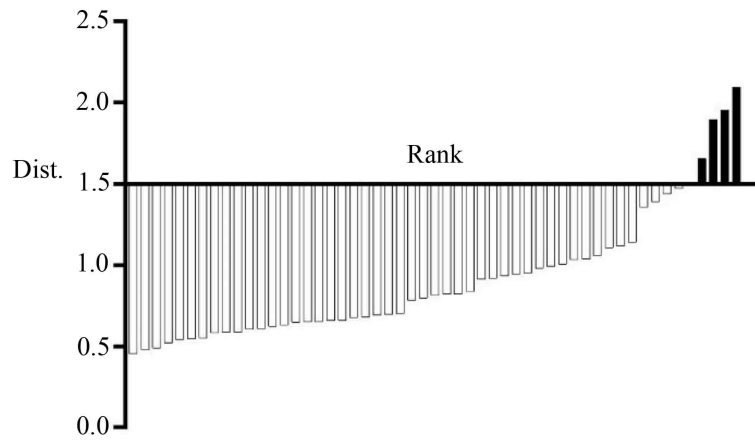
**Figure 4.** Spectrogram region of 10,000 - 4000  $\text{cm}^{-1}$  of dairy product adulterated with starch under transmission mode  
**图 4.** 透射方式掺淀粉乳制品体系 10,000~4000  $\text{cm}^{-1}$  光谱图



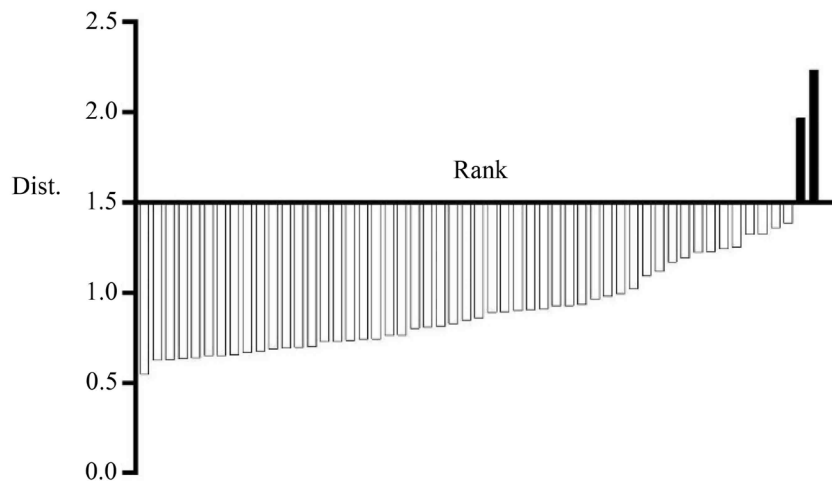
**Figure 5.** Analyzing abnormal points of dairy products adulterated with water under diffuse reflection mode  
**图 5.** 漫反射方式下掺水异常样品分析



**Figure 6.** Analyzing abnormal points of dairy products adulterated with water under transmission mode  
**图 6.** 透射方式下掺水异常样品分析



**Figure 7.** Analyzing abnormal points of dairy products adulterated with starch under diffuse reflection mode  
**图 7.** 漫反射方式下掺淀粉异常样品分析



**Figure 8.** Analyzing abnormal points of dairy products adulterated with starch under transmission mode  
**图 8.** 透射方式下掺淀粉异常样品分析

