

## 近红外光谱结合非负回归系数回归法(配方回归) 解析混合样品的组成比例

李雪莹<sup>1</sup>, 束茹欣<sup>2</sup>, 栾丽丽<sup>1</sup>, 李凯<sup>1</sup>, 杨凯<sup>2</sup>, 李军会<sup>1\*</sup>, 赵龙莲<sup>1</sup>, 张晔晔<sup>1</sup>, 张录达<sup>3</sup>

1. 中国农业大学信息与电气工程学院, 北京 100083
2. 上海烟草集团有限责任公司, 上海 200082
3. 中国农业大学理学院, 北京 100083

**摘要** 准确、快速解析混合样品的组成比例对食品、农产品加工过程的质量控制和配方设计具有重要作用。传统的解决方法多是利用大量代表性样品建立统计模型来实现, 耗时耗力。以醇类及酸类液体混合样品及其低浓度溶液混合样品(无近红外特征吸收的四氯化碳( $\text{CCl}_4$ )为溶剂介质)的透射光谱, 以及片状烟叶样品漫反射光谱为例, 采用导数和 S. G. 平滑, 结合非负回归系数回归法(配方回归)验证了在一定条件下解析混合样品组成比例的可行性。结果表明, 对醇酸类液体透射光谱, 根据非负回归系数回归法得到的解析比例更接近于醇酸类液体的实际摩尔比例, 与实际摩尔比例的误差在 8% 以内, 其低浓度溶液的解析效果更优, 与实际摩尔比例的误差在 4% 以内; 对片状烟叶的漫反射光谱, 根据非负回归系数回归法得到的解析比例与实际质量比例的误差在 10% 以内; 同时, 混合样品的实际光谱与理论解析光谱之间均具有高度一致性,  $F$  和  $t$  检验的结果均在 0.01 水平上, 无显著性差异, 从理论上分析了解析比例的可靠性。该方法只需已知几种纯样品的光谱数据, 即可解析出混合样品的纯样品组成比例, 具有较好应用前景。

**关键词** 近红外光谱; 非负回归系数回归法(配方回归); 比例解析

中图分类号: O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2016)04-0967-05

### 引言

近红外光谱含有丰富的物质信息, 通过对光谱特征的分析, 可以获得有关物质组成及含量的信息, 因此被广泛应用于许多领域<sup>[1-3]</sup>。目前应用近红外光谱的手段主要是通过大量已知属性样本来建立的定量或定性统计模型来实现<sup>[4-5]</sup>。近红外光谱定量分析是依据所得的近红外光谱数据, 建立相关组成成分含量模型, 方法一般采用偏最小二乘法(PLS)<sup>[6]</sup>等。近红外光谱定性分析即通过比较未知样与已知样或标准样品的光谱来确定未知样品的归属, 定性建模方法一般采用人工神经网络(ANN)<sup>[7-8]</sup>、定性偏最小二乘法(DPLS)<sup>[9-10]</sup>等有监督模式识别方法。

传统的定量、定性方法有一定的局限性, 都需要大量代表性样品建立统计模型, 耗时耗力。目前采用纯样品与混合样品近红外光谱直接进行解析组成比例的研究还鲜有报道。主要原因是对水或醇类等溶剂介质的透射光谱<sup>[11-12]</sup>在近红

外区吸收强度过大, 对固体样品的漫发射光谱<sup>[13]</sup>其散射作用差异过大, 均导致原始光谱信噪比下降, 并且应用无约束条件的回归模型<sup>[14-15]</sup>, 使得解析结果不理想, 以及一些液体物质混合后由于发生化学反应等原因使得混合样品光谱不再满足加和性质。

以醇类及酸类液体混合样品及其低浓度溶液混合样品(无近红外特征吸收的四氯化碳( $\text{CCl}_4$ )为溶剂介质)的透射光谱, 以及大小较一致的片状烟叶样品漫反射光谱为例, 采用导数光谱和 S. G. 平滑, 结合非负回归系数回归法(配方回归)验证了在一定条件下解析混合样品组成比例的可行性。通过配方回归能够快速、方便、准确的得到组成比例, 在食品、农产品等的质量控制和配方设计中具有较好的应用前景。

