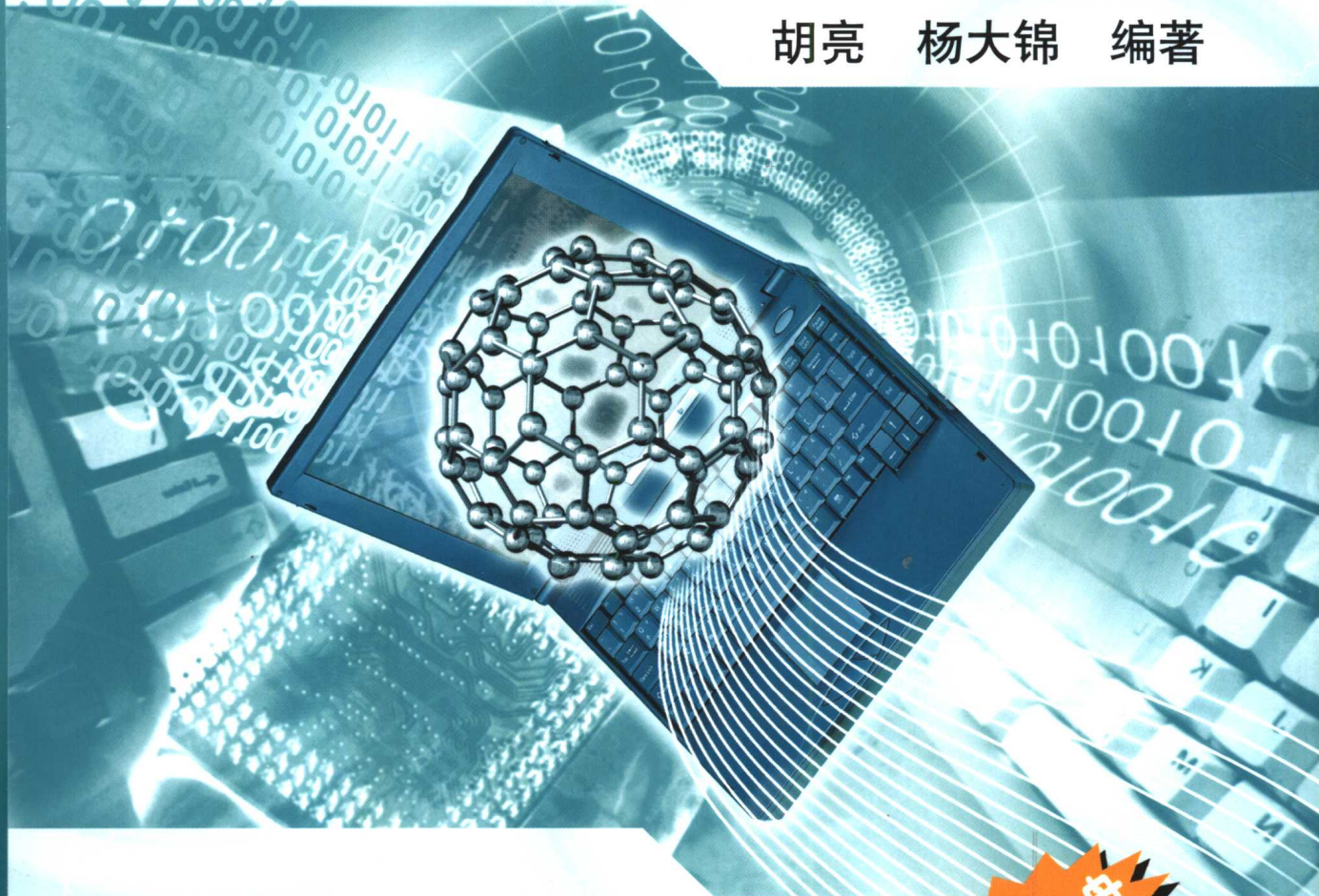


化学化工中计算机应用技术丛书

Excel与化学化工 试验数据处理

胡亮 杨大锦 编著



附赠光盘



化学工业出版社

化学与应用化学出版中心

化学化工中计算机应用技术丛书

Excel 与化学化工试验数据处理

胡 亮 杨大锦 编著

化学工业出版社

化学与应用化学出版中心

·北 京·

(京)新登字 039 号

图书在版编目(CIP)数据

Excel 与化学化工试验数据处理 / 胡亮, 杨大锦编著.
北京: 化学工业出版社, 2004.3
(化学化工中计算机应用技术丛书)
ISBN 7-5025-2992-6

I. E… II. ①胡… ②杨… III. 电子表格系统, Excel-
应用-化学工业-化学实验-数据处理 IV. TQ016-39

中国版本图书馆 CIP 数据核字(2004)第 016811 号

化学化工中计算机应用技术丛书
Excel 与化学化工试验数据处理

胡 亮 杨大锦 编著

责任编辑: 杜进祥

文字编辑: 操保龙

责任校对: 郑 捷

封面设计: 潘 峰

*

化 学 工 业 出 版 社 出版发行
化学与应用化学出版中心

(北京市朝阳区惠新里 3 号 邮政编码 100029)

发行电话: (010) 64982530

[http:// www.cip.com.cn](http://www.cip.com.cn)

