

西安石油大学优秀学术著作出版基金资助出版



Chemometrics in Excel

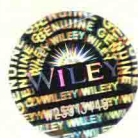
Excel



中的化学计量学

[俄] 阿列克谢 L. 波莫兰采夫 (Alexey L. Pomerantsev) 著

焦龙 译



化学工业出版社

Chemometrics
in Excel

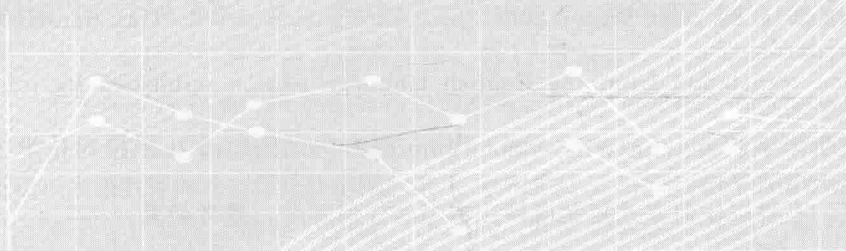
Excel



中的化学计量学

[俄] 阿列克谢 L. 波莫兰采夫 (Alexey L. Pomerantsev) 著

焦龙 译



化学工业出版社

· 北京 ·

全书共分4部分,第1部分对化学计量学进行了总体介绍;第2部分概述了阅读本书所需的相关数学和计算机基础知识;第3部分详细介绍了 Chemometrics Add-In 所包含化学计量学算法的原理和操作;第4部分是本书的附录章节,包括 Chemometrics Add-In 的扩展功能、使用 Chemometrics Add-In 所提供的化学计量学方法对光谱数据的动力学建模示例,以及 MATLAB® 的简单初学者指南等内容。

本书可供化学、物理、生物、计算机、数学等相关专业技术人员和研究人员使用,还可作为相关专业高年级本科生及研究生的专业课参考书。

图书在版编目(CIP)数据

Excel 中的化学计量学/[俄]阿列克谢 L. 波莫兰采夫(Alexey L. Pomerantsev)著;焦龙译.—北京:化学工业出版社,2018.1

书名原文:Chemometrics in Excel

ISBN 978-7-122-31099-6

I. ①E… II. ①阿… ②焦… III. ①表处理软件-应用-化学计量学 IV. ①O6-04

中国版本图书馆 CIP 数据核字(2017)第 293372 号

Chemometrics in Excel, 1st edition/by Alexey L. Pomerantsev

ISBN 978-1-118-60535-6

Copyright©2014 by John Wiley&Sons, Inc. All rights reserved. This translation published under license.

Authorized translation from the English language edition published by John Wiley&Sons, Inc.

本书中文简体字版由 John Wiley&Sons International Rights, Inc. 授权化学工业出版社独家出版发行。

未经许可,不得以任何方式复制或抄袭本书的任何部分,违者必究。

北京市版权局著作权合同登记号:01-2017-8298

责任编辑:张艳 刘军

文字编辑:向东

责任校对:王静

装帧设计:王晓宇

出版发行:化学工业出版社(北京市东城区青年湖南街13号 邮政编码100011)

印刷:三河市延凤印装有限公司

装订:三河市宇新装订厂

710mm×1000mm 1/16 印张18 $\frac{3}{4}$ 字数360千字 2018年1月北京第1版第1次印刷

购书咨询:010-64518888(传真:010-64519686)

售后服务:010-64518899

网址: <http://www.cip.com.cn>

凡购买本书,如有缺损质量问题,本社销售中心负责调换。

定 价:128.00 元

版权所有 违者必究

译者序

PREFACE

化学计量学是一门通过统计学或数学方法对化学体系的测量值与体系状态之间建立联系的学科。本书对 Excel 中的化学计量学进行了全面介绍。全书共分 4 部分、包括 15 章，主要介绍了化学计量学基础知识，常用化学计量学算法，以及在 Microsoft Excel® 软件中通过作者自编的 Chemometrics Add-In 这个插件运行这些化学计量学算法的途径。

第 1 部分是对化学计量学的总体概述，包括 4 章。第 1 章介绍了化学计量学的定义、研究主题和起源。第 2 章简要介绍了本书的主要内容。第 3 章说明了安装和使用 Chemometrics Add-In 的基本条件。第 4 章列出了化学计量学的主要学习资源。

