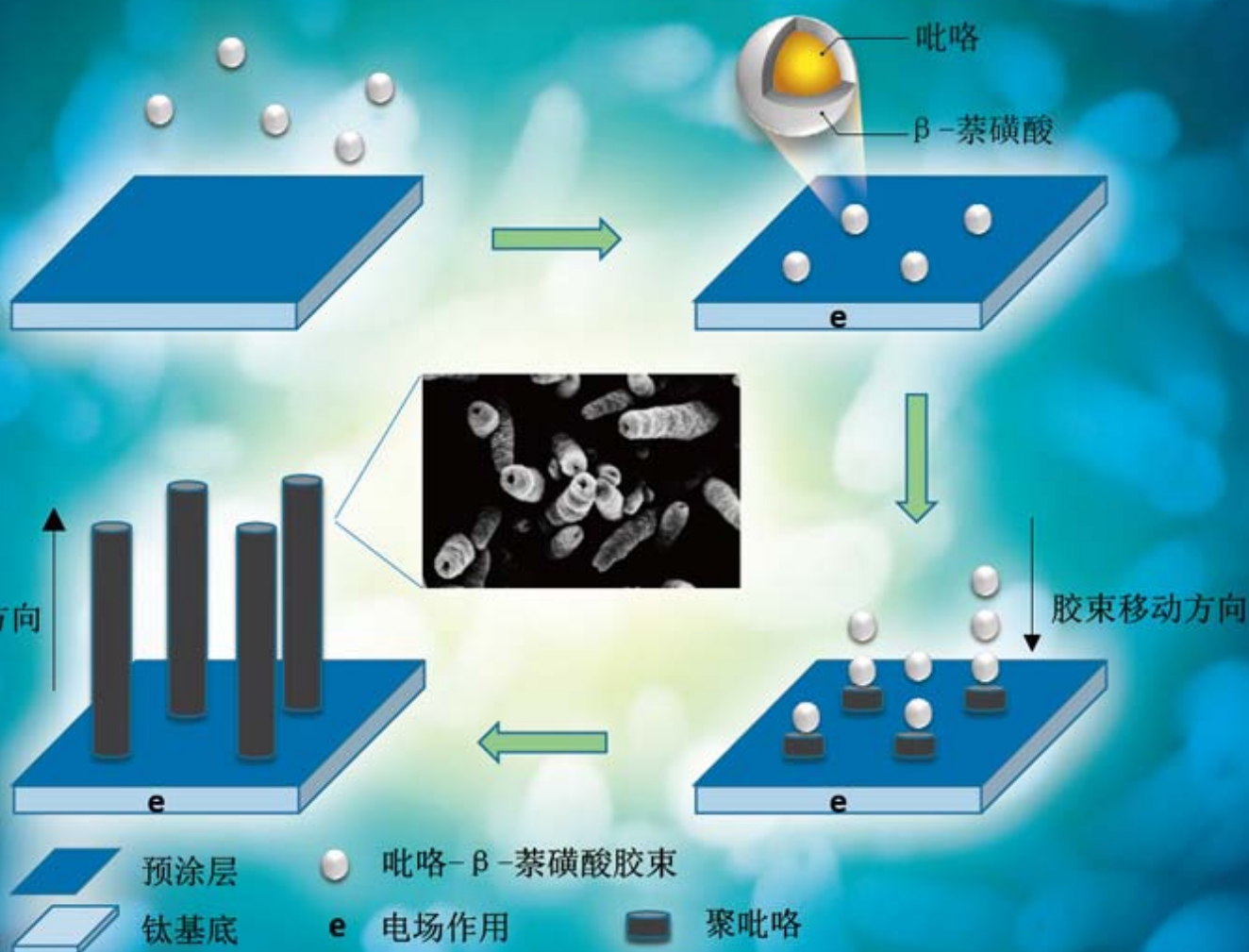


SCIENTIA SINICA Chimica

中国科学 化学

第44卷 第10期 2014年10月 1499-1684 ISSN 1674-7224 CN 11-5838/O6



中国科学院 主办
国家自然科学基金委员会

《中国科学》《科学通报》

荣誉总主编：周光召 总主编：朱作言

《中国科学：化学》编辑委员会

主 编：万立骏 中国科学院化学研究所

副主编：曹 镛 华南理工大学材料科学与工程学院 陈洪渊 南京大学化学化工学院 冯守华 吉林大学化学学院

李亚栋 清华大学化学系 林国强 中国科学院上海有机化学研究所

田 禾 华东理工大学化学与分子工程学院 田中群 厦门大学化学化工学院 薛子陵 University of Tennessee, USA

编 委：

无机化学

卜显和	南开大学化学学院	陈小明	中山大学化学与化学工程学院	高 松	北京大学化学与分子工程学院
郭子建	南京大学化学化工学院	洪茂椿	中国科学院福建物质结构研究所	刘 俊	Pacific Northwest National Laboratory, USA
任咏华	香港大学化学系	熊仁根	东南大学化学化工学院	张洪杰	中国科学院长春应用化学研究所
郑南峰	厦门大学化学化工学院	左景林	南京大学化学化工学院		

有机化学

丁奎岭	中国科学院上海有机化学研究所	黄培强	厦门大学化学化工学院	黄 震	Georgia State University, USA
蒲 林	University of Virginia, USA	索祖才	The Ohio State University, USA	王梅祥	清华大学化学系
席 真	南开大学化学学院	谢作伟	香港中文大学化学系	姚祝军	南京大学化学化工学院
张德清	中国科学院化学研究所	张礼和	北京大学医学部药学院	赵玉芬	厦门大学化学化工学院

高分子科学

安立佳	中国科学院长春应用化学研究所	陈国平	National Institute for Materials Science, Japan	高长有	浙江大学高分子科学与工程系
管治斌	University of California, Irvine, USA	李永舫	中国科学院化学研究所	乔金樑	中国石油化工股份有限公司北京化工研究院
唐本忠	香港科技大学化学系	汤华荣	Commonwealth Scientific & Industrial Research Organisation (CSIRO), Australia	杨 柏	吉林大学化学学院
宛新华	北京大学化学与分子工程学院				

物理化学

包信和	中国科学院大连化学物理研究所	韩布兴	中国科学院化学研究所	胡培君	Queen's University, UK
梁文平	国家自然科学基金委员会	刘智攀	复旦大学化学系	刘忠范	北京大学化学与分子工程学院
帅志刚	清华大学化学系	王 树	中国科学院化学研究所	王 野	厦门大学化学化工学院
杨学明	中国科学院大连化学物理研究所	张劲松	University of California, Riverside, USA	张金中	University of California, Santa Cruz, USA
赵新生	北京大学化学与分子工程学院	庄 林	武汉大学化学与分子科学学院		

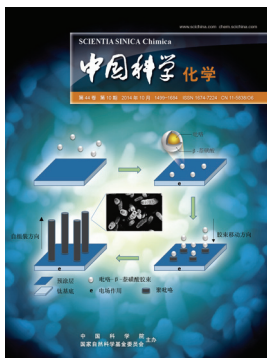
分析和环境化学

C. Amatore	Ecole Normale Supérieure, France	柴之芳	中国科学院高能物理研究所	陈 义	中国科学院化学研究所
R. Compton	University of Oxford, UK	鞠焯先	南京大学化学化工学院	乐晓春	University of Alberta, Canada
林金明	清华大学化学系	邵元华	北京大学化学与分子工程学院	谭蔚泓	湖南大学化学化工学院
汪海林	中国科学院生态环境研究中心	杨芾原	复旦大学生物医学研究院	张玉奎	中国科学院大连化学物理研究所
赵进才	中国科学院化学研究所	赵宇亮	国家纳米科学中心	朱 彤	北京大学环境科学与工程学院

化学工程

段 雪	北京化工大学理学院	何鸣元	中国石化石油化工科学研究院	陆小华	南京工业大学化学化工学院
骆广生	清华大学化学工程系	聂书明	Georgia Institute of Technology and Emory University, USA	孙立成	Royal Institute of Technology, Sweden
徐春明	中国石油大学(北京)	张锁江	中国科学院过程工程研究所	张 涛	中国科学院大连化学物理研究所

责任编辑：付利 许军舰 张学梅 郑建芬 封面设计：胡 煜

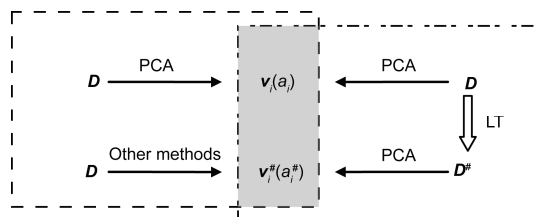


封面说明 通过调节制备过程中的掺杂剂浓度、单体浓度及温度等因素, 华南理工大学宁成云课题组实现了对聚吡咯纳米线形状及取向性的调控. 封面为该课题组采用电化学无模板法制备出聚吡咯纳米线材料的示意图. 该法具有工艺简单、反应时间短等特点, 其对形貌的可控性对改进聚吡咯材料的性能有一定的参考意义(见本期论文: 陈俊祺, 廖景文, 周正难, 谭帼馨, 宁成云. 导电聚吡咯纳米线的电化学无模板法可控合成与表征. p1570~1575).