*

新华书店北京发行所经销

北京密云红光印刷厂印刷

北京密云红光印刷厂装订

开本 787 毫米×1092 毫米 1/16 印张 18³/₄ 字数 460 千字

2004 年 4 月第 1 版 2004 年 4 月北京第 1 次印刷

ISBN 7-5025-2992-6/TP · 277

定 价: 39.00 元

版权所有 违者必究

该书如有缺页、倒页、脱页者, 本社发行部负责退换

前 言

由于计算机技术的迅速发展, 计算机应用得到广泛普及, 微机的用户数量与日俱增, 如何有效地利用微机已成为人们十分关注的问题。计算机可视化技术、所见即所得、面向对象方法给计算机应用与开发带来了极大的方便, 而这些技术和方法也得到了极大的发展和完善。试验设计与分析是科学工作者开展科学研究遇到的首要问题, 其基本原理和方法是科研人员必须掌握的法宝。古人云: “运筹帷幄之中, 决胜于千里之外”, 因此, 试验设计与分析方法的研究一直得到人们的重视。数学家华罗庚曾长期对优选法进行了推广普及。然而, 如何将现代计算机技术运用于试验设计与分析, 仍然是不断发展的新问题。微软 Excel 电子工作表格是当今最成功的应用软件之一, 用户遍及全球。Excel 是强大的可视化数据处理表格, 具有强大的图表和数值分析功能, 而 VBA 使其功能可得到延伸, 网络功能近年来也迅速增强。由于其典型的活动工作表格表现出非程序性“算法”特点, 并且有着丰富的工作表函数和数据分析工具, 特别适用于试验设计与分析。俗话说: “授予鱼不如授予渔”, 学习的目的主要是学习科学方法。试验设计与分析正是讲述科学研究的方法及其应用实例。

本书不仅要介绍 Excel 作为一个软件工具如何应用, 如何具体应用于化学化工试验数据分析处理, 而更重要的是要让读者去感受计算机信息时代一种全新的科学思维模式——电子表格模式。人类探索数据信息经历了漫长的历程, 从远古的结绳和刻划记数开始, 人们就在艰难地记录和管理数据信息了。文字、纸张和印刷术的发明, 有力地促进了人类传递、交换、分析处理和应用数据信息。曾几何时, 表格的出现为人们管理和运用数据提供了极大的方便。由于它可清楚地表达数据关系, 在社会各个方面得到了广泛应用, 无论是在物流管理、人力资源等社会科学管理与应用, 还是在自然科学数据管理与应用, 都能看见其身影。表格对于人类的文明与进步起了巨大的推动作用。计算机的出现为表格提供了新的活力, 表格从此有了新的载体和辅助分析处理工具。E. F. Codd 为庞大的表格数据库奠定了坚实的理论基础, 这使表格在规模化和规范化方面都得到了巨大发展和完善, 由此必然带来巨大的社会效益和经济效益。Dan Brick 在财务管理的商业需求推动下, 点燃了电子表格的星星之火, 他的初衷是为了使会计师在制作财务报表时, 能直接用电子表格代替计算器进行有关统计计算, 然而新的模式必然会照亮人类进步的前程。人们纷纷效法, 产生了各种电子表格, 微软利用其新视窗界面 Windows 推出了著名的电子表格 Excel。随后在商业需求和计算机技术进步的双重推动下, 微软将 Excel 推向了极至, 由此获得巨大成功。Excel 是当今电子表格模式的最先进工具。

本书第一篇介绍了 Excel 电子表格应用基础。在介绍了可视化操作界面、面向对象的可视化操作方式之后, 在第 3 章首先概述了数据与数据模型, 然后引出了数据的电子表格表述及其优越性。这使看似静态的数据具有了动态的特性, 其动态的产生可由事件和参数(数据)所驱使, 即信息的驱使。Excel 数据的动态处理已不仅是普通意义下的动态数据接收、处理与发送, 这种动态数据的动态特性是由数据本身的时序性所决定的。信息驱使的数据在未受使役时则是静止的, 一旦接收到外界信息, 就会做出相应的反应。由于表格数据间的相互引用, 使这种动态性具有了互动性。这种互动性甚至反映在图表中, 读者可用第 5 章操作题 1

做个试验，慢双击数据，用鼠标拖拽数据图形，看看表格中的数据如何变化。这就是电子表格模式所包含的辩证法思想——变化和广泛联系的观点和方法。这种动态的数据关系在其他应用软件中也有体现，最为明显的如各种三维设计软件，同样是由事件和参数所驱使的。参数驱使的过程使三维设计软件称之为参数化设计。在介绍了各种 Excel 工具之后，在第 4 章引入了应用软件在使用过程中的算法设计。这使我们可以主动地应用电子表格，而不是被动地应用。结构化、模块化、自顶向下逐步细化这些科学的方法论思想作为电子表格模式的支撑，使电子表格模式成为一种真正的科学思维模式，而 Excel 只是实现这种模式的工具。事实上，电子表格 ORIGIN 在科学数据处理方面也作了专业性的探索，也可以成为这种模式实现的辅助工具。随后介绍的图形化技术，使表格数据显得更加广泛而形象，网络化使电子表格模式插上了翱翔的翅膀。

