

# 基于 6 mm 检测皿近红外测量光谱 iPLS 模型识别 复原驼乳参伪水解动物蛋白的研究

苑柯岩<sup>1</sup>, 王 嵘<sup>2</sup>, 王翔翔<sup>2</sup>, 薛莉婷<sup>2</sup>, 余 丽<sup>2\*</sup>

1. 呼和浩特市检验检测中心, 内蒙古 呼和浩特 010018

2. 安徽建筑大学环境与能源工程学院, 安徽 合肥 230601

**摘 要** 驼乳因具有较高的营养价值和独特的保健作用, 已逐渐成为广大消费者所信赖的保健乳制品。由于驼乳产量小而其市场价值高, 为在驼乳中参杂参伪提供了利润上的操作空间。随着国家对乳制品中非法添加三聚氰胺的严打态势进一步加强, 劣质水解动物蛋白以其高蛋白含量、价格低廉和非法添加隐蔽性强等特点逐渐成为乳制品中参伪的新宠。防范和打击驼乳中参伪劣质水解动物蛋白成为了消费者和驼乳产业从业人员共同面临的巨大挑战, 如何快速、便捷、低成本检测驼乳当中参伪动物水解蛋白成为亟待解决的问题。随着近红外光谱分析技术的飞速发展, 其分析速度快、成本低、样品无需前处理、操作简便等优点逐渐在石油化工、食品、农业、医药等诸多领域得到广泛应用。采用 6 mm 测样皿的近红外光谱仪对驼乳参伪不同含量的动物水解蛋白进行测量获得原始光谱矩阵, 采用一阶导数、SNV、SG 卷积平滑法、一阶导数+SG 平滑法、一阶导数+SNV、SG+SNV 等方法对原始光谱进行预处理, 以全光谱 10 个主成分回归模型为评价, 通过比较原始光谱在不做任何预处理为本体系建立 iPLS 模型最佳光谱。通过调整主成分计算规模, 确定了最佳主成分计算规模为 10 个。通过调整区间划分数量, 以其对应的模型的  $R^2$  和 RMSECV 值为评价标准, 最终确定了最佳区间划分数量为 30。通过实验和计算, 在  $7\ 887.87\sim 7\ 590.87\ \text{cm}^{-1}$  区间得到了主成分数 6, 相关系数 0.945 1, RMSECV 值 0.200 1 为驼乳掺伪水解动物蛋白最佳预测模型。经内部交互验证, 该模型在本体系可以很好地预测复原驼乳中掺伪水解动物蛋白的情况, 为相关领域研究提供技术参考。

**关键词** 近红外光谱; 间隔偏最小二乘法; 复原驼乳; 水解动物蛋白

**中图分类号:** O657.33 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2022)10-3143-05

## 引 言

驼乳含有免疫球蛋白、乳铁蛋白、溶菌酶等多种营养物质和特有的保健功能, 使其逐渐成为广大消费者所信赖的保健乳制品原料。水解动物蛋白(HAP)以其较高的含氮量以及与驼乳同属动物氮源的特性, 可以作为提高驼乳及其制品的添加剂。目前, 乳制品当中掺伪水解动物蛋白的定量检测主要采用高效液相法、气相色谱法、电泳法、PCR 技术、免疫酶联法等检测技术<sup>[1]</sup>。

近红外光谱(NIR)是指波长 780~2 500 nm 范围内的电磁波, 主要测量分子中含氢官能团 X—H(X=C, N, O 和 S 等)振动的倍频及合频吸收, 可检测乳制品中的蛋白质、脂

肪、水分、淀粉、糖、等成分<sup>[2]</sup>。范瑞等采用 1 mm 测量皿对牛奶中掺伪水解动物蛋白的现象进行了偏最小二乘法的建模研究得到了相关系数为 0.983 定量预测模型<sup>[3]</sup>。魏玉娟等<sup>[4]</sup>采用近红外光谱技术结合模式识别方法对液态奶中违法添加三聚氰胺进行快速检测, 采用 PLS 判别法结合近红外光谱技术对牛奶中三聚氰胺不同掺假量的快速识别研究, 所构建的偏最小二乘判别分析(partial least-square discriminant analysis, PLSDA)方法模型对三聚氰胺不同掺假量牛奶样品训练集和预测集的近红外原始光谱的识别正确率分别达到 100%和 90.32%。Chen 等<sup>[5]</sup>研究了采用近红外光谱和偏最小二乘法检测牛奶中三聚氰胺掺假的可行性。本工作将近红外 6 mm 检测皿检测光谱结合间隔偏最小二乘法(iPLS)应用于复原驼乳中参伪水解动物蛋白的含量测定, 扩大了近红外

