

科学版研究生教学丛书

# 化学计量学

统计学与计算机在  
分析化学中的应用

〔德〕M. 奥托 著

邵学广 蔡文生 徐筱杰 译



科学出版社

[www.sciencep.com](http://www.sciencep.com)

图字:01-2001-3929 号

## 内 容 简 介

本书全面讲解了化学计量学的基本知识与应用,既包括统计学方法、平滑、滤波、回归校正等传统内容,又包括近期发展起来的遗传算法、人工神经网络、模糊理论等新方法。在各项内容的讲述中,作者强调了对基本概念和基础理论的描述并配有应用实例,而对于复杂的数学推导进行了省略或简化。此外,书中还配有大量的例题与习题,便于读者理解和掌握。

本书适合作为化学、药学、生物化学或生态学专业的研究生教材,也可供相关从业人员参考。

Originally published in the English language by WILEY-VCH Verlag GmbH, Pappelallee 3, D-69469 Weinheim, Federal Republic of Germany, under the title "Otto: Chemometrics". Copyright 1998 by WILEY-VCH Verlag GmbH.

### 图书在版编目(CIP)数据

化学计量学:统计学与计算机在分析化学中的应用/(德)奥托(Otto)著;  
邵学广,蔡文生,徐筱杰译. —北京:科学出版社,2003  
(科学版研究生教学丛书)

ISBN 7-03-010646-6

I. 化… II. ①奥…②邵…③蔡…④徐… III. 化学计量学 IV. O6-04

中国版本图书馆 CIP 数据核字(2002)第 054763 号

责任编辑:刘俊来 王志欣/责任校对:宋玲玲

责任印制:安春生/封面设计:曹 烨

科学出版社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencecp.com>

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

\*

2003年1月第一版 开本:B5(720×1000)

2003年1月第一次印刷 印张:17 1/2

印数:1—3 000 字数:333 000

定价:39.00 元

(如有印装质量问题,我社负责调换〈北燕〉)

## 译者的话

化学计量学是建立在多学科基础上的一门新兴学科,已成为化学学科的一个重要分支。“化学计量学(Chemometrics)”一词由瑞典科学家 S. Wold 教授首次于 20 世纪 70 年代初提出,并与美国科学家 B. R. Kowalski 教授一起于 1974 年在美国西雅图成立了国际化学计量学学会。化学计量学的初次定义为:“化学计量学是一门化学分支学科,它应用数学和统计学方法(借助计算机技术),设计和选择最优的测量程序和实验方法,并且通过解释化学数据获得最大限度的信息。在分析化学领域中,化学计量学通过应用数学和统计学方法,用最佳的方式获取关于物质系统的有关信息。”可见,化学计量学的基本任务是:应用和发展统计学方法及其他数学方法进行实验设计,并从实验测量数据中获取有用的化学信息。

在化学计量学的发展过程中,特别是 20 世纪 90 年代以后,化学计量学的内涵也得到了发展。S. Wold 曾重新界定了化学计量学的基本涵义,即“从化学测量数据中获取、表述、显示相关化学信息。”并指出:“化学计量学的主要特征是把化学问题构造成可以通过数学关系表达的数学模型。”S. D. Brown 则认为:“化学计量学是定量测量化学与应用统计学的一个交叉领域,通过数学与统计学方法从化学数据中提取信息。”总之,化学计量学是化学与计算机科学、数学、统计学的接口,它运用计算机上实现的数学与统计学方法,优化测量程序,并从化学测量数据中最大限度地提取有用的化学信息。

化学计量学对分析化学的发展具有非常重要的意义。利用化学计量学方法可从大量的实验数据中最大限度地提取有用信息,实现分析工作由“数据提供者”到“问题解决者”的飞跃。化学计量学已在许多方面取得了成功的应用,如它能帮助化学工作者确定最优的实验方案,提高分析测试的精密度、准确度、灵敏度以及选择性。现代信号处理技术的应用,使分析测试发生了深刻的变革。化学计量学还促进了分析仪器的自动化及智能化。另外,化学计量学还可帮助化学家发展新的测量方法。