专题论述

复杂化学体系测量数据的主因子数估计方法的分类及进展

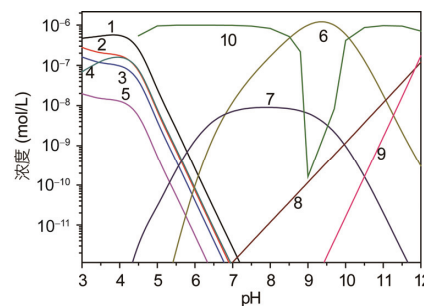
陆玮, 邵利民
 中国科学:化学, 2014, 44(10): 1499-1507



评述

地球化学模拟在高放废物地质处置中的应用与发展

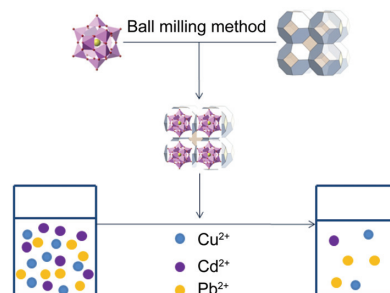
蒋美玲, 陈涛, 康明亮, 王祥云, 刘春立
 中国科学:化学, 2014, 44(10): 1508-1520



论文

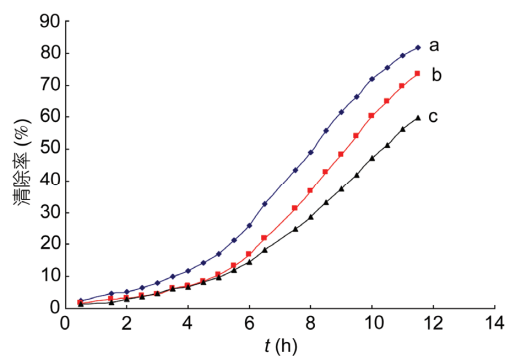
在温和条件下合成金属有机骨架与多酸复合材料及其性能研究

任晓倩, 李锐, 宋玉娜, 冯霄, 王博
 中国科学:化学, 2014, 44(10): 1521-1527



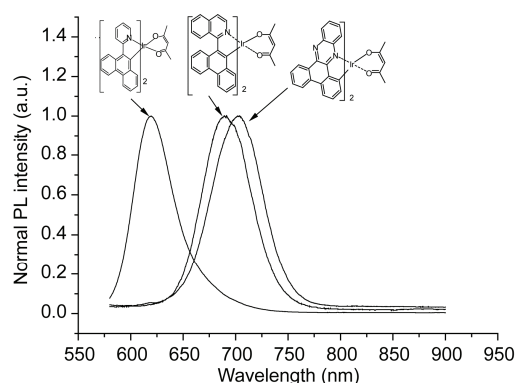
双水杨醛缩二乙烯三胺 Cu(II)配合物的合成
及催化染料降解性能

江银枝, 时永强, 史银瓶, 徐火英
中国科学:化学, 2014, 44(10): 1528–1535



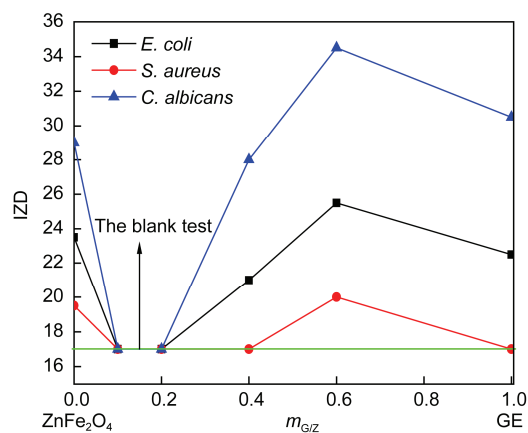
一类以菲衍生物为配体的新型红色到近红外磷光配合物的
合成及其光致和电致发光性质

骆开均, 陈艳芳, 苏祎伟, 邓先平, 朱卫国
中国科学:化学, 2014, 44(10): 1536–1543



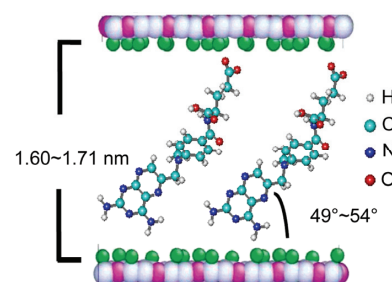
石墨烯/锌铁氧体复合物的制备及抗菌性能

马歌, 陈瑶瑶, 李良超, 蒋冬花, 诸雨蒙, 郑婷婷
中国科学:化学, 2014, 44(10): 1544–1552



甲氨蝶呤/锌铝层状双氢氧化物的合成以及生物效应研究

曾美桂, 李晓东, 李淑萍
中国科学:化学, 2014, 44(10): 1553–1561

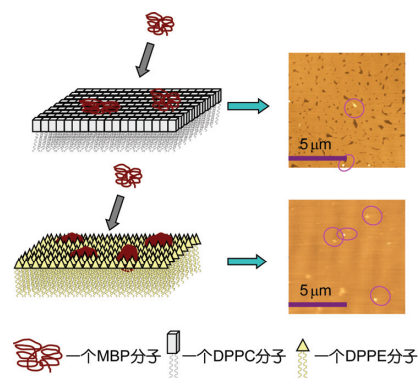


髓鞘碱性蛋白对细胞膜脂质分子 DPPC、DPPE

单层膜影响机理研究

张蕾, 孙润广, 郝长春, 杨静, 何光晓

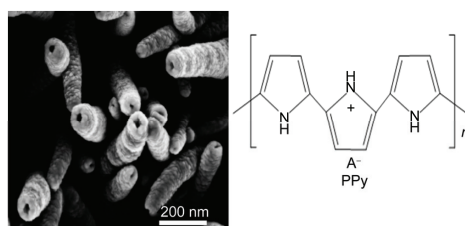
中国科学:化学, 2014, 44(10): 1562-1569



导电聚吡咯纳米线的电化学无模板法可控合成与表征

陈俊祺, 廖景文, 周正难, 谭帼馨, 宁成云

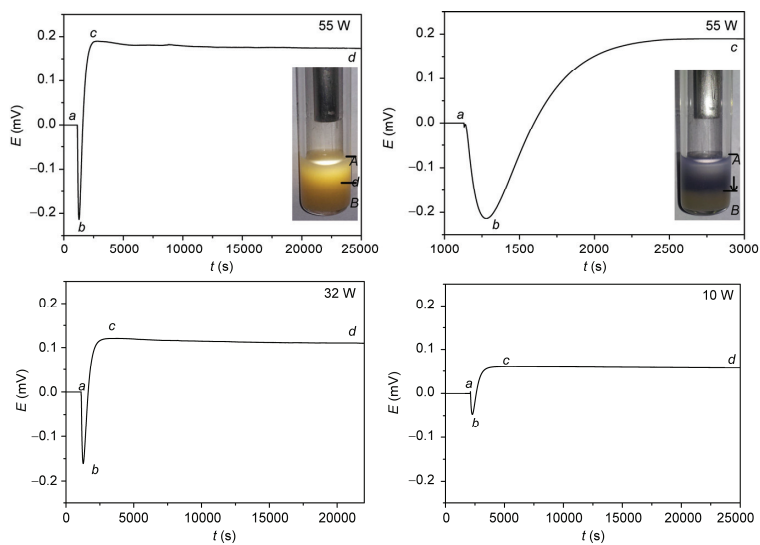
中国科学:化学, 2014, 44(10): 1570-1575



可见光驱动 Ag@AgCl 催化反应的原位微量热研究

李星星, 范高超, 马昭, 谭学才, 黄在银

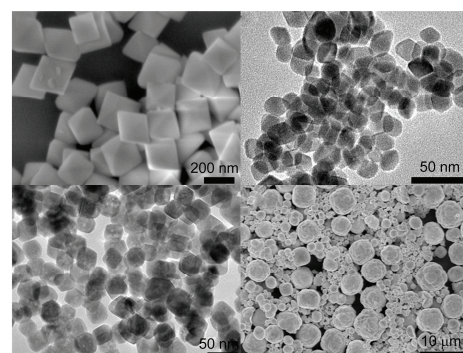
中国科学:化学, 2014, 44(10): 1576-1584



不同粒度八面体纳米钼酸镉的表面热力学性质

刘晓林, 范高超, 黄在银, 刘作娇, 王腾辉, 谭学才

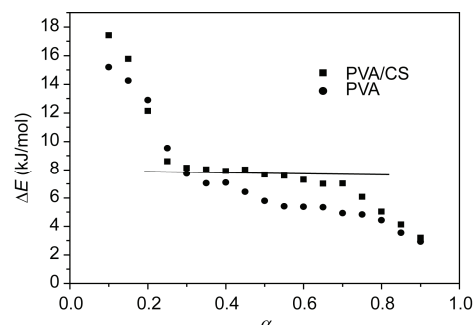
中国科学:化学, 2014, 44(10): 1585-1590



聚乙烯醇/壳聚糖复合水凝胶的物理化学性质

王娜, 王园园, 陈兴波, 白同春

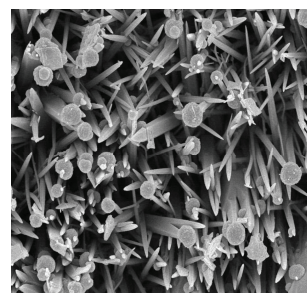
中国科学:化学, 2014, 44(10): 1591-1598



取向纳米 ZnO/Cu₂O 异质结阵列的制备及光电特性

宋亚男, 李梦轲, 张天民, 孙杰婷, 王楠, 王玉琼, 杨蕊, 付宏波, 冯秋菊, 顾吉林

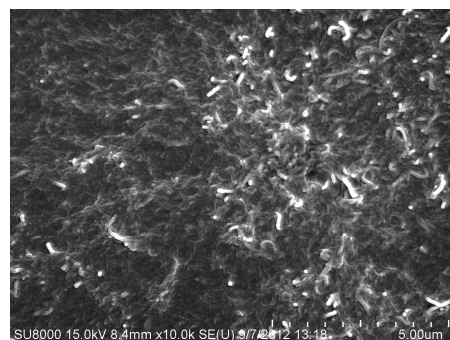
中国科学:化学, 2014, 44(10): 1599-1608



ZnO 纳米线基 MWNTs/PVDF 复合热电材料的制备及特性研究

王玉琼, 李梦轲, 张天民, 李雅楠, 杨蕊, 付宏波, 孙杰婷, 许宁宁, 王军艳

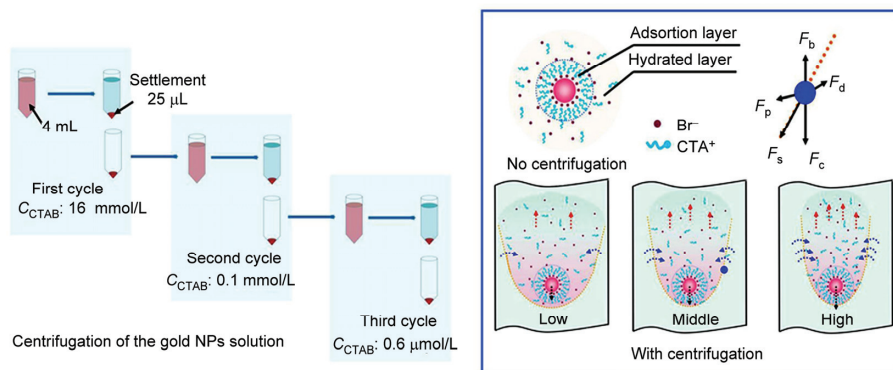
中国科学:化学, 2014, 44(10): 1609-1617



基于离心技术的金纳米粒子表面清洁方法和机理研究

江庆宁, 尹秉胜, 林海昕, 李幻, 杨凝, 刘德宇, 田中群

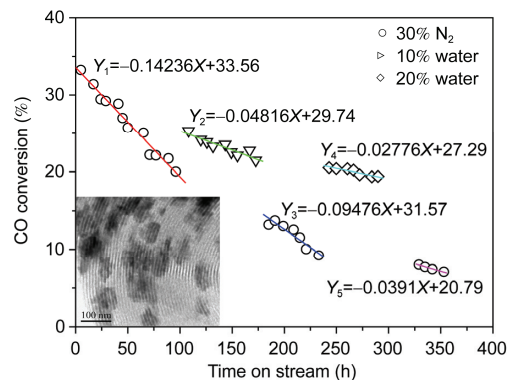
中国科学:化学, 2014, 44(10): 1618-1626



水对 SBA-15 负载的钴基费-托合成催化剂催化性能的影响

赵燕熹, 张煜华, 韦良, 李金林

中国科学:化学, 2014, 44(10): 1627-1632

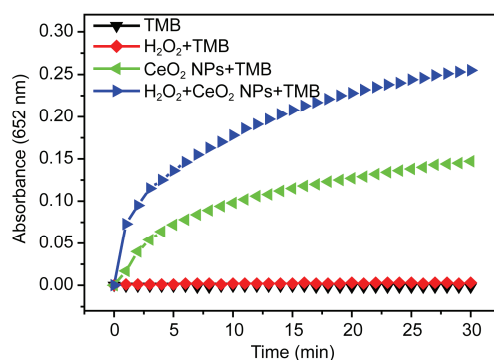


二氧化铈纳米微粒过氧化物酶活性及其在葡萄糖