收稿日期: 2021-08-03, 修订日期: 2022-03-22

基金项目: 国家自然科学基金项目(41601203, 51978003), 安徽省自然科学基金项目(2108085ME186, 1908085QE241, KJ2020A0467)资助

作者简介: 苑柯岩, 1990 年生, 呼和浩特市检验检测中心硕士研究生 e-mail: syfmri@vip.163.com

\* 通讯作者 e-mail: ronger@ahjzu.edu.cn

光谱的应用范围,同时为实现近红外光谱实现液态乳制品实时监测提供了相关实践经验。

## 1 实验部分

### 1.1 材料与样品

所用驼奶为所市售驼乳粉按照奶粉与水 1 : 7 溶解后制得的复原驼乳,驼乳粉为新疆生产的“原始黄金”品牌全脂驼乳粉,水解动物蛋白是由泉州盛达食品添加剂有限公司生产的食品级水解动物蛋白粉。

### 1.2 参伪驼乳样品配置

按照表 1 信息精确称取水解动物蛋白于 100 mL 称量瓶内,用配置好的复原驼乳定容到 100 mL,置于摇床室温振荡

30 min,确保水解动物蛋白完全溶解。

### 1.3 原始光谱的采集和样品集的划分

采用德国布鲁克光谱仪器公司生产的傅里叶近红外光谱仪(MPA)以空气为背景,6 mm 测量皿,分辨率  $8\text{ cm}^{-1}$ ,扫描次数为 64,检测器 INGaAs,波长范围:  $10\ 000\sim 4\ 000\text{ cm}^{-1}$  同一样品在相同的条件下重复测定 3 次<sup>[6]</sup>,逐一采集样品并保存光谱点数据(dpt 格式),共采集样品光谱图数据 129 条。取每个样品中间光谱,即第二条光谱构成待测样品光谱矩阵,得到  $43\times 1\ 555$  原始光谱矩阵。等距离选取矩阵中 33 份具有代表性的样品作为训练集,其余 10 份样品作为验证集,用于动物水解蛋白含量定量分析模型建立,结果如表 2 所示。验证集中水解动物蛋白含量在训练集含量范围内,此训练集和验证集可用于定量分析模型的建立<sup>[7-9]</sup>。

表 1 实验样品配置情况

Table 1 The configuration of experimental samples

实验号	水解动物蛋白/ [g · (100 mL) <sup>-1</sup> ]	实验号	水解动物蛋白/ [g · (100 mL) <sup>-1</sup> ]	实验号	水解动物蛋白/ [g · (100 mL) <sup>-1</sup> ]	实验号	水解动物蛋白/ [g · (100 mL) <sup>-1</sup> ]
1	0.01	12	0.50	23	1.05	34	1.60
2	0.03	13	0.55	24	1.10	35	1.65
3	0.05	14	0.60	25	1.15	36	1.70
4	0.10	15	0.65	26	1.20	37	1.75
5	0.15	16	0.70	27	1.25	38	1.80
6	0.20	17	0.75	28	1.30	39	1.85
7	0.25	18	0.80	29	1.35	40	1.90
8	0.30	19	0.85	30	1.40	41	1.95
9	0.35	20	0.90	31	1.45	42	2.00
10	0.40	21	0.95	32	1.50	43	0.00
11	0.45	22	1.00	33	1.55		

表 2 训练集测试集情况

Table 2 Training set test set condition

样品集	样品数	最低值/[g · (100 mL) <sup>-1</sup> ]	最高值/[g · (100 mL) <sup>-1</sup> ]	平均值/[g · (100 mL) <sup>-1</sup> ]
训练集	33	0.00	2.00	0.95
测试集	10	0.10	1.90	0.97

## 2 结果与讨论

### 2.1 光谱数据预处理

将原始光谱矩阵在 Origin 软件中进行 SG 平滑和一阶导数的处理,得到 SG 平滑光谱矩阵和一阶导数矩阵,分别将原始、SG 平滑、一阶导数光谱矩阵进行 SNV 处理分别得到原始 SNV、SG 平滑 SNV、一阶导数 SNV 矩阵。