经过 20 多年的发展,化学计量学的研究内容已相当广泛,如统计学方法、最优化方法、信号处理、因子分析、曲线分辨、数据校正、模型化与参数估计、结构与活性相关、数据库及其应用、模式识别、人工智能等等。新的化学计量学方法也不断出现,如小波分析、免疫算法等。化学计量学在分析化学中的应用也日趋广泛,已涉及到分析化学的各个领域,如光谱分析、电化学分析、色谱分析、质谱分析以及各种联用分析技术。

*Chemometrics: Statistics and Computer Application in Analytical Chemistry* 一书是 M. Otto 长期从事化学计量学教学和科研工作经验的结晶。该书的突出特点是全面、新颖、简捷。它非常全面地讲解了化学计量学的基本内容,既包括了传统的统计学方法、平滑滤噪、回归校正等化学计量学内容,又包括了近期发展起来的遗传算法、人工神经网络、模糊理论等新型方法。在各项内容的讲述中,强调了对基本概念和基础知识的描述并配有应用实例,而对于难点问题和复杂的数学推导进行了省略或简化。因此,该书对于自学的读者具有特别重要的意义。由于作者具有长期从事化学计量学方面的科研和教学经验,该书的章节内容及结构安排合理、难易程度适宜。同时,该书配有参考读物(general reading)、例题(examples)和习题(questions and problems),特别适合作为高等院校的化学计量学教材。另外,该书具有较广泛的参考价值,可作为化学、生物、药学、地质等相关学科研究人员的参考书。

本书的翻译工作由北京大学徐筱杰教授推荐,科学出版社委托。第 1、2、5、8、9 章主要由中国科学技术大学邵学广教授完成,第 3、6、7 章主要由中国科学技术大学蔡文生教授完成,第 4 章主要由北京大学徐筱杰教授完成。本书在翻译过程中得到了中国科学技术大学化学系化学信息学实验室诸位研究生的帮助,在此一并表示衷心的感谢!

原文中存在少量错误,译者在翻译过程中做了更正和注释。另外,在个别章节的翻译中译者根据国内读者的情况增加了少量帮助性的注释。

由于我们的水平所限,书中欠妥或错误之处在所难免,敬请专家和读者批评指正。

译 者

2002 年 4 月

## 序

化学计量学可以定义为“应用数学和统计学技术进行化学数据处理”。这是一个广义的定义,但它非常恰当,因为化学计量学的研究领域非常广泛。

20 世纪 60 年代中期,计算机化的分析仪器开始为化学分析提供大量的数据,给分析化学带来了所谓的“数据爆炸”,造成了由于数据的富有而窘迫的局面。过去,化学工作者仅依靠少量昂贵且难得的数据片段即可得出结论,但现在他们不得不面对大量便宜且又容易得到的冗长数据才能得出结论。在对这些数据进行分类并提取有用相关信息的过程中,化学工作者开始利用其他领域,特别是社会科学领域的数学和统计学技术,并发现这些技术非常有效,可以得到所需要的信息。随后,化学工作者又发现这些数学和统计学技术可以使他们得到比预期的还多的信息。这使化学工作者开始在这些数据中探索意想不到的内部结构。

20 世纪 60 年代末,我在普渡大学读研究生的时候,参加了 Buck Roger 每周一次的分析化学讨论班,听到了 Crawford 和 Morrison 关于根据质谱数据进行分类的讨论班汇报。他们将质谱数据投影到一个几何超球体表面上,充分利用了相似化合物之间微小的观测距离和不相似化合物之间较大的观测距离。这次讨论班给了我一个极棒的经历,使我知道化学数据中存在内部结构。这可能是非常有用的。

