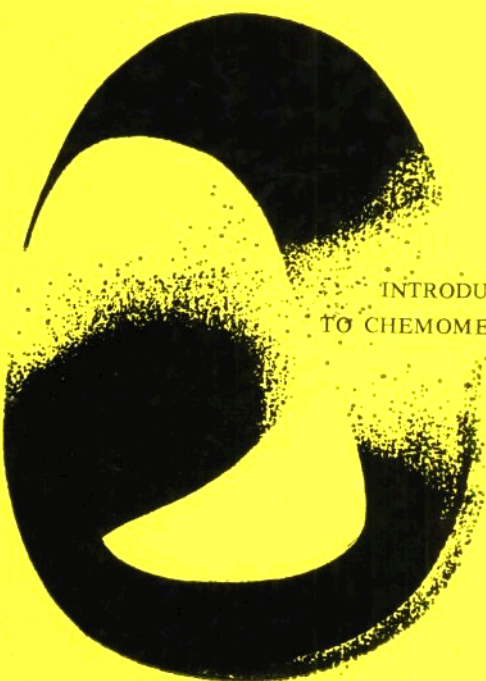


# 化学计量学导论

俞汝勤著



INTRODUCTION  
TO CHEMOMETRICS

湖南教育出版社

HUNAN EDUCATION PUBLISHING HOUSE

## 内 容 提 要

化学计量学是一门新兴的化学分支学科。它应用数学、统计学与计算机科学的工具与手段,设计或选择最优化学量测方法,并通过解析化学量测数据,最大限度地获取有关物质系统的化学信息及其他相关信息。化学计量学研究化学量测的基础理论与方法学。本书围绕与化学量测即分析测试过程相关的基础问题,论述有关化学计量学原理与方法,包括分析信息理论,分析采样、检测、校正理论,分析信号的滤波、平滑、求导、变换、卷积等处理方法,试验设计与优化,数字模拟技术等。在有关校正理论的部分,论述了如何通过多元校正,由多维化学量测数据获取有关物质定性、定量化学组成的信息。在有关化学模式识别、构效关系研究、人工智能化学专家系统等章节,论述了如何运用这些工具与手段,从多维化学量测数据中提取有关物质系统更深层次的隐含信息,以解决化学及各种相关领域中的实际问题。本书的读者对象是广大化学工作者,以及用到化学量测即分析测试手段的其他领域的科技人员,包括轻工、化工、环境保护、卫生防疫、药物设计、法医检验、地质勘探、质量控制与商品检验、各种合金、无机非金属材料、催化剂设计等领域的有关人员。本书还可作为有关院校开设化学计量学课程的教材或教学参考书。

33506/101

# 前 言

本书是以作者在湖南大学为化学、化工类专业的研究生开设的化学计量学学位课程内容为基础写成的。化学计量学这一新兴化学分支涉及的内容极为广泛,各种数理统计方法在化学及分析化学中的应用,都可归入化学计量学范畴。如将化学中应用的经典数理统计方法全部收入,则这些内容势必占据全书相当大的篇幅,而冲淡了有关近十余年发展的新概念、新方法的阐述。从教学角度看,最好将基础数理统计方法在化学中的应用的的教学,与化学计量学教学分阶段进行,循序渐进。本书中仍收入了部分属上述基础性质的内容。为了防止“淡化”效应,部分收入本书的这种基础性章节,例如化学试验设计与优化(第十一章),内容压缩到最低限度,并放在主要章节之后。属于化学计量学范畴的各种化学中应用的计算机技术与方法,内容亦极为广泛,不可能全部收入书中。例如,化学数据库与谱图检索这一重要内容即未收入本书。本书内容的取舍,基本上反映了作者实际教学中的取材范围。我们力图使本书包含能代表化学计量学这一新兴化学分支特征内容的典型材料,而尽量避免将化学计量学表述为化学中应用的各种数理统计方法的集合,或是化学中应用的计算机技术与方法的集合。“化学计量学”这一名词的提出者S.Wold当年不选用“化学数据分析”、“化学中的计算机”这类名词,而创造“化学计量学”这一名词,实际上即反映了这一指导思想。他在为一本编排极为详尽、内容多达27章的化学计量学教科书撰写书评时,论及如何在罗列材料时注意反映化学计量学的整体指导思想,抓住其特征与基本目标的问题[见S.Wold, *Chemometrics & Intelligent Laboratory Systems*, 5, (4), 276(1989)]。本书作为