### 1 实验部分

#### 1.1 样品

收稿日期: 2015-03-04, 修订日期: 2015-07-18

基金项目: 国家自然科学基金项目(61144012)和上海烟草集团科技项目资助

作者简介: 李雪莹, 女, 1990年生, 中国农业大学信息与电气工程学院硕士研究生

e-mail: ponneylx@163.com

\*通讯联系人 e-mail: caunir@cau.edu.cn

甲醇(北京化工厂生产,纯度:  $\geq 99.5\%$ ,密度(20 °C):  $0.791 \sim 0.793 \text{ g} \cdot \text{mL}^{-1}$ ,分子量: 32.04);乙醇(北京化工厂生产,纯度:  $\geq 99.5\%$ ,密度(20 °C):  $0.789 \sim 0.791 \text{ g} \cdot \text{mL}^{-1}$ ,分子量: 46.07);乙酸(北京化工厂生产,纯度:  $\geq 99.5\%$ ,密度(20 °C):  $1.049 \sim 1.051 \text{ g} \cdot \text{mL}^{-1}$ ,分子量: 60.05);四氯化碳(北京化工厂生产,纯度:  $\geq 99.5\%$ ,密度(20 °C):  $1.592 \sim 1.598 \text{ g} \cdot \text{mL}^{-1}$ ,分子量: 153.82);上、中、下三个不同部位的烟叶样品。

### 1.1.1 醇类及酸类液体透射光谱实验样品的制备

分别配置甲醇、乙醇、乙酸体积比为 2:6:2(pure1), 6:2:2(pure2), 1:1:1(pure3)三个实验样品,并按照式(1)和式(2)计算相应的摩尔比例。

### 1.1.2 醇类及酸类低浓度溶液透射光谱实验样品的制备

甲醇、乙醇、乙酸分别与四氯化碳按 1:9 的比例混合配成的稀溶液后,分别配置甲醇、乙醇、乙酸稀溶液体积比为 2:6:2(dilute1), 6:2:2(dilute2), 1:1:1(dilute3)三个实验样品,并按照式(1)和式(2)计算相应的摩尔比例。

### 1.1.3 片状烟叶漫反射光谱实验样品的制备

将上、中、下烟叶样品在 55 °C 下烘 3 h 后磨成直径 0.3 ~ 3 cm 大小的片状,分别配置上、中、下不同部位的片状烟叶质量比为 2:6:2(tobacco1), 6:2:2(tobacco2), 1:1:1(tobacco3)三个实验样品。

$$n = \frac{W\rho V}{M} \quad (1)$$

$$c_1 : c_2 : c_3 = \frac{n_1}{n_1 + n_2 + n_3} : \frac{n_2}{n_1 + n_2 + n_3} : \frac{n_3}{n_1 + n_2 + n_3} \quad (2)$$

式(1)中  $W$  为纯溶液的纯度,  $\rho$  为纯溶液的密度,  $V$  为总溶液的体积,  $M$  为纯溶液的分子量;式(2)中  $n_1$ ,  $n_2$  和  $n_3$  分别为按照式(1)计算的甲醇、乙醇、乙酸的相对物质的量。

## 1.2 仪器与测试方法

仪器为 Bruker 型傅里叶变换近红外光谱仪(德国布鲁克光谱仪器公司)。工作参数:光谱采集范围  $12\,000 \sim 4\,000 \text{ cm}^{-1}$ ;光谱分辨率:  $8 \text{ cm}^{-1}$ ;扫描次数: 64 次。

纯液体、低浓度溶液及各种不同组分比例的液体实验样品采用仪器前光路透射方式进行测量,其中纯液体及其不同组分比例的液体实验样品池的光程为 1 mm,低浓度溶液及其各种不同组分比例的液体实验样品池光程为 2 mm。

不同类型及组分比例的片状烟叶样品采用积分球漫反射方式进行光谱测量。

### 1.3 含有线性约束及非负回归系数的回归算法(配方回归)

含有线性约束及非负回归系数的回归算法,又称配方回归算法,由我国著名应用数学专家方开泰等设计提出,并在工业中<sup>[16-18]</sup>得到成功运用,该算法描述模型如下<sup>[19-20]</sup>

$$Y = X\beta + \epsilon \quad (3a)$$

$$\text{PR: } I_p'\beta = 1 \quad (3b)$$

$$\beta \geq 0 \quad (3c)$$

模型 PR 中  $Y$  为  $n \times 1$ ,  $X$  为  $n \times p (n \geq p)$ ,  $\beta$  为  $p \times 1$ ,  $I_p$  为元素全为 1,  $\epsilon$  为  $n \times 1$ 。如用  $\beta$  的最小二乘估计就是在约束式(3b)和式(3c)之下使