第 2 部分概述了阅读本书所需的相关数学和计算机基础知识，包括 4 章。第 5 章介绍了矩阵和向量的基础知识。第 6 章介绍了相关的统计学基础知识。第 7 章主要介绍了 Excel 中的基本矩阵运算。第 8 章讲述了在 Excel 中通过 Chemometrics Add-In 进行投影方法的运算。

第 3 部分详细介绍了 Chemometrics Add-In 所包含化学计量学算法的原理和操作，仍然包括 4 章。第 9 章是对主成分分析方法的介绍。第 10 章讲述了校正方法。第 11 章是对多种分类方法的介绍。第 12 章则介绍了各种多元曲线分辨方法。

第 4 部分是本书的附录章节，包括 3 章。第 13 章介绍了 Chemometrics Add-In 的扩展功能。第 14 章简要介绍了使用 Chemometrics Add-In 所提供的化学计量学方法对光谱数据的动力学建模示例。第 15 章则包括了 MATLAB® 的简单初学者指南。

本书可供化学、物理、生物、计算机、数学等相关专业技术人员使用，还可作为相关专业高年级本科生、研究生的专业课参考书。

感谢西安石油大学优秀学术著作出版基金、陕西省青年科技新星计划项目 (No. 2016KJXX-16)、中国石油科技创新基金研究项目 (No. 2015D-5006-0407)、西安石油大学科研创新团队 (No. 2013QNKYCXTD01)、西安石油大学校级拔尖人才计划对于本书出版的资助与支持。

感谢西安石油大学近年来为作者工作提供的支持。

感谢西安石油大学化学化工学院硕士研究生秦玉翠、王彦昭为本书翻译工作所做的工作。

焦龙
2017 年 10 月

前言

FOREWORD

化学计量学是一门非常实用的学科。要学习它，不仅要理解众多的化学计量学方法，还应懂得这些方法的实际应用。这本书可以帮助读者完成这项艰巨的任务。本书主要面向那些对实验数据分析感兴趣的化学家、物理学家、生物学家和其他专家，也可以作为多元数据分析初学者的化学计量学入门读物。

化学计量学培训的常规方法是使用专门的程序（Unscrambler、SIMCA 等）或者 MATLAB®。我们提出自己特有的（化学计量学培训）方法，这就是利用世界上最普及的程序 Microsoft Excel® 提供的功能。但是，使用基本 Excel 系统很难实现最主要的一些化学计量学算法，例如主成分分析（PCA）、偏最小二乘（PLS）这类投影方法。因此，我们开发了对标准 Excel 版本的一个专门补充，称为 Chemometrics Add-In，这个插件可用于执行这些化学计量学算法的计算。

这本书不仅是一个多元数据分析技术的使用手册，也是一个多元数据分析技术的实用指导。几乎每章都附带有专用的 Excel 工作簿，其中包含了与所讨论技术有关的计算。为了更好地理解本书的素材，读者应该经常查阅这些工作簿并尽力重复执行这些计算。在 Wiley Book Support 网站（<http://booksupport.wiley.com>）收集了带有实例的工作簿以及 Chemometrics Add-In。安装指南在第 3 章中介绍。我们强烈建议学习这本书的时候安装和使用该软件（即 Chemometrics Add-In）。

我的同事 Oxana Rodionova 博士在本书的编写中发挥了非常重要的作用。她设计和完善了 Chemometrics Add-In 软件，还帮助撰写了本书的多个章节。因此，可以毫不夸张地说，没有她的帮助这本书就不能完成。第 15 章是与我的学生 Yevgeny V. Mikhailov 一起撰写的。

非常感谢编辑 Svetlana 和 Pieter Rol 为筹备本书做出的大量贡献。

我们计划持续地修改和补充本书。欢迎读者通过邮箱 rscs@chph.ras.ru 给我们发送您的意见和建议。

第 1 部分 导论	001
第 1 章 什么是化学计量学?	002
1.1 化学计量学的主题	002
1.2 历史上的题外话	004
第 2 章 这本书是关于什么的?	007
2.1 有用的提示	007
2.2 本书大纲	008
2.3 标记法	009
第 3 章 Chemometrics Add-In 的安装	010
3.1 安装	010
3.2 基本信息	013
第 4 章 关于化学计量学的更多读物	014
4.1 书籍	014
4.1.1 基础知识	014
4.1.2 化学计量学	015
4.1.3 补充读物	015
4.2 互联网	016
4.2.1 教程	016
4.3 期刊	017
4.3.1 化学计量学	017
4.3.2 分析化学	017
4.3.3 数学	018
4.4 软件	018
4.4.1 专用程序包	018
4.4.2 通用统计学软件包	018
4.4.3 免费软件	019
第 2 部分 基础知识	020
第 5 章 矩阵和向量	021
5.1 基础知识	021
5.1.1 矩阵	021