第二篇逐步引入了试验数据产生方式、统计学分析方法。由于电子表格的应用，使试验数据分析处理的传统方法变得清晰可见。最优回归选择的完成，体现了 Excel 算法实现的能力。第三篇 Excel 试验数据分析应用，体现了 Excel 对于复杂数据表格模型的建立能力，试验数据分析处理的强大威力，使复杂的研究工作变得科学、简捷、可靠和轻松直观。

感谢昆明理工大学科研发展基金会的支持，感谢云南省应用基础研究基金会(98B004R)、云南省科学技术开发专项基金会(2000KFZX-14、2001KFZX-41)的资助。

衷心感谢化学工业出版社责任编辑对本书策划、组织工作的热心支持。

衷心感谢合作组负责人、云南省冶金集团技术中心冶金环保部主任、高级工程师杨大锦给予的积极支持，他在科研与工程技术的百忙之中仍参与了本书的撰写，为全书的组织提出了富有启发性的建议，亲自撰写了第 5 章图形技术，合写了第 7 章。

参加编写工作的除封面署名外，还有苏毅教授、张若愚研究员、牛存镇教授、徐亚飞高级工程师、刘雅刚、唐光阳、廖元双、张彬、梁可、张润虎、张慧、杨刘祥、胡心钰、王润元、彭庆华，他们协助做了大量的工作，在此表示衷心的感谢。

由于涉及内容广泛，时间仓促，难免有疏漏之处，恳请读者提出宝贵意见。

胡 亮

2003 年 12 月于昆明

内 容 提 要

微软公司的 Excel 是集文字、数据、图表处理于一体的，并能进行数据计算和统计分析的电子表格软件。其操作简单，所见即所得。

本书简要介绍 Excel 电子表格应用基础知识，并在此基础上结合大量化学化工实例，重点介绍 Excel 试验数据分析及应用，能满足化学化工工作者进行试验设计和数据处理工作的需要。

附赠光盘中根据书中出现的实例和有关 Excel 电子表格图示，按照篇章整理出了相应的 Excel 电子表格，方便与书中实例对照学习。同时以 Excel 电子表格的形式，摘录了部分实验用表，方便参考。

本书适用于化学化工科技工作者和科研人员阅读，也可作为化学化工相关专业教师、学生用教材。

目 录

引言	1
----------	---

第一篇 Excel 电子表格应用基础

第 1 章 Excel 简介	8
1.1 概述	8
1.2 Excel 工作环境	13
1.3 Excel 基本操作	18
1.3.1 启动和退出 Excel	18
1.3.2 菜单操作	18
1.3.3 程序窗口设置	21
1.3.4 工作簿的创建和管理	23
1.4 Excel 帮助	34
1.5 思考与练习	35
第 2 章 数据表格的建立	37
2.1 数据类型	37
2.2 数据的输入	38
2.2.1 向单元格输入数据	38
2.2.2 单元格区域的数据输入	39
2.2.3 多个工作表中输入数据	40
2.3 表格的编辑与格式化	41
2.3.1 单元格和单元格区域的编辑	41
2.3.2 整个工作表的编辑	43
2.3.3 格式化表格	43
2.4 Excel 模板与窗体	49
2.4.1 基本概念	49
2.4.2 模板的创建和使用	49
2.4.3 窗体操作	51
2.5 思考与练习	59
第 3 章 Excel 数据处理基本功能	61
3.1 基本概念	61
3.1.1 信息与数据	61
3.1.2 数据模型	63
3.1.3 实验数据的表示方法	64
3.1.4 Excel 数据表述与处理	65
3.2 公式与函数	66

3.2.1	公式的构造	66
3.2.2	数据区域引用	67
3.2.3	使用函数	71
3.2.4	编辑公式	75
3.2.5	公式的调试	75
3.3	数据管理与利用	82
3.3.1	Excel 表格中实现数据库管理	83
3.3.2	数据的排序	85
3.3.3	筛选数据	86
3.3.4	分类汇总	88
3.3.5	数据的合并	89
3.3.6	数据库函数的应用	90
3.4	思考与练习	90
第 4 章	Excel 数据处理高级工具	92
4.1	数据透视表	92
4.1.1	建立数据透视表	92
4.1.2	数据透视表的操作	94
4.2	模拟运算与方案设计	96
4.2.1	“模拟运算表”工具的使用	96
4.2.2	“方案管理器”的使用	99
4.3	目标搜索与规划分析	101
4.3.1	单变量求解	102
4.3.2	规划求解	103
4.4	数据分析工具库	104
4.5	现代化高效算法的实现	105
4.5.1	算法及其表示方法	105
4.5.2	算法的结构化	106
4.6	思考与练习	107
第 5 章	图形技术	109
5.1	Excel 数值图形化	109
5.1.1	图表图形绘制	109
5.1.2	图表图形修改	110
5.2	图形对象的创建与操作	118
5.2.1	基本图形绘制	119
5.2.2	自选图形绘制	121
5.2.3	其他图形的导入	122
5.2.4	图形文字的输入	122
5.2.5	图形操作	123
5.3	思考与练习	124
第 6 章	信息共享与网络化	127