检测中的应用

石文兵, 刘希东, 张小丹, 黄玉明

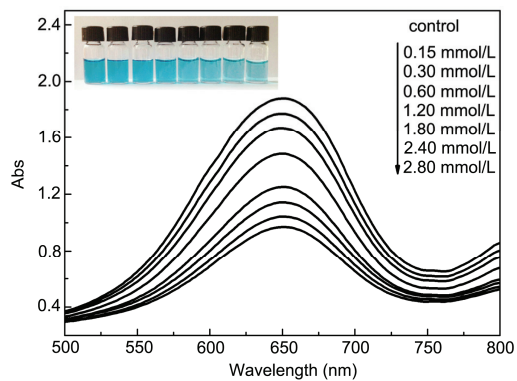
中国科学:化学, 2014, 44(10): 1633-1640



DNA 增强金纳米颗粒过氧化物模拟酶活性检测 K⁺

邹柏舟, 刘跃, 王健, 黄承志

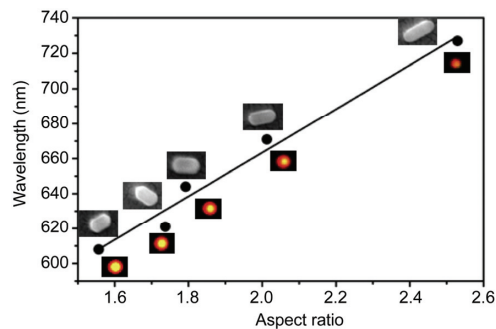
中国科学:化学, 2014, 44(10): 1641-1646



单金纳米棒的光散射分析化学: 径向比与分析灵敏度的关系

马军, 刘跃, 黄承志

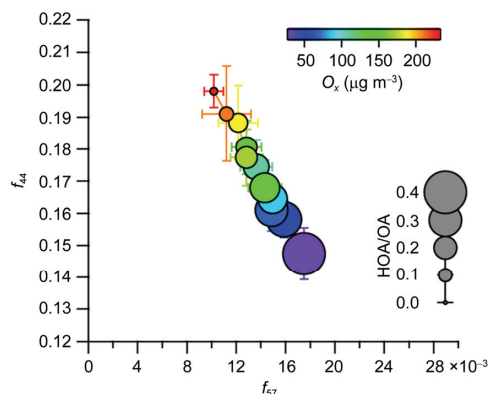
中国科学:化学, 2014, 44(10): 1647-1653



南京夏季大气有机气溶胶老化过程在线观测研究

张运江, 汤莉莉, 王壮, 于红霞, 张祥志, 周宏仓,
陈杨, 江蓉馨

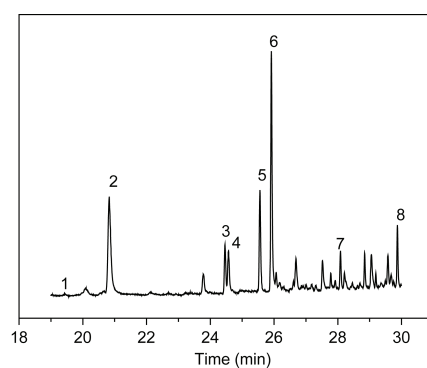
中国科学:化学, 2014, 44(10): 1654-1664



金属氧化物对十氯联苯的降解研究

刘莎, 石瑞芳, 黄学敏, 黄林艳, 刘焯焯, 黎烈武, 芦会杰,
杨帆, 苏贵金

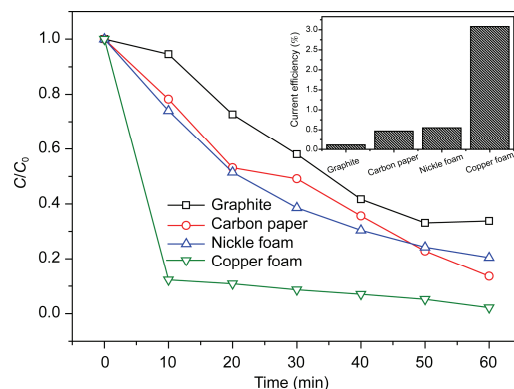
中国科学:化学, 2014, 44(10): 1665-1674



泡沫铜电极电还原去除溴酸盐研究

李昂臻, 冒冉, 赵旭

中国科学:化学, 2014, 44(10): 1675-1681



亮点介绍

纳米颗粒化学大幅提升贵金属催化剂的催化活性界面

李亚栋

中国科学:化学, 2014, 44(10): 1682-1683

超细的显微结构使超硬材料变得更硬

邹广田

中国科学:化学, 2014, 44(10): 1684



专题论述

复杂化学体系测量数据的主因子数估计方法的分类及进展

陆玮, 邵利民*

中国科学技术大学化学系, 合肥 230026

*通讯作者, E-mail: lshao@ustc.edu.cn

收稿日期: 2013-10-28; 接受日期: 2013-12-02; 网络版发表日期: 2014-05-04

doi: 10.1360/N032013-00051

摘要 复杂化学体系的数据解析中, 估计主因子数是必要步骤, 同时也是一个难题. 目前存在多种各具特点的估计方法. 对这些方法进行整理和归纳可为解决这一难题提供全局视角下的综合性信息, 有助于进一步研究. 本文结合本课题组的相关研究对主因子数估计问题进行了系统深入的分析; 整理了近年来出现的各种估计方法, 并将之分为三类, 即经验方法、数学原理完备的方法以及统计方法, 并对每类方法的特点和共性进行了分析和说明; 归纳出第二类方法的基本原理.

关键词
数据矩阵
降维
主因子数
主成分分析

1 引言

仪器的发展使矩阵形式的二维测量数据在分析化学中渐成主流, 如 GC-MS 数据和 HPLC-DAD 数据, 或一系列一维测量数据的组合^[1, 2]. 这种数据包含了丰富的化学信息, 但同时也存在诸多干扰和冗余, 导致数据解析方面的困难. 数据解析的关键在于去除冗余, 提取主要信息, 即对数据矩阵进行“降维”(dimensionality reduction). 降维的重要结果之一是信息主要贡献者的数目——主因子数(number of primary/significant factors)^[3, 4].

主要因子向数据矩阵贡献了独立的、彼此无关的信息, 所以主因子数在概念上类似于线性代数中矩阵的“秩”, 在理想情况下等于被分析样品的化学组分数. 但是在实际分析中, 干扰不可避免. 尽管这些干扰属于次要因子(secondary/insignificant factors), 但仍导致测量数据矩阵满秩, 其秩一般远大于主因子数. 主因子数与矩阵的秩有内在联系却不同, 这也是存在其他表述方式的原因, 如化学秩(chemical rank)^[5]和伪秩(pseudo-rank)^[6]等. 这些表述方式虽然

没有统一, 但本质相同. 本文采用“主因子数”这一名称.

多元校正分析中存在相似概念, 指建立最佳模型所需要的潜变量数, 一般通过交叉验证^[7, 8], 或借助独立验证集来获得^[9].

通过主因子数可以了解被分析样品的化学组分数, 这一信息对于复杂样品的分析具有重要价值, 因为复杂样品中微量组分的化学信息不一定可见^[10]. 此外, 主因子数还可以用于研究测量数据中的非化学信息. 仪器测量过程中的某些影响因素(噪声、基线漂移和仪器脉动等)以及数据分析中的某些预处理手段(如平滑)^[11, 12], 均会向测量数据中引入特殊因子. 这些因子并未包含被分析样品的化学信息, 所以称为仪器因子, 或“鬼”因子(ghost factor)^[13]. 如果仪器因子的影响显著, 相应数据矩阵的主因子数会增大^[3], 仪器因子数越大, 仪器性能则越差, 测量数据的可靠性越低. 所以, 通过主因子数, 能够评价分析仪器的性能以及测量数据的可靠性.

