

合成计量学 与化学化工 系统优化

周声勋 梁 亮
梁逸曾 编著

湖南大学出版社

合成计量学 与 化学化工系统优化

周声励 梁 亮 梁逸曾 编著

湖南大学出版社

1996年·长沙

内 容 简 介

本书共分九章,内容包括:概述、化学化工试验条件研究、有机反应性相关分析与定量构效关系研究、筛选试验、试验设计的化学计量学方法、配方问题的试验设计、化学化工试验优化方法、试验变量与量测响应的定量关系等。本书采用循序渐进的启发方式,分别介绍在化学化工研究的全过程中,怎样利用多变量分析方法进行试验设计和过程优化,并在理论分析的基础上,比较详细地讨论这些方法各自的特点,以使读者对其能有较系统的了解。

本书可作为大专院校化学化工有关专业教学参考书,也可供从事化学化工研究的专业人员参考。

合成计量学与化学化工系统优化

Hecheng Jilixue He Hualue Huagong Xitong Youhua

周声功、梁逸曾、梁逸曾 编著

责任编辑 俞 海

湖南大学出版社出版发行

长沙岳麓山 邮政编码 410082

湖南省新华书店经销。湖南省望城县湘江印刷厂印装

☆

850×1168 32开 7.75印张 188千字

1996年8月第1版 1996年8月第1次印刷

印数:1—1 000

ISBN 7-81053-041-0/TQ·1

定价:13.00元

(湖南大学版图书凡有印装差错,请向承印厂调换)

序

随着科学技术和国民经济的迅速发展,新产品、新方法、新工艺不断产生,怎样以最小的能源、水资源、原材料消耗以及最小的环境损害来获取最高的产率和性能最优的产品,是化学化工及其相关产业如材料、生化、医药、食品、农业乃至环境保护等各个领域普遍关注的问题。要实现上述目标,就必须调整各种化学试验条件和化工工艺相关参数,使过程最优化。

化学历来就是一门多变量的研究学科。任何一个化学反应或化工过程都涉及到多种变量,如温度、压力、PH 值、反应试剂及其浓度或不同原料的配比及催化剂等;此外,化学反应和化工过程的结果表征也是多变量的,如反应的产率、产物纯度、化学或生物活性、物理性能等。然而,很多化学家却只习惯于单变量的研究方法,常采用变动一个,固定其余的方法来设计试验和进行优化。这种方法明显是低效和不合事实的。本书的宗旨就是介绍一些多变量解析方法,使我们可以同时考虑和控制多种影响体系的因素,找到它们之间的相互关系,从而指导化学化工试验和生产过程,实现最优控制。

近年来,随着计算机科学和仪器量测技术的迅猛发展,不但使一个化学反应或化工过程可获得的信息量大幅增长,更重要的是,计算机技术大大地解放了很多实用的数学方法,使原先很多由于计算困难而难以在实际中得到应用的数学分支,如多元统计、模式识别、试验设计等,得到飞速发展,较系统地为化学家们提供了解析复杂多变量化学体系的数学工具。化学计量学

(Chemometrics)是这样一门新化学分支学科,其主要目标就是研究怎样利用数学、统计学以及计算机科学的方法来解决化学化工中的实际问题。它产生于70年代末,至今已在分析化学、有机化学、化学工程和环境化学中得到了广泛的应用。还产生了所谓合成计量学(Synthetometrics)环境计量学(Envirometrics)等新分支,这些事实说明,以多变量解析方法来研究复杂化学体系正在为更多的化学家们所接受。

为适应上述发展和应用的需要,我们把向有关专业研究生和本科生授课的讲稿及我们在近年来的工作成果和相关文献整理成书,奉献给读者,为促进用多变量方法来研究化学过程,通过多变量进行试验设计和优化的发展潮流聊尽绵薄之力。本书共分九章,其中第四、五、六章,和第一、七章的部分内容由周声励执笔;第二、三章,和第八章部分内容由梁亮执笔;第九章,和第一、七、八章的部分内容由梁逸曾执笔,全书由周声励和梁逸曾商定统稿。由于作者的学识和时间所限,书中内容难免疏漏错误,希望专家和读者不吝指正。

