

化学计量学

史永刚 冯新泸 李子存 编著

中国石化出版社

The cover features a vibrant green and yellow background. On the left, there is a cluster of laboratory glassware, including test tubes and beakers, some containing colored liquids (yellow, blue, red). A pipette is shown dispensing liquid into one of the tubes. On the right, a DNA double helix structure is depicted, with a green leaf and a small globe integrated into its structure. At the bottom, there is a horizontal scale bar with tick marks.

化学计量学

史永刚 冯新泸 李子存 编著

中国石化出版社

内 容 提 要

化学计量学是化学的一个新的分支学科,它应用数学、统计学、计算机科学、化学等学科的理论和方法,研究化学量测理论和方法,设计和选择最优的化学量测方法,并通过对化学数据的解析,最大限度地获取有关物质系统的化学信息。

本书将从化学量测的基本问题出发,阐述化学计量学的基本理论和方法,介绍常用的化学计量学方法,内容涉及化学量测的统计描述、分析采样理论、化学实验的设计与优化、分析信号的检测与处理、化学信号的分辨、化学校正技术、化学模式识别、人工神经网络在化学中的应用等。

本书可作为化学及相关学科的研究生和高级大学生的教材和参考书,也可作为化学、化工、工程等领域的科技工作者和高等学校教师的参考书。

图书在版编目(CIP)数据

化学计量学/史永刚,冯新沪,李子存编著.
—北京:中国石化出版社,2002
ISBN 7-80164-299-6

I.化… II.①史… ②冯… ③李… III.化学计量学 IV.06-04

中国版本图书馆 CIP 数据核字(2002)第 086151 号

中国石化出版社出版发行

地址:北京市东城区安定门外大街 58 号

邮编:100011 电话:(010)84271850

<http://www.sinopec-press.com>

E-mail:press@sinopec.com.cn

北京精美实华图文制作中心排版

河北省徐水县印刷厂印刷

新华书店北京发行所经销

*

850×1168 毫米 大 32 开本 9.375 印张 242 千字 印 1—2000

2003 年 1 月第 1 版 2003 年 1 月第 1 次印刷

定价:22.00 元

前 言

过去的几十年中，分析化学、分析仪器和计算机技术得到了迅速的发展，智能仪器和实验室自动化加速了分析化学向信息科学的转化。目前，分析化学及化学工作者能在很短的时间里获取大量的复杂化学量测数据，这就要求分析工作者必须具备化学、统计学、数学和计算机科学方面的技能，进行数据挖掘(Data Mining)，以便从复杂的数据中提取尽可能多的有效信息，化学计量学(Chemometrics)便应运而生。

传统的分析化学教育关注的是仪器分析和化学分析的基本原理及其应用。但在整个分析的全过程中，量测仅是其中的一个步骤。要做好一个分析，还必须十分重视与量测相关的其他分析过程，如采样、数据处理及分析结果中化学信息的提取等。不正确或不适当的采样，对分析结果的不合理解释必将导致错误的结论。化学计量学正是研究化学量测全过程中每一步所涉及基础理论及方法的一门化学分支学科，包括分析采样理论、试验设计与优化、分析信号的检测与处理、化学信号的分辨、化学校正技术、化学模式识别等。

化学计量学的研究内容是相当广泛的，本书将系统地介绍化学计量学的基础理论和方法，确保知识的逻辑性和完整性，除重点介绍比较成熟和应用广泛的理论和方法外，也适当地介绍发展快、相对较为成熟的新理论和新方法，如人工神经网络等。

本书共分九章，第一章论述化学计量学的概念、历史及发展；第二章描述化学量测数据的常用统计学描述方法；第三章重点论述分析采样理论；第四章阐述实验的设计与优化；第五章介绍分析信号的检测及处理；第六章讨论化学模式识别；第七章重点介绍化学校正技术；第八章探讨化学数据库；第九章介绍知识处理与软模型。