### 2.2 光谱预处理方法的选择

将 6 个光谱矩阵,分别计算其主成分个数,并以相应主成分建立全局光谱偏最小二乘法模型,以校正均方差(RMSECV)、相关系数( $r$ )为指标筛选建模方法。由表 3 可知,原始光谱在 1 个主成分情况下全局建模情况最好,相关系数为 0.854 4, RMSECV 为 0.322 0; SG 法与原始光谱相差不大,相关系数为 0.849 1, RMSECV 为 0.326 0。选择原始光谱矩阵进行 iPLS 模型的建立为最优。

表 3 预处理方法全局建模参数

Table 3 Preprocessing method global modeling parameters

光谱预处理方法	主成分数	$r$	RMSECV
一阶倒数	1	0.119 9	0.670 3
SG 平滑	1	0.849 1	0.326 0
SNV	1	0.407 7	0.429 0
一阶导数+SG	1	0.234 4	0.708 6
一阶导数+SNV	1	0.163 4	0.466 7
SG+SNV	1	0.399 9	0.429 8
原始光谱	1	0.854 4	0.322 0

### 2.3 基于原始光谱矩阵 iPLS 模型的初步建立

参考文献<sup>[7-9]</sup>,对特征区间进行筛选建模,计算步骤如下:

(1) 计算全局光谱 10 个主成分的 RMSECV 值,得到图 1。由图 1 可以看出全局主成分 RMSECV 的最小值为 1,选取 1 个主成分在全光谱范围内即全局波谱建立待测样品的

偏最小二乘回归模型，相关系数为 0.877 3，RMSECV 值 0.297 7(见图 2)，以此模型参数为参比。

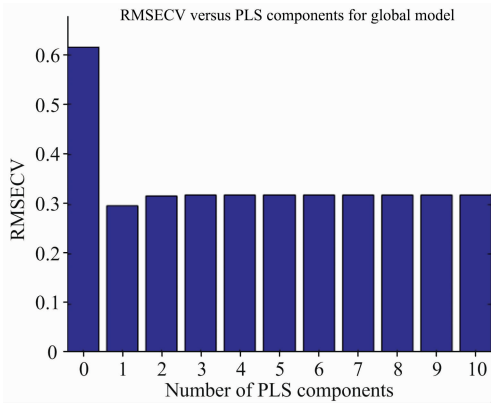


图 1 计算主成分 RMSECV 值

Fig. 1 Calculate RMSECV value of principal component

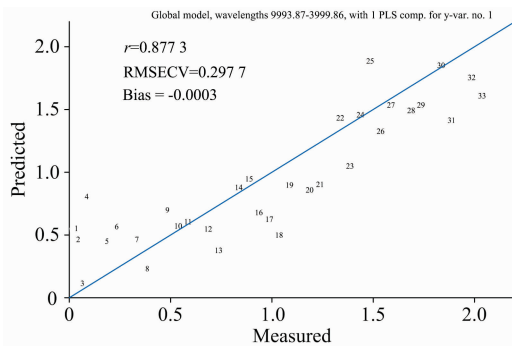


图 2 全局 1 个主成分 PLS 回归曲线

Fig. 2 Global 1 principal component PLS regression curve

(2)将全光谱区域划分为多个等宽的子区间，初步设定 20 个。在每个子区间上计算主成分并进行偏最小二乘回归，建立水解动物蛋白浓度的局部回归模型，得到 20 个局部回归模型。由图 3 可知，8 号区间 RMSECV 值低于全局值为 0.208 1，相关系数为 0.941 3，以此模型参数为区间调整参比标准。

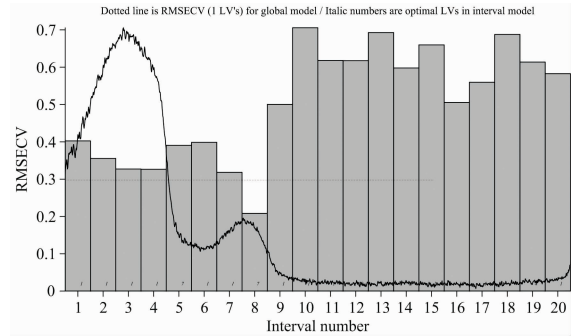


图 3 区间划分情况

Fig. 3 Interval division

### 2.4 建模间隔的调整

将区间间隔分别设置为 15, 25, 30 和 40 建立 iPLS 模型，相关参数记录于表 4。从表 4 可以看出，在整个间隔波段选择过程中，过于稀疏或过于密集的采样对于模型的建立产生较大影响，相关性和精确度均下降并且未出现多个可用区间。在区间划分数为 30 时，出现最优子区间模型，即 12 号，其相关系数最高 (0.945 1)，且 RMSECV 值最低 (0.200 1)，见图 4 和图 5。