不久,Bruce Kowalski 访问了普渡并举行了关于线性学习机应用的讨论班。更加深了我对这些知之甚少的数学和统计学技术重要性的认识。这时,Svante Wold 提出了“化学计量学(Chemometrics)”这一术语,使这个领域有了自己的名字。这个名字使该领域进入了一个兴旺发达的时期。许多化学工作者,特别是分析化学工作者,开始将他们并不十分清楚的数学和统计学技术(化学计量学方法)应用于他们的数据处理并得到了令人惊讶的信息。实验设计也迅速成为有用的化学计量学工具,使所设计的实验非常有效,既能得到所需的信息又能保证最小的资源开支。

在化学计量学发展的初级阶段,我们这些化学计量学工作者利用并尝试了可以得到的任何东西,如主成分分析、序贯单纯形优化、正则相关分析、 $k$ -最近邻法等等。在很多情况下,我们给这些方法找到了合适的应用,实现了数学和统计学技术与某些化学数据之间的良好配合。也有某些情况,在我看来,这种配合并不太好,正像我们努力将一条方木桩插进一个圆洞一样。但我们做得非常带劲!因为各种各样的理由,我们有时对这些边缘的方法吹嘘过度,宣传过多,但使用者发现这些方法并不能很好地解决他们的实际问题。失望使化学计量学失去了光彩。这是化

学计量学的羞辱,如果使用适当的话,化学计量学方法是非常强大和有用的。

目前,化学计量学的发展已接近成熟,失去光彩的年代已经过去。老的技术正在被适当地应用,新的数学和统计学技术正在为解决特定的化学问题而发展,结果经常令人惊骇。事实证明,化学计量学是一个值得研究的领域。

阅读 Matthias Otto 的《化学计量学》一书令人振奋。在这里,在同一本书中,有大多数化学计量学方法的清晰描述并配有相关实例,也有化学计量学的发展远景。这里没有天花乱坠的吹嘘,只有对化学计量学方法清晰、合理、有根据的描述以及它们的最佳使用方法。

Otto 是一位有实际经验的化学计量学工作者、勤奋的学者和杰出的教师。他的风格新颖并且清晰。因此,我怀着极大的热情推荐这本有价值的书。

Stanley N. Deming

休斯敦大学化学系

1998 年 8 月

## 前 言

撰写一本化学计量学教科书的想法源于大学生和研究生的教学工作。在佛莱贝格(Freiberg)矿业技术大学,我曾主讲和指导化学计量学方面的上机训练 10 余年,并且在维也纳技术大学等其他研究机构做访问教授时也有类似的教学经验,如基于计算机的分析化学(COBAC)。

从那些教育经历,我了解到学生对于化学计量学充满热情。目前,几乎不用担心计算机训练初期的困难,因为很多学生进入大学时或多或少都有些计算机嗜好。应注意的问题在于化学测量结果的评价,因为这需要很多的统计数学知识。不幸的是,在大多数国家,化学工作者所受到的基础统计学和基础数学教育与物理工作者相比相对较弱。

因此,在我的化学计量学教科书中,讲解了对于化学,特别是分析化学测量结果评价非常重要的统计数学主题,它将适合于实验观测的评价而不是理论化学。

本书分为 9 章。第 1 章对化学计量学学科及其应用领域进行了介绍。第 2 章提供了描述化学数据和进行统计检验所需要的统计学基础知识。第 3 章的主题是数字滤波和时间序列表征中的信号处理方法。第 4 章讲述了基于实验设计和优化的高效实验设计方法,概括地说明了这些方法可以同样很好地用于化学合成、分析步骤或药物制备等过程的优化。第 5 章介绍了模式识别方法以及分类问题中的数据归属方法。该部分由无监督学习和有监督学习构成,在介绍了数据预处理方法之后,概述了用于多维数据分析的代表性化学计量学方法。第 6 章介绍了包括直线回归、多元回归及非线性回归在内的关系建模方法。第 7 章讨论了分析化学数据库,即化学结构和光谱的计算机可操作的表示方法,包括 LIMS(实验室信息与管理系统)的使用。第 8 章考虑了化学计量学的最新进展,除人工智能的基础知识外,还讨论了专家系统、神经网络、模糊理论以及遗传算法的应用。作为统计学方法在化学实验室中应用的热门主题,第 9 章包括了用于实验室内部和实验室之间的质量保证、质量验证、可靠性认证以及实验室管理规范等方面的最重要方法。