编 著 者

1996年6月于湖南大学

符 号 说 明

$A, B, C \cdots$	矩阵
$A^t, B^t, C^t, \cdots$	转置矩阵
a, b, c	列矢量
a^t, b^t, c^t	行矢量
$A_{p \times q}$	$(p \times q)$ 阶矩阵 A , 即矩阵 A 具有 p 行 q 列
I	单位矩阵
$\mathbf{1}$	所有分量皆为 1 的列矢量
$\mathbf{0}$	所有分量皆为 0 的列矢量
$\det(A)$	A 的行列式
$\text{tr}(A)$	方阵 A 的迹
$\text{Cond}(A)$	矩阵 A 的条件数
$\text{Rank}(A)$	矩阵 A 的秩
$\text{diag}(\cdot)$	对角矩阵
A^+	矩阵 A 的广义逆或称 Moore-Penrose 逆
dA/dt	矩阵 A 的导数矩阵
$\ a\ $	矢量 a 的长度
$\ A\ $	矩阵 A 的范数
$\ A\ _F$	矩阵 A 的 Frobenius 范数
$E(a)$	随机变量或随机矢量 a 的均值
$\text{Var}(a)$	随机变量 a 的方差
$\text{Cov}(a)$	随机矢量 a 的协方差阵
$N(\mu, \sigma^2)$	均值为 μ , 方差为 σ^2 的正态分布
$N(\mu, \sigma^2 I)$	均值为 μ , 协方差阵为 $\sigma^2 I$ 的多元正态分布
$\text{span}\{a_1, a_2, \cdots, a_n\}$	为矢量组 $\{a_1, a_2, \cdots, a_n\}$ 张成的子空间
$R(A)$	矩阵 A 的值域子空间
$N(A)$	矩阵 A 的零子空间

目 次

Contents

符号说明

1 概 论(Introduction)

- 1.1 多变量化学过程与化学计量学(Multi-variable Chemical Procedure and Chemometrics) (1)
- 1.2 合成计量学与化学化工中的配方问题(Synthetic Metrics and Formula Problem in Chemistry and Chemical Engineering) (3)
- 1.3 化学试验与化工过程优化(Optimization in Chemistry and Chemical Engineering Proceed) (5)
- 1.4 本书各章内容简介(Outline for Book Contents) (6)

2 化学化工试验条件研究(Study of Experimental Conditions for Chemistry and Chemical Engineering)

- 2.1 化学合成过程(Chemical Synthesis Procedure) (8)
- 2.2 试验设计(Experiment Design) (15)
- 2.3 试验变量、试验区间和响应值(Independent Variables, Experiment Domain and Responses) (17)
- 2.4 重要(有效)变量的筛选与优化(Screening of Important Variables and Optimization) (19)
- 2.5 合成化学与定量模型(Synthetic Chemistry and Quantitative Models) (23)

3	有机反应性相关分析与定量构效关系研究 (Correlation Analysis of Organic Reactivity and Quantitative structure and activity relationship)	
3.1	有机反应性相关分析及化学试剂选择 (Correlation analysis of organic reactivity and selection of chemicals)	(57)
3.2	定量构效关系研究新近进展简介 (Introduction of quantitative structure and activity relationship)	(64)
3.3	分子连接性法及化学反应活性 (Molecular connectivity method and chemical reactional activity)	(72)
4	筛选试验 (Screening Experiment)	
4.1	筛选试验原理 (Principles of Screening Experiment)	(91)
4.2	一步筛选法 (One-step Screening Method)	(94)
4.3	多步分组筛选法 (Multi-step Screening in Groups)	(97)
5	试验设计的化学计量学方法 (Chemometric Methods for Experimental Design)	
5.1	因子设计与正交设计 (Factorial design and orthogonal design)	(107)
5.2	均匀设计 (Uniform design)	(117)
5.3	D-最优设计 (D-optimal design)	(120)
6	配方问题的试验设计 (Mathematical Modeling for Mixing Problem in Chemical Engineering)	
6.1	一般配方问题 (General Formula Problem)	(128)
6.2	无附加约束的配方问题 (Formula Problem without Constraints)	(132)
6.3	有附加约束的配方问题 (Formula Problem with Some Constraints)	(142)

6.4	特殊问题的配方模型(Other Special Formula Models)	
		(153)
7	化学化工试验优化方法(Experimental Optimization for Chemistry and Chemical Engineering)	
7.1	单纯形优化法(Simplex Optimization Method)	(170)
7.2	化学模式识别方法(Chemical Pattern Recognition)	(175)
7.3	人工神经网络(Artificial Neural Network)	(187)
8	试验变量与量测响应的定量关系(Quantitative Relations between Observed Responses and Experimental variables)	
8.1	多元线性回归(Multiple Linear Regression)	(199)
8.2	主成分分析与主成分回归(Principal Component Analysis and Principal Component Regression)	(204)
8.3	偏最小二乘方法(Partial Least Squares)	(205)
9	统计学和线性代数基础知识及一些常用算法(Necessary Knowledge on Statistics and Linear Algebra and Some Common Algorithms in Chemometrics)	
9.1	必要线性代数基础知识(Necessary Knowledge on Linear Algebra)	(217)
9.2	必要的统计学基础知识(Necessary Knowledge on Statistic)	(230)