作为供化学工作者学习和参考的化学计量学著作，没有必要对各种算法的数学原理做过于详细的描述。计算机技术的发展，已经逐步使科学工作者从过去全部算法程序必须由自己编写过渡到以使用大量出现的各种功能强大的应用软件包和网上计算为主。因此，本书将重点介绍各种化学计量学最基本算法的原理、应用特点和方法，使具有一般化学和数学基础的人员能比较轻松地阅读这本书，学会如何进行化学计量学方法的选择和如何理解化学计量学的分析结果。

由于计算机技术和信息技术的飞跃发展，化学计量学也产生了巨大的变化，已经成为化学工作者必不可少的工具。希望本书能够帮助化学工作者理解化学计量学，并对化学计量学的普及与推广起到一定的作用。

本书的一、四、五、六、九章由史永刚编著，七、八章由冯新沪编著，二、三章由李子存编著。由于作者水平有限，书中的缺点和错误在所难免，恳请广大读者批评指正。

在本书出版之际，感谢后勤工程学院及油料应用工程系的关心和支持；感谢后勤工程学院油品测试评定中心在资金、时间上的大力支持和鼓励。

编著者



史永刚 男，陕西省人，1964年生。毕业于中国人民解放军后勤工程学院，1998年获应用化学专业工学博士学位。现为国家科学核心期刊《光谱实验室》编辑委员会委员，中国人民解放军后勤工程学院测试计量技术及仪器专业硕士研究生指导教师，主要从事测试新技术新方法、化学信息学、化学计量学及其在石油产品分析中应用研究，获国家科技进步三等奖（1998）、军队科技进步二等奖（1997）和四等奖（1993）、第十届全国发明展览会金奖各一项，发表学术论文50余篇，专利3项。



冯新沪 男，河北省人，1953年生，教授。1985年毕业于成都科技大学，获高分子专业硕士学位。现任重庆市光谱学会副理事长，中国人民解放军后勤工程学院测试计量技术及仪器专业硕士研究生指导教师，主要从事石油产品测试新技术、快速分析技术、实验数据可视化及专家系统的研究，获国家科技进步三等奖、军队科技进步二等奖和第十届全国发明展览会金奖各一项，发表学术论文数十篇，专利3项。

目 录

第一章 绪论	(1)
1.1 化学计量学的定义	(1)
1.2 化学计量学的研究范畴	(2)
1.3 化学计量的历史沿革	(5)
1.4 化学计量学实验中常用工具	(9)
第二章 实验数据统计分析基础	(11)
2.1 何谓数据	(11)
2.2 实验数据的统计描述	(12)
2.2.1 随机分布	(13)
2.2.2 置信区间	(22)
2.2.3 误差的传递	(23)
2.3 统计检验	(26)
2.3.1 标准正态分布	(26)
2.3.2 假设检验	(29)
2.3.3 异常数据的剔除	(39)
2.4 方差分析(ANOVA)	(44)
2.4.1 单因素方差分析	(44)
2.4.2 双因素方差分析	(46)
第三章 分析采样理论	(50)
3.1 正确采样的标准	(50)

3.2	总体的类型	(51)
3.3	固体散状物料的采样	(51)
3.4	分层采样	(56)
3.5	抽样检验	(62)
3.5.1	计量抽样检验	(63)
3.5.2	计数抽样检验法	(64)
3.6	动态过程中的采样	(66)
第四章	化学实验设计与优化	(68)
4.1	化学实验设计基础	(68)
4.1.1	试验指标	(69)
4.1.2	因素和水平	(69)
4.1.3	同时试验和序贯试验	(70)
4.1.4	试验最优化和解析最优化	(70)
4.1.5	有效实验存在的条件	(71)
4.1.6	实验设计的基本原理	(71)
4.1.7	实验设计的步骤	(72)
4.2	析因实验设计	(74)
4.2.1	析因设计表	(74)
4.2.2	析因设计的步骤	(75)
4.2.3	一般的二水平析因设计[$FD_n(2^m)$]	(78)
4.3	正交实验设计	(82)
4.3.1	正交表及其特点	(82)
4.3.2	正交实验设计	(82)
4.4	均匀实验设计	(88)
4.4.1	均匀设计表	(88)
4.4.2	均匀实验设计	(89)
4.5	随机区组	(92)
4.6	响应曲面法	(94)