在附录中,读者可以找到统计学表格、推荐的软件以及线性代数的简介。利用本书每章开头的学习目的、大约 60 个工作实例以及每章末的练习题,学习化学计量学方法应用应该是比较容易的。

本教科书不仅为化学课程中的化学计量学教学,而且为化学、药学、矿物学、地质学、生物学以及相关学科工作者的自学而编写。本书对于工业领域中的同仁也有参考价值。同时,运行近红外光谱仪、质量保证中应用统计学检验或研究定量结

构-活性关系(QSAR)等也可以参考本书的某些内容,如多变量方法。

同样,没有化学计量学同行的意见和建议,本书是不可能完成的。我开始从事计算机在分析化学中的应用源于多年前 Heinz Zwanziger 的帮助,后来的许多化学计量学课程是与 Wolfhard Wegscheider 一起进行的,借此机会谨表示衷心的感谢。

Matthias Otto

1997 年 2 月

## 英文版前言

我非常高兴,在本书的德文版面世不久,英文版也完成了。因此,我有了一个改正某些错误和增加某些方法(如 Kalman 滤波或 Bayesian 判别等)的机会。另外,我还在英文版中增加了学习目的以及练习题。

Matthias Otto

1998 年 6 月

## 缩 写

|                  |                                                                |              |
|------------------|----------------------------------------------------------------|--------------|
| ACE              | Alternating Conditional Expectations                           | 交错条件期望法      |
| ADC <sup>①</sup> | Analog-to-Digital Converter                                    | 模-数转换器       |
| ANOVA            | Analysis of Variance                                           | 方差分析         |
| CPU              | Central Processing Unit                                        | 中心处理单元       |
| CRF              | Chromatographic Response Function                              | 色谱响应函数       |
| DAC              | Digital-to-Analog Converter                                    | 数-模转换器       |
| EPA              | Environmental Protection Agency                                | 环保局          |
| FA               | Factor Analysis                                                | 因子分析         |
| FFT              | Fast Fourier Transformation                                    | 快速傅里叶变换      |
| FHT              | Fast Hadamard Transformation                                   | 快速哈达玛变换      |
| FT               | Fourier-Transformation                                         | 傅里叶变换        |
| GC               | Gas Chromatography                                             | 气相色谱         |
| HORD             | Hierarchically Ordered Ring Description                        | 多层有序环描述      |
| HOSE             | Hierarchically Ordered Spherical Description<br>of Environment | 多层有序球形环境描述   |
| HT               | Hadamard Transformation                                        | 哈达玛变换        |
| I/O              | Input/Output                                                   | 输入/输出        |
| IND              | Indicator function                                             | 因子指示函数       |
| IR               | Infra Red                                                      | 红外           |
| ISO              | International Organization for Standardiza-<br>tion            | 国际标准组织       |
| JCAMP            | Joint Committee on Atomic and Molecular<br>Data                | 原子与分子数据联合委员会 |
| KNN              | $k$ -nearest neighbor method                                   | $k$ -最近邻法    |
| LAN              | Local Area Network                                             | 局域网          |
| LDA              | Linear Discriminant Analysis                                   | 线性判别分析       |
| LIMS             | Laboratory-Information-and-Management-<br>System               | 实验室信息与管理信息系统 |
| LISP             | List Processing Language                                       | LISP 语言      |
| LLM              | Linear Learning Machine                                        | 线性学习机        |
| MANOVA           | Multidimensional ANOVA                                         | 多维方差分析       |