除了上述直接应用外, 主因子数还是绝大多数化学因子分析方法的一个必要参数, 对化学因子分

析的结果具有决定性作用^[13]。

主因子数的估计一直是一个难题, 其充分反映在多种估计方法并存的现状。研究者对这些方法进行了分类, 不仅是对诸多方法的梳理, 更重要的是能够归纳出同类方法的共性; 因不同研究者的分类角度不同, 因此, 可以从多方面揭示各种方法的性质和特点。这些信息对于改进现有估计方法或研究新方法, 甚至深入理解化学因子分析具有重要的价值。

Meloun 从仪器误差是否已知的角度, 将主因子数估计方法分为精确方法和近似方法^[14]。精确方法是指在仪器误差已知的情况下, 将方法的指标函数和仪器误差进行比较, 以此估计主因子数。比较典型的精确方法包括剩余标准差(residual standard deviation, RSD)、均方根误差(root mean square error, RMS)和平均误差(average error criterion, AE)等^[13, 15]。若仪器误差不可知或不准确, 则需要采用近似方法, 包括碎石检验(scree test)^[16]、嵌入误差函数(imbedded error, IE)和因子指示函数(factor indicator function, IND)^[17]等。

梁逸曾等^[18]曾将主因子数估计方法分为两类, 分别是基于特征值和特征向量的方法。前者的典型代表有 Malinowski F 检验和 IND, 后者主要倾向于分析特征向量的频域部分^[19], 因为主因子所代表的光谱信号频率较低, 而噪声因子的频率较高。

Dable 和 Booksh^[4]将主因子数估计方法分成三类: 直观方法、经验方法和统计方法。直观方法是指通过观察不同因子水平的得分向量(主成分向量)和特征向量, 找出特征向量呈现随机特征时对应的因子水平; 经验方法是指运用数学函数和经验总结来估计主因子数; 统计方法则是以统计学原理为基础。经验方法和统计方法均会引入一些并非适用于所有数据的假设。例如, 经验方法的假设是随机噪声服从某种分布; 大部分统计方法的假设是随机噪声服从正态分布^[13]。事实上, 一些实际情况并不满足上述假设, 那么应用这些判据则无法对主因子数进行正确估计。

Malinowski^[3]曾将主因子数的估计方法分为三类: 经验方法、伪统计方法和统计方法。Malinowski 认为, 经验方法的实际效果很好; 伪统计方法基于统计方面的理论近似, 其中的准统计假设存在不合理性; 而统计方法是以成熟的统计学原理为基础, 并未使用不合理假设。

本文整理了近年来各种主因子数的估计方法,

并将之分为三类, 即经验方法、数学原理完备的方法以及统计方法, 并对每类方法的特点和共性进行了分析和说明, 归纳出第二类方法的基本数学原理。在阐述三类方法之前, 首先简要介绍了一些必要的理论要点以及基础概念。

2 估计主因子数的理论基础及概念

设有 m 个化学组成相同的被分析样品, 对每个样品在 n 个分析通道上进行测量。将测量数据按行排列, 则获得 m 行 n 列的二维数据矩阵, 以 $D_{m \times n}$ 表示。尽管仪器分析手段不同, 但是其测量数据具有共性, 可视为组分浓度不同的混合物的谱学数据, 因此, $D_{m \times n}$ 的行向量和列向量分别对应于混合组分的光谱和浓度分布。此外, 根据谱学中的比尔定律, $D_{m \times n}$ 具有双线性的特性^[1, 2]。若被分析样品的组分数为 p , 而且以下三个条件成立: (1) $D_{m \times n}$ 仅为组分信息, 不包含噪声等干扰; (2) 各组分的浓度分布向量线性无关, 光谱向量也线性无关; (3) m 和 n 均大于 p , 则 $D_{m \times n}$ 的秩等于 p 。这是估计主因子数的数学基础之一^[2, 20]。现代仪器分析中, 样品数 m 和分析通道数 n 一般均大于组分数 p , 所以 $D_{m \times n}$ 是亏秩矩阵。

矩阵 $D_{m \times n}$ 的行向量处于一个线性空间(以下称“线性空间 1”), 此空间由纯组分的光谱向量组张成; 列向量处于另一个线性空间(以下称“线性空间 2”), 此空间由纯组分的浓度分布向量组张成。若能够从矩阵 $D_{m \times n}$ 的行向量组构造出最大线性无关向量组, 则此向量组为“线性空间 1”的一个基底, 其中向量的数目等于矩阵 $D_{m \times n}$ 的秩。类似地, 若能够从矩阵 $D_{m \times n}$ 的列向量组构造出最大线性无关向量组, 则此向量组为“线性空间 2”的一个基底, 其中向量的数目等于矩阵 $D_{m \times n}$ 的秩。这一事实是估计主因子数的数学基础之二。