“化学计量学导论”，期望能引导读者将化学计量学作为化学量测的基础理论与方法学这样一个整体进行研究，而不是将它认作许多互不关联的概念与方法的集合。

化学计量学的大量算法都离不开编制计算机程序。计算机硬件与软件发展十分迅速，化学中应用的很多传统数理统计方法往往借助于可编程序计算器即能实现。在化学计量学发展过程中得到广泛应用的BASIC、FORTRAN等传统语言，面临被一些新兴的符号运算语言取代的挑战。在北美一些化学计量学研究中心，上述传统高级语言已渐少有人问津，这是一个可喜的趋势，它预示化学家在试图使用化学计量学算法时，将较少遇到编制冗繁的计算机程序的困扰。在计算机硬件逐步更新换代、协处理器应用日益广泛、新的软件与高级语言趋于普及之后，化学量测中使用一些较复杂艰深的化学计量学算法，将会像50年代用对数表或手摇计算器计算分析结果、60年代用袖珍计算器计算误差、70年代用可编程序计算器进行回归等运算一样容易。有理由预言，在不久的将来，现代化学计量学算法必将成为化学、分析化学工作者日常工作中不可或缺的工具。鉴于上述发展趋势，本书中除极个别情况下为说明有关算法引述了一些程序片断外，未提供任何具体计算机源程序。虽然计算机硬件与软件在不断更新，但化学计量学的基本原理与内容，仍必须循序渐进、逐步提高才能掌握。用MATLAB这类高级语言仅书写一条或数条语句，即能完成多元线性回归、奇值分解、主成分回归等典型的化学计量学基本运算，但初学者仍应力求弄懂这些算法的基本原理，掌握了基本原理，可根据实际的条件与可能，用任何一种合适的计算机语言实现有关运算。

化学计量学的兴起与发展，是以计算机应用为标志的科学技术大变革在化学、分析化学领域的重要体现。学习化学计量学，是引导化学工作者、分析化学工作者步入计算机时代的一个最佳门径。作者期望本书能起到抛砖引玉的作用，有助于吸引更多的

化学工作者了解、应用与研究化学计量学基本理论与方法，并望读者们对本书进行批评指正。

本书叙及的作者实验室部分研究成果，系国家自然科学基金资助的工作。本书在成稿与勘校工作中，得到本实验室博士研究生吴海龙、黄杉生、孙立贤、宋新华、谢玉琰、李梦龙、张增荣、曾志恽及硕士研究生柳爱平、陈茁、于正英等同志的协助，谨致诚挚的谢忱。

**作 者**

1990年9月于岳麓山

# 目 录

<b>第一章 绪论</b> .....	( 1 )
§ 1—1 化学计量学的研究对象与发展概况.....	( 1 )
§ 1—2 化学计量学——化学量测的基础理论与方法学 .....	( 3 )
§ 1—3 本书各章内容简介.....	( 6 )
参考文献.....	( 11 )
<b>第二章 分析信息理论</b> .....	( 12 )
§ 2—1 化学量测的“不确定度”与熵.....	( 12 )
§ 2—2 分析量测的信息量.....	( 16 )
§ 2—3 分析仪器的供信能力.....	( 20 )
§ 2—4 分析信息量的其他量度.....	( 22 )
参考文献.....	( 24 )
<b>第三章 分析采样理论</b> .....	( 26 )
§ 3—1 概述.....	( 26 )
§ 3—2 固体物质采样.....	( 27 )
§ 3—3 分层采样与二步采样.....	( 30 )
§ 3—4 从动态过程物质流采样.....	( 35 )
§ 3—5 生产过程连续监控中的采样.....	( 36 )
§ 3—6 质量检验中的采样.....	( 42 )
参考文献.....	( 48 )
<b>第四章 分析检测理论</b> .....	( 50 )
§ 4—1 分析信号的检出.....	( 51 )