**Table 2.** Statistical analysis of the modeling results of the adulterated dairy products system under diffuse reflection mode  
**表 2.** 漫反射方式下掺伪乳制品体系的建模结果统计

| 掺伪物质 | 光谱的预处理方法 |           | 建模集    |             | 预测集    |             |
|------|----------|-----------|--------|-------------|--------|-------------|
|      | 光谱的格式    | 平滑方法      | RMSEC  | Corr. Coeff | RMSEP  | Corr. Coeff |
| 水    | 原始光谱     | 无平滑       | 0.0403 | 0.9796      | 0.0667 | 0.9539      |
|      |          | S-G 平滑    | 0.0403 | 0.9795      | 0.0667 | 0.9538      |
|      | 一阶导数谱    | 无平滑       | 0.0289 | 0.9895      | 0.0460 | 0.9787      |
|      |          | S-G 平滑    | 0.0284 | 0.9899      | 0.0470 | 0.9779      |
|      |          | Norris 平滑 | 0.0453 | 0.9742      | 0.0644 | 0.9566      |
|      |          | 无平滑       | 0.0200 | 0.9950      | 0.0499 | 0.9754      |
|      | 二阶导数谱    | S-G 平滑    | 0.0276 | 0.9905      | 0.0597 | 0.9647      |
|      |          | Norris 平滑 | 0.0254 | 0.9920      | 0.0410 | 0.9855      |

Continued

|    |       |           |        |        |        |        |
|----|-------|-----------|--------|--------|--------|--------|
| 淀粉 | 原始光谱  | 无平滑       | 5.0900 | 0.9790 | 8.3000 | 0.9560 |
|    |       | S-G 平滑    | 5.0800 | 0.9798 | 8.1000 | 0.9585 |
|    | 一阶导数谱 | 无平滑       | 3.9900 | 0.9876 | 13.000 | 0.8924 |
|    |       | S-G 平滑    | 4.1000 | 0.9869 | 13.200 | 0.8875 |
|    |       | Norris 平滑 | 4.7400 | 0.9825 | 15.400 | 0.8544 |
|    |       | 无平滑       | 2.0000 | 0.9969 | 8.2500 | 0.9588 |
|    | 二阶导数谱 | S-G 平滑    | 2.9800 | 0.9931 | 9.8100 | 0.9358 |
|    |       | Norris 平滑 | 6.4700 | 0.9670 | 23.200 | 0.6051 |

掺水样品无论采用哪种光谱格式以及平滑方法,预测集对应的误差均方根都较小(均没有超过 0.0680),相关系数都很大(均超过 0.9530)。对比可知,掺水样品在二阶导数谱格式下采用 Norris 平滑方法得到的模型最理想,适用性和预测性最好,其预测集相关系数和误差均方根分别为 0.9855 和 0.0410。

掺淀粉样品在一阶导数谱格式下,无论采用哪种平滑方法,对应的误差均方根都较大(均大于 10.00),相关系数都很小(均没有超过 0.9000),所以掺淀粉样品不适合在此光谱格式下建模;该组样品在原始光谱-S-G 平滑方法和二阶导数谱 - 无平滑方法得到的模型均较为理想,其中适用性和预测性能更高的模型是在二阶导数谱格式下 - 无平滑方法,其预测集相关系数和误差均方根分别为 0.9588 和 8.25。

### 3.4. 透射方式下的建模结果分析

见表 3 分析得:

**Table 3.** Statistical analysis of the modeling results of the adulterated dairy products system under transmission mode  
**表 3.** 透射方式下掺伪乳制品体系的建模结果统计

