

化学计量学方法 及 MATLAB 实现

史永刚 栗斌 田高友 等编著

中国石化出版社

[HTTP://WWW.SINOPEC-PRESS.COM](http://www.sinopec-press.com)

化学计量学方法及 MATLAB 实现

史永刚 栗斌 田高友 等编著

中国石化出版社

内 容 提 要

化学计量学是一个新的化学分支学科,它应用数学、统计学、计算机科学和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,最大限度地获取关于物质系统的化学信息。

本书从化学量测的基本问题出发,阐述化学计量学基本理论和方法,介绍常用的化学计量学方法及 MATLAB 实现,内容涉及化学量测数据的统计描述、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、化学模式识别等。

本书可作为化学及相关学科的研究生和高年级大学生的教学用书和参考书,也可作为化学及化工等相关领域的科技工作者和教师的参考书。

图书在版编目 (CIP) 数据

化学计量学方法及 MATLAB 实现 / 史永刚等编著.
—北京:中国石化出版社,2010.4
ISBN 978-7-5114-0243-1

I. ①化… II. ①史… III. ①化学计量学②化学计量学-计算机辅助计算-软件包, MATLAB IV. ①06-04

中国版本图书馆 CIP 数据核字 (2010) 第 051294 号

未经本社书面授权,本书任何部分不得被复制、抄袭,或者以任何形式或任何方式传播。版权所有,侵权必究。

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

读者服务部电话:(010)84289974

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

E-mail:press@sinopec.com.cn

河北天普润印刷厂印刷

全国各地新华书店经销

*

850×1168 毫米 32 开本 11.625 印张 296 千字

2010 年 4 月第 1 版 2010 年 4 月第 1 次印刷

定价:30.00 元

过去的五十多年中，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。在很短的时间内获取关于物质系统的大量的复杂化学量测数据是一件十分容易的事情，这就使得现代化学工作者面临一个重大挑战，即如何处理与分析这些数据。因此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，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以及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，它涉及分析采样理论、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相当广泛。本书主要介绍一些比较常用的、重要的化学计量学方法，如实验设计、化学信号处理、化学校正技术和化学模式识别等。为了保持内容的完整性和逻辑性，将由易到难地介绍常用的化学计量学方法。

作为化学工作者学习和参考的化学计量学书籍，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，并给出常用化学计量学方法的 MATLAB 实现程序，从而使具有一般化学和数学基础的人员能比较轻松地阅读，学会如何进行化学计

量学方法的选择、理解化学计量学的计算结果和方便地使用这些化学计量学方法。

化学计量学从诞生到现在，发生了巨大的变化，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。

在编写的过程中，参考了大量的科学文献，在此对书中提及的和未提及的文献作者表示衷心的感谢。本书由史永刚主编，栗斌、田高友、李华峰、龚海峰和化岩等参加了部分章节的编写。在编写过程中，得到解放军后勤工程学院科研部和油料应用与管理工程系的大力支持和鼓励，在此表示真挚的感谢。

由于作者对化学计量学的理解不够深入和透彻，难免存在错误与不当之处，敬请读者批评指正。

第1章 概述	(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 化学计量学的发展历史	(5)
1.4 化学计量学的发展	(10)
1.5 化学计量学发展的动力	(11)
第2章 实验数据统计分析	(13)
2.1 数据及数据类型	(13)
2.2 实验数据的统计描述	(14)
2.2.1 随机分布	(15)
2.2.2 置信区间	(23)
2.2.3 误差的传递	(25)
2.3 统计推断	(27)
2.3.1 标准正态分布	(27)
2.3.2 χ^2 分布	(30)
2.3.3 F 分布	(31)
2.3.4 t 分布	(32)
2.3.5 假设检验	(34)
2.4 异常数据的剔除	(49)
2.4.1 粗大误差规则	(49)
2.4.2 Dixon 检验	(50)
2.4.3 Grubbs 检验规则	(53)
2.4.4 肖维勒准则	(55)
2.5 方差分析	(56)
2.5.1 单因素方差分析	(56)