§ 4—2	分析信号的检测特性函数	( 54 )
§ 4—3	分析方法的检测限	( 56 )
	参考文献	( 58 )
<b>第五章</b>	<b>校正理论I 化学量测体系的基础校正方法</b>	( 59 )
§ 5—1	化学量测中的校正作为系统参量估计问题	( 59 )
§ 5—2	单一组分测定的线性校正模式	( 61 )
§ 5—3	多元线性回归(MLR)基本校正方法	( 67 )
§ 5—4	通用标准加入法	( 69 )
§ 5—5	非线性校正	( 71 )
	参考文献	( 75 )
<b>第六章</b>	<b>校正理论II 基于因子分析的多元校正方法</b>	( 77 )
§ 6—1	多元线性校正模式的进一步考察	( 77 )
§ 6—2	分析信号阵的特征值及特征矢量; NIPALS算法	( 80 )
§ 6—3	分析试样中组分数确定	( 84 )
§ 6—4	交互校验法	( 88 )
§ 6—5	分析体系中存在的组分的鉴别; 目标转换因子分析	( 90 )
§ 6—6	主成分回归法(PCR)	( 93 )
§ 6—7	偏最小二乘法(PLS)	( 95 )
§ 6—8	结语: “软”“硬”校正模式	( 97 )
	参考文献	( 98 )
<b>第七章</b>	<b>分析信号处理</b>	( 100 )
§ 7—1	分析信号的滤波	( 100 )
§ 7—2	分析信号的平滑与求导	( 105 )
§ 7—3	光谱分析的多重性效益与Hadamard变换	( 110 )
§ 7—4	实现光谱多重性效益的Fourier编码	( 113 )
§ 7—5	FFT及其在光谱与电化学量测中的应用	( 120 )
§ 7—6	卷积运算用于分析信号解析	( 129 )

参考文献	(134)
<b>第八章 化学模式识别I 借化学量测辨识物质的隐含性质</b>	(136)
§ 8—1 化学模式空间	(136)
§ 8—2 模式空间的距离与相似性量度	(141)
§ 8—3 线性学习机	(142)
§ 8—4 K-最近邻法	(146)
§ 8—5 原始化学量测数据的预处理	(149)
§ 8—6 参量选择	(151)
§ 8—7 降维与显示技术	(155)
§ 8—8 无监督的分类: 聚类分析用于化学量测数据解 析	(158)
§ 8—9 化学模式识别的应用	(163)
参考文献	(166)
<b>第九章 化学模式识别II 定量构效关系研究</b>	(169)
§ 9—1 构效关系研究中的全局“硬”模式与局部“软” 模式	(169)
§ 9—2 化学模式识别与QSAR的4个水平	(172)
§ 9—3 SIMCA法I 水平2的化学模式识别与QSAR	(175)
§ 9—4 SIMCA法II 水平3,4的化学模式识别与QSAR .....	(181)
§ 9—5 QSAR中表征化合物结构的参量	(184)
§ 9—6 分子连接性指数	(188)
参考文献	(196)
<b>第十章 人工智能与化学专家系统</b>	(198)
§ 10—1 人工智能用于化学计量学研究	(198)
§ 10—2 人工智能的基本要素	(200)
§ 10—3 化学专家系统	(209)
§ 10—4 人工神经网络与自适应化学模式识别	(218)



参考文献	(225)
<b>第十一章 化学试验设计与优化</b>	(227)
§ 11—1 化学量测与试验中的设计与优化问题	(227)
§ 11—2 析因试验	(230)
§ 11—3 正交试验设计	(233)
§ 11—4 单纯形优化法	(238)
参考文献	(243)
<b>第十二章 化学反应及化学量测误差的数字模拟</b>	(244)
§ 12—1 模拟用于化学反应机理及量测误差的研究	(244)
§ 12—2 随机变量及随机数的模拟与产生方法	(245)
§ 12—3 化学量测误差传递的 Monte Carlo 模拟	(249)
§ 12—4 化学反应动力学的 Monte Carlo 模拟	(251)
§ 12—5 电化学过程的数字模拟	(257)
参考文献	(264)
<b>CONTENTS (英文目录)</b>	(266)

