

高等学校教材

# 化学计量学

*Chemometrics*

梁逸曾 俞汝勤 编著



高等教育出版社

高等学校教材

# 化学计量学

Chemometrics

梁逸曾 俞汝勤 编著

高等教育出版社

## 内 容 提 要

本书是教育部立项招标和“九五”规划教材。在编排上,本书力图包括化学计量学的大部分内容,并主要以基本概念、基本方法为主线介绍。

本书的主要内容有:概论,分析采样理论和方法,化学实验设计与优化方法,分析检测理论与信号处理方法,多元校正与多元分辨,化学模式识别,计算机数字模拟法,化学构效关系的研究方法,人工智能与化学专家系统方法,以及统计学与线性代数基础知识。

本书可作为化学专业的本科及研究生教材使用。

## 图书在版编目(CIP)数据

化学计量学/梁逸曾,俞汝勤编著. —北京:高等教育出版社,2003.5

ISBN 7-04-011968-4

I.化… II.①梁…②俞… III.化学计量学-高等学校-教材 IV.06-04

中国版本图书馆 CIP 数据核字(2003)第 000415 号

---

出版发行	高等教育出版社	购书热线	010-64054588
社 址	北京市东城区沙滩后街 55 号	免费咨询	800-810-0598
邮政编码	100009	网 址	<a href="http://www.hep.edu.cn">http://www.hep.edu.cn</a>
传 真	010-64014048		<a href="http://www.hep.com.cn">http://www.hep.com.cn</a>
经 销	新华书店北京发行所		
印 刷	北京人卫印刷厂		
开 本	787×960 1/16	版 次	2003 年 5 月第 1 版
印 张	21.5	印 次	2003 年 5 月第 1 次印刷
字 数	360 000	定 价	27.50 元

---

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

**版权所有 侵权必究**

## 前 言

化学计量学(chemometrics)在我国发展已有二十多年的历史,是一门化学与统计学、数学、计算机科学交叉所产生的新兴的化学学科分支。它运用数学、统计学、计算机科学以及其他相关学科的理论与方法,优化化学量测过程,并从化学量测数据中最大限度地提取有用的化学信息。化学界有时将化学计量学与基于量子化学的计算化学(computational chemistry)列入同一学科分支领域,但化学计量学与通常理解的计算化学有显著差异,化学计量学是以化学量测为其基点,实质上是化学量测的基础理论与方法。

化学计量学为化学量测提供基础理论和方法,为各类化学波谱及化学量测的数据解析,为化学化工过程的机理的研究和优化提供新途径,它涵盖了化学量测的全过程,包括了采样理论与方法、实验设计与化学化工过程的优化控制、化学信号处理、分析信号的多元校正与分辨、化学模式识别、化学过程和化学量测过程的计算机模拟、化学定量构效关系、化学数据库与化学专家系统等主要分支,是一门内涵相当丰富的化学学科分支。化学计量学的发展为化学各分支学科,其中特别是分析化学、有机化学、高分子化学、物理化学、无机化学、化学工程、环境化学、药物化学、生物化学等,提供了不少解决问题的新思路、新途径和新方法,是一门实用性很强的化学学科分支。近年来,由于计算机和计算技术的飞速发展,化学计量学的内涵还在不断扩大,怎样将这门新兴的化学分支学科的基本内容及最新发展成果总结为一本可为化学和化学工程本科学生和研究生使用的教材,以适应现代化学学科的发展,的确是一件值得做好的工作。

化学计量学是以计算机和近代计算技术为基础的一门新兴交叉学科,所用的方法大都是基于矩阵运算的方法,对于数学基础学得不够的化学专业的学生来说,的确是有很大的难度的。如按常规教学方法(即依靠BASIC、FORTRAN等经典高级语言为工具,以讲算法为内容)来进行的话,不但需要很多学时,甚至还很难组织好它的教学,只会使学生感到该课程支离破碎,不成系统。然而,化学计量学虽的确是以计算机和近代计算技术为其基础,但它是具有其深刻的化学内涵的;而且,如果化学计量学离开了它得以生长的土壤——化学学科,就失去了它存在的必要性。笔者认为,进行化学计量学研究的首先必须是化学家,研究者所关心的问题