表 4 间隔划分参数情况表

Table 4 Table of interval partitioning parameters

区间数量	小于全局 RMSECV 区间个数	主成分数	子区间号	对应波数	相关系数	区间 RMSECV 值
15	1	14	6	7 988.16~7 590.87	0.913 1	0.250 5
20	1	7	8	7 887.87~7 590.87	0.941 3	0.208 1
25	1	6	10	7 882.30~7 587.01	0.929 0	0.227 1
30	1	6	12	7 787.56~7 590.87	0.945 1	0.200 1
35	1	6	14	7 737.44~7 567.73	0.939 3	0.211 7
40	1	5	16	7 737.44~7 590.87	0.928 5	0.228 8

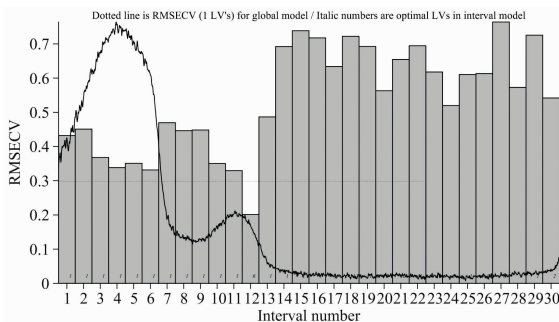


图 4 区间划分为 30 的区间选择情况

Fig. 4 Interval selection when interval is divided into 30

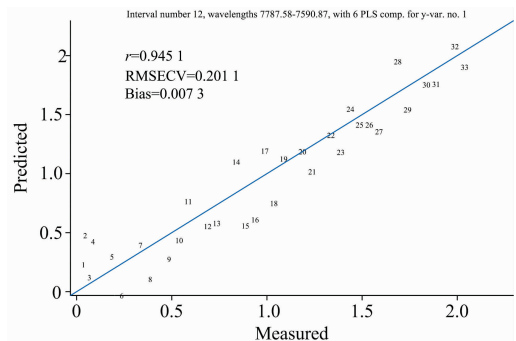


图 5 第 12 区间模型回归情况

Fig. 5 Regression of the 12th interval model

### 3 结 论

(1)通过计算全局光谱不同规模的主成分 RMSECV 值,得出全局主成分规模为 10 个最佳,在遇到建模主成分达到规模上限时再扩大主成分计算规模,可以有效的减少模型冗余计算提高模型计算效率。

(2)通过不同方法预处理后的光谱矩阵全局偏最小二乘回归模型,对比各模型预处理后的模型参数,可知原始光谱

矩阵不做任何预处理做 iPLS 建模为最佳。

(3)在对划分光谱区间数量的考察实验中,发现区间数为 30 时能够将子区间确定在最佳范围即子区间为  $7\ 787.56 \sim 7\ 590.87\ \text{cm}^{-1}$ ,并建立本实验最优 iPLS 模型,相关系数为 0.945 1, RMSECV 值为 0.200 1。

采用间隔偏最小二乘法可以对 6 mm 检测皿近红外光谱实现特征提取,并在该体系建立复原驼乳中掺伪动物水解蛋白快速定性表征和定量分析,可以液态乳及液态乳制品中的掺伪行为提供新的分析鉴别思路和方法。