| 掺伪物质 | 光谱的预处理方法 |           | 建模集    |             | 预测集    |             |
|------|----------|-----------|--------|-------------|--------|-------------|
|      | 光谱的格式    | 平滑方法      | RMSEC  | Corr. Coeff | RMSEP  | Corr. Coeff |
| 水    | 原始光谱     | 无平滑       | 0.0165 | 0.9968      | 0.0172 | 0.9968      |
|      |          | S-G 平滑    | 0.0167 | 0.9967      | 0.0172 | 0.9969      |
|      | 一阶导数谱    | 无平滑       | 0.0193 | 0.9957      | 0.0192 | 0.9960      |
|      |          | S-G 平滑    | 0.0194 | 0.9956      | 0.0198 | 0.9959      |
|      |          | Norris 平滑 | 0.0202 | 0.9952      | 0.0204 | 0.9956      |
|      |          | 无平滑       | 0.0131 | 0.9980      | 0.0369 | 0.9953      |
|      | 二阶导数谱    | S-G 平滑    | 0.0226 | 0.9940      | 0.0311 | 0.9806      |
|      |          | Norris 平滑 | 0.0218 | 0.9944      | 0.0214 | 0.9952      |
| 淀粉   | 原始光谱     | 无平滑       | 4.3700 | 0.9854      | 3.8300 | 0.9898      |
|      |          | S-G 平滑    | 3.9200 | 0.9883      | 4.5500 | 0.9852      |
|      | 一阶导数谱    | 无平滑       | 4.6100 | 0.9838      | 4.1000 | 0.9890      |
|      |          | S-G 平滑    | 4.6300 | 0.9836      | 4.0200 | 0.9895      |
|      |          | Norris 平滑 | 4.4700 | 0.9847      | 3.6000 | 0.9913      |
|      |          | 无平滑       | 5.1600 | 0.9796      | 4.9100 | 0.9839      |
|      | 二阶导数谱    | S-G 平滑    | 4.1600 | 0.9834      | 4.3800 | 0.9875      |
|      |          | Norris 平滑 | 4.3900 | 0.9853      | 3.6900 | 0.9906      |

掺水样品无论采用哪种光谱格式及平滑方法, 预测集对应的误差均方根都较小(均没有超过 0.0400), 相关系数均接近于 1。对比可知, 掺水样品在原始光谱下采用 S-G 平滑方法得到的模型适用性和预测性最高, 其预测集相关系数和误差均方根分别为 0.9969 和 0.0172。

掺淀粉样品无论使用何种预处理方法, 预测集的相关系数均在 0.9900 左右, 模型的预测性能较高。在一阶导数谱格式下采用 Norris 平滑的预处理方法, 掺淀粉样品定量模型最理想, 适用性最好, 其预测集相关系数和误差均方根分别为 0.9913 和 3.60。

### 3.5. 两种采集方式下的最优建模结果对比分析

将漫反射和透射采集方式下的最优建模结果进行统计, 掺水样品的理想模型是在透射方式下, 采用原始光谱-S-G 平滑的预处理方法, 其模型的预测集相关系数和误差均差根分别为 0.9969, 0.0172; 掺淀粉样品的理想模型是在透射方式下, 采用一阶导数谱-Norris 平滑的预处理方法, 其预测集相关系数和误差均差根分别为 0.9913, 3.60。

### 3.6. 定量模型因子数确定

试验在进行定量回归时, 模型因子数是影响 PLS 模型效果的因素之一: 当因子数过小时, 会导致模型的误差较大; 当因子数过多时, 可造成定量模型的过度拟合, 即影响模型预测新样品。优化模型时, 一般第一个极小值点出现时的因子数为模型最适的因子数。通过分析, 可确定 A 组的因子数为 5, B 组的因子数为 4, 见图 9、图 10 所示。

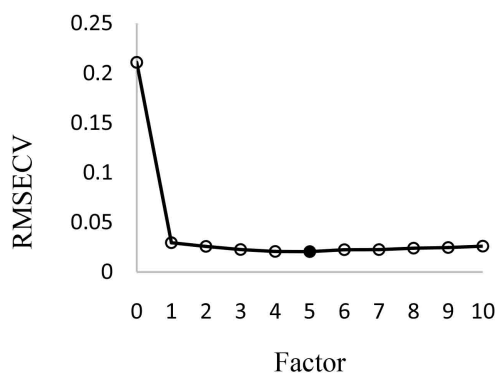


Figure 9. RMSECV changed with change of factor in dairy products adulterated with water