有些方法可以从矩阵 $D_{m \times n}$ 的行向量组(或者列向量组)构造出一个线性无关向量组, 其中最常用的是主成分分析(principal component analysis, PCA)。PCA 对数据矩阵 D 的处理示意图如图 1 所示, 结果是一系列特征向量 v_i 和一系列主成分向量 p_i ($p_i = Dv_i$), p_i 和 v_i 均为线性无关向量组。主成分向量组或特征向量组中非零向量的数目即为矩阵 $D_{m \times n}$ 的秩, 此结论与上述“数学基础之二”等价。PCA 的另一个性质是能够使主成分最大地解释原始数据的方差, 所以根据方差

$$D_{m \times n} \xrightarrow{\text{PCA}} \begin{cases} \mathbf{v}_i, \mathbf{p}_i, a_i (1 \leq i \leq q) & \leftarrow \text{代表主要因子} \\ \mathbf{v}_j, \mathbf{p}_j, a_j (q < j \leq \min(m, n)) & \leftarrow \text{代表次要因子} \end{cases}$$

图1 主成分分析处理数据矩阵 D 得到的一系列特征向量 \mathbf{v}_i 、主成分向量 \mathbf{p}_i 以及相应的特征值 a_i . $1 \leq i \leq q$, q 为矩阵 D 的秩(理想情况)或主因子数(实际情况)

的大小, PCA 的结果可以分为主要因子和次要因子(图 1). 理想情况下, 代表次要因子的主成分向量和特征向量均为零向量.

但是, 实际测量数据不可避免地包含噪声等干扰, 结果是矩阵 $D_{m \times n}$ 的秩等于 $\min(m, n)$, 即噪声等干扰使 $D_{m \times n}$ 由亏秩矩阵变为满秩矩阵. 这种情况下, 估计主因子数实际上是区分化学组分与噪声等干扰对矩阵秩的贡献. 这种区分可以借助于 PCA 方差特性. 对于有效的测量数据, 化学信息一直比噪声等干扰占优势, 表现为具有较大的方差. 所以, 经过 PCA, 化学信息因其方差较大而集中于主要因子, 噪声等干扰则集中在次要因子. 这一事实是估计主因子数的数学基础之三. 若有办法区分 PCA 结果中的主要因子和次要因子, 则可以估计出主因子数, 而不同的区分方法就形成了不同的主因子数判据.

从上面的介绍可以发现, 对于实际测量矩阵, 估计主因子数是如何合理而且清晰地区分主要因子和次要因子; 即使实际数据中包含干扰, 估计主因子数的基本数学原理依然有效, 只是不太严密. 这造成了经验估计方法以及具有完备数学原理的估计方法并存的局面.

3 三类主因子数估计方法

3.1 经验估计方法

经验估计方法是通过对大量数据分析实例的总结, 归纳出经验判据, 用以区别主要因子和次要因子, 从而估计出主因子数. 经验判据一般是关于特征值的某种组合运算, 尽管没有经过严密的数学推证, 但在实际应用中有效^[3]. 经验判据的实际有效性基于的事实为: 特征值在数值上等于主成分的方差, 而主要因子和次要因子在方差上存在差异. 如果特征值的某种组合运算能够突出这种差异, 则该组合即为一种有效判据.

最早的经验方法是 Cattell^[16]于 1966 年提出的剩余方差碎石检验方法. 该方法绘制剩余方差百分比-

因子水平曲线, 在曲线上找出转折点, 对应的因子水平即为主因子数. 曲线的转折点需要通过观察判断, 具有一定的主观性, 所以, 当混合物中组分含量较低或噪声水平较大时, 转折点较难判断, 导致结果可靠性降低.

Malinowski^[17]根据因子分析误差理论提出了一些判据, 包括嵌入误差(IE)和因子指示函数(IND)等. IE 和 IND 的数值随因子水平的增加而变化, 在最小值处给出主因子数. IE 和 IND 的计算比较简单, 结果判断不受主观因素的影响, 而且处理模拟数据可以给出相当准确的结果, 所以得到了广泛应用. 尤其是 IND, 尽管其属于早期判据, 但目前仍在使用, 在很多新判据的设计中, 仍然作为参考. 有实验证明, 当混合物中的微量组分相对浓度为 0.5%~1%时, 使用 IND 仍然可以检测出此微量组分^[4]. 但是, 如果实际数据中的随机误差不服从均匀分布, 或还存在过失误差和系统误差时, IE 效果很差. Malinowski 在提出 IND 判据时假设噪声具有等性方差(即同方差噪声)且服从均匀分布, 所以, IND 抵抗不等性方差噪声的能力较弱.

特征值比(eigenvalue ratio, ER)是何锡文等^[21]提出的一种经验判据, ER 为相邻两个因子水平下的特征值之比(前一个因子水平的特征值与后一个因子水平的特征值之比), 主因子数为 ER 出现极大值时的因子水平. ER 计算简单, 在数据处理实例中比较灵敏.

方差比例(variance percentage to variance sum, VPVRS)与 ER 的基本原理相近, 首先计算特征值的差分, 然后计算相邻因子水平下差分的比值, 即为 VPVRS^[22]. 主因子数为 VPVRS 出现极大值时的因子水平. 研究表明, 在对抗不等性方差噪声时, IND 能力最弱, ER 次之, VPVRS 最强^[23]. 研究者推测, 其可能原因是 ER 和 VPVRS 分别为特征值和特征值差分的比值, 一定程度上抵消了噪声对判据中分子和分母的影响, 因而表现出一定的抗噪声能力.

子矩阵分析(submatrix analysis)是 Cartwright^[24]提出的一种方法, 对原始矩阵的子集进行初步分析,

通过比较子矩阵的特征值来估计矩阵的主因子数. 当混合物中的主因子数受一些可调参数的控制时, 即当溶液 pH、压力、配体浓度、温度或溶剂极性发生变化时, 数据矩阵的主因子数会随之变化, 可以采用此方法. 此方法可用于估计溶液中水解产物、离解产物或络合产物的种类以及体系的平衡常数.

根据免疫算法的基本思想, 邵利民等^[25]提出了一种基于免疫识别 (immune recognition, IR) 的主因子数估计方法. 此方法模拟生物学上的抗体对抗原的特异性结合, 通过比较抗体对残余矩阵的识别作用 (以抗体对抗原的亲和力表示), 来估计主因子数. 研究结果表明, 该判据能够克服较严重的干扰.

尽管绝大多数经验判据是关于特征值的某种组合, 特征向量也可用于估计主因子数. Rossi 和 Warner^[19]发现, 特征向量经过傅里叶变换后, 具有特定的频率分布, 包含化学信息的特征向量显示出低频特征, 而包含噪声的特征向量则显示了高频特征, 通过比较特征向量的频谱特征, 可以分辨主要因子和次要因子, 进而估计出主因子数. 在估计三维荧光数据矩阵的主因子数的研究中^[5], 原理上也是对特征向量进行频域分析.

经验判据是对大量数据处理实例归纳和总结的结果, 在一定范围内无疑是有效的. 但是经验判据的数学原理尚不明确, 因此, 其有效性往往局限于某类数据. 例如, 相当一部分经验判据是在处理色谱数据 (HPLC-DAD) 时设计的, 因而对色谱数据比较有效; 对于其他类型数据, 尤其是当数据特征或数据量明显不同于一般色谱数据时, 经验判据结果的准确性会降低. 这种情况下, 需要考虑其他经验判据, 或使用数学原理较完备的判据.

3.2 数学原理完备的估计方法

分析化学中, 大多数矩阵形式测量数据具有双线性的结构特征, 而且各化学组分的光谱向量线性无关^[2]. 这些特性使得线性代数的某些理论可以用于估计主因子数. 基于这一策略的估计方法具有较完备的数学原理, 一般对测量数据没有显性或隐性的要求, 因此适用范围较广而且结果比较可靠. 此外, 此类判据的数学原理或实现方式也能够为新判据的设计提供参考.

在理论部分已经介绍过, PCA 可以将化学组分的光谱信息集中于主要因子的特征向量. 这一事实的

数学等价是: 化学组分的光谱向量组与主要因子的特征向量组张成同一线性空间, 即对于光谱向量组张成的线性空间, 主要因子的特征向量组是其一个基底. 这一数学原理为主因子数的判定提供了坚实的理论基础, 即如果某特征向量组被确定是这样一个基底, 则其特征向量的数目即为主因子数. 但是, 化学组分的光谱向量未知, 所以获得上述线性空间还需另外一个线性无关向量组. 如何从原始数据 D 获得 (不同于 PCA 特征向量组的) 一个向量组则成为关键问题, 不同的获取方法形成了不同的判据. 通过对目前此类判据的分析, 可以归纳出两种获取模式, 下面分别介绍.