### References

- [1] Bouhaddaou I S, Chabi R R, Errachid I F, et al. The Scientific World Journal, 2019, 2019: 2517293.
- [2] RONG Han, GAN Lu-jing, WANG Lei(荣 茜, 甘露菁, 王 磊). China Condiment(中国调味品), 2019, 44(12): 144.
- [3] FAN Rui, SUN Xiao-kai, YANG Chen, et al(范 睿, 孙晓凯, 杨 晨, 等). Food Industry(食品工业), 2017, 38(16): 253.
- [4] WEI Yu-juan, LI Lin, YANG Xiao-ya, et al(魏玉娟, 李 琳, 杨笑亚, 等). China Dairy Industry(中国乳品工业), 2016, 44(10): 48.
- [5] Chen H, Tan C, Lin Z, et al. Spectrochim Acta A, 2017, 173: 832.
- [6] ZHANG Hang, LIU Guo-hai, JIANG Hui, et al(张 航, 刘国海, 江 辉, 等). Progress in Laser and Optoelectronics(激光与光电子学进展), 2017, (2): 314.
- [7] Frimani P, De Luca S, Bucci R, et al. Food Control, 2019, 100: 292.
- [8] Tejerina David, Contador Rebeca, Ortiz Alberto. Food Chemistry, 2021, 356: 129733.
- [9] Chen Hui, Tan Chao, Lin Zan, et al. Computers in Biology and Medicine, 2013, 43(7): 865.

## Identification and Restoration of Pseudo-Hydrolyzed Animal Protein of Lacteus Camelus Based on iPLS Model of Near-Infrared Measurement Spectrum of 6 mm Detection Plate

YUAN Ke-yan<sup>1</sup>, WANG Rong<sup>2</sup>, WANG Xiang-xiang<sup>2</sup>, XUE Li-ping<sup>2</sup>, YU Li<sup>2\*</sup>

1. Huhhot City Inspection and Testing Center, Huhhot 010018, China

2. School of Environment and Energy Engineering, Anhui Jianzhu University, Hefei 230601, China

**Abstract** Camel milk has gradually become a health care dairy product trusted by consumers because of its high nutrition and unique health care effects. However, due to the small output of camel milk and its high market value, this provides a profitable operating space for the hybridization of camel milk. With the further strengthening of the state's crackdown on the illegal addition of melamine in dairy products, inferior hydrolyzed animal protein has gradually become a new favorite for counterfeiting in dairy products due to its high protein content, and low price and strong concealment of illegal addition. Preventing and cracking down on fake and inferior hydrolyzed animal protein in camel milk has become a huge challenge faced by consumers and practitioners in the camel milk industry. How to detect fake and low-cost animal hydrolyzed protein in camel milk has become an urgent problem to be developed. With the rapid development of near-infrared spectral analysis technology in the past ten years, near-infrared spectral analysis technology has gradually become widely used in many fields such as petrochemical, food, agriculture, medicine, etc. widely used. In this paper, the near-infrared spectrometer with a 6 mm sample dish was used to measure the animal hydrolyzed protein of camel milk ginseng with different contents to obtain the original spectral matrix. The original spectra were preprocessed by order derivative + SNV, SG + SNV and other methods, and the 10 principal component regression models of the global spectrum were used for evaluation. By adjusting the calculation scale of principal components, the optimal calculation scale of principal components is determined to be 10. By adjusting the number of interval divisions and using the  $R^2$  and RMSECV values of the corresponding model as evaluation criteria, the optimal number of interval divisions is finally determined to be 30. Through experiments and calculations, the principal component score of 6 was obtained in the range of  $7\ 887.87 \sim 7\ 590.87\ \text{cm}^{-1}$ , the correlation coefficient was 0.945 1, and the RMSECV value was 0.200 1, was the best prediction model for camel milk adulterated hydrolyzed animal protein. After internal interactive verification, the modified model can well predict the situation of adulterated and hydrolyzed animal protein in recovered camel milk in this system, which can provide technical reference for research in related fields.

**Keywords** Near-infrared spectroscopy; Interval partial least squares; Restored camel milk; Hydrolyzed animal protein

\* Corresponding author

(Received Aug. 3, 2021; accepted Mar. 22, 2022)

## 《光谱学与光谱分析》投稿简则

《光谱学与光谱分析》是由中国科协主管,中国光学学会主办,钢铁研究总院、中国科学院物理研究所、北京大学、清华大学共同承办的专业学术期刊。国内外公开发行,从 2004 年起为月刊,大 16 开本,每期 332 页。《光谱学与光谱分析》主要报道我国光谱学与光谱分析领域内具有创新性科研成果,及时反映国内外光谱学与光谱分析的进展和动态;发现并培育人才;推动和促进光谱学与光谱分析的发展。为科教兴国服务。读者对象为从事光谱学与光谱分析的科研人员、教学人员、分析测试人员和科研管理干部。