① 原文误为 ADU。——译者注

|        |                                                |                       |
|--------|------------------------------------------------|-----------------------|
| MARS   | Multivariate Adaptive Regression Splines       | 多变量自适应回归样条函数          |
| MS     | Mass Spectrometry                              | 质谱                    |
| MSDC   | Mass Spectrometry Data Center                  | 质谱数据中心                |
| MSS    | Mean Sum of Squares                            | 均方和                   |
| NIPALS | Nonlinear Iterative Partial Least Squares      | 非线性迭代偏最小二乘            |
| NIR    | Near Infra Red                                 | 近红外                   |
| NIST   | National Institute of Standards and Technology | 国家标准与技术研究所            |
| NLR    | Nonlinear Regression                           | 非线性回归                 |
| NMR    | Nuclear Magnetic Resonance                     | 核磁共振                  |
| NPLS   | Nonlinear Partial Least Squares                | 非线性偏最小二乘              |
| OLS    | Ordinary Least Squares                         | 常规最小二乘                |
| PCA    | Principal Component Analysis                   | 主成分分析                 |
| PCR    | Principal Component Regression                 | 主成分回归                 |
| PLS    | Partial Least Squares                          | 偏最小二乘                 |
| PRESS  | Predictive Residual Sum of Squares             | 预测残差平方和               |
| PROLOG | Programming in Logic                           | PROLOG 语言, 逻辑编程       |
| RAM    | Random Access Memory                           | 随机存取存储器               |
| RDA    | Regularized Discriminant Analysis              | 正则判别分析                |
| RE     | Real Error                                     | 真实误差                  |
| ROM    | Read Only Memory                               | 只读存储器                 |
| RR     | Recovery Rate                                  | 回收率                   |
| RSD    | Relative Standard Deviation                    | 相对标准偏差                |
| RSM    | Response-Surface Method                        | 响应曲面法                 |
| SEC    | Standard Error of Calibration                  | 校正标准误差                |
| SEP    | Standard Error of Prediction                   | 预测标准误差                |
| SIMCA  | Soft Independent Modeling of Class Analogies   | 软独立建模分类法 <sup>①</sup> |
| SS     | Sum of Squares                                 | 平方和                   |
| SVD    | Singular Value Decomposition                   | 奇异值分解                 |
| TTFA   | Target-Transformation Factor Analysis          | 目标变换因子分析              |
| UV     | Ultraviolet                                    | 紫外                    |
| VIS    | Visible                                        | 可见                    |

<sup>①</sup> SIMCA 也可译为簇类的独立软模式, 有时也理解为 Simple Classification Algorithm(简单分类法)或 Statistical Isolinear Multiple Component Analysis(统计均线性多元分析)。——译者注