2.5.2 双因素方差分析	(60)
第3章 采样理论及方法	(65)
3.1 正确采样的统计学标准	(65)
3.2 总体的类型	(66)
3.3 随机采样	(66)
3.4 分层采样	(67)
3.5 固体散状物料采样	(74)
3.6 抽样检验	(78)
3.6.1 计量抽样检验	(79)
3.6.2 计数抽样检验法	(80)
3.7 采样操作方法	(82)
3.8 动态过程中的采样	(83)
第4章 化学实验设计与优化	(85)
4.1 化学实验设计基础	(85)
4.1.1 试验指标	(86)
4.1.2 因素和水平	(86)
4.1.3 同时试验和序贯试验	(87)
4.1.4 试验最优化和解析最优化	(88)
4.1.5 有效实验存在的条件	(88)
4.1.6 实验设计的基本原理	(89)
4.1.7 实验设计的步骤	(90)
4.2 析因实验设计	(91)
4.2.1 析因设计表	(91)
4.2.2 析因设计的步骤	(92)
4.2.3 中心复合设计和 Box - Behnken 设计	(101)
4.3 正交实验设计	(103)
4.3.1 正交表及其特点	(103)
4.3.2 实验设计	(103)
4.4 均匀实验设计	(110)

4.4.1	均匀设计表	(110)
4.4.2	实验设计	(112)
4.5	随机区组	(115)
4.6	响应曲面法	(117)
4.6.1	形状已知的响应界面法	(117)
4.6.2	形状未知的响应界面法	(118)
4.6.3	单纯形法	(121)
4.7	D-最优实验设计	(129)
第5章	化学量测信号的检测与处理	(136)
5.1	分析信号检测	(138)
5.2	检测限	(139)
5.2.1	定义	(139)
5.2.2	检测限的点估计	(144)
5.2.3	检测限的t检验法估计	(151)
5.2.4	检测限的精度	(155)
5.3	分析信号处理	(156)
5.3.1	数字平滑与滤波	(156)
5.3.2	曲线拟合	(177)
5.3.3	谱峰面积的估计	(185)
5.4	分析信号的变换	(189)
5.4.1	傅立叶变换(FT)	(189)
5.4.2	Hadamard 变换(HT)	(190)
5.4.3	FT 和 HT 的应用	(191)
第6章	化学模式识别	(195)
6.1	数据预处理	(195)
6.1.1	丢失数据的弥补	(195)
6.1.2	中心化变换	(196)
6.1.3	归一化处理	(196)
6.1.4	正规化处理	(197)

6.1.5	标准化处理	(197)
6.2	方差 - 协方差矩阵、相关矩阵和距离	(197)
6.2.1	方差 - 协方差矩阵	(197)
6.2.2	相关矩阵	(198)
6.2.3	距离	(198)
6.3	特征选取	(200)
6.3.1	化学模式识别中常用的特征	(200)
6.3.2	特征选择的方法	(201)
6.4	主成分分析	(202)
6.5	无监督模式识别	(216)
6.5.1	聚类分析	(217)
6.5.2	显示技术	(219)
6.6	有监督模式识别	(219)
6.6.1	Fisher 线性分类法	(219)
6.6.2	Bayes 方法	(224)
6.6.3	K 最邻近法	(229)
6.6.4	SIMCA 法	(232)
第7章	化学校正理论	(237)
7.1	单变量校正	(238)
7.1.1	一元线性回归	(238)
7.1.2	相关系数和显著性检验	(243)
7.1.3	权重回归	(243)
7.1.4	曲线回归	(244)
7.2	多元线性回归分析	(246)
7.2.1	基本原理	(246)
7.2.2	K - 矩阵法和 P 矩阵法	(257)
7.3	逐步线性回归方法	(260)
7.4	主成分回归分析	(265)
7.5	偏最小二乘法	(270)

7.6	回归模型的诊断	(276)
7.6.1	残差与预测误差	(277)
7.6.2	界外值	(277)
7.6.3	大效应观测点	(278)
7.7	通用标准加入法(GSAM)	(279)
第8章	人工神经网络	(283)
8.1	人工神经网络基础	(283)
8.1.1	神经元	(284)
8.1.2	传递函数	(285)
8.1.3	人工神经网络学习方式	(289)
8.1.4	学习规则	(289)
8.1.5	网络构架	(290)
8.1.6	人工神经网络模型	(291)
8.2	MATLAB 神经网络工具箱	(296)
8.2.1	工具箱的应用方法	(296)
8.2.2	工具箱函数	(297)
8.2.3	nntool 的使用方法	(300)
8.3	应用举例	(306)
8.3.1	曲线拟合	(306)
8.3.2	化学校正	(308)
8.3.3	试验条件优化	(310)
8.3.4	化学模式识别	(312)
附录 I	线性代数基础	(315)
1	矢量和矩阵运算	(316)
2	线性相关	(325)
3	范数	(327)
附录 II	MATLAB 基础	(329)
1	简介	(329)
2	MATLAB 的启动与基本概念	(330)