图 2 显示了数学原理完备判据的两种实现模式: 第一种模式的特点是“同一数据, 两种方法”, 即通过 PCA 和另外一种方法处理原始数据矩阵 D , 分别获得特征向量组 v_i 和另一线性无关向量组 $v_i^\#$; 第二种模式的特点是“两种数据, 同一方法”, 即先对原始数据 D 进行线性变换得到 $D^\#$, $D^\#$ 与 D 的形式不同, 但包含的信息完全一致, 然后使用 PCA 分别处理 D 和 $D^\#$, 得到特征向量组 v_i 和另一线性无关向量组 $v_i^\#$.

按照特征值从大到小的顺序, 分别对特征向量组 v_i 和另一线性无关向量组 $v_i^\#$ 进行排列, 化学组分的光谱信息主要集中于前几个向量, 后面的向量则主要包含噪声. 然后, 逐渐增大 i , 并且比较相应的两个向量组是否张成同一线性空间, 如果是, 说明这两个线性无关向量组均为线性空间的基底, 其向量的数目即为主因子数. 需说明的是, 上述判定过程中的比较也可以使用特征值.

子空间比较 (subspace comparisons) 是由沈海林等^[26, 27]提出的一种方法, 属于图 2(a) 所示的模式, 其

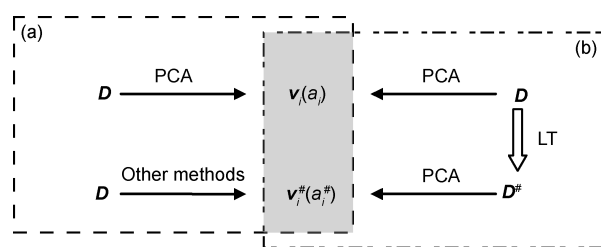


图 2 主因子数估计方法的基本数学原理. 其中(a)和(b)分别表示两种实现模式. 阴影部分是原理的核心; D 表示原始数据矩阵; v_i 表示 PCA 处理 D 后得到的一系列特征向量; a_i 是相应的特征值, 若通过其他方式获得, 加上标#以示区别; LT 表示线性变换

中“other method”表示易用迭代自模混合信号分析 (simple-to-use iterative self-modeling mixture analysis, SIMPLISMA)^[28]、正交投影方法(orthogonal projection approach, OPA)^[29]和简化卑根方法(the simplified Borgen method, SBM)中的一种. 由此获得两组 Schmidt 正交关键变量, 以关键变量为基底张成两个子空间, 然后使用子空间差异函数比较这两个子空间的差异. 在体系测量谱相似度较高的情况下, 子空间比较法可以准确地估计矩阵的主因子数. 但是, 判断结果很大程度上取决于 Schmidt 正交关键变量的获取方法以及测量数据矩阵的结构, 因此主因子数的估计容易偏低.

陈增萍等^[30]设计了 RESO 判据(the ratio between eigenvalues in smoothed PCA and their counterparts in ordinary PCA), 原理上属于图 2(a)所示的模式, 其中“other method”表示具有滤噪作用的平滑 PCA (smoothed PCA, SPCA)^[11, 31]. 主要因子的特征向量主要包含化学组分的光谱信息, 受 SPCA 滤噪作用的影响较小, 而次要因子的特征向量主要是噪声, 受 SPCA 滤噪作用的影响较大. 因此, SPCA 的主要因子与 PCA 的主要因子相似, 相应的特征值 a_i 与 $a_i^\#$ 比较接近, 但是次要因子不同, 相应的特征值 a_i 与 $a_i^\#$ 差别很大. 因此, $a_i^\#$ 与 a_i 之比可用于估计主因子数: 与化学信息有关的 RESO 值接近 1, 与噪声信息有关的 RESO 值远小于 1. RESO 能够在成分含量偏低, 光谱中共线性现象较严重, 或一定范围内异方差噪声存在等复杂情况下, 获得比较准确的结果. 有研究表明, 在混合物中微量组分相对浓度为 0.5%~1%时, 使用 RESO 判据仍能预测到此微量组分^[14]. RESO 要求组分具有平滑光谱, 否则(如条形质谱数据)结果的准确性降低.

应用 RESO 判断主因子数的标准是 RESO 是否接近 1, 但是接近程度的衡量存在主观性. 梁逸曾等^[18]曾对三组分模拟系统进行 SPCA, 因子水平为 3 时, RESO 值介于 0.8~0.9 之间; 当 RESO 值大于 0.8 时, 由于结果可能存在偶然性, 对应的特征值并不一定属于化学信号.

为了解决上述偶然性问题, 梁逸曾等^[18]又提出了泛函主成分分析中的噪声扰动判据(noise perturbation in functional principal component analysis, NPFPCA), 该判据在原理上与 RESO 相似, 区别在于判别特征向

量发生的变化. NPFPCA 通过蒙特卡罗模拟^[32], 研究添加不同的随机噪声时特征向量所发生的变化, 以使结果更加准确. Hu 等^[33]曾在二维相关能谱的分析中使用 NPFPCA 方法, 克服了噪声干扰问题, 取得了较好效果.

谭涌溪等^[34]提出了基于化学子空间对线性变换稳定的秩估计方法, 原理上属于图 2(b)所示的模式, 采用随机方式对原始数据矩阵 D 进行线性变换, 然后对原数据矩阵和变换后的数据矩阵分别进行 PCA. 使用投影残差法和子空间夹角法两个指标来衡量两个向量组张成的子空间的差异. 这一方法解决了子空间比较法中, 关键变量获得方法对判断结果的影响问题.

Henry 等^[35]提出了一种因子数(number of factors)的估计方法, 并命名为 NUMFACT 方法, 原理上属于图 2(b)所示的模式, 其中的线性变换是重采样技术(resampling technique), 然后对原数据矩阵和重采样后的数据矩阵分别进行 PCA. NUMFACT 计算特征向量 $v_i^\#$ 在特征向量 v_i 上的投影, 投影值最大时对应的因子水平即为主因子数. NUMFACT 方法应用广泛, 已经成功处理了多种类型的数据, 包括气相色谱、红外光谱, 以及列车时刻、英文词频.

在某些情况下, 原始数据矩阵之外还存在副本数据矩阵(replicate matrix), 此副本数据矩阵与原始数据矩阵的形式不同, 但包含的信息完全一致. 这种情况下, 可以使用典型相关方法(canonical correlation)来估计矩阵的主因子数^[36]. 这种方法的实质是图 2(b)所示的模式, 副本数据矩阵与原始数据矩阵的关系即为图中的 $D^\#$ 与 D , 只是不再需要通过纯数学手段(如线性变换)获得 $D^\#$. 对原始数据矩阵以及副本数据矩阵分别进行 PCA, 然后通过空间夹角来判断两个特征向量组张成的子空间的相关性: 包含化学信号的特征向量间相关性比较大, 而噪声信号的特征向量间相关性很小.

Shrager 和 Hendler^[37]将自相关用于主因子数的估计, 通过使用自相关系数来区别分别包含化学信号和噪声的特征向量. 这种方法并没有按照图 2 所示的模式进行, 只是对原始数据矩阵进行 PCA, 得到一系列特征向量. 然后对特征向量进行自相关分析, 主要包含化学信息的特征向量之间的自相关系数较大, 主要包含噪声信息的特征向量之间的自相关系数较小.

2011年 Malinowski^[38]提出了一种通过增广矩阵的秩估计方法(determination of rank by augmentation, DRAUG). DRAUG 的基本思路是构造一个与原始数据矩阵同样大小、秩为 1 的矩阵, 并与原始数据矩阵相加, 得到增广矩阵. 外加信息干扰了原始数据矩阵中的主因子信息, 而与随机误差有关的次要因子信息未受影响. 对原始矩阵和增广矩阵分别进行 PCA, 然后在不同因子水平下比较各自残余矩阵的方差, 如果不存在显著性差异, 说明残余矩阵中几乎全部为次要因子信息, 此时的因子水平即为主因子数. 可以发现, DRAUG 的核心思想与图 2(b)所示模式在原理上一致, 不同的只是次要因子信息的比较(其他方法是主要因子的比较, 文献[34]的方法为典型代表). 需要说明的是, DRAUG 使用显著性检验来比较残余矩阵的方差, 所以, Malinowski 将其归于统计类方法.