$$Q = (Y - X\beta)(Y - X\beta) \rightarrow \min \quad (4)$$

得到  $\beta^*$  值,利用矩阵的消去变换,解得  $\beta^*$ 。

$Y$  为混合样品光谱数据向量( $n \times 1$ ,  $n$  为光谱点数),  $X$  为纯样品光谱数据矩阵( $n \times p (n \geq p)$ ),  $p$  为样本数,  $n$  为光谱点数,按照配方回归算法模型解得的配比向量  $\beta^*$ ,即为混合样品中各纯样品的组成比例。

## 2 结果与讨论

为消除由于样品松紧度、器皿状态、环境温度等对光谱产生的系统偏差,提高分析结果的可靠程度,采用一阶导数与 S-G 平滑处理后的光谱进行混合样品的组成比例解析。

### 2.1 醇类及酸类液体混合样品透射光谱的组成比例解析

甲醇、乙醇及乙酸纯液体样品导数光谱及其低浓度溶液样品导数光谱分别见图 1、图 2。按 1.1.1 及 1.1.2 中配置的不同比例纯液体混合样品的实际体积比例、实际摩尔比例及应用配方回归求解所得的解析比例分别见表 1、表 2。

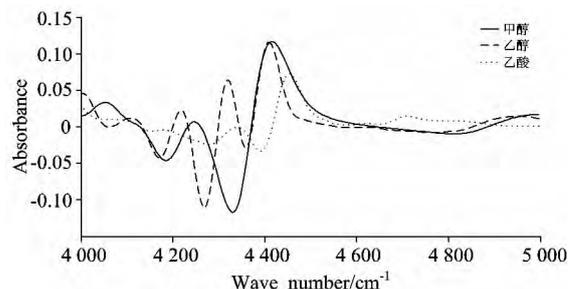


图 1 三种醇酸类液体导数光谱

Fig 1 The derivative spectra of three kinds of alcohol and acids liquid samples

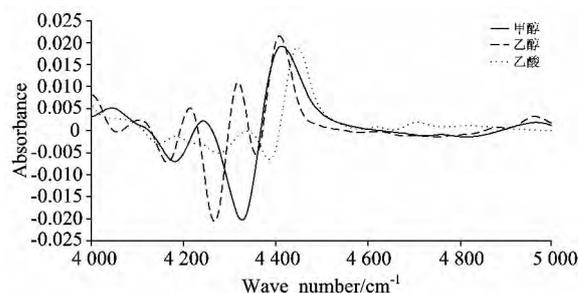


图 2 三种低浓度溶液样品导数光谱

Fig 2 The derivative spectra of three kinds of dilute solution samples

由表 1、表 2 可以看出,纯液体混合样品及低浓度溶液混合样品解析出的组成比例均更接近于实际摩尔比例,其中纯液体混合样品解析比例与实际摩尔比例的误差在 8% 以内,低浓度溶液的误差在 4% 以内,优于纯液体混合样品解析比例的结果。

对纯液体及低浓度溶液液体实际混合样品光谱与解析光谱之间计算得出的显著性检验方法中的  $F$  值、 $t$  值以及相关系数  $r$  值统计分析结果分别见表 3、表 4,结果表明在  $\alpha =$

0.01 水平下, 均无显著性差异。

表 1 醇酸类液体样品实际体积比例、实际摩尔比例与解析比例结果

Table 1 The result of actual volumetric proportion, actual molar proportion and analytic proportion in alcohol and acids liquid samples

	pure1	pure2	pure3
实际体积比例	0.20 : 0.60 : 0.20	0.60 : 0.20 : 0.20	0.33 : 0.33 : 0.33
实际摩尔比例	0.26 : 0.55 : 0.19	0.68 : 0.16 : 0.16	0.42 : 0.29 : 0.29
解析比例	0.23 : 0.54 : 0.23	0.63 : 0.17 : 0.20	0.36 : 0.27 : 0.37