# 第一章

## 绪 论

### § 1—1 化学计量学的研究对象与发展概况

1971年,瑞典化学家S.Wold在为—项基金项目定名时,从“化学数据分析”、“化学中的计算机”及“化学计量学”几个名词中选出后者。美国B.R.Kowalski赞赏此—名词,后与wold共同创建国际化学计量学学会。化学计量学这—新兴的化学分支领域,在70—80年代发展十分迅速,它对化学与分析化学的发展产生了深远的影响。化学计量学作为化学的分支学科,应用数学、统计学与计算机科学的工具与手段设计或选择最优化学量测方法,并通过解析化学量测数据最大限度地获取化学及其他相关信息。在分析化学这—化学表征与量测科学领域,化学计量学面临的课题是获取有关物质系统的成分、结构等相关的信息,它研究的是有关化学量测的基础理论与方法学。简言之,化学计量学研究运用数学、统计学与计算机科学的方法,进行化学量测试验设计与数据的处理、分类、解析和预测。可以认为,化学计量学是化学、分析化学与数学、统计学及计算机科学之间的“接口”。化学计量学这—名词与经济计量学等类似名词对应,在化学术语中,不可与研究化合物组成的经典的“化学计量关系”概念相混淆。化学文献中曾出现过“化学统计学”等其他名词,由于化学计量学涉及的并不局限于统计学方法,而包括逻辑、拓扑、图论等许多其他数学方法,以及诸如人工神经网络智能化学系统等计算机科学方法,

采用“化学统计学”这种名词将过于受限，实际上化学计量学是一门通过学科交缘发展起来的新的化学分支学科领域，而不单纯是统计方法在化学中的应用。

化学计量学的发展，可追溯到较早的年代。在发展前期，即“化学计量学”这一名词出现之前的时期，各种数理统计方法即在化学中应用，研究化学及其他量测误差的Student分布，就是化学家W.S.Gossett作出的贡献。有机化学家早期研究的线性自由能关系，可认作化学计量学的重要分支——定量构效关系研究的前身。分析化学家广泛应用统计学方法考察分析结果的误差，这些统计学方法多属“描述型”的，如计算分析结果的标准差、置信区间等<sup>[1-1]</sup>。60年代末期，计算机应用日益广泛，为70年代的新发展准备了条件。70年代是化学计量学兴起与发展的时期，这时出现了“化学计量学”这一名词。化学工作者，特别是分析化学工作者，为这一新兴的化学学科分支的发展作出了重要贡献。化学工作者并不局限于应用现成的数学与统计学方法，他们根据化学学科的需要创造了一系列化学量测数据的处理、分类、解析与预测方法。这一发展的背景，是计算机的普及应用，以及各种强有力的数学方法借助计算机得以在化学量测中实际应用。化学计量学在80年代有了较大的发展，各种新的化学计量学算法的基础及应用研究取得了长足进展，编制了许多化学计量学软件，有的软件成为现代化学量测仪器主要是分析仪器的有机组成部分。化学计量学课程进入化学教学大纲，分析化学被重新认识为化学信息科学，专门刊登化学计量学论文的学术期刊Journal of Chemometrics, Chemometrics & Intelligent Laboratory Systems等问世，检阅化学计量学研究成果的国际学术会议陆续召开，从事化学计量学研究的化学与分析化学工作者，正步入90年代新的发展时期。由于计算机及软件技术的更新与发展，符号处理高级语言的普及应用，使许多过去认为过于复杂难于普及的化学计量学算法逐步推广应用。愈来愈多的化学工作者认识到，化学计量