### 栏目设置和要求

1. 研究报告 要求具有创新性的研究成果,一般文章以 8000 字(包括图表、参考文献、作者姓名、单位和中文、英文摘要,下同)为宜。
2. 研究简报 要求在前人研究的基础上有重大改进或阶段性研究成果,一般不超过 5000 字。
3. 评述与进展 要求评述国内外本专业的发展前沿和进展动态,一般不超过 10000 字。
4. 新仪器装置 要求介绍新型光谱仪器的研制、开发、使用性能和应用,一般不超过 5000 字。
5. 来稿摘登 要求测试手段及方法有改进并有应用交流价值,一般以 3000~4000 字为宜。

### 稿件要求

1. 投稿者请经本刊编委(或历届编委)一人或本专业知名专家推荐,并附单位保密审查意见及作者署名顺序,主要作者介绍。文章有重大经济效益或有创新者,请说明,同时注明受国家级基金或国家自然科学基金资助情况。
2. 来稿要观点明确、数据真实可靠、层次分明、言简意明、重点突出。来稿必须是网上在线投稿(含各种符号和外文字母大写、小写、正体、斜体;希腊字母、拉丁字母;上角、下角标位置应标清楚)。中文摘要以 800 字为宜,英文摘要(建议经专业英语翻译机构润色)与中文摘要要对照;另附关键词。要求来稿应达到“齐、清、定”,中文、英文文字通顺,方可接受送审。
3. 为了进一步统一和完善投稿方式、缩短论文发表周期,本刊只接收网上在线投稿,不接收以邮寄方式或 e-mail 方式的投稿,严禁“一稿多投”,对侵权、抄袭、剽窃等学术不端行为,一经发现,取消三年投稿资格。
4. 文中插图要求完整,图中坐标、线条、单位、符号、图注等应标注准确、完整。如作者特殊要求需出彩色插图者,必须在投稿时事先加以说明,并承担另加的彩印费用。图幅大小:单栏图 7.5cm(宽)×6cm(高);双栏图:14cm(宽)×6cm(高);图中数字、图题、表题全部用中文、英文对照,图中数字、中文、英文全用 6 号字。电子文档中除实物图外,曲线图要用 Matlab, Excel, Visio 或 Origin 等软件制作,稿件中图片的原图并转成相应(可编辑)的文件格式(.fig, .xls, .vsd, .opj),非“.jpg”格式的文档,随电子版修改稿一同发送到本刊的修改稿专用邮箱。
5. 文中出现的单位必须按“中华人民共和国计量标准”及有关 GB 标准规定缮写。物理量符号一律用斜体,单位符号和词头用正体字母。
6. 名词术语,请参照全国科学技术名词规定缮写。
7. 参考文献,采用顺序编码制,只列主要文献;以 15~20 条为宜。内部资料、私人通讯、未经公开发表的一律不能引用。日文、俄文等非英文文献,请用英文表述;中文文献和中文图书采用中、英文对照表述,文献缮写格式请参照本刊。
8. 请在投稿第一页左下角写明投稿联系人的电话和两个 e-mail,以便及时联系。

### 稿件处理

1. 自收到稿件之日起,一个月内作者会收到编辑部的稿件处理意见。请根据录用通知中所提出的要求认真修改,希望修改稿在 30 天内寄回编辑部,并作为作者最终定稿(当作者接到校样时,以此修改稿为准进行校对,请勿再做大的改动),若二个月内编辑部没收到修改稿,将视为自行撤稿处理。
2. 有重大创新并有基金资助者可优先发表;不录用的稿件,编辑部将尽快通知作者,底稿一律不退,请自留底稿。
3. 来稿一经发表将酌致稿酬并送样刊 2 册。
4. 遵照《中华人民共和国著作权法》,投稿作者须明确表示,该文版权(含各种媒体的版权)授权给《光谱学与光谱分析》期刊社。国内外各大文献检索系统摘录本刊刊出的论文;凡不同意被检索刊物无稿酬摘引者,请在投稿时事先声明,否则,本刊一律认为已获作者授权认可。
5. 修改稿请寄:100081 北京市海淀区学院南路 76 号(南院南门),《光谱学与光谱分析》期刊社(收)  
电话:010-62182998 或 62181070 传真:010-62181070  
e-mail: chngpxygpfx@vip.sina.com; 修改稿专用邮箱: gp2008@vip.sina.com 网址: http://www.gpxygpfx.com