概 论

1.1 多变量化学过程与化学计量学

化学化工(包括材料)生产是我国国民经济的一个庞大而重要的组成部分,它的效率,它的产品质量和成本对国民经济有着至关重要的影响。怎样以最少的能源、水资源、原材料消耗以及最小的环境损害来获取最高的产率和性能最优的产品,是人们普遍关注的问题。要实现上述目标,就必须调整各种试验条件或工艺参数,使过程最优化。而化学历来就是一门多变量的研究学科。任何一个化学反应或化工过程都涉及到多种变量,如温度、压力、pH、反应试剂或不同原料的配比及催化剂等。此外,化学反应和化工过程的结果表征也是多变量的,如反应的产率、产物纯度、化学或生物活性、物理性能等,构成一个完整的多变量空间(图 1.1)。然而,很多化学家却只习惯于单变量的研究方法,常采用变动一个,固定其余变量的方法来设计试验和进行优化。这种方法明显是低效和不合事实的。首先,这种单变量试验方法很难真正找到最优试验条件;此外,这种方法还隐含了一个重要假设,即假设这些变量都是互不相关的,然而,这一假设一般与实际情况不符。用多变量方法来研究化学过程显然优于单变量方法。多变量方法是这样一种方法,即通过多变量的试验设计,较全面地进行影响因素的评价和筛选,进而经数据分析,推导出

某些结论和假设。本书的宗旨就是介绍这样的解析方法,使我们
可以得到多种因素是如何同时影响分析体系的知识,并找到它
们之间的相关关系,从而指导化学化工试验和生产过程,实现最
优控制。

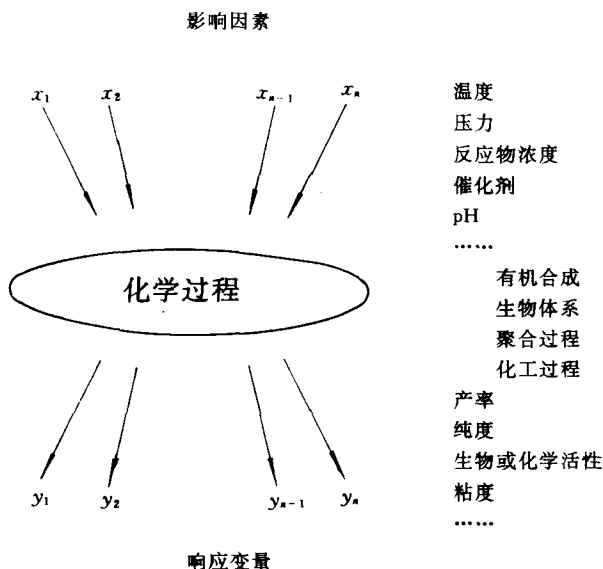


图 1.1 化学过程示意图

近年来,随着计算机科学和仪器量测技术的迅猛发展,不但
使一个化学反应或化工过程可获得的信息量大幅增长,而且更
重要的是,计算机技术大大地解放了很多实用的数学方法。不但
使原先很多由于计算困难而难以在实际中得到应用的数学分
支,如多元统计、模式识别、试验设计等,得到飞速发展,而且还
产生了以计算机为基础的计算数学、控制论等数学新分支。较系
统地为化学家们提供了解析复杂多变量化学体系的数学工具。
化学计量学(Chemometrics)就正是这样一门新化学分支学科。
它的主要目标就是研究怎样利用数学、统计学以及计算机科学

的方法来解决化学中的实际问题。它产生于 70 年代末,至今已在分析化学、有机化学、化学工程和环境化学中得到了广泛的应用。还产生了所谓的合成计量学(Synthetometrics)、环境计量学(Envirometrics)的新分支,这些事实说明,以多变量解析方法来研究复杂化学体系正在为更多的化学家所接受。