学方法是其日常工作中不可或缺的基本工具，借助这些工具，才能有效地处理化学试验中得到的大量量测数据，提取其中蕴含的丰富信息。

## § 1—2 化学计量学——化学量测的基础理论与方法学

化学计量学探讨各种类型化学量测过程的共性问题，包括化学量测试验的设计、测试数据的处理及有用信息的提取，也就是说，化学计量学研究的对象是化学量测的基础理论与方法学。这就决定了它与分析化学之间特殊密切的关系，因为分析化学正是化学表征与量测科学。分析化学这一对任何化学家都十分重要的基础化学分支，有着悠久的发展历史。在化学发展初期，分析化学曾发挥了特殊的作用，在大量化学元素的发现、化学现象的定量解析、化学反应及化合物组成的基本定量关系的确立等许多方面，分析化学作出过重大的贡献，它促进了化学上升为定量科学的过程，故有人称分析化学为化学之父。百余年前，分析化学引进物理化学有关溶液平衡等原理，构造了以化学分析法为主体的经典分析化学基础理论，这促进了分析化学由一种精湛的实验技艺发展成为具有定量基础理论的科学。约一个世纪以后，分析化学最突出的发展是分析工作的仪器化，为了完成现代科学技术向分析化学提出的基本任务——通过化学量测获取有关物质的化学组成、结构等相关信息，分析工作者引进了各种化学的、物理的、生物的及其他新技术，建立了各种基于光、电、声、热、生物等原理与现象的新型分析仪器与方法，即使是继续沿用的经典分析方法如滴定分析等，亦逐步仪器化、自动化。各种新型分析方法的涌现丰富了分析化学这一化学表征与量测科学的内容。但也为分析学科提出了新的课题。经典分析化学曾把着眼点主要放在分析数据的获取上面，用纯粹化学的方法取得分析数据曾经是要求高超实验

技巧的艰巨繁重的工作。W. A. Lampadius在一部无机物分析大全经典著作中写道：“谁要是缺乏耐心准备以几个星期或几个月时间取得分析结果，谁就根本不必去开始一项分析工作”。现代分析化学已取得如此巨大的进步，许多分析仪器能在数秒钟时间内输出成打的分析数据，因此，如何处理这些大量的原始分析数据，就成为十分突出的问题。另一方面，现代科学技术的发展给分析化学提出了越来越高的要求，人们已不能满足于取得原始分析数据，而要求从中最大限度地提取有用的信息，以解决各种生产实际问题与科研课题，数据处理与信息提取便成为化学量测中的首要问题。化学计量学就是在这种背景下发展起来的。当然，客观需要是一个方面，实际可能性是另一个方面。由于计算机科学的发展，各种过去无法使用的复杂数学方法可在计算机上实现，为解决复杂的数据处理与信息提取问题提供了实际可能性。化学计量学的发展为现代化学研究提供了新的手段，大大丰富了化学表征与量测科学——分析化学的基础理论与工具库，它标志着现代分析化学已被重新认识为化学信息科学，分析化学在化学科学舞台上扮演的角色已由单纯提供分析数据上升为提供化学信息与直接参与实际课题的解决。这可能是自化学发展初期分析化学起过特殊重要作用以来，分析化学在现代化学科学中一次新的崛起与复兴。

化学计量学的奠基人之一B. R. Kowalski<sup>[1-2]</sup>在80年代初曾讨论分析化学与应用数学、统计学结合的必然趋势。高鸿<sup>[1-3]</sup>曾预言分析化学与数学密切结合的时代来临。实际上，数学理论对任何自然科学的重要性早已为哲学家们所认识。康德曾写道：“在自然科学各个分支中，只有那些能以数学表述的学科，才是真正的科学”。以分析化学家为主体的化学计量学家，对分析化学的发展作出了重要的贡献，那就是以现代数学语言，构造与发展了严谨的化学量测基础理论，这些理论正成为现代分析化学基础理论的重要组成部分<sup>[1-4]</sup>。前面提及的经典分析化学基础理论，已不能适应现代分析化学新的极其丰富的内涵，从不同的化学分支及