6.1 工作表的打印输出	127
6.1.1 打印准备	127
6.1.2 预览工作表	129
6.1.3 打印设置	130
6.2 工作表数据的链接	131
6.2.1 数据链接方式	131
6.2.2 数据链接工具	132
6.3 共享工作簿	135
6.4 Excel 的国际互联网功能与 Web 电子表格	135
6.4.1 在网络上发布 Excel 数据	136
6.4.2 从网络上获取数据	138
6.5 思考与练习	140

第二篇 试验数据处理基础

第 7 章 试验数据的科学来源	142
7.1 试验优化设计	142
7.1.1 试验设计的含义和内容	142
7.1.2 常用术语	144
7.2 试验研究工作程序	147
7.3 试验误差与试验设计原则	149
7.4 试验数据采集	153
7.4.1 数据采集传感器	153
7.4.2 数据采集技术	154
7.5 试验数据整理	159
7.6 习题	162
第 8 章 数理统计基础	163
8.1 基本概念	163
8.2 直方图与分布函数	165
8.2.1 直方图	165
8.2.2 统计分布函数	167
8.3 随机变量的统计描述与参数估计	170
8.4 假设检验	173
8.4.1 假设检验的基本思想	173
8.4.2 Excel 假设检验工具	175
8.5 习题	177
第 9 章 相关与回归分析	179
9.1 相关分析	179
9.1.1 协方差与相关系数	179
9.1.2 相关系数与相关检验	179
9.1.3 协方差矩阵或相关矩阵	180

9.1.4 全相关系数	181
9.2 线性回归分析	183
9.2.1 线性回归模型	183
9.2.2 线性回归的显著性检验	184
9.3 非线性回归分析	185
9.3.1 一元非线性回归	186
9.3.2 多元非线性回归分析	188
9.4 最优回归的选择	188
9.5 习题	193

第三篇 Excel 试验数据分析应用

第 10 章 简单试验的数据分析	195
10.1 单因素试验数据分析	195
10.1.1 试点设计	195
10.1.2 简单对比试验法	196
10.1.3 完全随机化试验法	197
10.1.4 单因素方差分析	197
10.2 简单多因素试验数据分析	200
10.2.1 全面试验法	200
10.2.2 随机化区组试验	202
10.2.3 拉丁方与正交拉丁方试验法	203
10.3 优选法	207
10.4 习题	210
第 11 章 正交试验及其数据分析	212
11.1 正交试验的基本概念	212
11.1.1 正交表	212
11.1.2 $L_m^N(m^k)$ 正交表基本列	214
11.1.3 正交表 $L_m^N(m^k)$ 的有限域运算特性	215
11.1.4 $L_m^N(m^k)$ 正交表交互列	217
11.1.5 正交表的正交性	219
11.2 正交表的构造	220
11.2.1 水平数为素数的正交表构造	221
11.2.2 水平数为素数幂的正交表构造	221
11.2.3 混合型正交表的构造	223
11.3 正交试验步骤	226
11.4 正交试验数据的直观分析	227
11.5 正交试验数据的方差分析	232
11.6 正交试验的效应计算与指标值预估	236
11.7 正交表的灵活运用	238
11.7.1 混杂技巧与赋闲列设计	238

11.7.2 水平数不同的正交试验	239
11.7.3 正交试验序贯法	245
11.7.4 配方与工艺条件的正交试验	248
11.8 习题	253
第 12 章 均匀试验及其数据分析	258
12.1 概述	258
12.2 均匀设计表及其用法	259
12.3 均匀设计表原理及 Excel 构造	260
12.4 均匀试验数据分析处理	263
12.5 习题	267
第 13 章 VBA 编程处理试验数据	268
13.1 概述	268
13.2 Excel 应用中遇到的一些问题	269
13.3 VBA 程序特点	271
13.4 程序调试	272
13.5 VBA 编程方法的试验数据处理应用	274
13.5.1 自定义函数的应用	274
13.5.2 单纯形优化法数据处理	275
13.6 习题	284
主要参考文献	286

引言

一、试验与试验数据处理

人们在科研、生产和经营管理活动中，经常要通过试验来获取知识和寻求问题的解决办法。试验（或实验，Experiment）由一组有计划的操作（Operation, Trial, Test, Investigation, 即具体的试验检测）组成，每个操作又需要一组对应的观测（Observation）。由这些观测得到一系列观测数据，并构成一项试验的个别试验结果。试验研究可分为两种类型，一种是被动地进行调查观测，抽样检验；另一种是主动地规划、影响、控制和操纵研究条件所进行的观测。后者又可分为两类，一类是依据研究对象的确切机制，从特定过程的各项基本关系入手，对各种参量加以研究。例如，与化学过程有关的各种化学反应的速率和平衡常数，连同与之有关的各项传递系数，渗透速度等。然而，往往所研究的问题非常复杂，想要研究问题的潜在原因的工作量大得不可能完成，因此就必须经常采用另一类经验方法，即直接研究特定过程中各种变化的效应，用统计学方法加以分析。