1.2 合成计量学与化学化工中的配方问题

化学合成可以说是化学研究的一个最重要的领域,在有机化学中尤其占有特殊的地位。如何有效地进行化学合成和分子设计,一直是化学研究中的一个热点和难点。量子化学从对分子结构的计算中以求得分子的反应特性来进行分子设计;有机化学则从大量的试验中获取经验,以寻求新化合物的合成方法;合成计量学作为化学计量学的一个分支,则是研究怎样利用数学、统计学以及计算机科学的方法来解决化学合成中的具体问题,其中特别是利用多变量解析方法来进行合成方法的试验设计和过程优化的问题。合成计量学以化学合成过程作为其研究对象,其基点就是采用多变量的研究方法对化学合成过程进行有效地试验设计,继而筛选出其中对反应过程有着重要作用的变量,通过对这些变量的合理调整来达到对整个化学反应过程的优化,进而经数据分析而推导出某些结论和假设。化学反应过程是一个典型的多变量过程,图 2.2 简要地示出了反应全过程的相互关系。从图中可以看出,任何一个化学反应过程可由三个变量空间组成,即所谓的反应空间、试验空间和结果空间。首先,一个化学反应需要有反应物质、基体物质、溶剂或催化剂等,它们中间的任何一种物质发生变化都会引起整个反应过程的变化,它们就构成了合成的反应空间。如何结合各类不同物种的化学反应性质,构造和评价它们对反应体系结果的影响,显然是化学

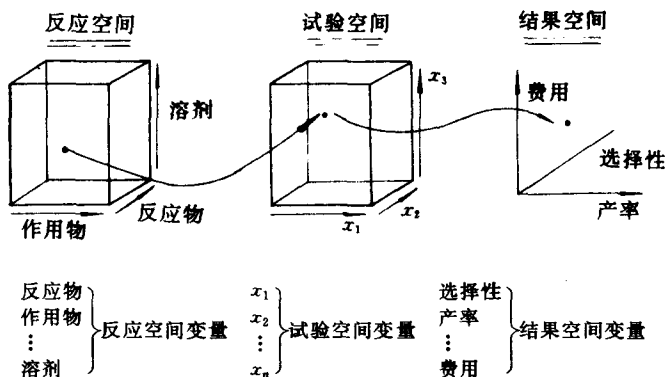


图 1.2 合成变量空间示意图

家们感兴趣的研究课题。在化学计量学中,化学结构与化学物质性能的定量构效关系(Quantitative structure and activity relationship, QSAR)将对反应空间的选择提供有益的参考,应属于合成计量学研究的重要内容。此外,当化学反应选定了反应物种,进行反应的试验条件就成为十分重要的前提了。不同的试验条件,如反应温度、反应压力以及反应物的不同配比等都将产生不同的反应结果。这就是图中所示的试验空间,显然它也是一个多变量空间,对其中不同试验点的选择,就意味着不同的试验设计。怎样采用多变量的研究方法对化学合成过程进行有效的试验设计,即用最少的试验次数而获取最大的有关反应空间的有用信息,继而从中筛选出对反应过程有着重要作用的变量,而不是只采用变动一个,固定其余变量的方法来进行试验和优化,也是合成计量学所研究的一个重要内容。值得提出的是,对结果的优化也是一个多变量问题。化学反应结果好坏的表征也是多变量的,即表征一个结果需用多个指标,如反应产率、反应所需费用以及产物的生物活性及物理特性等,它们构成了反应的结果空间,必须综合考虑。研究这三个空间的相互关系,通过试验

设计来进行多变量结果空间的多目标优化就成了合成计量学所研究的主要内容。

化工配方设计是化工研究,特别是精细化工研究的一个重要内容。随着化学计量学研究的深入发展,怎样通过合理调整配方以改善产品的质量,将越来越受到化学化工研究工作者的重视。可以认为,化工配方设计是一种特殊的试验设计方法,有其自身特点。可是从研究方法和研究手段来看,它又与一般试验设计有着共同的地方,即都必须采用多变量的研究方法。故本书在讨论一般设计的基础上,专门列出一章来对化工中的配方设计问题给出特别的介绍,以期引起同行们的重视。

1.3 化学试验与化工过程优化

化学与化工过程优化一直是化学计量学中研究的一个热点。如何有效地采用多变量研究方法,特别是近年来发展起来的模式识别(Pattern Recognition)和人工神经网络(Artificial Neural Network)来进行化学化工过程的优化,一直受到化学计量学家的特别关注。随着计算机技术的发展和自动化程度的提高,这一领域的研究可望得到更多化学和化工工作者的关注。

经典的回归分析一直是化学化工过程研究的一种优化方法,对于一些相对简单的化学化工过程,它至今仍不失为一种有效的方法。然而,近年来化学计量学的发展引进了不少多变量的研究新方法,其中著名的有主成分回归(Principal Component Regression, PCR)和偏最小二乘法(Partial Least Squares, PLS)。值得提出的是,偏最小二乘方法因其可用于多目标的预测,近年来受到化学计量学家们的特别关注,被成功地应用于很多化学化工领域。用模式识别和人工神经网络来进行化学化工过程的优化是近年来化学计量学研究的另一个热点领域。我国