是化学问题,而不是计算机和数学问题,所以,如何从化学的角度来介绍化学计量学,怎样将化学计量学的全貌介绍给学生,使他们通过学习这门课程,充分理解化学计量学是一门什么样的化学分支,它为化学家们提供了什么样的新思路,可以解决什么样的化学问题。这样,才会使化学学科的本科生和研究生认识到学习这门课程的重要性,才会激发起化学学科的学生对这门新兴学科的学习热情。所幸的是,由于计算机科学、统计学和应用数学的发展,出现了像基于计算器运算似的新一代高级语言 Matlab(Matrix Laboratory,矩阵实验室)。它的出现,为广大需要计算的从事科研和工程技术人员提供了一套全新的工具,可将大量计算机编程简化,直接用简洁矩阵计算式来进行计算,而毋需理会一些编程的细节,为专业人员节省了大量时间。可以这么说,现在在国际化学计量学学界,90%都是使用 Matlab 语言来编程的,用 Matlab 语言来进行交流已成为国际惯例。更为重要的是,使用这种语言编程,只要对简洁的矩阵运算式的数学意义有了理解,就可顺利地进行各种化学计量学的运算实验,对学生来说,将起到大大降低学习难度的作用,使学生可将主要精力放在方法的化学意义和基本思路之上,而没有必要将精力放在编程技巧之上。这样,化学计量学这门课程的学习内容也可增加,尽量做到使学生对化学计量学的全貌有所了解。所以,本书就是以 Matlab 语言为其工作语言来进行编排的。

我们在此书的编排上采用了下述策略:

1. 力图在本书包括化学计量学的大部分内容,以求学生对化学计量学有一个整体概念,使学生对化学计量学在化学学科中的地位有一个正确的理解。

2. 本书主要强调化学计量学中的基本概念和化学计量学方法的基本思路,而对于一些方法的数学推导都是以简洁的矩阵运算形式给出,只要学生熟悉了矩阵运算的符号系统,对它们的理解将比以常规标量形式给出的推导更容易接受。本书第十章提供了必要的有关矩阵运算的基础知识。教师如何做到先让学生理解和熟悉矩阵运算的符号系统及其化学和物理意义是学习这门课程的至关重要的一步。

3. 对一些需要较深数学基础但又十分重要的新方法,本书以小字编排给出,教师不必全部讲授,可由学生自己根据学习的实际情况取舍。但是,这一部分内容对于研究生的学习却是十分重要,如研究生较熟练掌握了这些内容,即可直接进入化学计量学的研究工作。

化学计量学是一门很年轻的交叉学科,现还正处在快速发展的时期,本书虽力图涵盖它目前的大部分内容,给出该学科的全貌,但仍会有不到

之处;另外,由于我们对化学计量学的理解也绝非完全正确,所以书中的疏漏之处和错误还必定不少,希望得到化学计量学同行的批评指正。另外,此书的出版,得到了教育部和高等教育出版社的大力支持,在此谨致谢忱。