## 符号说明

|          |                                  |           |                 |
|----------|----------------------------------|-----------|-----------------|
| $\alpha$ | 显著性(临界)水平, 分离因子                  | $J$       | 雅可比矩阵           |
| $A$      | 面积                               | $k$       | 峰度, 峭度          |
| $b$      | 宽度, 回归系数                         | $k'$      | 容量因子            |
| $A^2$    | $\chi^2$ -分布的分位数                 | $K_A$     | (酸)的电离常数        |
| cov      | 协方差                              | $\lambda$ | 特征值, Poisson 参数 |
| $C$      | 方差-协方差矩阵                         | $L$       | 载荷矩阵            |
| $\delta$ | 误差                               | $\mu$     | 群体平均            |
| $d$      | 距离, Kolmogorov-Smirnov 检验<br>统计量 | $m$       | 隶属函数            |
| $D$      | 差异检验统计量                          | $m_r$     | 分布的矩            |
| $D_k$    | Cook 检验统计量                       | $nf$      | 近邻函数            |
| $\eta$   | 学习速率常数                           | $N$       | 分析分离度, 塔板数      |
| $e$      | 误差                               | $N(\nu)$  | 噪声              |
| $E(.)$   | 期望值                              | $P$       | 概率              |
| $E$      | 残差矩阵                             | $Q$       | 四分位数, Dixon 统计量 |
| $f$      | 自由度, 函数                          | $r$       | 相关系数, 半径        |
| $f(x)$   | 概率密度函数                           | $R$       | 极差区间            |
| $F$      | Fisher 分布的分位数                    | $R$       | 相关矩阵            |
| $F(.)$   | 频率域中的函数                          | $R_s$     | 色谱分离度           |
| $f(t)$   | 时间域中的函数                          | $R^2$     | 决定系数            |
| $F$      | 得分矩阵                             | $\sigma$  | 标准偏差            |
| $G$      | 几何平均                             | $s$       | 标准偏差估计, 偏度      |
| $G(.)$   | 频率域的平滑函数                         | $s_r$     | 相对标准偏差估计        |
| $g(t)$   | 时间域的平滑函数                         | $S$       | 相似形量度           |
| $h$      | 公因子方差                            | $\tau$    | 时间延迟量           |
| $H$      | 调和平均                             | $t$       | 学生分布的分位数        |
| $H$      | 帽子矩阵, 哈达玛变换矩阵                    | $T$       | Grubbs 检验的检验量   |
| $H_0$    | 虚假设( $H_0$ 假设)                   | $T$       | 主成分得分矩阵, 变换矩阵   |
| $H_1$    | $H_0$ 假设的互补假设( $H_1$ 假设)         | $U$       | 左特征向量矩阵         |
| $H(.)$   | 频率域的滤波函数                         | $V$       | 右特征向量矩阵         |
| $h(t)$   | 时间域的滤波函数                         | $w$       | 奇异值, 权重         |
| $I$      | 单位矩阵                             | $x$       | (自)变量           |
| $I_{50}$ | 四分位数极差区间                         | $x$       | 自变量向量           |
|          |                                  | $X$       | 自变量矩阵           |

---

|           |            |     |                 |
|-----------|------------|-----|-----------------|
| $\bar{x}$ | 数值平均       | $z$ | 标准正态分布偏移量, 信号位置 |
| $y$       | (因)变量      |     | 置, 目标函数         |
| $y^*$     | 变换后(滤波后)的值 |     |                 |

# 目 录

|                      |            |
|----------------------|------------|
| <b>1 什么是化学计量学</b>    | <b>1</b>   |
| 1.1 基于计算机的实验室        | 2          |
| 1.2 统计与数据解释          | 8          |
| 1.3 基于计算机的信息系统与人工智能  | 9          |
| 参考读物                 | 9          |
| 习题                   | 10         |
| <b>2 基础统计学</b>       | <b>11</b>  |
| 2.1 描述统计学            | 11         |
| 2.2 统计检验             | 22         |
| 2.3 方差分析             | 36         |
| 参考读物                 | 42         |
| 习题                   | 42         |
| <b>3 信号处理与时间序列分析</b> | <b>44</b>  |
| 3.1 信号处理             | 44         |
| 3.2 时间序列分析           | 62         |
| 参考读物                 | 68         |
| 习题                   | 69         |
| <b>4 优化与实验设计</b>     | <b>70</b>  |
| 4.1 目标函数与因子          | 71         |
| 4.2 实验设计与响应曲面法       | 77         |
| 4.3 序贯优化:单纯形法        | 95         |
| 参考读物                 | 100        |
| 习题                   | 100        |
| <b>5 模式识别与分类</b>     | <b>101</b> |
| 5.1 数据预处理            | 103        |
| 5.2 无监督方法            | 105        |
| 5.3 有监督方法            | 138        |
| 参考读物                 | 148        |
| 习题                   | 148        |
| <b>6 建模</b>          | <b>150</b> |