的化学计量学工作者在这一方面作了大量工作。在此特别值得提出的一点是,模式识别和人工神经网络采用“黑箱”式工作方式,故可被用来进行离线调优,即可用以前已积累的现存数据来对化工过程进行优化,而不必进行大量现场试验。当然,这样找到的最优试验条件不能保证一定是试验空间的全局最优点,但却可能在现有的条件下尽可能地提高化学化工过程的效益。对以上所讨论的这些方法,本书将给出详细的介绍,以期读者对此有一个较全面的了解。

1.4 本书各章内容简介

全书共分为九章,采取循序渐进的启发式,分别介绍在化学化工研究的全过程中,怎样利用多变量分析方法来进行试验设计和过程优化。并在理论分析的基础上,比较详细地讨论这些方法各自的特点,以使读者对它们能有较系统的了解。另外,因本书所涉及的数学内容较广,为使读者便于理解,作者还特意将散见于各专著中的一些必要的数学和统计学基础并为一章,使本书成为一自封闭系统,读者可根据各自不同需要选择阅读。

化学化工试验条件研究(第二章)以化学过程与试验设计,自变量、试验区间和响应值,重要变量的筛选与优化以及合成化学与定量模型为主要线索,逐步介绍一般化学化工过程试验条件研究的基本方法和思路。为以后各章介绍的具体方法提供必要的基础知识。

有机反应性相关分析与定量构效关系研究(第三章)对有机反应性相关分析与化学试剂选择的关系进行简要介绍。在此基础上,对近年来发展迅猛的定量构效关系的研究也给出必要的讨论,我们还特别对分子连接性法作了较具体的介绍。

筛选试验(第四章)以筛选试验的原理、步骤和设计选择为

线索展开。主要一般地介绍变量筛选的原理和试验步骤。在此基础上,适当地讨论变量筛选试验的设计选择,以便读者对变量筛选的全过程有一个清晰的了解。

试验设计的化学计量学方法(第五章)分别具体介绍不同的试验设计方法。本章主要介绍因子设计与正交设计、D-最优设计以及均匀设计。在比较详细地讨论这些方法的各自特点及适应性的基础上,读者对这些方法将有较系统的了解,从而为正确选择这些方法提供坚实的理论基础。

配方问题的试验设计(第六章)对这一化工研究的特殊问题作了较详细的讨论,分别对无附加约束的配方问题、有附加约束的配方问题以及其他配方模型作出介绍。在讨论不同方法的基础上,尽量给出一些实例加以说明。

化学化工试验优化方法(第七章)以介绍不同的优化方法为线索,主要介绍单纯形优化方法(Simplex Optimization Method)、化学模式识别(Chemical Pattern Recognition)和人工神经网络(Artificial Neural Network)方法,重点介绍化学模式识别方法。

试验变量与量测响应的定量关系(第八章)介绍一些重要的定量方法。主要是多元线性回归、主成分分析与主成分回归以及偏最小二乘方法。对偏最小二乘方法的介绍是本章的重点。总的说来,用化学计量学来研究化学化工试验设计和过程优化,实质上就是以多变量分析方法为其主要特征,而对多变量分析方法的熟练掌握,不但对数学和多元统计的基础知识有较高要求,而且,在实际应用中还要求读者对计算机编程有一定基础。为使读者在学习本书后能直接将介绍的新方法用于实际,故将它们编制成可方便使用的软件供读者参考。感兴趣的读者可直接与作者联系。

化学化工试验条件研究

2.1 化学合成过程

化学合成的实质是寻找一种有效的方法,用简单的起始原料构造一个复杂的分子。一个理想的化学合成过程应如图 2.1 所示。

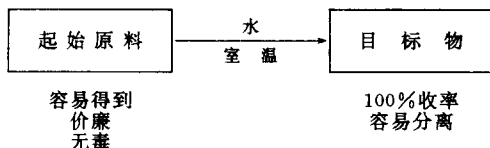


图 2.1 理想的化学合成过程

然而,在实际过程中这样的合成很少,最普遍的情况是:目标物是通过顺序反应,即合成路线,连结分子的不同部位而建造成一个所需要的分子。选择适合的合成路线的标准,往往由合成反应的总收率和成本来决定。在纯化学合成研究领域,关心的是达到目标物合成的总收率,收率的好坏成为唯一的标准。而在工业合成中,还必须考虑其它因素,如目标物的生产费用、毒性及所涉及的溶剂、环境保护及生产规模等。

我们知道,在有机合成领域中,解决不同复杂程度的问题有不同的战略。这些战略水平可以按图 2.2 分层排列。在这样的方法中,各层之间是相互关联的,即上层水平的制约或控制受下