要想做出好的决策，就要从所做的工作中获取尽可能多的可靠信息。为此必须兼顾试验设计和试验数据分析处理，这两个方面密切相关，相辅相成。试验设计必须保证试验完成后就能对所需解决的问题给出明确的答案，要能从试验数据中提取出全部信息，因此试验设计方案必须与分析方法配套。如果找不到有效而可行的分析方法，那么所做的试验方案便是不妥当的，任何高超的数据分析处理都无法补救拙劣的试验设计方案。

二、计算机数据处理的发展概况

不管何种类型的试验，要使问题圆满地得到快速准确的解决，就要使用计算工具对各种数据进行分析 and 处理。人类使用计算工具的历史可以追溯到一千多年以前，我们的祖先在宋朝发明的算盘是人类最早的计算工具。至今，算盘还在许多国家普遍使用。随着生产的发展，人们所需要解决的计算问题越来越复杂，技术的进步也使计算工具不断地得到改进、更新，先后出现了计算尺、手摇计算机等计算工具。

1642 年，法国数学家帕斯卡（B. Pascal）发明了世界上第一台机械式手摇计算机。它用八套齿轮相互咬合，通过人工手摇来实现数值运算和自动进位操作。

1823 年，英国著名数学家巴贝奇（C. Babbage）开始致力于设计一种可以存储数字、能按照事先安排的操作步骤自动进行计算并能自动输出计算结果的分析机。这种机器的设计构思已和现代计算机的组成十分类似。但由于受当时科技水平的限制，也由于缺乏资金上的支持，巴贝奇的设想未能得到实现。

20 世纪 40 年代中期，在导弹、火箭和原子弹的研究过程中，需要解决一些十分复杂的数学和工程问题。由于传统的计算工具计算速度慢、精度差，已无法满足要求，因此，迫切需要研制计算速度快、精度高、能自动控制运算进程的新型计算工具。同时，电子管的出现，电子学和自动控制技术的发展，已为研制电子计算机提供了技术基础。

1946 年，由宾夕法尼亚大学电工系工程师埃克特（J. P. Eckert）博士和物理学家莫奇利（J. W. Mauchly）博士领导的研制小组，用电子管代替继电器作为基本物理元件，研制成功了世界上第一台由程序控制的电子数字计算机，命名为 ENIAC（Electronic Numerical Integrator

and Calculator)。该机的研制费用将近百万美元，整机使用了 18000 多只电子管，1500 多只继电器；占地面积 150m²，重约 30t；该机每秒钟可以完成加法运算 5000 次。虽然这台计算机很笨重，性能也不完善，但它却标志着计算工具进入了一个崭新的时代——电子计算机时代。

从世界上第一台电子数字计算机 ENIAC 诞生以来，电子计算机以十分惊人的速度发展着。计算机的运算速度、存储容量、外围设备的配置、软件功能和数量、计算机价格、重量或体积等都在发生着日新月异的变化。计算机的应用推动着各部门、各学科的发展，谁充分利用了计算机，谁就占领了制高点。如果说 18 世纪蒸汽机的发明引起了工业革命，使低效的体力劳动被高效的机器所延伸和取代，人类进入快速的工业化发展阶段，那么计算机的发明则引起了信息大革命，催生了知识经济时代的到来，使繁重的脑力劳动得到解放，犹如给机器、人们的脑力和体力插上了飞翔的翅膀，人类从此进入快速飞行的信息化时代。

计算机是人类智慧的结晶，是人类高度智能化的产物，它促进了人类文明高速发展。尤其是微机及其软件的质量和性能的迅速提高。如微机硬盘容量可与过去数据库容量相媲美，同时价格却迅速下降，性价比提高，使计算机应用得到空前普及。20 世纪 90 年代以来，由于计算机和网络的发展，以数字革命为特征的信息化浪潮滚滚而来，遍及全球，人类社会进入了信息时代。现代信息社会把信息作为最重要的资源和产业，它和材料、能源一起并列成为社会物质和精神活动的三大重要资源。据估算，有效地利用信息、合理地提取其中的有用成分能产生数倍的附加经济效益。计算机信息技术的发展将对社会进步、科技发展、经济增长等起到越来越重要的作用，已成为世界各国在科技战线上争夺的战略制高点。在人类社会进入信息时代的同时，也进入了知识经济时代。知识是人们对客观事物及其规律的认识，科学和技术是知识的核心。知识属于信息被认识了的部分。由于知识对经济的推动作用以及经济增长对知识的依赖，知识成为全新的经济增长点。随着社会对知识的需求日益增加，出现了知识商品化和产业化。在知识经济时代，天然资源已经不是一个国家经济发展的关键，知识成为推动经济增长的主动力。人们迫切希望快速高效地获得有价值的知识，要想达到此目的，就必须要有科学的方法，这是因为事物的本质和规律总是隐藏在事物表象背后，人们需要透过现象看本质，去粗取精，去伪存真，对表象信息经过加工处理才能获得知识。掌握科学的研究方法是人才素质的基本组成要素。科学地进行试验研究是获取知识的有效方法，而科学试验研究则离不开行之有效的试验数据处理技术，它能加快获取知识成果的速度。计算机的应用加大了知识的产生速度和产生数量，导致现代社会的知识大爆炸，同时也带给了人们处理巨量信息的高效手段。