5.1.2	矩阵的简单运算	022
5.1.3	矩阵乘法	023
5.1.4	方阵	024
5.1.5	迹和行列式	024
5.1.6	向量	025
5.1.7	向量的简单运算	026
5.1.8	向量的积	026
5.1.9	向量的范数	027
5.1.10	向量之间的夹角	027
5.1.11	矩阵的向量表示	028
5.1.12	线性相关向量	028
5.1.13	矩阵的秩	028
5.1.14	逆矩阵	029
5.1.15	伪逆	029
5.1.16	矩阵与向量的乘积	030
5.2	进阶信息	030
5.2.1	线性方程组	030
5.2.2	双线性型和二次型	031
5.2.3	正定矩阵	031
5.2.4	Cholesky 分解	031
5.2.5	极分解	031
5.2.6	本征值和本征向量	032
5.2.7	本征值	032
5.2.8	本征向量	033
5.2.9	等价性和相似性	034
5.2.10	对角化	034
5.2.11	奇异值分解	034
5.2.12	向量空间	035
5.2.13	空间基	036
5.2.14	几何解释	036
5.2.15	基的不唯一性	036
5.2.16	子空间	037
5.2.17	投影	037
第 6 章	统计学	038
6.1	基础知识	038
6.1.1	概率	038

6.1.2	随机值	039
6.1.3	分布函数	039
6.1.4	数学期望	039
6.1.5	方差和标准差	040
6.1.6	矩	040
6.1.7	分位数	040
6.1.8	多变量分布	041
6.1.9	协方差和相关性	041
6.1.10	函数	041
6.1.11	标准化	042
6.2	主要的分布	042
6.2.1	二项分布	042
6.2.2	均匀分布	043
6.2.3	正态分布	044
6.2.4	卡方分布	045
6.2.5	Student 分布	047
6.2.6	F 分布	048
6.2.7	多元正态分布	049
6.2.8	伪随机数	050
6.3	参数估计	050
6.3.1	样本	050
6.3.2	异常值和极端值	051
6.3.3	统计总体	051
6.3.4	Statistics	052
6.3.5	样本均值和方差	052
6.3.6	样本协方差和相关性	053
6.3.7	顺序统计量	053
6.3.8	经验分布和直方图	054
6.3.9	矩方法	055
6.3.10	极大似然方法	056
6.4	估计量的性质	056
6.4.1	一致性	056
6.4.2	偏置	057
6.4.3	有效性	057
6.4.4	稳健性	057
6.4.5	正态样本	058

6.5	置信估计	058
6.5.1	置信区域	058
6.5.2	置信区间	059
6.5.3	置信区间的示例	059
6.5.4	正态分布的置信区间	059
6.6	假设检验	060
6.6.1	假设	060
6.6.2	假设检验	060
6.6.3	第一类错误和第二类错误	060
6.6.4	示例	061
6.6.5	皮尔逊卡方检验	061
6.6.6	F 检验	062
6.7	回归	064
6.7.1	简单回归	064
6.7.2	最小二乘法	064
6.7.3	多元回归	065
第7章	Excel 中的矩阵计算	067
7.1	基本信息	067
7.1.1	地区及语言	067
7.1.2	工作簿、工作表和单元格	069
7.1.3	地址	070
7.1.4	区域	072
7.1.5	简单运算	073
7.1.6	函数	073
7.1.7	重要函数	075
7.1.8	公式中的错误	078
7.1.9	公式拖拽	079
7.1.10	创建图表	080
7.2	矩阵运算	081
7.2.1	数组公式	081
7.2.2	数组公式的创建和编辑	082
7.2.3	最简单的矩阵运算	083
7.2.4	访问矩阵的一部分	083
7.2.5	一元运算	085
7.2.6	二元运算	086
7.2.7	回归	087