4.6.1 形状已知的响应界面法	(95)
4.6.2 形状未知的响应界面法	(96)
4.6.3 单纯形法	(99)
第五章 化学量测信号的检测与处理	(104)
5.1 分析信号检测	(105)
5.2 检测限的点估计	(111)
5.3 检测限的点估计—— t 检验法	(113)
5.3.1 信噪比	(114)
5.3.2 以量测数据($x_A - x_B$)为基础的 t 检验法	(116)
5.4 检测限的精度	(117)
5.5 分析信号的处理	(118)
5.5.1 数字平滑与滤波	(118)
5.5.2 曲线拟合	(127)
5.5.3 谱峰面积的估计	(130)
5.6 Fourier 变换和 Hadamard 变换	(133)
5.6.1 Fourier 变换	(133)
5.6.2 Hadamard 变换	(134)
5.6.3 FT 和 HT 的应用	(135)
第六章 化学模式识别	(139)
6.1 数据预处理	(141)
6.1.1 丢失数据的弥补	(141)
6.1.2 中心化变换	(141)
6.1.3 归一化处理	(142)
6.1.4 正规化处理	(143)
6.1.5 标准化处理	(143)
6.2 方差—协方差矩阵和相关矩阵	(144)
6.2.1 方差—协方差矩阵	(145)
6.2.2 相关矩阵	(145)

6.3 特征选取	(146)
6.3.1 化学模式识别中常用的特征	(146)
6.3.2 特征选择的方法	(147)
6.4 非监督模式识别	(149)
6.4.1 因子分析	(149)
6.4.2 聚类分析	(172)
6.4.3 显示技术	(181)
6.5 有监督模式识别	(185)
6.5.1 线性学习机	(185)
6.5.2 K 最邻近法	(188)
6.5.3 SIMCA 方法	(189)
第七章 化学校正理论	(195)
7.1 单变量校正	(195)
7.1.1 校正曲线法	(195)
7.1.2 标准加入法	(210)
7.2 多元线性回归	(211)
7.2.1 最小二乘回归	(211)
7.2.2 有偏参数估计——PCR 和 PLS	(214)
7.2.3 多元校正模型在多组分分析中的应用	(220)
7.2.4 回归诊断	(225)
7.3 通用标准加入法	(228)
7.4 非线性校正方法	(230)
7.4.1 非线性回归	(231)
7.4.2 非参数方法	(234)
第八章 化学数据库	(236)
8.1 分析信息的表示	(237)
8.1.1 信息的类型	(237)
8.1.2 数据库的结构	(237)

8.2 数据库检索	(242)
8.2.1 顺序检索	(242)
8.2.2 光谱的相似性量度	(244)
8.3 光谱的模拟	(245)
第九章 知识处理与软模型	(248)
9.1 人工智能与专家系统	(248)
9.1.1 人工智能	(248)
9.1.2 符号知识处理	(249)
9.1.3 专家系统	(256)
9.2 人工神经网络	(259)
9.2.1 人工神经网络基础	(259)
9.2.2 一种简单的神经网络——BAM	(262)
9.2.3 人工神经网络学习范例	(266)
9.2.4 学习规则	(266)
9.2.5 网络构架	(268)
9.2.6 人工神经网络模型	(268)
9.3 模糊理论	(274)
9.3.1 基础知识	(274)
9.3.2 模糊集合的运算	(278)
9.3.3 应用	(279)
9.4 遗传算法	(280)
9.4.1 基本原理和算法	(280)
9.4.2 遗传算法的应用	(284)
9.5 非回归拓扑技术	(285)

第一章 绪 论

1.1 化学计量学的定义

如 Oldroyd 所述，知识建立的过程存在两类科学方法论，即首先通过对可观察对象的分析得出基本原理(“First Principles”)，然后用基本原理指导现象的解释，也就是所说的推理(“Inference”)。正是这种方法学形式使 Oldroyd 有了知识架构的概念。