2.1	启动	(330)
2.2	帮助	(331)
2.3	基本术语	(332)
3	数组/矩阵的创建及运算	(333)
3.1	创建	(333)
3.2	MATLAB 的基本数学运算	(334)
3.3	矩阵的运算	(340)
4	绘图	(344)
4.1	基本二维绘图命令	(344)
4.2	基本三维绘图命令	(348)
5	数据类型及输入输出	(349)
5.1	数据类型	(349)
5.2	数据的输入与输出	(350)
6	程序设计	(350)
6.1	MATLAB 程序的基本设计原则	(350)
6.2	MATLAB 程序的基本组成结构	(351)
6.3	M 文件的编辑及 MATLAB 工作路径的设置	(352)
6.4	MATLAB 的程序类型	(352)
6.5	声明子程序(函数程序)变量	(353)
6.6	字符串及其宏命令	(353)
6.7	常用的编程命令	(354)
6.8	关系与逻辑运算符	(354)
6.9	程序流程控制	(357)
6.10	小贴士	(359)

第1章 概 述

1.1 化学计量学的定义

归纳和推理是获得知识常用的两类科学方法论。建立知识体系时，人们首先通过对可观察对象的分析获得基本原理，然后用基本原理指导科学实践并进行现象的解释。这就构成 Oldroyd 所说的知识架构的概念^[1]，如图 1-1 所示。

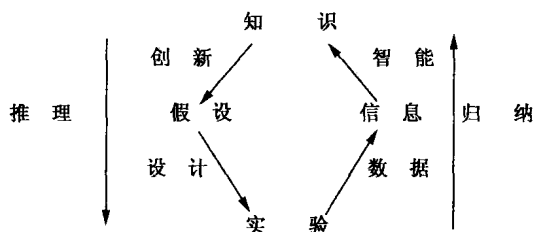


图 1-1 知识架构

化学计量学的产生与发展也遵循这种知识产生模式。Massart 等将化学计量学定义为^[2]：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。

化学工作者面对一个化学问题，要做的第一件事就是给定一个假设。为考察假设和验证假设的有效性，就要确定该做哪些实验、怎样做试验及如何优化实验[化学计量学定义中的(a)]。化学计量学借助以数学和统计学方法为基础的实验设计理论，可很好地完成此项工作。实验产生的数据是对研究对象的客观与真实反映。为了深入了解研究对象，必须从试验数据中提取尽可能多

的信息[化学计量学定义中的(b)],如应用化学校正技术建立测量值与实验变量间的关系模型、进行信号的分析与处理以及分析信号的识别等。在对分析信号有了较为全面的理解后,化学工作者就会应用这些信息给出关于研究系统的更多的知识[化学计量学定义中的(c)]。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,获取更多的数据和信息,从而更为深入地理解系统,最终建立一个完善的知识架构。

化学计量学主要涉及实验设计、信息提取与分析处理。关于知识的归纳,其他化学学科提得更多一些。在化学研究过程中,数据质量至关重要。应用化学计量学理论和方法,人们便能够更好地获取有效的数据,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^[3,4]。

1.2 化学计量学的研究范畴

实验是化学研究的基础。实验中,化学过程研究与试验结果分析是化学工作者关注的重点。早期,由于人们能够使用的实验与测试技术有限、方法提供数据的能力不足、数据包含的信息量较低和对实验数据处理的要求不高,较少涉及数据处理、实验方法选择与过程优化,这往往使人们不能以最优的研究尽可能多地提取有用的化学信息,丰富化学知识体系。

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。因此,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该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:实验设计和优化、样品测试和数据分析处理。

如果我们要对某石油产品进行评价,那么首先必须选择和建

立方法，即如何测定石油产品流变学性能、抗氧化安定性、摩擦学性质等；其次是怎样进行这样的测定，即建立分析方法，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，而且要充分利用化学计量学的理论与方法。

样品分析的第二步是测试样本的制备和量测。首先，涉及的是分析样本的采取。分析采样理论应用统计理论，确保采取的样本在必要测量精度满足测量需要。应用采样理论，我们就能确定分析所需的最低限度的样本量。采样理论的正确使用关系到产品质量的控制水平，也是发达国家确保其产品质量高和稳定的诀窍之一。获得分析样本后，就进行样品的制备和量测。目前，人们可以选用的分析手段多种多样，确定测定方法和手段后，就必须依据方法和手段的要求，合理制样，并应用现代分析仪器得到关于样品的信息。