上述主因子数估计方法的数学原理完备, 因此结果比较可靠. 这些数学原理完备的估计方法在处理模拟数据时, 具有良好的效果, 但是在处理某些实际数据时的效果不佳, 所以不能完全取代经验方法.

3.3 基于统计学的估计方法

基于统计学的主因子数估计方法多数以 F 检验为基础, 而且要求测量数据中的随机误差服从某种分布.

Malinowski^[39]将约化特征值(reduced eigenvalue, REV)作为估计主因子数的判据. 由于 REV 与标准偏差成比例关系, Malinowski 提出了通过 F 检验估计主因子数, 即 Malinowski F 检验^[40].

Faber 和 Kowalski^[41]发现, Malinowski F 检验存在一定的问题, 通过模拟数据的解析发现, 次要因子的 REV 值并不为常数. 基于 Mandel 的创始工作, 他们提出了 Faber-Kowalski F 检验的估计方法, 重新定义了 Malinowski F 检验中的自由度和表达式. 结果表明, Faber-Kowalski F 检验比 Malinowski F 检验更加灵敏, 在显著性水平为 1% 时即可以分辨主因子和噪声因子, 而且计算时间短. 之后, Malinowski 发现, Faber-Kowalski F 检验在处理实验数据时, 主因子数容易偏高, 建议使用 2 倍 Mandel 自由度, 此判据即为修正 Faber-Kowalski F 检验^[12, 42].

Vogt 和 Mizaikoff^[43]从多元回归的角度提出了一种基于 F 检验的动态维数确定方法(dynamic determination of dimension, DDD), 又称 Vogt-Mizaikoff

F 检验. 在该方法中, 首先以原始数据矩阵的某一列作为目标向量, 将目标向量移除后得到缩减矩阵. 然后, 使用主成分回归处理缩减矩阵, 在不同因子水平下对目标向量进行预测. 最后计算出目标向量与预测向量的残余方差, 通过 F 检验来判断此方差与噪声方差有无显著差异. 数据分析结果表明, Vogt-Mizaikoff F 检验能够检测出数据中光谱强度较小但是属于光谱重要特征的信号. Vogt-Mizaikoff F 检验要求数据矩阵的行数大于列数.

Malinowski^[44]对 Vogt-Mizaikoff F 检验进行了修正, 得到了修正 Vogt-Mizaikoff F 检验(adaptation of Vogt-Mizaikoff F-test, AVM 或 VMmad). 此方法实际上是进行多次 Vogt-Mizaikoff F 检验, 每次选取数据矩阵的一列作为检验向量, 当所有列向量均使用过后, 即得到一系列主因子数估计值. 对于这些主因子数估计值, 通过绝对偏差中位数 (median absolute deviation, MAD) 检测并修正其中的离群值, 最后以平均值作为主因子数.

Malinowski^[3]将 MAD 与 PCA 结合, 提出了通过绝对偏差中位数的秩估计(determination of rank by median absolute deviation, DRMAD)方法. 在 DRMAD 方法中, 通过 PCA 计算出处于不同因子水平的残余标准偏差(residual standard deviation, RSD), 应用 MAD 检测出 RSD 中的离群值, 最后以离群值的数目作为主因子数. DRMAD 对随机误差的分布不作要求, 算法简单、速度较快.

这类主因子数估计方法以统计分析为基础, 原理完备而且严密. 但是, 统计判据对随机误差的分布比较敏感, 当实际数据中随机噪声的分布异于统计判据的要求时, 会出现结果准确性降低, 甚至判据失效的情况. 所以, 处理模拟数据时性能较好, 而处理实际数据时性能平平甚至急剧下降的现象, 在基于统计学的主因子数估计方法中尤为突出.

4 结语

在复杂化学体系测量数据的解析中, 主因子数的估计是必要步骤, 同时也是一个难题. 目前存在多种各具特点的估计方法, 但是没有一种通用方法适用于所有类型的数据. 在本文的分类体系中, 三类方法各有特点: 经验方法具有实际有效性; 数学原理完备的方法和统计方法具有比较明确的理论作为基础,

因而结果的可靠性较高, 且其理论基础对于新判据的设计以及化学因子分析均具有重要的参考作用. 但是后两者均需满足一定的前提条件, 尤其是统计方法对随机噪声的分布非常敏感, 所以并不比经验

判据具有明显的优势. 尽管目前还没有一种通用方法, 但多种方法的并存一方面可以互相验证, 另一方面也为主因子数估计问题提供了多个研究角度.

致谢 本工作得到国家自然科学基金(21175123)和教育部新世纪优秀人才支持计划(NCET-11-0878)资助, 特此一并致谢.

参考文献

- 1 梁逸曾, 许青松. 复杂体系仪器分析: 白、灰、黑分析体系及其多变量解析方法. 北京: 化学工业出版社, 2012
- 2 邵利民. 分析化学数据解析中的微机应用. 合肥: 中国科学技术大学出版社, 2011
- 3 Malinowski ER. Determination of rank by median absolute deviation (DRMAD): a simple method for determining the number of principal factors responsible for a data matrix. *J Chemom*, 2009, 23: 1–6
- 4 Dable BK, Booksh KS. Selecting significant factors by the noise addition method in principal component analysis. *J Chemom*, 2001, 15: 591–613
- 5 Yu SH, Zhang YJ, Wang HB, Xiao X. An efficient method for determining the chemical rank of three-way fluorescence data arrays. *Chemom Intell Lab Syst*, 2010, 103: 83–89
- 6 Kritchman S, Nadler B. Determining the number of components in a factor model from limited noisy data. *Chemom Intell Lab Syst*, 2008, 94: 19–32
- 7 Wold S. Cross-validatory estimation of the number of components in factor and principle components models. *Technometrics*, 1978, 20: 397–405
- 8 Xu QS, Liang YZ. Monte Carlo cross validation. *Chemom Intell Lab Syst*, 2001, 56: 1–11
- 9 Liu ZC, Ma X, Wen YD, Wang Y, Cai WS, Shao XG. A practical approach for near infrared spectral quantitative analysis of complex samples using partial least squares modeling. *Sci China Ser B*, 2009, 52: 1021–1027
- 10 Shao LM, Griffiths PR. Obtaining qualitative information on trace species in continuous open-path fourier transform spectroscopic measurements using target factor analysis and related techniques. *Anal Chem*, 2007, 79: 2118–2124
- 11 Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. *J R Statist Soc B*, 1991, 53: 233–243
- 12 Malinowski ER. Abstract factor analysis of data with multiple sources of error and a modified Faber-Kowalski F-test. *J Chemom*, 1999, 13: 69–81
- 13 Malinowski ER. *Factor Analysis in Chemistry*. 3rd ed. New York: John Wiley and Sons, 2002
- 14 Meloun M, Capek J, Miksik P, Brereton RG. Critical comparison of methods predicting the number of components in spectroscopic data. *Anal Chim Acta*, 2000, 423: 51–68
- 15 Kaiser HF. The application of electronic computers to factor analysis. *Educ Psych Meas*, 1960, 20: 141–151
- 16 Cattell RB. The scree test for the number of factors. *Multiv Behav Res*, 1966, 1: 140–146
- 17 Malinowski ER. Theory of error in factor analysis. *Anal Chem*, 1977, 49: 612–617
- 18 Xu CJ, Liang YZ, Li Y, Du YP. Chemical rank estimation by noise perturbation in functional principal component analysis. *Analyst*, 2003, 128: 75–81
- 19 Rossi TM, Warner IM. Rank estimation of emission excitation matrixes using frequency analysis of eigenvectors. *Anal Chem*, 1986, 58: 810–815
- 20 Xu L, Tang LJ, Cai CB, Wu HL, Shen GL, Yu RQ, Jiang JH. Chemometric methods for evaluation of chromatographic separation quality from two-way data: a review. *Anal Chim Acta*, 2008, 613: 121–134
- 21 何锡文, 李洪, 史慧明. 用因子分析-光度法测定混合物的组份. *分析化学*, 1986, 14: 34–40
- 22 Alex S, Savoie R. Evaluation of nonzero eigenvalues in factor-analysis. *Spectroscopy*, 1989, 34: 27–33
- 23 Chen ZP, Liu Z, Cao YZ, Yu RQ. Efficient way to estimate the optimum number of factors for trilinear decomposition. *Anal Chim Acta*, 2001, 444: 295–307
- 24 Cartwright HM. Determination of the dimensionality of spectroscopic data by submatrix analysis. *J Chemom*, 1987, 1: 111–120