由于科学技术的飞速发展，试验方法、试验技术不断创新和发展，尤其是计算机应用于试验研究后，不但使试验仪器设备得到改进和创新，而且使试验方法和试验数据分析处理得到改进和创新。计算机的应用加快了实验室成果的产生和工业化，试验数据计算机处理在科学研究中发挥着重要的作用。面对时代的巨变，掌握试验数据处理技术具有特别重要的意义，积极地探索和研究新方法、新技术十分必要。

三、计算机化学的产生与发展

化学是一门古老的实验科学，化学的研究与发展为人类的进步做出了巨大贡献。回顾化学的发展史可以发现，把整个科学领域产生的新概念引入到化学研究当中，尤其是与其他学科的交叉融合都会使化学产生重大进展。人们最早研究的是无机化学。当化学家去研究生物体中的物质时产生了有机化学、生物化学；与检测技术结合后产生了分析化学；把物理方法

引入到化学后,产生了物理化学、化学物理学;甚至与文科结合也开辟了许多新的研究领域,如法医化学、化学史、化学哲学等。同样计算机与化学结合产生了计算机化学,为化学的发展,试验测试技术以及数据处理等开辟了一个崭新的领域。有了计算机这个数据处理工具,就可以用统计学方法严格处理试验数据和分析各种因素,由此发展了化学计量学。多元统计分析、人工神经网络、遗传算法、小波分析等先进的数据挖掘技术因此得以在化学的许多领域成功应用。而化学化工与信息科学、计算机技术的交叉,则产生了化学信息学,它包括化学化工信息的计算机表示、管理、网络化传输与检索,化学化工信息的解析、处理与推演,化工制图与化学信息的可视化和虚拟技术,远程教育。

近年来,可视化和虚拟技术、分子建模和计算化学、分子的物理性质和光谱技术、化工仿真技术取得了巨大成功,影响着化学、化工信息产品的发展,例如 Chemoffice、WebLab Viewer、Gaussian、MOPAC、AspenPlus、Winsim 等。有些软件在 Internet 上甚至可以免费得到。典型的化学信息产品有:网络与文献信息产品,分子建模图形与设计软件,化工仿真与工艺设计软件,化学化工专家系统,化学化工数据库(化学与物性数据库,化学反应数据库,晶体结构数据库,图谱数据库,基因、蛋白质等生物大分子数据库),各种化学应用软件,化学教学课件与远程教育系统。

计算机化学将在以下领域取得进展。

- 化学 3D 结构的计算机表示、存储与检索
- 合成反应的自动分析
- 化学数据信息的网络传输
- 大型化学化工数据库
- 与数据库相结合的化学化工数据挖掘技术
- 高效率的非局域优化技术
- 化学信息的可视化技术与虚拟技术
- 分子动力学从头算方法
- 生物大分子建模、构象分析与第一原理计算
- 基因、蛋白质分子的生物信息解析
- 溶液结构
- 分子器件、纳米组装的模拟与计算
- 材料结构与物性
- 相图计算与化工过程仿真技术
- 计算机网络远程教育系统

计算机给化学带来了崭新的面貌,产生了许多新概念、新学科,人们必须从各个方面重新去审视化学这门古老的学科。今天化学领域的知识并非都来自于传统实验室的经验测量,其中部分是根据已有知识和定理通过计算机推演、大规模计算预测与模拟得到的。美国化学会在期刊 *Journal of Chemical Information and Computer Sciences* (化学信息与计算机科学) 上增加了副标题 *Includes Chemical Computation and Molecular Modeling* (包含化学计算和分子建模)。

我国的现代计算机化学与技术研究起步于 20 世纪 70 年代,南京大学忻新泉教授运用计算机开始在数值处理和非数值运算方面展开工作,并率先在南京大学开设了计算机化学类课程,推动了我国计算机辅助化学教学的研究活动;中国科技大学张懋森教授开展计算分析和

专家系统领域的研究工作,并成为 1984 年我国的第一本计算机化学期刊《计算机与应用化学》的主编;中国科学院上海冶金研究所陈念贻教授将化学模式识别和计算机模拟方法应用于材料研究;中国科学院北京化工冶金研究所许志宏教授提出建立化学化工数据库。国内还有许多学者在计算机化学和计算化学的重要领域中做出了贡献:中国科学院上海有机化学研究所郑崇直教授、湖南大学俞汝勤教授、中国科学院长春应用化学研究所许禄教授、中国科学院药物研究所稽汝运教授、中国药科大学安登魁教授。清华大学胡鑫尧教授、北京化工大学麻德贤教授、南开大学林少凡教授等。