样品分析的第三步为数据的处理与分析。这也是化学计量学应用最为广泛、深入及富有成果的领域。分析检测理论是化学量测数据定性检测的基础，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有重要参考价值。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分析仪器，能瞬间提供大量数据。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^[5]。

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。化学计量学则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，从而使化学计量学成为现代分析化学基础理论和方法学的重要组成部分。化学计量学属于交叉学科，是化学与数学、计算机科学和现代信息科学的结合，是化学的一个重要分支学科，既是与分析化学紧密相关的“软件”学科，又具有很强的理论性和实践性。

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正以惊人的速度发展。化学计量学使现代分析化学不再单纯提供化学量测数据，而且能够提供化学信息并直接参与化学实践和化学问题的解决，这也使得分析化学在现代科学中再次崛起和复兴。

表 1-1 列出了化学计量学的基本研究领域。

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
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、正交试验设计、响应面分析、单纯形优化、仪器参数的优化
3	分析采样理论	样本类型与取样方法、分层采样和分步采样、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

续表

序号	研究领域	基本理论和方法
5	分析信号的分辨	主成分分析、因子分析、信号的微分、重叠信号的分解、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、模糊校正和支持向量机等
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 法和支持向量机等
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 化学计量学的发展历史

化学计量学的历史可追溯到 1969 年。当时，Kowalski 和 Isenhour 将线性学习机应用于低分辨质谱的分析，在《Analytical Chemistry》上发表了一系列的学术论文^[6]。这些论文引入了一种全新的思维方式，即通过某种模式将大量的分析数据转换成有用的信息。产生这种新的研究方式的动力在于“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术，最终舍弃了这些数据”^[7]。

将 SIMCA 算法引入多元数据处理时，Wold^[8]进一步发展了这种思维模式，并将其称之为“软模型”。这些新技术的产生引起了其他科学工作者的注意，人们开始积极地应用这些“现代”算法，运用计算机所提供的强大的运算能力从复杂的数据中提取化学信息或解决疑难的最优化问题。

Dijkstra^[9]将信息理论用于质谱库的压缩，Massart^[10]则应用运筹学原理来进行色谱方法的优化。这些研究与分析化学所研究的基本科学问题是一致的，研究表明数学(如模式识别、信息理论、运筹学等)与分析化学的基本问题是密切相关的，如评估、优化、选择、分类、过程的组合与指定。换句话说，就是与涉及

决定该采用什么样的分析程序的所有过程密切相关^[11]。同时，在药物定量构效关系研究领域，模式识别的应用同样引起了人们的注意^[12]。

化学计量学 (Chemometrics) 是瑞典化学家 S. Wold^[13] 在 1971 年为一项基金命名时首次提出的，并在 1972 年瑞典的科学杂志上发表了化学计量学的相关论文。这一命名得到 B. R. Kowalski 的支持，并将研究小组所在实验室称为化学计量学实验室 (Laboratory of chemometrics)。

1974 年，Wold 和 Kowalski 合作创建了国际化学计量学学会 (ICS)，并创办了学会刊物《J. of Chemical Information and Computer Science》。一年后，国际化学计量学学会将化学计量学定义为：化学计量学是化学的一个分支学科，它应用数学和统计学方法设计和选择的测量过程和实验，并通过解析化学数据最大限度地获取化学信息^[14]。

随着化学计量学软件包的普及，越来越多的分析化学工作者采用模式识别和多元数据分析技术对化学量测数据进行探索性分析。1976 年由美国化学会举办了题为“Chemometric Theory and Application”的学术会议，并出版了首册关于化学计量学的专著《Chemometrics: Theory and Applications》，书中包含了化学计量学的主要研究内容：实验设计、最优化和多元数据分析。

1978 年后，在 Amsterdam 举办了“Computers in Analytical Chemistry (CAC)”国际会议。Analytica Chimica Acta 系列报道了“Computer Techniques and Optimization”。随后，美国化学会 (ACS) 出版首本化学计量学教科书，其他化学计量学教材也相继出版，如 1978 年 Massart 等的《Evaluation and Optimization of Laboratory Methods and Analytical Procedures》，1981 年 Kateman 和 Pi-jpers 的《Quality Control in Analytical Chemistry》、1988 年 Massart 的《Chemometrics: A Textbook》、1982 年 Lewi 的《Multivariate Data Analysis in Industrial Practice》和 1986 年 Sharaf 等的《Chemomet-