作者 于岳麓山下  
2000年元月

# 目 录

<b>第一章 概论(Introduction)</b> .....	1
§ 1-1 化学计量学的研究对象与发展概况(Research Object and Developing Situation of Chemometrics) .....	1
§ 1-2 化学计量学——化学量测的基础理论与方法学(Chemometrics: the Fundamental Theory and Methodology of Chemical Measurements) .....	2
§ 1-3 本书各章内容简介(Brief Description of Content of Each Chapter) .....	4
§ 1-4 使用说明(Illustration of Usage) .....	7
习题.....	8
参考文献.....	8
<b>第二章 分析采样理论和方法(Analytical Sampling Theory and Methods)</b> .....	10
§ 2-1 采样的基本概念和理论(Basic Concepts and Theory of Sampling) .....	10
2-1-1 随机采样(Random Sampling) .....	10
2-1-2 系统采样(Systematic Sampling) .....	11
2-1-3 分层采样(Stratified Sampling) .....	12
2-1-4 代表性采样(Representative Sampling) .....	12
2-1-5 最小采样数目的估计(Estimation of Minimum Number of Samples) .....	14
2-1-6 采样常数(The Sampling Constants) .....	15
§ 2-2 非均匀体系建模方法及大批物质的采样误差(Modeling the Heterogeneity Pattern and Sampling Errors for Bulk Materials).....	16
2-2-1 大批固体物质的采样理论和方法(Sampling Theory and Methods for Bulk Materials) .....	16
2-2-2 颗粒性质因子:Gy 理论(Particle Property Factor:Gy's Theory) .....	17
2-2-3 动态过程中物质流的采样理论和方法(Sampling From Streams of Material in Dynamic Processes) .....	18
§ 2-3 质量检验的采样方法(Sampling Methods for Quality Control) .....	19
2-3-1 计量抽样检验(Sampling of Variables) .....	20
2-3-2 计数抽样检验(Sampling of Attributes) .....	22
习题 .....	24
参考文献 .....	25
<b>第三章 化学试验设计与优化方法(Chemical Experiment Design and</b>	

Optimization Methods) .....	26
§ 3-1 因子设计及其析因分析方法(Factorial Design and Its Rationale Analysis) .....	28
3-1-1 主效应的估价(Estimation of the Main Effect) .....	28
3-1-2 交叉效应的估价(Estimation of the Interactive Effect) .....	30
3-1-3 效应及残差正态图(Effect and Residual Normal Plot).....	33
§ 3-2 部分因子设计(Fractional Factorial Design) .....	35
3-2-1 半因子设计法(Half-Fraction Factorial Design) .....	36
3-2-2 半因子设计的产生方法及四分之一因子设计法(Generating Method of Half-Fraction and Quarter-Fraction Factorial Design) .....	41
§ 3-3 正交试验设计和正交设计表(Orthogonal Design and Orthogonal Arrays) .....	43
3-3-1 正交表及其交互效应表(Orthogonal Arrays and Their Interaction Tables) .....	44
3-3-2 正交设计表的线性图及其应用(Linear Graphs of Orthogonal Array and Their Applications) .....	46
§ 3-4 均匀试验设计及均匀设计表(Uniform Experimental Design and Uniform Design Table).....	47
3-4-1 均匀设计表的构造(Construction of Uniform Design Tables) .....	47
3-4-2 均匀性准则和使用表的产生(Criterion for Uniformity and Production of Accessory Tables).....	50
3-4-3 拟水平均匀设计(Uniform Design for Pseudo-Level) .....	53
§ 3-5 $D$ -最优试验设计( $D$ -Optimal Experimental Design) .....	55
§ 3-6 单纯形试验设计法(Simplex Experimental Design Method) .....	59
§ 3-7 化学中常用优化方法(Commonly Used Optimization Methods in Chemistry) .....	62
3-7-1 局部优化算法(Local Optimization Method) .....	64
3-7-2 全局优化算法(Global Optimization Methods) .....	65
习题 .....	73
参考文献 .....	74
<b>第四章 分析检测理论与信号处理方法(Analytical Detection Theory and Signal Treatment Methods) .....</b>	<b>77</b>
§ 4-1 分析方法的检测限(Detection Limit of Analytical Method) .....	77
§ 4-2 分析数据的统计检验方法(Statistic Test Methods of Analytical Data) .....	81
4-2-1 分析结果的 $t$ 检验( $t$ Test for Analytical Results) .....	81
4-2-2 分析结果的方差检验(Varisance Test for Analytical Results).....	82
§ 4-3 分析信号的平滑方法(Smoothing Methods of Analytical Signals).....	85