图 9. 掺水样品 RMSECV 随因子数的变化图

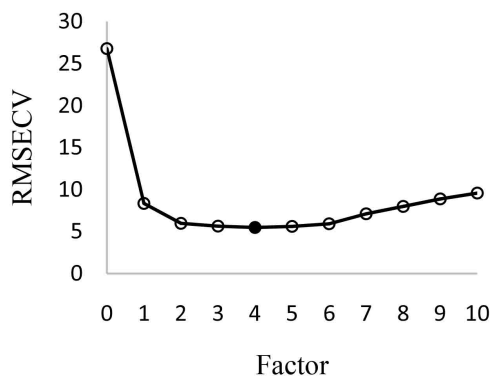


Figure 10. RMSECV changed with change of factor in dairy products adulterated with starch

图 10. 掺淀粉样品 RMSECV 随因子数的变化图



### 3.7. 不同光程矫正的模型优化

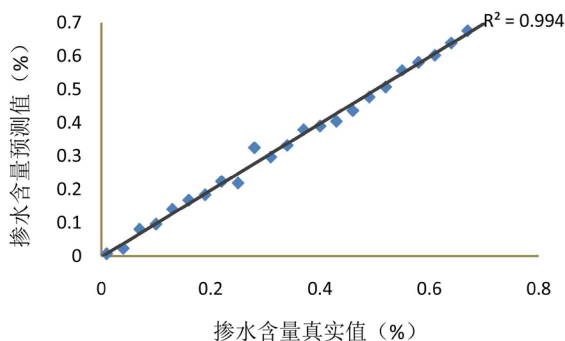
在定量模型设置好因子数后，试验进一步研究了不同光程矫正方式对模型建立的影响，得到的结果见表 4 所示。掺水样品在原始光谱、S-G 平滑方法的预处理下采用附加散射矫正方式时，模型的适用性和预测性更为理想，其预测集的相关系数和误差均方根分别为 0.9975，0.0155；掺淀粉样品在一阶导数谱、Norris 平滑方法的预处理下采用光程不矫正时模型最理想，其预测集的相关系数和误差均方根分别为 0.9913，3.6000。

**Table 4.** Modeling results of different optical path corrections under transmission mode  
**表 4.** 透射方式下不同光程矫正方式建模结果

| 掺伪物质 | 光谱预处理方法         | 光程矫正方式 | 建模集    |             | 预测集    |             |
|------|-----------------|--------|--------|-------------|--------|-------------|
|      |                 |        | RMSEC  | Corr. Coeff | RMSEP  | Corr. Coeff |
| 水    | 原始光谱-S-G 平滑     | 不矫正    | 0.0167 | 0.9967      | 0.0172 | 0.9969      |
|      |                 | 附加散射矫正 | 0.0163 | 0.9969      | 0.0155 | 0.9975      |
|      |                 | 标准正则变换 | 0.0172 | 0.9960      | 0.0164 | 0.9962      |
| 淀粉   | 一阶导数谱-Norris 平滑 | 不矫正    | 4.4700 | 0.9847      | 3.6000 | 0.9913      |
|      |                 | 附加散射矫正 | 4.4500 | 0.9848      | 3.6100 | 0.9912      |
|      |                 | 标准正则变换 | 4.4500 | 0.9848      | 3.6100 | 0.9912      |

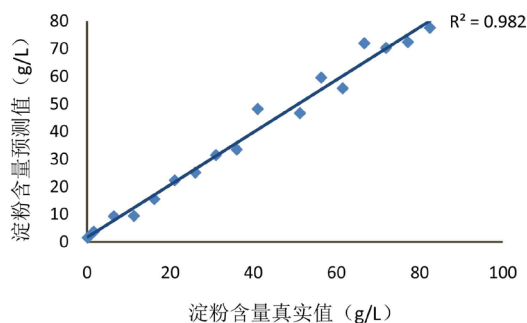
### 3.8. 掺伪乳制品定量模型的验证

基于 Excel 软件对两组掺伪乳制品定量分析模型进行了验证，见图 11、图 12 所示，两组模型图的  $R^2$  均较高，说明两组掺伪乳制品是可以用此模型来测定的，模型的预测准确性和适用性较高。