四、化学化工试验数据处理的发展概况

在计算机化学飞速发展的今天,我们再次审视化学化工试验数据处理这门知识时,发现有许多概念都发生了变化,如果不加以界定,就会感到整足和语言上的矛盾。

首先,试验可以单独用计算机在虚拟环境中完成,由此获取知识或解决问题。计算机不仅可作为控制试验过程的仪器和数据处理工具,而且可作为单独完成试验的工具。这一点有些人可能会不理解,那就请看其发展过程。1979 年,McKay, Beckman 与 Conover 在杂志 *Technometrics* 中发表 *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*, 提出拉丁超立方体抽样 (Latin Hypercube Sampling, LHS) 方法,从此出现了电算试验 (Computer Experiments)。我国数学家吴文俊教授采用数学机械化方法解决了数学问题,使数学思维方式有了革命性转变,这从数学的角度证明了用计算机是可以获取新知识的。化学工作者可以在计算机上完成特定的化学化工试验工作,从而节省经费,甚至达到单纯实验室研究无法完成的工作。这种计算机试验与纯粹的理论分析有所不同,需要一系列的计算机操作才能完成,同样需要认真的试验设计。计算机化学试验与传统试验相比,在试验目的、过程、数据处理、结果分析等方面都十分相似。上述计算机化学的许多分支都可看作计算机化学试验,如计算化学、化学信息解析、大分子建模、构象分析等。计算机化学试验可以用统计学方法进行试验设计。蒙特卡罗 (Monte Carlo) 方法是常用的模拟分析方法,例如我国高分子化学家杨玉良、吴大诚等采用蒙特卡罗法获得了高分子材料科学研究的突出成就。另外日本田口玄一博士等人依据正交表提出的产品三次设计 (three stage design) 可对产品的整体稳定性择优,并在产品质量与成本之间进行决策。当产品研究项目是可计算性项目时,三次设计也就可作为计算机试验应用的一个实例,它的应用已经取得了巨大的经济效益。当产品性能指标同各有关组件的参数之间具有已知的函数关系时,属于可计算性的项目。可计算性项目所占比例还很小,如何扩大可计算性的领域,把目前尚不可计算的项目发展成可计算项目,是该技术领域和数学相结合的边缘学科应予解决的重大科研任务。

其次,试验数据概念有了更丰富的内容,这是由于计算工具所能处理的数据范围有了显著性改变。除了传统的具有统计结构的试验数据以外,计算机可储存和处理图形图像、声音、动画等。事实上,计算机所能处理的数据已具有了相当大的广义性,正是数据信息概念的发展造就了今天辉煌的信息时代。

再者,由于计算机的方便性和高效性,我们不可能将试验数据处理工作仅仅局限于试验结果数据的处理和分析,自然会将计算机用于整个试验过程的数据处理。这个过程包括现有资料的收集整理,试验的设计、记录、结果分析、预测、验证等。例如著名的化学软件 Chemoffice 就考虑了分子特性的收集与检索,分子动力学模拟,化学试验记录分析等。如此庞大的实验数据处理内容,在此根本不可能面面俱到,只能结合电子表格软件 Excel 对传统的试验数据

处理技术进行一些探索,以达到抛砖引玉的作用。过去计算工具相当程度上满足不了试验数据处理的要求,即便是试验结果数据的处理都很难达到,而今人们在实验研究中却跟不上计算机发展的步伐,甚至在观念上都需要迅速地更新。例如,人们在进行试验安排以便快速获取有效数据时,常常需要被动地翻阅试验设计书籍,就像过去查阅《常用对数表》一样查阅各种试验安排表格,理解使用方法。由于试验设计理论的复杂性,往往令普通人望而生畏,将之束之高阁,很难谈得上自觉地应用,更谈不上成为一种试验习惯。

五、试验设计方法的发展及对数据处理的影响

试验数据处理方法取决于试验数据类型和数据处理目的。试验数据主要来源于试验结果,试验结果因试验方案而不同,试验方案又是由特定的试验技术和方法制定的。试验设计方法是数理统计学的应用方法之一。一般的数理统计方法主要是对已经获得的数据资料进行分析,对所关心的问题做出尽可能精确的判断。试验设计则是研究如何合理而有效地获得数据资料的方法,它的主要内容是讨论如何合理地安排试验、取得数据,然后进行综合的科学分析,从而达到尽快获得最优方案的目的。一项优良的试验设计能以最少的试验工作量取得所需的资料,反之试验次数增加则造成浪费,或数据不足得出虚假的结论,难以达到预期结果,甚至导致失败。

试验设计方法起源于概率论与数理统计学的研究,在 20 世纪 20 年代由英国生物统计学家、数学家费歇尔(R. A. Fisher, 1890~1962 年,英国罗隆姆斯台特试验农场工程师)创立了最早的试验设计方法——方差分析法,从此开创了试验设计的先河。如今各种试验设计方法指导着我们的试验活动,为人类探索自然、获取知识,创造物质和精神财富发挥着重要作用。例如日本战后工业生产迅速发展的重要原因之一就是在各工业领域里普遍推广和应用试验设计法,日本把试验设计技术誉为他们的国宝。

我国从 20 世纪 50 年代开始研究这门学科,但并没有引起广泛的重视。大约在 20 世纪 70 年代初,我国著名数学家华罗庚教授向全国推广应用试验设计方法的一个分支——优选法。到了 20 世纪 70 年代中期,优选法已在全国各行各业取得了巨大成果,效果十分显著。因此在 20 世纪 80 年代后期我国一些著名学者呼吁全国各行各业掌握并加以应用试验设计方法。试验设计方法,不仅在技术性领域应用广泛,在非技术性领域,如生产计划、产品销售、经营管理业务上也有应用。普及和推广试验设计方法将对我国的经济的发展起到很大的推动作用。