4-3-1 窗口移动平均法(Window Moving Average Method) .....	85
4-3-2 窗口移动多项式最小二乘平滑法(Window Moving Polynomial Least Squares Smoothing Method) .....	86
4-3-3 窗口移动中位数平滑法(Window Moving Median Smoothing Method) .....	88
§4-4 分析信号的求导方法(Derivative Methods of Analytical Signals) .....	91
4-4-1 简单差分法(Simple Difference Method) .....	91
4-4-2 窗口移动多项式最小二乘拟合法(Window Moving Polynomial Least Squares Fitting Method) .....	91
§4-5 分析信号的变换方法(Transformation Methods of Analytical Signals) .....	96
4-5-1 卷积运算的物理意义(Physical Meaning of Convolution Algorithm) .....	96
4-5-2 光谱多重性效益与 Hadamard 变换法(Multichannel Advantage in Spectroscopy and Hadamard Transformation) .....	98
4-5-3 傅里叶变换法及其用于分析信号处理(Fourier Transformation Method as Applied to Treatment of Analytical Signals) .....	100
4-5-4 小波多分辨变换法及其用于分析信号处理(Wavelet Multiresolution Transformation Method as Applied to Treatment of Analytical Signals) .....	107
4-5-5 小波变换在分析信号处理中的应用(Wavelet Transformation as Applied to the Treatment of Analytical Signals) .....	111
习题.....	113
参考文献.....	113
<b>第五章 多元校正与多元分辨(Multivariate Calibration and Multivariate Resolution) .....</b>	<b>116</b>
§5-1 白色分析体系的多元校正方法(Multivariate Calibration Methods for White Analytical Systems) .....	117
5-1-1 直接校正方法(Direct Calibration Methods) .....	117
5-1-2 间接校正方法(Indirect Calibration Methods) .....	122
5-1-3 通用标准加入法(Generalized Standard Addition Method) .....	127
5-1-4 人工神经网络法(Artificial Neural Network Method) .....	129
5-1-5 岭回归法(Ridge Regression) .....	135
§5-2 灰色分析体系的多元校正方法(Multivariate Calibration Methods for Grey Analytical Systems) .....	139
5-2-1 矢量校正方法(Vector Calibration Methods) .....	139
5-2-2 矩阵校正方法(Matrix Calibration Methods) .....	142
§5-3 黑色分析体系的多元分辨方法(Multivariate Resolution Methods for	

Black Analytical Systems) .....	151
5-3-1 基于主成分分析的体系组分数确定方法(Methods Based on PCA for Estimating Number of Chemical Components in System) .....	152
5-3-2 矩阵分辨方法(Matrix Resolution Methods) .....	157
5-3-3 张量分辨方法(Tensor Resolution Methods) .....	183
习题.....	185
参考文献.....	187
<b>第六章 化学模式识别(Chemical Pattern Recognition)</b> .....	191
§ 6-1 模式空间的几种距离与相似性度量(Several Measures of Distance and Similarity in Pattern Space) .....	191
§ 6-2 特征抽取方法(Feature Extraction Methods) .....	194
§ 6-3 模式识别的数据预处理方法(Pretreatment Methods for Pattern Recognition) .....	195
§ 6-4 有监督的模式识别方法:判别分析法(Supervised Pattern Recognition: Discrimination Methods).....	196
6-4-1 距离判别法(Distance Discrimination Method) .....	197
6-4-2 线性学习机(Linear Learning Machine) .....	198
6-4-3 $K$ -最近邻法( $K$ -Nearest Neighbors Discrimination Method) .....	199
6-4-4 势函数判别法(Potential Function Discrimination).....	200
6-4-5 人工神经网络判别法(Artificial Neural Networks) .....	202
§ 6-5 无监督的模式识别方法:聚类分析法(Unsupervised Pattern Recognition Methods:Clustering Analysis Methods) .....	203
6-5-1 最小生成树方法(Minimum Spanning Tree Method) .....	203
6-5-2 $K$ -均值聚类法( $K$ -Means Clustering Method).....	204
6-5-3 基于全局寻优的聚类法(Clustering Methods Based on Global Optimization) .....	206
§ 6-6 基于特征投影的降维显示方法(Pattern Recognition by Latent Projections) .....	210
6-6-1 基于主成分分析的投影显示法(Projection Discrimination Based on Principal Component Analysis) .....	210
6-6-2 基于主成分分析的 SIMCA 分类法(SIMCA Classification Based on Principal Component Analysis).....	214
6-6-3 基于偏最小二乘的投影显示法(Projection Discrimination Based on Partial Least Squares) .....	219
习题.....	226
参考文献.....	227
<b>第七章 计算机数字模拟法(Computer Numerical Simulation)</b> .....	229
§ 7-1 基于统计机理的 Monte Carlo 数字模拟法(Monte Carlo Numerical	