- 25 邵利民, 邵学广, 林祥钦. 一种基于免疫识别的主因子数判断方法. 高等学校化学学报, 2003, 24: 595–598
- 26 沈海林, 李晓宁, 梁逸曾. 化学体系组分数确定的新方法: 子空间比较法. 科学通报, 2000, 45: 587–592
- 27 Shen HL, Liang YZ, Kvalheim OM, Manne R. Determination of chemical rank of two-way data from mixtures using subspace comparisons. *Chemom Intel Lab Syst*, 2000, 51: 49–59
- 28 Windig W, Guilment J. Interactive self-modeling mixture analysis. *Anal Chem*, 1991, 63: 1425–1432
- 29 Sanchez FC, Toft J, Bogaert BVD, Massart DL. Orthogonal projection approach applied to peak purity assessment. *Anal Chem*, 1996, 68: 79–85
- 30 Chen ZP, Liang YZ, Jiang JH, Li Y, Qian JY, Yu RQ. Determination of the number of components in mixtures using a new approach incorporating chemical information. *J Chemom*, 1999, 13: 15–30
- 31 Silverman BW. Smoothed functional principal components analysis by choice of norm. *Ann Stat*, 1996, 24: 1–24
- 32 Binder K, Heermann DW. *Monte Carlo Simulation in Statistical Physics: An Introduction*. Berlin: Springer, 2010
- 33 Hu Y, Li BY, Sato H, Noda I, Ozaki Y. Noise perturbation in functional principal component analysis filtering for two-dimensional correlation spectroscopy: its theory and application to infrared spectra of a poly(3-hydroxybutyrate) thin film. *J Phys Chem A*, 2006, 110: 11279–11290
- 34 谭涌溪, 蒋健晖, 谢洪平, 梁逸曾, 俞汝勤. 基于化学子空间对线性变换稳定的秩估计方法. 高等学校化学学报, 2001, 22: 1319–1322
- 35 Henry RC, Park ES, Spiegelman CH. Comparing a new algorithm with the classic methods for estimating the number of factors. *Chemom Intell Lab Syst*, 1999, 48: 91–97
- 36 Tu XM, Burdick DS, Millican DW, McGown LB. Canonical correlation technique for rank estimation of excitation-emission matrixes. *Anal Chem*, 1989, 61: 2219–2224
- 37 Shrager RI, Hendler RW. Titration of individual components in a mixture with resolution of difference spectra, pKs, and redox transitions. *Anal Chem*, 1982, 54: 1147–1152
- 38 Malinowski ER. Determination of rank by augmentation (DRAUG). *J Chemom*, 2011, 25: 323–358
- 39 Malinowski ER. Theory of the distribution of error eigenvalues resulting from principal component analysis with applications to spectroscopic data. *J Chemom*, 1987, 1: 33–40
- 40 Malinowski ER. Statistical F-tests for abstract factor analysis and target testing. *J Chemom*, 1988, 3: 49–60
- 41 Faber K, Kowalski BR. Modification of Malinowski's F-test for abstract factor analysis applied to the quail roost II data sets. *J Chemom*, 1997, 11: 53–72
- 42 Faber K, Kowalski BR. Critical evaluation of two F-tests for selecting the number of factors in abstract factor analysis. *Anal Chim Acta*, 1997, 337: 57–71
- 43 Vogt F, Mizaikoff B. Dynamic determination of the dimension of PCA calibration models using F-statistics. *J Chemom*, 2003, 17: 346–357
- 44 Malinowski ER. Adaptation of the Vogt-Mizaikoff F-test to determine the number of principal factors responsible for a data matrix and comparison with other popular methods. *J Chemom*, 2004, 18: 387–392

Methods to estimate the number of principal factors of data matrices of complex chemical systems: a review and a categorization

LU Wei, SHAO LiMin*

Department of Chemistry, University of Science and Technology of China, Hefei 230026, China

*Corresponding author (email: lshao@ustc.edu.cn)

Abstract: Determining the number of principal factors is a prerequisite step in the analyses of matrix-type experimental data, and it is also a difficult problem, which is reflected by the existence of many methods. A review and a categorization of those methods would provide a panoramic view, and make it possible to generalize important and practically useful information for this problem. In this paper, combined with our relevant work, we presented a systematic and in-depth analysis on determining the number of principal factors, and categorized current methods into three categories which are tagged empirical, mathematically rigorous, and statistical. From the methods of category two, a general mathematical model was proposed. The results of this paper are helpful in various ways, such as improving current methods, designing new methods, and even helping to understand theories of chemical factor analysis.

Keywords: data matrix, dimensionality reduction, number of principal factors, principal component analysis

中国科学化学

SCIENTIA SINICA Chimica

第 44 卷 第 10 期 2014 年 10 月出版

变更声明

本刊自 2010 年起, 中文版刊名由《中国科学 B 辑: 化学》变更为《中国科学: 化学》(对应的外文名称为: *SCIENTIA SINICA Chimica*); 英文版刊名由 *Science in China Series B: Chemistry* 变更为 *SCIENCE CHINA Chemistry*.

中文版 ISSN 号由 ISSN 1006-9240 变更为 ISSN 1674-7224; CN 号由 CN 11-1788/N 变更为 CN 11-5838/O6.

英文版 ISSN 号由 ISSN 1006-9291 变更为 ISSN 1674-7291; CN 号由 CN 11-1789/N 变更为 CN 11-5839/O6.

特此声明.

地 址: 北京东黄城根北街 16 号
电 话: (010) 64016732 (编辑部)
(010) 64019709 (发行部)
(010) 64008316 (广告部)
传 真: (010) 64016350

邮政编码: 100717
电子信箱: chemistry@scichina.org (编辑部)
sales@scichina.org (发行部)
ads@scichina.org (广告部)

版权所有, 未经许可, 不得转载

主 管	中 国 科 学 院	出 版	《中国科学》杂志社
编 辑	中 国 科 学 院 《中国科学》编辑委员会 北京(100717)东黄城根北街 16 号	印刷装订	北京中科印刷有限公司
		总发行处	北京报刊发行局
主 编	万 立 骏	订 购 处	全国各邮电局 《中国科学》杂志社发行部

刊号: $\frac{\text{ISSN } 1674-7224}{\text{CN } 11-5838/O6}$ 代号: $\frac{\text{国 外 BM40B}}{\text{国内邮发 } 80-202}$

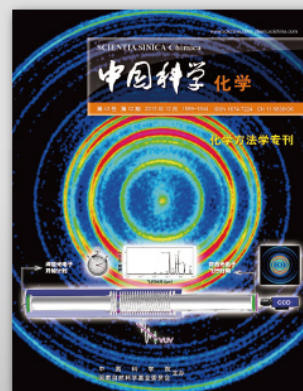
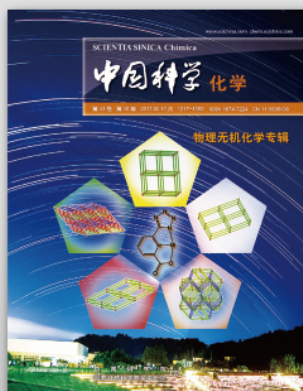
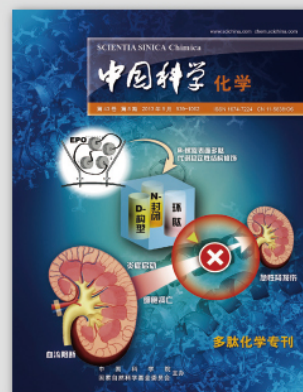
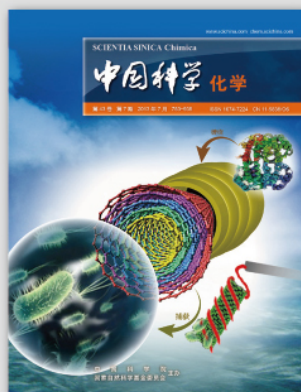
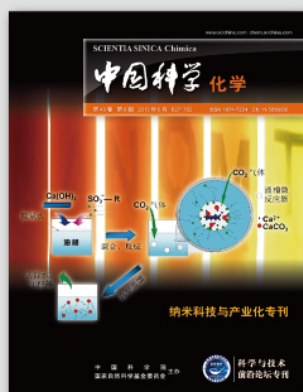
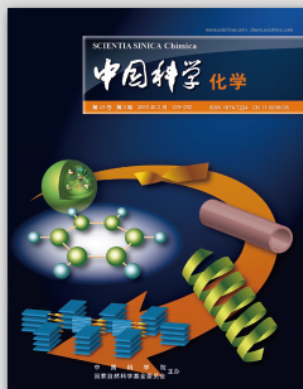
每期定价: 138.00 元 全年定价: 1656.00 元

广告经营许可证: 京东工商广字第 0429 号



COVERS IN 2013

主编：万立骏（中国科学院化学研究所）



主办单位：中国科学院 国家自然科学基金委员会

SCIENCE CHINA PRESS

ISSN 1674-7224



10>

9 771674 722147