其他学科引入的新原理、新方法、新技术，其相关的理论均自然地进入现代分析化学基础理论，成为它的组成部分，但分析化学从总体来看，不论采用何种方法与手段，它所涉及的问题是进行化学量测，获取有关物质化学成分、结构及相关的信息，化学计量学正是从总体上考察这些问题，这就很自然地使化学计量学研

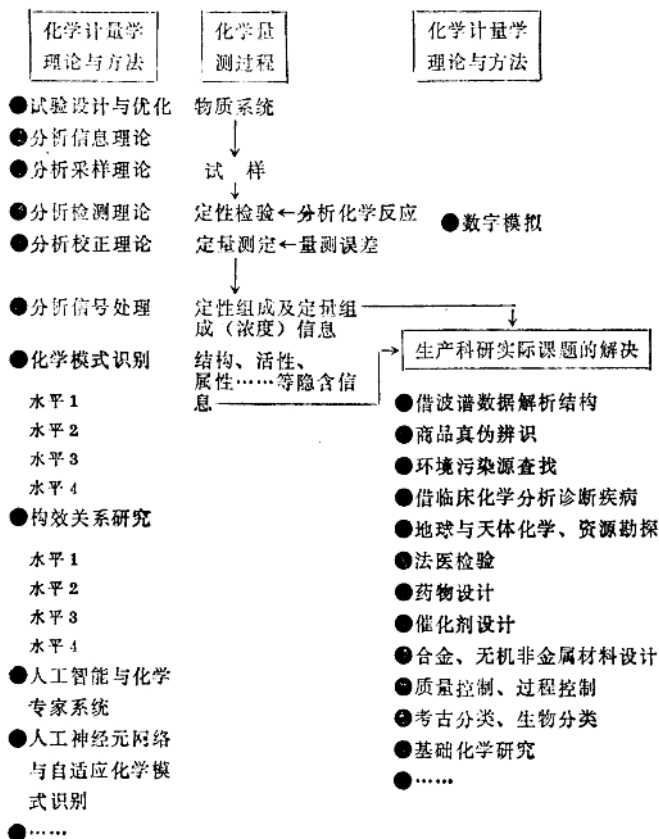


图1-1 化学量测过程与化学计量学

究的规律,成为现代分析化学基础理论与方法学的重要组成部分。

化学量测从实验设计、采样、定性检测、定量校正到信息提取,无一不涉及有关化学计量学的理论与方法。图1-1示意列出后者与化学量测过程的关系及其对解决生产、科研实际问题的意义。这些化学计量学原理与方法,将在本书各章中进行叙述与讨论,下节将简要介绍各章内容的梗概。

### § 1—3 本书各章内容简介

分析信息理论(第二章)从信息理论的角度研究化学量测过程。化学量测的目的,是获取有关物质系统的化学成分、结构及其他相关信息。在进行分析测试之前,缺乏有关物质系统的这些信息,或知之不多,即存在“不确定度”。化学量测能消除或减少这种“不确定度”。本章讨论用熵等概念定量表征这种“不确定度”(即化学量测所获得的信息量),并对分离、定性鉴定、定量测定等进行信息理论分析。显然,获取最大的信息量是选择与设计一个优良的化学量测方法的基本指标。现代分析化学使用越来越多的复杂的分析仪器,这不是出于进行化学量测的化学家的偏爱,而是由于这些分析仪器具有强大的提供信息的功能。

分析采样理论(第三章)涉及的是所有化学量测的第一步——测试样本的采集。分析采样理论是指试样采集的数学统计理论,对于静态的均相试样,例如静止的均匀液体或气体试样,取很小的局部样本即足以代表总体,而对于非均相的试样,则情况较为复杂,为保证必要的量测精度最低限度需采集的样本量,可借统计理论计算求出。分层采样与二步采样是检验大宗试样常用的采样方法。对于动态体系,需考虑试样成分随时间变化的规律,特别是动态过程的实时监测,需运用时间序列分析这一数学方法进行考察,采样频率与分析滞后将影响过程控制反馈的实时性。生产过程中的质量检验,采用计量抽检或计数抽检等采样方法。