Simulation Based on Statistic Mechanism) .....	229
7-1-1 伪随机数的产生方法(Generating Methods of Quasi-Random Numbers) .....	230
7-1-2 化学动力学系统的 Monte Carlo 模拟算法(Monte Carlo Simulation Algorithm for Chemical Kinetic Systems) .....	232
7-1-3 误差分析的 Monte Carlo 模拟算法(Monte Carlo Simulation Algorithm for Error Analysis) .....	238
§7-2 基于微分方程数字解法的计算机模拟方法(Computer Digital Simulation Based on Numerical Integration of Differential Equations) .....	240
7-2-1 化学动力学系统的模拟算法(Modeling Algorithms for Chemical Kinetic Systems) .....	240
7-2-2 电化学体系的数字模拟方法(Digital Simulation Method for Electrochemical System) .....	244
习题 .....	248
参考文献 .....	248
<b>第八章 化学构效关系的研究方法(Research Methods for Chemical Structure-Activity and/or Structure-Property Relationships)</b> .....	<b>250</b>
§8-1 有机反应性相关分析方法简介(Brief Description of Correlation Analysis of Reactivity of Organic Compounds) .....	251
8-1-1 线性自由能及经典 Hammett $\sigma$ 常数(Linear Free-Energy and Classic Hammett $\sigma$ Constant) .....	251
8-1-2 取代基电子效应常数( $\sigma$ 常数)的应用与拓展(Applications and Its Extensions of Substituent Electronic Effect Constant) .....	252
8-1-3 取代基的立体效应常数(Steric Effect Parameters of Substituent) .....	255
8-1-4 取代基电子效应常数与立体参数的协同效应(Synergetic Effect of Electronic and Steric Constants) .....	256
§8-2 化合物结构表征方法(Expressing Methods of Chemical Structures) .....	257
8-2-1 Wiener 拓扑数(Wiener Topological Index) .....	258
8-2-2 Randic 分支指数和分子连接性指数(Randic Branching Index and Molecular Connectivity Index) .....	258
8-2-3 分子 ID 指数(Molecular Identification Number, MIDN) .....	265
8-2-4 苏尔兹分子拓扑指数(Schultz Molecular Topological Indices, MTI) .....	268
8-2-5 回归距离和及回归顶点价(Regressive Distance Sums, RDS,	

and Regressive Vertex Degrees, RVD) .....	273
§ 8-3 构效关系建模方法(Modeling Methods of Structure-Activity Relationship) .....	275
8-3-1 基于回归分析的建模方法(Modeling Methods Based on Regression) .....	275
8-3-2 基于模式识别的建模方法(Modeling Methods Based on Pattern Recognition) .....	277
8-3-3 基于人工神经网络的建模方法(Modeling Methods Based on Artificial Neural Networks) .....	282
习题 .....	283
参考文献 .....	283
<b>第九章 人工智能与化学专家系统方法(Artificial Intelligence and Chemical Expert Systems) .....</b>	<b>287</b>
§ 9-1 概述(Introduction) .....	287
§ 9-2 启发式分类与搜索方法(Heuristic Classification and Search Methods) .....	288
9-2-1 广度优先搜索(Breadth-First Search) .....	289
9-2-2 深度优先搜索(Depth-First Search) .....	289
9-2-3 启发式搜索方法(Heuristic Search Method) .....	290
§ 9-3 知识表达技术(Knowledge Expressing Methods) .....	291
9-3-1 逻辑表达方法(Logical Expressing Method) .....	291
9-3-2 语义网络表达法(Semantic Net Expressing Method) .....	292
9-3-3 产生式规则表达法(Geerating Rule Expressing Method) .....	292
§ 9-4 化学专家系统简介(Brief Description of Chemical Expert Systems) ..	294
9-4-1 DENDRAL 质谱、核磁共振谱图解析专家系统(DENDRAL Expert System for Elucidation of Mass and Nuclear Magnetic Resonance Spectra) .....	296
9-4-2 PLATO 数据解析专家系统(PLATO Data Analysis Expert System) .....	298
9-4-3 高效液相色谱专家系统(High Performance Liquid Chromatography Expert System) .....	300
9-4-4 ESESOC 有机化合物结构解析专家系统(ESESOC Expert System for the Elucidation of Structure of Organic Compounds) .....	301
习题 .....	302
参考文献 .....	302
<b>第十章 统计学和线性代数基础知识(Necessary Fundamental Knowledge of Statistics Linear Algebra) .....</b>	<b>304</b>
§ 10-1 必要的统计学基础知识(Necessary Fundamental Knowledge of	