化学计量学遵从这种总的模式，这在化学计量学的定义中表现得尤为明显。Massart 等对化学计量学作出如下定义：化学计量学是化学的分支学科，它应用数学、统计学和形式逻辑(a)设计或选择优化实验过程；(b)通过分析化学数据提供最大限度的相关化学信息；(c)获得关于化学系统的知识。

图 1-1 给出了一个典型的知识构架图。对一个化学问题，化学工作者首先做一个假设。为考察假设和验证它的有效性，化学工作者首先要做的就是确定该做哪些实验[化学计量学定义中的(a)]。化学计量学借助数学和统计学方法，

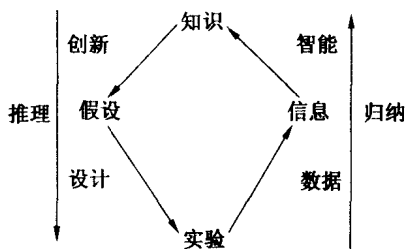


图 1-1 知识的架构

即利用实验设计理论，可很好地完成此项工作。实验产生数据，化学计量学工作者利用这些数据提取尽可能多的信息[化学计量学定义中的(b)]，如应用回归分析技术建立测量值与实验变量间的关系模型。然后，化学工作者应用这些信息和化学知识给出关

于研究系统的更多的知识[化学计量学定义中的(c)]。例如,如果研究的是化学反应,化学工作者也许就会得出该反应是二级反应的化学反应动力学的结论。随着关于系统的知识的增加,化学工作者就能提出更加完善的实验设计方案,获取更多的数据、更多的信息及更深入的对系统的理解,最终建立一个完善的知识构架。

化学计量学更多涉及的是实验的设计和信息的提取,关于知识的归纳其他化学学科的提得更多。在化学计量学研究的过程中,数据的质量占有至关重要的地位。应用化学计量学理论和方法,人们能够更好地获取有效的数据,进而指导产品的生产和过程的控制。化学计量学的最终目标就是改善、优化和监控产品的质量和生产过程。

1.2 化学计量学的研究范畴

实验是化学研究的基础,化学工作者十分注意实验过程的研究,而较少涉及方法选择、数据处理及实验结果的解释,因此往往不能以最优的研究、尽可能多地提取有用的化学信息。传统化学研究中,人们能够使用的实验测试技术有限,对实验数据处理的要求不高,而且方法提供数据的能力不强,加之数据包含的信息量较低,这就必然导致上述的以实验过程研究为重点的化学研究方法。

电子技术、计算机技术及现代仪器制造技术的发展,使现代分析仪器能快速提供大量的多维数据,因而对实验数据的处理、分类、解析和预测提出了更高的要求,计算机的广泛应用使这样的任务的完成成为可能。化学计量学正是以化学试验设计与优化、数据处理、信息提取及结果解释为主要研究内容的一门新兴化学分支学科。

对于一个样品的分析可以分为三个部分:实验方法的设计及

优化、样品的分析测定和分析数据的处理。

如果要对一个润滑油样品进行评价，那么首先必须选择和建立方法，即如何测定润滑油的粘性、抗氧化安定性、摩擦学性质等。怎么进行这样的测定，这又包括实验条件的选择。方法建立后还要对方法进行优化和评价，这不仅要依靠化学工作者的实践经验，而且也要充分使用化学计量学的理论与方法。

对样品分析的第二步就是测试样本的制备和测试。这里首先涉及的是分析样本的采取。分析采样理论是试样采取的数学统计理论，是确保必要测量精度的理论基础。应用采样理论，就能确定分析所需的最低限度的样本量。因此，采样理论的正确使用关系到质量的控制水平，也是发达国家确保其产品质量的诀窍之一。当取得分析样本后，接下来就是进行样品的制备和测定。目前人们可以选用的分析手段多种多样，当确定测定方法和手段后，就必须依据方法和手段，合理地制样，然后用现代分析仪器进行分析，得到关于样品的复杂的信息。