表 2 低浓度溶液样品实际体积比例、实际摩尔比例与解析比例结果

Table 2 The result of actual volumetric proportion, actual molar proportion and analytic proportion in dilute solution samples

	dilute1	dilute2	dilute3
实际体积比例	0.20 : 0.60 : 0.20	0.60 : 0.20 : 0.20	0.33 : 0.33 : 0.33
实际摩尔比例	0.26 : 0.55 : 0.19	0.68 : 0.16 : 0.16	0.42 : 0.29 : 0.29
解析比例	0.26 : 0.59 : 0.15	0.72 : 0.15 : 0.13	0.41 : 0.30 : 0.29

表 3 醇酸类液体样品实际混合光谱与解析光谱之间统计分析

Table 3 The statistical analysis between actual spectra and analytic spectra in alcohol and acids samples

	pure1 实际光谱与解析光谱	pure2 实际光谱与解析光谱	pure3 实际光谱与解析光谱	检验临界值
F 值	1.135 9	1.083 3	1.224 7	1.371 8
t 值	0.021 4	0.007 6	0.012 9	2.334 9
R 值	0.988 2	0.993 0	0.979 3	

注: 解析光谱为根据醇酸类液体混合样品的解析比例与纯醇酸类样品光谱按照线性加和计算得到光谱。

Note: Analytic spectra are calculated by linear addition according to analytic proportion and the spectra of alcohol and acids

表 4 低浓度溶液样品实际混合光谱与解析光谱之间统计分析

Table 4 The statistical analysis between actual spectra and analytic spectra in dilute solution samples

	dilute1 实际光谱与解析光谱	dilute2 实际光谱与解析光谱	dilute3 实际光谱与解析光谱	检验临界值
F 值	1.193 4	1.223 8	1.271 6	1.371 8
t 值	0.126 2	0.180 8	0.121 5	2.334 9
R 值	0.993 3	0.991 7	0.987 4	

注: 解析光谱为根据低浓度溶液混合样品的解析比例与纯低浓度溶液样品光谱按照线性加和计算得到光谱。

Note: Analytic spectra are calculated by linear addition according to analytic proportion and the spectra of dilute solution

## 2.2 不同类型片状烟叶混合样品漫反射光谱的组成比例解析

上、中、下三部位片状烟叶样品导数光谱见图 3。

按 1.1.3 中配置的各不同比例片状烟叶混合样品的实际质量比例及应用配方回归求解所得的解析比例见表 5。

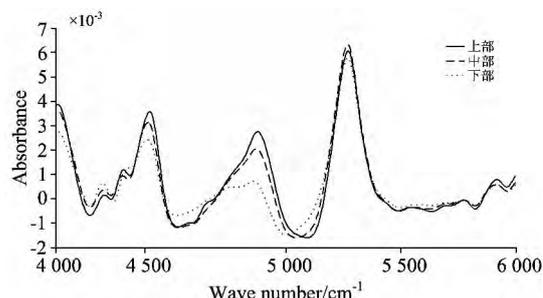


图 3 三种片状烟叶样品导数光谱

Fig 3 The derivative spectra of three kinds of sheet tobacco leaf samples

表 5 片状烟叶混合样品实际质量比例与解析比例结果

Table 5 The result of actual mass proportion and analytic proportion in sheet tobacco leaf samples

	tobacco1	tobacco2	tobacco3
实际质量比例	0.20 : 0.60 : 0.20	0.60 : 0.20 : 0.20	0.33 : 0.33 : 0.33
解析比例	0.20 : 0.58 : 0.22	0.51 : 0.24 : 0.25	0.23 : 0.37 : 0.40

由表 5 可以看出, 片状烟叶混合样品解析出的组成比例接近于实际质量比例, 解析比例与实际质量比例的误差在 10% 以内, 准确度较好; 对片状烟叶混合样品光谱与解析光谱之间计算得出的显著性检验方法中的 F 值、t 值以及相关系数 r 值统计分析结果见表 6, 结果表明在  $\alpha=0.01$  水平下, 无显著性差异。

表 6 片状烟叶混合样品实际光谱与解析光谱之间结果分析

Table 6 The statistical analysis between actual spectra and analytic spectra in sheet tobacco leaf samples

	tobacco1 实际光谱与解析光谱	tobacco 2 实际光谱与解析光谱	tobacco 3 实际光谱与解析光谱	检验临界值
F 值	1.002 9	1.001 3	1.002 4	1.237 8
t 值	0.011 8	0.012 3	0.009 9	2.330 2
R 值	0.999 9	0.999 8	0.999 9	