|     |                         |     |
|-----|-------------------------|-----|
| 6.1 | 单变量线性回归·····            | 151 |
| 6.2 | 多元线性回归·····             | 164 |
| 6.3 | 非线性方法·····              | 183 |
|     | 参考读物·····               | 191 |
|     | 习题·····                 | 192 |
| 7   | 分析化学数据库·····            | 193 |
| 7.1 | 分析化学信息的表示·····          | 194 |
| 7.2 | 数据库检索·····              | 201 |
| 7.3 | 光谱模拟·····               | 206 |
|     | 参考读物·····               | 207 |
|     | 习题·····                 | 207 |
| 8   | 知识处理与软计算·····           | 208 |
| 8.1 | 人工智能与专家系统·····          | 208 |
| 8.2 | 神经网络·····               | 216 |
| 8.3 | 模糊理论·····               | 228 |
| 8.4 | 遗传算法·····               | 236 |
|     | 参考读物·····               | 240 |
|     | 习题·····                 | 241 |
| 9   | 质量保证与实验室管理规范·····       | 242 |
| 9.1 | 验证与质量控制·····            | 242 |
| 9.2 | 可靠性认证与实验室管理规范(GLP)····· | 246 |
|     | 参考读物·····               | 247 |
|     | 习题·····                 | 247 |
|     | 附录·····                 | 248 |
|     | 附录1 统计分布·····           | 248 |
|     | 附录2 数字滤波器·····          | 256 |
|     | 附录3 实验设计·····           | 258 |
|     | 附录4 矩阵代数·····           | 262 |
|     | 附录5 软件·····             | 264 |

# 1 什么是化学计量学

## 学习目的

- 定义化学计量学
- 学习如何利用二进制位(bits)进行计数以及如何利用计算机进行算术和逻辑运算
- 了解主要的计算机术语以及机器人和自动化的含义

化学计量学学科的发展与计算机在化学中的应用有很大关系。早在 20 世纪 70 年代一些分析化学研究组就开始利用统计学和数学方法,即现在的化学计量学方法,其研究与大型计算机的使用有直接关系。

化学计量学这一术语由瑞典的 Svante Wold 和美国的 Bruce R. Kowalski 于 1972 年提出,1974 年国际化学计量学协会的成立第一次对这一学科进行了描述。此后,又举行了一系列的国际会议,如 COMPANA (computer application in analytics), COBAC (computer-based analytical chemistry) 以及 CAC (chemometrics in analytical chemistry),一些学术期刊也为化学计量学主题开辟了专栏。后来,出现了新的化学计量学专业期刊,如 *Journal of Chemometrics* (Wiley) 和 *Chemometrics and Intelligent Laboratory Systems* (Elsevier)。

化学计量学目前的定义为:

化学计量学是一个化学分支,它利用数学和统计学方法进行设计和选择最优的测量程序和实验方案,并通过对化学数据的分析提供最大限度的化学信息。

化学计量学学科产生于化学领域,化学计量学方法的典型应用是发展定量的结构-活性关系或评价化学分析的数据。分析化学工作者开展化学计量学方法应用研究的一个重要原因是现代分析仪器产生大量数据。在分析化学中,化学计量学利用数学和统计学方法获取物质系统的有关信息。

20 世纪 80 年代初,随着个人计算机的普及,化学数据的获得、处理和解释进入了一个新的时代。事实上,目前科学家使用的软件要么与数学方法有关,要么用于知识的处理。科学家的需要使这些方法得到了更深入的理解。

化学工作者在数学和统计学方面所受到的教育往往不能令人满意,因此,从化学计量学的开始阶段,其目的之一就是如何使复杂的数学方法在实际工作中变得

容易使用。同时,统计和数字计算的商业软件也起到了一定作用,使重要的化学计量学方法可以通过适当的计算机演示进行教学。

除统计数学方法以外,化学计量学还包括实验室的计算机化问题、化学或光谱数据库的处理方法以及人工智能方法等。

另外,化学计量学家致力于所有这些方法的发展,但通常这些方法往往是由于某种特殊需要而得到发展,如色谱分离条件的优化或化合物生物活性的预测。

## 1.1 基于计算机的实验室

当前,计算机是科学研究必不可少的工具之一,计算机与分析仪器相连接进行数据采集、文字处理、数据库操作以及质量保证等。另外,计算机是现代通讯技术的基础,如 E-mail、电视会议等。为了对重要的计算机原理有所了解,这里包括了某些基础知识,如数字信息的编码与处理、计算机的基本构成、计算机编程语言以及自动化过程等。