样品分析的第三步为数据分析，这也是化学计量学应用最为广泛的领域。分析检测理论提供了化学量测数据定性检测的数学统计学基础，使人们能依据化学量测信号，判断试样中是否存在某种微量组分，并可以确定仪器或方法的检出限。化学校正技术是从化学量测数据中取得有关物质系统成分的定量信息的基本手段，是化学计量学研究的中心课题之一。应用多维数据分析技术，人们可以从复杂的量测数据中提取有效的化学信息。分析信号的处理技术用于分析检测、校正、干扰消除、分析信号的分辨等过程；化学模式识别是将分析数据转化为有用信息的重要化学计量学手段，对研究和预测物质的性质很有参考价值。多维化学测量技术的发展为化学模式识别的推广提供了条件。现代分析化学广泛使用复杂的分析仪器，能在瞬间提供大量数据，因此化学计量学理论和方法的应用就显得更为重要。

由此可见，化学计量学适用于分析的全过程。它的任务是应用数学、统计学、计算机科学的理论和方法，获取有关物质系统的成分、结构等相关的定性或定量信息，其理论和方法贯穿并服务于分析测量的全过程。

现代分析化学，不论是采用物理的、化学的，还是生物学的方法，基本任务都是获取物质系统的化学成分、结构及相关的信息，涉及的是化学量测问题。化学计量学正是从总体的角度考察现代分析涉及的采样、实验设计与优化、信号检测和分析及复杂数据的分析处理等问题，这自然地使化学计量学研究的规律成为现代分析化学基础理论和方法学的重要组成部分。因此，化学计量学属于边缘学科，是化学与数学的结合，化学与计算机科学的结合、化学与现代信息科学的结合，是化学的分支学科，既是与分析化学紧密相关的“软件”学科，又具有很强的实用性和实践性。

经过三十多年的发展，化学计量学的研究范畴已远远超出了分析化学范畴。它在环境化学、材料化学、医药与临床化学、地球化学，特别是在过程分析化学中的应用以惊人的速度发展。化学计量学使现代分析化学不再单纯提供化学量测数据，而且提供化学信息并直接参与化学实践和化学问题的解决，这也是分析化学在现代科学中又一次崛起和复兴。

化学计量学是一门年轻且正在迅猛发展的学科，其研究范畴正在不断的扩大，表 1-1 列出了化学计量学的基本研究领域，这对了解化学计量学的内容是十分有益的。

表 1-1 化学计量学的基本研究领域

序号	研究领域	基本理论和方法
1	分析数据处理	经典统计分析、非参数统计分析、回归与相关分析、时间序列分析、多元统计分析、方差分析、界外值检验

续表

序号	研究领域	基本理论和方法
2	试验设计与优化	析因设计、均匀试验设计、拉丁方试验设计、正交与试验设计、响应面分析、单纯形优化分析法优化、仪器参数的优化
3	分析采样理论	样本类型与取样方法、分层采样和采样最小取值量的估计、计量检验、动态采样与连续采样、计数检验
4	分析信号的处理	滤波、平滑和求导、曲线拟合、分析信号的检测理论、多样性光谱技术
5	分析信号的分辨	主成分分析、因子分析、信号的微分、重叠信号的去积、曲线拟合等
6	分析校正技术	线性校正、非线性校正、标准加入法、模糊校正
7	化学模式识别	降维与显示技术、特征选取、决策与分类、线性学习机、K-最邻近法、SIMCA 法
8	分析质量控制	质量控制和质量保证系统、方法的可靠性评价、灵敏度和检测限再现性和耐变性、选择性
9	人工智能和专家系统	数据库和谱图检索、计算机辅助教学、人工神经网络、化学专家系统、智能实验室

1.3 化学计量的历史沿革

化学计量学的历史可追溯到 1969 年, 当时 Kowalski 和 Isen-haur 将线性学习机应用于低分辨质谱的分类, 并在《Analytical Chemistry》上发表了一系列的学术论文。这些论文引入了一种新的思维方式, 用于将大量的分析数据转换成有用的信息。在分析化学中产生这种新的研究方式的原动力在于“很多年以来, 实验科学工作者记录了大量的数据, 但由于缺乏恰当的数据解析技