注: 解析光谱为根据片状烟叶混合样品的解析比例与纯片状烟叶样品光谱按照线性加和计算得到光谱。

Note: Analytic spectra are calculated by linear addition according to analytic proportion and the spectra of sheet tobacco leaf

## 3 结 论

应用近红外导数光谱结合非负回归系数回归法(配方回归)解析混合样品的组成比例, 只需已知几种纯样品的光谱数据, 无需采用大量数据建模, 即可解析出混合样品的组成

比例,在食品、农产品等的质量控制和配方设计中具有较好的应用前景。

所应用的液体透射样品均具有极性互溶性质,所应用的

固体漫发射样品为片状结构,对于其他类型样品的透射或漫反射光谱是否能够得到理想的解析结果还需要进一步验证。

## References

- [ 1 ] Rinnan A, van den Berg F, Engelsen S B. *TrAC-Trends in Analytical Chemistry*, 2009, 28(10): 1201.
- [ 2 ] Qian G, Wang Z Y. *Chemistry-an Asian Journal*, 2010, 5(5): 1006.
- [ 3 ] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al(严衍禄, 赵龙莲, 韩东海, 等). *Foundation and Application of NIR Spectra Analysis(近红外光谱分析基础与应用)*. Beijing: China Light Industry Press(北京: 中国轻工业出版社), 2005. 13.
- [ 4 ] Guillaume D, Casetta C, Bicchì C, et al. *Journal of Chromatography A*, 2010, 1217(44): 6882.
- [ 5 ] Bulat FA, Toro-Labbe A, Brinck T, et al. *Journal of Molecular Modeling*, 2010, 16(11): 1679.
- [ 6 ] Gu Haiwei, Pan Zhengzheng, Xi Bowei, et al. *Analytica Chimica Acta*, 2011, 688(1-2): 57.
- [ 7 ] Martelo-Vidal M J, Vazquez M. *CyTA-Journal of Food M*, 2015, 13(1): 32.
- [ 8 ] Tenorio-Borroto E, Rivas C G P, Chagoyan J C V, et al. *Bioorganic & Medicinal Chemistry*, 2012, 20(20): 6181.
- [ 9 ] Galtier O, Abbas O, Le Dreau Y, et al. *Vibrational Spectroscopy*, 2011, 55(1): 132.
- [ 10 ] Tong Hong, Tang Jun, Zhang Zhengfang. *China Condiment*, 2014, 39(4): 54.
- [ 11 ] Zude M, Pflanz M, Spinelli L, et al. *Journal of Food Engineerin*, 2011, 103(1): 68.
- [ 12 ] Chen X G, Li W W, Wu J D, et al. *Materials Research Bulletin*, 2012, 47(1): 111.
- [ 13 ] Liu W, Wang Z Z, Qing J P, et al. *Pharmacognosy Magazine*, 2014, 10(40): 441.
- [ 14 ] Tabari H, Sabziparvar A A, Ahmadi M. *Meteorology and Atmospheric Physics*, 2011, 110(3-4): 135.
- [ 15 ] WANG Ju-xiang, LI Hua, XING Zhi-na, et al(王菊香, 李 华, 邢志娜, 等). *Chinese Journal of Analytical Chemistry(分析化学)*, 2011, 39(6): 846.
- [ 16 ] Wang Jianguo, Jang Shishang, Wong David Shan-Hill. *Control Engineering Practice*, 2013, 21(9): 1157.
- [ 17 ] Pan Changchun, Bai Jie, Yang Genke. *Journal of Process Control*, 2012, 22(9): 1637.
- [ 18 ] FANG Kai-tai, HE Shu-dong(方开泰, 贺曙东). *Mathematica Numerica Sinica(计算数学)*, 1985, 8: 237.
- [ 19 ] FANG Kai-tai, WANG Dong-qian, WU Guo-fu(方开泰, 王东谦, 吴国富). *Mathematica Numerica Sinica(计算数学)*, 1982, 2: 57.
- [ 20 ] DONG Han-zhong, ZHANG Xiao-wei(董汉忠, 张小薇). *Journal of Computer Applications(计算机应用)*, 1983, 4: 40.