当今自然科学、社会科学、工业与商业中不断涌现出新问题,这些问题带有各种复杂性,例如多元性、非正态性、非线性、实时动态性、有异常数据与有丢失数据等,这些复杂性要求我们不但要灵活应用已有的试验设计方法,而且需要发展更为强有力的方法。我国科技工作者为此做出了不懈的努力,并取得了可喜的成绩。数学工作者方开泰、王元建立了均匀试验设计方法,该方法得到了广泛的应用,解决了大量的实际问题。均匀设计的产生充分显示了纯粹数学数论方法的威力。朱伟勇等在计算机最优试验设计领域做出了贡献,提出了新的 D-最优设计的迭代构造方法:对称构造法、对称序列发生法及对称对偶方法等。这些新的方法已在实际工程中得到广泛应用,并取得显著经济效益。利用计算机研究和构造最优设计是美国的 Kiefer 于 1973 年首先开创的,随后计算机研究和构造最优设计逐渐形成一个新的发展方向。钟大放等用伪蒙特卡罗方法制定了计量型制剂均匀度检验新方法,该方法已被《中华人民共和国药典》使用,其效率居于世界领先地位,标志着我国蒙特卡罗方法研究领域已进入其主流。何为将鲍威尔(Powell)方法的原理应用到化学及材料科学试验设计当中,取

得了新的研究成果。Powell 方法的成功体现了试验设计的最优化特性，最优化方法对试验设计具有直接作用。

特定的试验设计方法要求一系列特定的试验数据处理方法与之配套，计算机则是数据处理最为强有力的现代化工具。人们在试验研究活动中常常忽视试验设计工作，这主要是传统的试验研究习惯和计算工具的落后，但是随着所需解决问题的复杂化及计算机技术的迅猛发展和普及，试验的合理设计以及在此基础上的计算机数据分析处理就变得非常必要。

20 世纪 80 年代，日本创造了经济强胜的神话。美国科学促进协会于 1982 年特邀日本东京产品质量研究院田口玄一主任到会做了《日本的质理工程》的专题报告。人们总结经验认为，日本引进专业性强的系统设计方案，利用参数设计及容差设计进行创新，充分利用统计学方法搞产品设计，使各工业领域领先于世界。由此可见，数理统计分析方法对经济发展是十分重要的。如今，统计分析方法已越来越受到人们的广泛关注，许多有识之士都预言未来社会的各企业、各单位将需要大量的数据统计分析人才，数据库数据的分析利用是建立数据库的目的。西方各国对数据统计分析软件的利用已经盛行，并且出现了各种著名的数据分析软件品牌，如 SAS、SPSS、BMDP、STATISTICA 等。未来需要利用统计分析创造知识财富，我国具有几千年讲求谋略取胜的传统，近代已不乏统计学新方法的发明者，应该充分应用统计分析方法发愤图强。

尽管数据分析处理与试验设计方法密切相关，但是目前统计分析软件仍采用一般数理统计方法模式，即主要是对已经获得的数据资料进行分析，并且主要满足商业应用。虽然近来也出现了正交设计，但是其最主要的任务是广泛应用于市场研究的结合分析（Conjoint Analysis）快速生成正交表格。结合分析在国外已广泛应用；它是一种市场调查研究方法，属于第一类被动型试验研究方法。结合分析要求消费者对各种属性水平不同组合的产品进行偏好程度的排序，再从中分析出各个属性的相对重要性，以及各种水平组合的受欢迎程度。它可以预测产品市场占有率，以及市场变化对产品的影响。为了减少消费者需要评价的属性组合数，在进行结合分析之前往往需要采用正交设计来得到最佳属性组合。结合分析中正交表的应用表明试验设计技术的通用性。在德国海德堡，Rosch 等在其著作《试验设计方法手册》（I）中应用了一种试验设计软件 CADEM，但是该软件在分析数据时又必须应用专业分析软件，如作者采用的是 SAS，很不方便。国内广西大学的 LGS 软件工作室开发的软件 ChemExcel 初步涉及了化学试验的正交设计及其数据分析，但是仅建立了少数无交互作用的正交设计表。由此可见，系统完整的试验数据处理技术的计算机应用开发还有许多工作需要去做，它的研究、开发、应用、普及，必将对知识经济时代产生重要作用。

六、Excel 数据分析处理能力

Microsoft Excel 是微软公司在 Windows 环境下的电子表格软件，具有强大的数据管理和分析处理功能，其工作表格对数据的处理非常灵活方便，同时具有丰富的统计函数和统计分析工具，很适合试验数据的处理工作。

研究试验方法及其数据分析方法的目的就是要为大家提供方便易用，切实可行的有效方法，用以完成工作任务。而 Excel 是一个有效的数据分析工具，包含有各种数据分析处理的解决方案，必然会给试验带来方便和高效。Excel 主要包括如下数据分析处理功能。

（1）强大的制表功能

可以直接输入数据形成人们所熟悉的数据表格，也可以导入数据来生成数据表格，导入数据包括数据库、文本数据清单。表格数据包括多种数据类型，其显著特点是可包含公式和