用序贯采样法能在保证较小的第一类错误（合格产品被拒收）与第二类错误（不合格产品被接受）概率的前提下，减少检验工作量。有关质量控制的采样理论，对生产过程中产品质量的严格控制极为重要，这是发达国家确保其产品质优的基本诀窍。普及采样检验知识，废除目前仍在不少部门沿用的缺乏科学依据的不合理采样方法，将有力地促进产品质量的提高。

分析检测理论（第四章）是化学量测中定性检验的数学统计理论基础。根据化学量测信号，需判断试样中是否存在某种微量组分，例如药物中某种有害的杂质是否存在。由于分析仪器存在噪声（仪器噪声），试样中除待检测的组分外，其他共存物质亦会造成干扰（化学噪声），因此，测出了分析信号不等于某一杂质一定存在，正确的检测应使两类错误的概率尽可能小，第一类错误是杂质并不存在，误将其“检出”，即将噪声信号认作杂质组分的真实分析信号；第二类错误恰好相反，在杂质确实存在时漏检。分析检测理论采用有严谨统计学依据的准则，判别化学量测信号是来自待测杂质或是来自噪声干扰，从而确定有关分析仪器或分析方法的检测特性，给出该仪器或方法检测限的严谨统计学定义。

校正正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，它是化学计量学基础研究的中心课题之一，在第五、六两章中有较详细讨论。第五章讨论化学量测体系的基础校正方法，将校正作为分析系统的参量估计问题处理，区分线性校正模式与非线性校正模式，并考察校正模式中的参量是否可以辨识。单组分线性校正体系采用常规线性回归方法处理，这方面内容实际是传统分析测试中常用的校正方法的总结。多数化学计量学校正方法是以多分析通道量测、多组分同时测定为基础的，因而提供信息的功能较传统分析测试方法更强。进行多元体系分析时，如需测定全部组分，采用直接校正或全校正；如仅需测定其中某一种或数种组分，采用间接校正或部分校正。为了克服基体效应



的影响,传统的标准加入法扩充为适用于多组分体系的通用标准加入法。非线性化学量测系统可通过某些变换转化为线性系统,或采用某些处理多变量系统的方法进行校正运算。

从单组分校正系统扩充到多组分系统及多通道量测是化学计量学提高获取化学信息功能的第一步。第五章讨论的多元线性回归方法提取信息的功能,可通过采用因子分析方法进一步提高。第六章集中讨论基于因子分析的这一大类多元校正方法。通过主成分分析,将化学量测数据中真正包含有关化学成分信息的部分,与量测噪声尽量分离开来。因子分析为确定未知体系中的组分数以及这些组分的光谱等特性提供了有力的手段,目标转换因子分析即是这种手段。本章中将第五章讨论的多元线性回归(MLR)扩充为主成分回归(PCR)和偏最小二乘回归(PLS)。PLS方法能充分地 从化学量测及浓度数据中提取有用信息,有效地克服由于校正模式偏离理想状态和存在各种干扰与不确定因素而带来的影响。这些方法也是在更深层次提取化学信息(化学模式识别,第八、九章)时有用的工具。

分析信号处理(第七章)包括各种化学量测信号的处理方法,它们是分析检测、校正、干扰消除、重叠信号分辨等的重要手段。数字滤波可用于有用分析信号与噪声的分离;Kalman滤波这一递归型的线性状态估计方法,可用于多元化学量测系统的干扰消除、信号分辨与动态状态估计等。从非平衡化学量测数据估计平衡时的状态,不但能显著缩短测试时间,而且使许多按传统方式无法使用的分析反应或测试方法的应用成为可能。分析信号的平滑,实际上是熟知的通过实验测得的数据点绘制一条光滑曲线这种方法的扩展,在给定拟合函数情况下,可进行任何数量数据点的平滑与求导。为了提高光谱化学量测的信噪比,获取“光谱多重性效益”,可使用Hadamard或Fourier变换。本章较详细地讨论了在化学量测中应用极广的快速Fourier变换(FFT)。FFT不但在光谱化学量测中有广泛应用,对电化学量测、化学模式识别数据预处理