---

Statistics) .....	304
10-1-1 随机事件的概率公式(Probability Formula of Random Events) .....	304
10-1-2 随机变量及其分布(Random Variable and Its Distributions) .....	307
10-1-3 随机变量的数值特征(Numerical Feature of Random Variable) .....	312
§ 10-2 必要的应用数学基础知识(Necessary Fundamental Knowledge of Applied Mathematics) .....	313
10-2-1 矢量及其运算(Vector and Its Calculation) .....	313
10-2-2 矩阵及其运算(Matrix and Its Calculation) .....	316
10-2-3 独立性,正交性和子空间(Independence, Orthogonality and Subspace).....	323
10-2-4 矢量范数和矩阵范数(F Number of Vector and Matrix).....	324
10-2-5 张量的概念(Concept of Tensor) .....	326
习题.....	326
参考文献.....	327

# 第一章 概 论

(Introduction)

## § 1-1 化学计量学的研究对象与发展概况

(Research Object and Developing  
Situation of Chemometrics)

二十多年前,瑞典人 Wold 提出“化学计量学”(chemometrics)<sup>[1-1]</sup>一词,他建议类比于生物计量学(biometrics)与经济计量学(econometrics),将研究从化学实验产生的数据中提取相关化学信息的学科分支称之为化学计量学。这一建议得到美国的 Kowalski 的响应。他们于 1974 年发起成立国际化学计量学会。化学计量学运用数学、统计学、计算机科学以及其他相关学科的理论与方法,优化化学量测过程,并从化学量测数据中最大限度地获取有用的化学信息,可以说是一门化学量测的基础理论与方法学<sup>[1-2]</sup>。

Howery<sup>[1-3]</sup>在 20 世纪 80 年代初叙述化学计量学发展历史时,将它划分为几个阶段。

60 年代及之前是第一阶段。这实际上就是 chemometrics 一词出现之前的时期。在这个时期,许多统计学与数学方法在化学与分析化学中逐步获得了应用。实际上,传统的“分析化学中的数理统计方法”课程或如前述 70 年代及之前发表的有关综述正是较集中地反映了这一发展阶段累聚的成果。颇耐人寻味的是 Howery 对这一发展阶段的评论,他认为(分析)化学家在这一发展时期尽管表面上十分重视数据分析,但基本上停留在一些描述型的统计量计算上,如均值、标准差、置信区间的计算等。相比之下,形为科学、工程科学等领域在应用数理统计方法方面取得的成果更大。Howery 惊叹数据分析方法在一些难于作定量分析的领域反而出人意外地被接受。如后来在化学计量学发展中成为骨干技术的因子分析、模式识别方法,最早是相应由心理学家及工程科学家构建运用的。在化学领域,Howery 比较赞赏物理有机化学家在这一时期作出的贡献,即线性自由能关系的研究,将大量有机化学数据与分子的特性进行了关联。这实际是后来迅猛发展的化学计量学定量构效关系(quantitative