## Study on the Analysis of the Proportion of Mixed Samples with Near Infrared Spectroscopy and Non-Negative Coefficients Regression

LI Xue-ying<sup>1</sup>, SHU Ru-xin<sup>2</sup>, LUAN Li-li<sup>1</sup>, LI Kai<sup>1</sup>, YANG Kai<sup>2</sup>, LI Jun-hui<sup>1\*</sup>, ZHAO Long-lian<sup>1</sup>, ZHANG Ye-hui<sup>1</sup>, ZHANG Lu-da<sup>3</sup>

1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Technology Center of Shanghai Tobacco (Group) Corporation, Shanghai 200082, China

3. College of Science, China Agricultural University, Beijing 100083, China

**Abstract** In the process of practical production, it is important to accurately analyze the proportion of mixed samples with high speed, which plays a great role for quality control and formulation design in food and agricultural processing. Traditional solution is to build statistical model with a large number of representative samples, which is both labor-intensive and time-consuming. In this paper, the proportion of alcohol and acids mixed samples, and their dilute solution mixed samples(used carbon tetrachloride (CCl<sub>4</sub>) which has no near-infrared absorption characteristics as the solvent medium), as well as sheet tobacco leaf mixed samples are respectively analyzed by using near infrared spectroscopy, SG smooth and non-negative coefficients regression, which verifies the feasibility of analyzing the proportion of the mixed samples. The results show that, the analytic proportion of transmission spectra of alcohol and acids according to non-negative coefficients regression is closer to actual molar proportion with result error less than 4%. The result of the dilute solution is much better with error less than 4%. The analytic proportion of diffuse reflectance spectra of sheet tobacco leaf according to non-negative coefficients regression is highly accurate with error less than 10%. In the meantime, it has a highly consistency between actual spectra and analytic spectra of mixed samples; and the result of F-test and T-test shows that there is no significant difference between them and the confidence level is 0.01. It has the reliability of

analytical proportion in theory. With the spectral data of several pure samples, the proportion of mixed samples can be thus analyzed, which has a good application prospect for quality control and formulation design in food and agricultural processing.

**Keywords** Near infrared spectroscopy; Non-negative coefficients regression; Proportion analyzing

(Received Mar. 4, 2015; accepted Jul. 18, 2015)

\* Corresponding author

欢迎投稿      欢迎订阅      欢迎刊登广告

## 《光谱学与光谱分析》2016年征订启事

国内邮发代码：82-68

国外发行代码：M905

《光谱学与光谱分析》1981年创刊，国内统一刊号：CN 11-2200/O4，国际标准刊号：ISSN 1000-0593，CODEN 码：GYGFED，国内外公开发行人，大 16 开本，306 页，月刊；是中国科协主管，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学术性刊物。北京大学出版社出版，每期售价 55.00 元，全年 660 元。刊登主要内容：激光光谱测量、红外、拉曼、紫外、可见光谱、发射光谱、吸收光谱、X 射线荧光光谱、激光显微光谱、光谱化学分析、国内外光谱化学分析领域内的最新研究成果、开创性研究论文、学科发展前沿和最新进展、综合评述、研究简报、问题讨论、书刊评述。

《光谱学与光谱分析》适用于冶金、地质、机械、环境保护、国防、天文、医药、农林、化学化工、商检等各领域的科学研究单位、高等院校、制造厂家、从事光谱学与光谱分析的研究人员、高校有关专业的师生、管理干部。

《光谱学与光谱分析》为我国首批自然科学核心期刊，中国科协优秀科技期刊，中国科协择优支持基础性、高科技学术期刊，中国科技论文统计源刊，“中国科学引文数据库”，“中国物理文摘”，“中国学术期刊文摘”，同时被国内外的 CSCD，SCI，AA，CA，Ei，AJ，MEDLINE，Scopus 等文献机构收录。根据国家科技部信息研究所发布信息，中国科技期刊物理类影响因子及引文量《光谱学与光谱分析》都居前几位。欢迎国内外厂商在《光谱学与光谱分析》发布广告（广告经营许可证：京海工商广字第 8094 号）。

《光谱学与光谱分析》的主编为高松院士。

欢迎新老客户到全国各地邮局订阅，若有漏订者可直接与《光谱学与光谱分析》期刊社联系。

联系地址：北京市海淀区学院南路 76 号（南院），《光谱学与光谱分析》期刊社      邮政编码：100081

联系电话：010-62181070，62182998

电子信箱：chnghpxygpfx@vip.sina.com；修改稿专用邮箱：gp2008@vip.sina.com

网 址：http://www.gpxygpfx.com