7.2.8	Excel 2003 中的严重漏洞	090
7.2.9	虚拟数组	090
7.3	Excel 功能的扩展	091
7.3.1	VBA 编程	091
7.3.2	示例	093
7.3.3	宏的示例	093
7.3.4	用户自定义函数示例	096
7.3.5	插件	096
7.3.6	插件的安装	097
第 8 章	Excel 中的投影方法	099
8.1	投影方法	099
8.1.1	概念和符号	099
8.1.2	主成分分析	100
8.1.3	PLS	101
8.1.4	数据预处理	102
8.1.5	教学示例	103
8.2	Chemometrics Add-In 的应用	103
8.2.1	安装	103
8.2.2	概述	103
8.3	主成分分析	104
8.3.1	ScoresPCA	104
8.3.2	LoadingsPCA	105
8.4	PLS	106
8.4.1	ScoresPLS	106
8.4.2	UScoresPLS	107
8.4.3	LoadingsPLS	109
8.4.4	WLoadingsPLS	109
8.4.5	QLoadingsPLS	110
8.5	PLS2	111
8.5.1	ScoresPLS2	111
8.5.2	UScoresPLS2	112
8.5.3	LoadingsPLS2	114
8.5.4	WLoadingsPLS2	115
8.5.5	QLoadingsPLS2	115
8.6	附加函数	116
8.6.1	MIIdent	116

8.6.2	MIdentD2	117
8.6.3	MCutRows	118
8.6.4	MTrace	118
第3部分 化学计量学		120
第9章 主成分分析		121
9.1	基础知识	121
9.1.1	数据	121
9.1.2	直观方法	122
9.1.3	降维	123
9.2	主成分分析	124
9.2.1	格式规范	124
9.2.2	算法	124
9.2.3	主成分分析和奇异值分解	125
9.2.4	得分	125
9.2.5	载荷	126
9.2.6	特殊类型的数据	127
9.2.7	误差	128
9.2.8	验证	130
9.2.9	分解的“质量”	131
9.2.10	主成分数	131
9.2.11	主成分的不唯一性	132
9.2.12	数据预处理	133
9.2.13	杠杆和偏差	133
9.3	人群和国家	133
9.3.1	示例	133
9.3.2	数据	134
9.3.3	数据探索	135
9.3.4	数据预处理	136
9.3.5	得分和载荷的计算	137
9.3.6	得分图	137
9.3.7	载荷图	139
9.3.8	残差分析	140
第10章 校正		143
10.1	基础知识	143
10.1.1	问题陈述	143
10.1.2	线性和非线性校正	144

10.1.3	校正和验证	146
10.1.4	校正“质量”	146
10.1.5	不确定度、精密度和准确度	148
10.1.6	欠拟合和过拟合	149
10.1.7	多重共线性	150
10.1.8	数据预处理	152
10.2	模拟数据	152
10.2.1	线性准则	152
10.2.2	“纯”光谱	152
10.2.3	“标准”样品	152
10.2.4	X数据的创建	153
10.2.5	数据中心化	154
10.2.6	数据概览	154
10.3	经典校正	155
10.3.1	单变量(单通道)校正	155
10.3.2	Vierordt 法	158
10.3.3	间接校正	160
10.4	逆向校正	162
10.4.1	多元线性校正	163
10.4.2	逐步校正	164
10.5	潜变量校正	167
10.5.1	投影方法	167
10.5.2	潜变量回归	170
10.5.3	潜变量校正的执行	172
10.5.4	主成分回归(PCR)	172
10.5.5	潜结构投影-1(PLS1)	174
10.5.6	潜结构投影-2(PLS2)	177
10.6	方法比较	179
第11章	分类	184
11.1	基础知识	184
11.1.1	问题陈述	184
11.1.2	分类的类型	185
11.1.3	假设检验	185
11.1.4	分类中的错误	186
11.1.5	单类别分类	186
11.1.6	训练和验证	187