structure-activity relationship, QSAR)研究的前导。

在随后的第二发展阶段,即 chemometrics 一词诞生的 70 年代,化学家、分析化学家不但将统计学、数学乃至包括行为科学、经济计量学以及诸多工程学科等相邻领域已发展的数据与信号分析方法在内的许多方法用于化学研究,而且根据化学科学的特定要求发展了许多新的数据与信号分析方法。化学计量学发展成为化学、分析化学学科的一个独特分支。两个重要的条件与因素推动了这方面发展。首先,化学与分析化学中大量涌现的现代化学量测仪器,使化学与分析化学家比以往任何时候都更容易获得大量化学量测数据。这种情况,在过去是难以想像的。如果我们追溯到经典分析化学发展的早期,则更是不可思议。像 1801 年出版的 Lampadius 著的《无机物分析大全》一书,作者在书中这样写道:“谁要是缺乏耐心等待几个星期或几个月以取得分析结果,他就根本不必去开始一项分析工作”!到 20 世纪 70 年代,在分析测试或化学量测中,人们第一次发现,取得数据甚至大量数据已不是最困难的一步。最难解决的瓶颈问题是这些数据的解析及如何从中提取所需的有用化学信息。化学家与分析化学家首次遇到类似行为科学家或经济学家所遇到的大量数据如何处理的问题。化学家与分析化学家比较幸运。因为大量现代分析测试仪器出现带来“数据爆炸时代”,也正是计算机普及的时代。这就构成了化学计量学发展的第二个条件:为了对极为复杂的化学量测数据(其中负载着在分子水平上表征物质世界的信息)进行解析,化学家、分析化学家有可能利用在计算机上实现的许多强有力的数学方法,包括一些相关学科发展的数据与信号处理新方法,从多维化学量测数据中提取有用的相关信息。

Howery 在这篇成稿于 80 年代初的文章中预测了化学计量学下二个发展阶段:从 80 年代起化学计量学进入化学课堂并在大学生及研究生中普及的阶段;以及化学计量学真正成为化学家手中的常规手段的下一发展阶段。80 及 90 年代的发展说明,这些预测是大致准确的。

## § 1-2 化学计量学——化学量测的 基础理论与方法学

(Chemometrics: the Fundamental Theory and  
Methodology of Chemical Measurements)

化学计量学以化学量测的基础理论与方法学为研究对象。根据 Valcarcel<sup>[1-4]</sup>建议的分析化学作为计量科学的定义,指出分析化学的任

务是发展、优化、应用量测过程,以获取全局或局部性的化学品质信息,解决所提出的量测课题。从这一定义可更清楚地看出化学计量学与作为化学信息科学的分析化学之间的特殊关系。化学计量学所涉及的问题,很多是分析化学的基础性问题,可以说它构成了分析化学第二层次的基础理论的重要组成部分<sup>[1-5]</sup>。

作为化学量测的基础理论与方法学,化学计量学正在向各个必须进行化学量测的分支学科渗透,包括环境化学<sup>[1-6]</sup>、食品化学<sup>[1-7-1-9]</sup>、农业化学<sup>[1-10]</sup>、医药化学<sup>[1-11]</sup>等以及化学工程学科<sup>[1-12-14]</sup>,而其中向化学工程学科的渗透尤为引人注目。