**Figure 11.** The verification result of quantitative model of dairy products adulterated with water

**图 11.** 掺水样品定量模型的验证结果图



**Figure 12.** The verification result of quantitative model of dairy products adulterated with starch

**图 12.** 掺淀粉样品定量模型的验证结果图

## 4. 结论

基于 PLS 法通过剔除异常点并利用不同的光谱预处理方法建立了掺伪乳制品鉴别模型。通过确定因子数、不同的光程矫正方式对模型进行优化后, 针对掺水乳制品, 在透射方式下, 因子数为 5, 采用原始光谱、S-G 平滑、附加散射矫正方式时定量分析模型最为理想, 其预测集相关系数为 0.9975, 相关性可达到 0.9949; 针对掺淀粉乳制品, 在透射方式下, 因子数为 4, 采用一阶导数谱、Norris 平滑、光程不矫正时模型最为理想, 其预测集相关系数为 0.9913, 相关性可达到 0.9827。以上研究不仅可为掺伪乳制品体系的快速鉴别提供理论依据, 而且可为乳制品品质的监控提供一定的参考。

## 基金项目

国家自然科学基金(201701021); 大学生创新创业计划训练项目(201710448081)。山东省教科所 BYGI2017003 创新创业教育与专业教育融合研究。

## 参考文献

- [1] 明双喜, 张然. 几种常见食品掺伪及其检测技术介绍[J]. 食品研究与开发, 2014, 35(20): 17-22.
- [2] 谢昕. 几类食品掺伪的化学检验[J]. 化学教育(中英文), 2014, 35(22): 30-32.
- [3] 魏晓敏. 乳与乳制品掺伪鉴别和检验技术[J]. 科学技术创新, 2016(6): 124-124.
- [4] 彭珊珊, 张俊艳. 食品掺伪鉴别检验[M]. 北京: 中国轻工业出版社, 2014.
- [5] 宋志强, 张恒, 郑晓, 等. 近红外光谱技术在食用植物油检测中的应用[J]. 武汉工业学院学报, 2013, 32(2): 1-5.
- [6] 徐喆. 近红外光谱技术在油品检测中的应用[J]. 中国石油和化工标准与质量, 2015(24): 13-15.
- [7] 陆婉珍, 袁洪福, 徐广通, 等. 现代近红外光谱技术在石油产品分析中的应用[C]//中国分析测试协会科学技术奖发展回顾. 2015.
- [8] 李桂峰. 近红外光谱技术及其在农业和食品检测中的应用[J]. 农业与技术, 2007, 27(5): 91-94.
- [9] 饶毅, 魏惠珍, 方少敏, 等. 近红外光谱技术在中药制药过程质控中的应用[J]. 中成药, 2011, 33(1): 126-130.
- [10] 王莉, 李增芳, 何勇, 等. 应用近红外光谱技术快速检测果醋糖度[J]. 光谱学与光谱分析, 2008, 28(8): 1810-1813.
- [11] 张萍, 闫继红, 朱志华, 等. 近红外光谱技术在食品品质鉴别中的应用研究[J]. 现代科学仪器, 2006(1): 60-62.
- [12] 李凯歌, 韩东润, 孙明. 纯牛奶中还原奶的近红外检测判别分析[J]. 农机化研究, 2008, 8: 145-147.
- [13] 程文字, 管晓, 刘静. 近红外光谱技术检测液态奶中微量三聚氰胺的可行性研究[J]. 食品与机械, 2015(1): 71-74+81.
- [14] Han, D., Lu, C. and Liu, Y. (2006) Nondestructive Detection of Brown Core in the Chinese Pear Yali by Transmission Visible-NIR Spectroscopy. *Dairy Guide*, 4, 39-41.
- [15] 李振庆, 黄梅珍, 倪一, 等. 改进偏最小二乘法在近红外牛奶成分测量中的应用[J]. 光学技术, 2009, 35(1): 70-73.

**知网检索的两种方式：**

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择：[ISSN]，输入期刊 ISSN：2330-1724，即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：[ojns@hanspub.org](mailto:ojns@hanspub.org)