11.1.7	有监督和无监督训练	187
11.1.8	维数灾难	187
11.1.9	数据预处理	187
11.2	数据	189
11.2.1	示例	189
11.2.2	数据子集	189
11.2.3	工作簿 Iris.xls	190
11.2.4	主成分分析	191
11.3	有监督分类	192
11.3.1	线性判别分析	192
11.3.2	二次判别分析	197
11.3.3	偏最小二乘判别分析	199
11.3.4	SIMCA	203
11.3.5	k -最近邻	208
11.4	无监督分类	210
11.4.1	再一次的主成分分析(再访问)	210
11.4.2	K 均值聚类	211
第 12 章	多元曲线分辨	215
12.1	基础知识	215
12.1.1	问题陈述	215
12.1.2	解的不确定性	217
12.1.3	可解性条件	218
12.1.4	两种类型的数据	219
12.1.5	已知光谱或分布曲线	219
12.1.6	主成分分析	220
12.1.7	主成分分析和多元曲线分辨	221
12.2	模拟数据	222
12.2.1	示例	222
12.2.2	数据	222
12.2.3	主成分分析	223
12.2.4	直观推导式演进特征投影图	225
12.3	因子分析	226
12.3.1	Procrustes 分析	226
12.3.2	渐进因子分析	227
12.3.3	窗口因子分析	231
12.4	迭代方法	233

12.4.1 迭代目标转换因子分析	233
12.4.2 交替最小二乘	235
第4部分 附录	238
第13章 Chemometrics Add-In 的扩展	239
13.1 使用虚拟数组	239
13.1.1 模拟数据	239
13.1.2 虚拟数组	241
13.1.3 数据预处理	241
13.1.4 分解	241
13.1.5 残差的计算	242
13.1.6 本征值的计算	243
13.1.7 正交距离的计算	244
13.1.8 杠杆值的计算	245
13.2 使用 VBA 编程	246
13.2.1 VBA 的优点	246
13.2.2 真实数组的虚拟化	247
13.2.3 数据预处理	247
13.2.4 残差的计算	248
13.2.5 本征值的计算	249
13.2.6 正交距离的计算	249
13.2.7 杠杆值的计算	250
第14章 光谱数据的动力学建模	252
14.1 “灰色”建模方法	252
14.1.1 问题陈述	252
14.1.2 示例	254
14.1.3 数据	254
14.1.4 软交替最小二乘法 (Soft-ALS)	255
14.1.5 硬交替最小二乘法 (Hard-ALS)	257
14.1.6 使用 Solver 插件	258
第15章 MATLAB®: 初学者指南	262
15.1 基础知识	262
15.1.1 工作区	262
15.1.2 基本计算	263
15.1.3 回显	264
15.1.4 工作区的保存: MAT 文件	264
15.1.5 日志	264

15.1.6 帮助	265
15.2 矩阵	266
15.2.1 标量、向量和矩阵	266
15.2.2 访问矩阵的元素	267
15.2.3 矩阵的基本运算	267
15.2.4 特殊矩阵	268
15.2.5 矩阵计算	269
15.3 Excel 和 MATLAB 的集成	270
15.3.1 配置 Excel	270
15.3.2 数据交换	271
15.4 编程	272
15.4.1 M-文件	272
15.4.2 脚本文件	272
15.4.3 函数文件	274
15.4.4 绘图	275
15.4.5 图的打印	276
15.5 示例程序	276
15.5.1 中心化和缩放	277
15.5.2 奇异值分解/主成分分析	277
15.5.3 主成分分析/NIPALS	277
15.5.4 PLS1	278
15.5.5 PLS2	280
后记: 第四范式	283

第 1 部分

导 论

第 1 章

什么是化学计量学？

前导章节

依存章节

矩阵能力

低

统计学能力

低

Excel 能力

低

化学计量学能力

低

Chemometrics Add-In

不使用

附带工作簿

无

化学计量学就是化学计量学家正在做、正在研究的科学。

——一个通俗的观点

化学计量学家就是那些喝啤酒并从统计学家那里窃取思想的人。

——Svante Wold

1.1 化学计量学的主题

“什么是化学计量学”以及“化学计量学是做什么的”，可以有很多种不同的方式对这两个问题进行解释，或简单或复杂。但到目前为止，还没有能被普遍接受的定义，似乎永远也不会有了。最常用的是 D. Massart 的定义——化学计量学是运用了数学、统计学或其他采用形式逻辑的方法来设计或选择最优量测过程和实验，并通过分析化学数据最大程度提取有用化学信息的一门学科 (Chemometrics is the chemical discipline that uses mathematical, statistical and other methods employing formal logic to design or select optimal measurement procedures and experiments, and to provide maximum relevant chemical information by analyzing chemical data)。这个定义遭到了很多批评；于是，有人