化学计量学为化学量测提供理论和方法,所以,它的发展主要表现在下述两个方面:①发展化学数据解析的新理论和方法;②化学计量学解析方法在各个化学分支学科的新应用研究。近年来,计算机科学、统计学、应用数学及信息科学皆得到长足发展,它们的发展为化学计量学注入了新鲜血液,如各类人工神经网络(artificial neural networks)新技术、基于自然计算的全局最优算法如模拟退火(simulated annealing)和遗传算法(genetic algorithm)、信息科学中的小波分析(wavelet analysis)及图像分析(image analysis)方法及近年来统计学中研究热烈的稳健方法(robust methods)等,都引起了化学计量学家的浓厚兴趣,所以,90年代以来大量的化学计量学方法研究很多都集中于上述领域<sup>[1-15-1-16]</sup>。当然,对于化学计量学特有的且已得到深入应用的多元校正和多元分辨及化学模式识别等方法,如偏最小二乘法、SIMCA 分类法、秩消失因子分析法、渐近因子分析方法等,在新方法的理论和算法研究上也得到了长足发展,为解决化学领域,其中特别是分析化学领域的难点问题和基础理论开辟了崭新的通路<sup>[1-17-1-21]</sup>。另一方面,化学计量学在各化学分支学科的应用研究也取得很多重要成果,例如在环境化学、食品化学、农业化学、医药化学、石油化学等及化学工程学科就得到了相当广泛而深入的应用,为环境化学中的污染源识别、环境质量预测<sup>[1-22-1-26]</sup>;食品、农业、医药化学中试验设计和复杂样品分析<sup>[1-27-1-29]</sup>;医药化学中的分子设计、新药发现及结构性能关系(QSAR)研究<sup>[1-30-1-32]</sup>;石油化学中的化学模式识别、波谱与物质特性的关系<sup>[1-33-1-36]</sup>;化学工程学科中的过程分析、工艺过程诊断、控制和优化都提供了新思路和新方法<sup>[1-37-1-39]</sup>,使得化学计量学在这些领域中得到广泛认同。

化学计量学诞生至今,已有二十余年历史,其发展前景亦是一个令人关注的问题。从前述分析化学与化学计量学之关系可以看出,化学计量学的发展已对分析化学产生了深刻的影响,构成了分析化学第二层次基



基础理论和方法学的重要组成部分,特别值得提出的是,化学计量学的发展还将为分析仪器的智能化提供新理论和新方法,为新型高维联用仪器的构建提供新思路和新方法,是21世纪分析仪器软件主体化发展的新突破口。此外,随着微型计算机的飞速发展,对于化学波谱库的建立与检索及化学人工智能和专家系统研究也取得长足进步。在采用计算机网络技术将多种波谱仪器连接的基础上,将数值化计算技术(近年来化学计量学方法学发展的主体)与基于经验的逻辑推理方法的有机结合,可望解决化合物结构自动解析的难题,并使得长期困扰分析化学家的混合物波谱同时定性定量解析成为可能。在分析化学领域中,化学计量学的发展前景十分诱人。当然,化学结构解析、定量构效关系研究、合成设计、过程优化等问题对于无机化学、有机化学和物理化学等各门基础化学分支学科都是十分重要的。另外,化学计量学与其他化学分支学科,如环境化学、食品化学、农业化学、医药化学、化学工程学科等,也正在产生更密切的联系,得到更广泛的应用。Lavine在最近美国两年一度的“化学计量学”评论中指出:“统计过程控制已成为这一评论时期的热点应用领域,组合化学将期望成为下一个热点应用领域。因为在组合化学中面临的数据结构和形式将不同于其他领域,新的数据解析方法将应运而生”<sup>[1-16]</sup>。随着各化学分支学科的发展,可以预期,化学计量学也将继续得到更蓬勃的发展。

### § 1-3 本书各章内容简介

#### (Brief Description of Content of Each Chapter)

全书共分为十章,是依照化学量测的全过程,从采样开始,经化学量测的试验设计、信号预处理、定性定量分析的多元校正和多元分辨、再到有用决策信息的提取,包括化学模式识别、机理研究的数字模拟方法、化学构效关系研究直至人工智能与化学专家系统来进行编排的,试图覆盖化学计量学的基本内容。本书将主要以基本概念、基本方法为主线来介绍化学计量学。在讨论了化学计量学的基本概念和基本方法的基础上,尽量同时给出相应的计算机程序和参考文献,使读者通过阅读此书后即可根据本书所提供的方法进行试验,并可进一步将其运用于实际问题的解决。

第一章主要讨论化学计量学的基本定义,发展历史和概要情况。通过这章的学习,力求学生对化学计量学全貌有一个大致了解。

第二章讨论分析化学中采样理论和方法。采样理论是指如何进行试样采集的数学统计理论。本章对常用的采样理论和方法分别加以介绍,