### 1.1.1 模拟信号与数字信号

与使用模拟信号相比,使用数字信号有一些优势。数字信号对噪声不敏感,唯一的噪声来源是由于表示数字的位数限制而带来的舍入误差。这种误差远比电子干扰要小,并且与数字计算机兼容。

通常,原始信号要么以连续的模拟信号方式产生,要么以不连续的数字信号方式产生(如图 1-1)。例如,用光电池监测光的辐射强度将产生连续信号,但是,弱辐射则需要采用光电倍增管对单个光子进行监测。

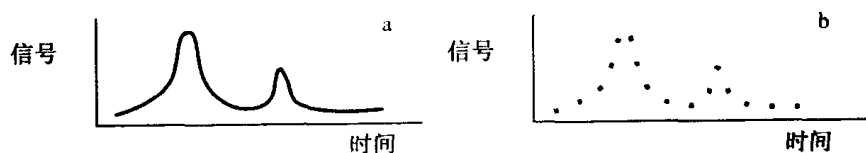


图 1-1 模拟信号检测器(a)和数字信号检测器(b)的检测信号与时间的关系

模拟信号一般采用模数转换器转换为数字信号,有关内容将在后面进行介绍。

### 1.1.2 二进制数与十进制数的转换

数字测量一般是对特定边界条件下脉冲数目的计数,最简单的计数方式是将脉冲表示为二进制数,因为这种方式只需要两种电平(高电平和低电平)状态,而表示 0 到 9 的十进制数将需要 10 种不同的状态。用于表示二进制数 0 和 1 的典型电压信号一般分别为 0.5V 和 5V。二进制数表示为 2 的不同次方,因此二进制系

统可表示任何数值。

### 例 1-1: 二进制数的表示

十进制数的 77 表示为二进制数为 1 001 101, 即:

$$\begin{array}{ccccccc}
 1 & 0 & 0 & 1 & 1 & 0 & 1 \\
 1 \times 2^6 & 0 \times 2^5 & 0 \times 2^4 & 1 \times 2^3 & 1 \times 2^2 & 0 \times 2^1 & 1 \times 2^0 = \\
 64 & + 0 & + 0 & + 8 & + 4 & + 0 & + 1 = 77
 \end{array}$$

表 1-1 提供了更多关于二进制数和十进制数之间的关系。每一个二进制数均由二进制位构成, 最右端的位称为最不重要的位, 而最左端的位称为最重要的位。

表 1-1 二进制数和十进制数之间的关系

| 二进制数 | 十进制数 | 二进制数  | 十进制数 | 二进制数      | 十进制数 |
|------|------|-------|------|-----------|------|
| 0    | 0    | 101   | 5    | 1 010     | 10   |
| 1    | 1    | 110   | 6    | 1 101     | 13   |
| 10   | 2    | 111   | 7    | 10 000    | 16   |
| 11   | 3    | 1 000 | 8    | 100 000   | 32   |
| 100  | 4    | 1 001 | 9    | 1 000 000 | 64   |

如何利用二进制数进行计算? 其运算方法与十进制数相似, 但比十进制数更简单。例如加法, 有 4 种情况:

$$\begin{array}{cccc}
 0 & 0 & 1 & 1 \\
 + 0 & + 1 & + 0 & + 1 \\
 \hline
 0 & 1 & 1 & 10
 \end{array}$$

注意: 二进制数 1 和 1 相加时需要进位, 即升到 2 的加一次方。

### 例 1-2: 二进制数的计算

21 + 5 的十进制数情况(a)和二进制数情况(b):

$$\begin{array}{rcl}
 \text{a.} & 21 & \text{b.} \quad 10 \ 101 \\
 + & 5 & \quad \quad 101 \\
 \hline
 & 26 & \quad 11 \ 010
 \end{array}$$

在计算机中除算术运算以外, 逻辑运算在某些算法(如专家系统)中也是必要的。表 1-2 列出了二进制数的逻辑运算操作。