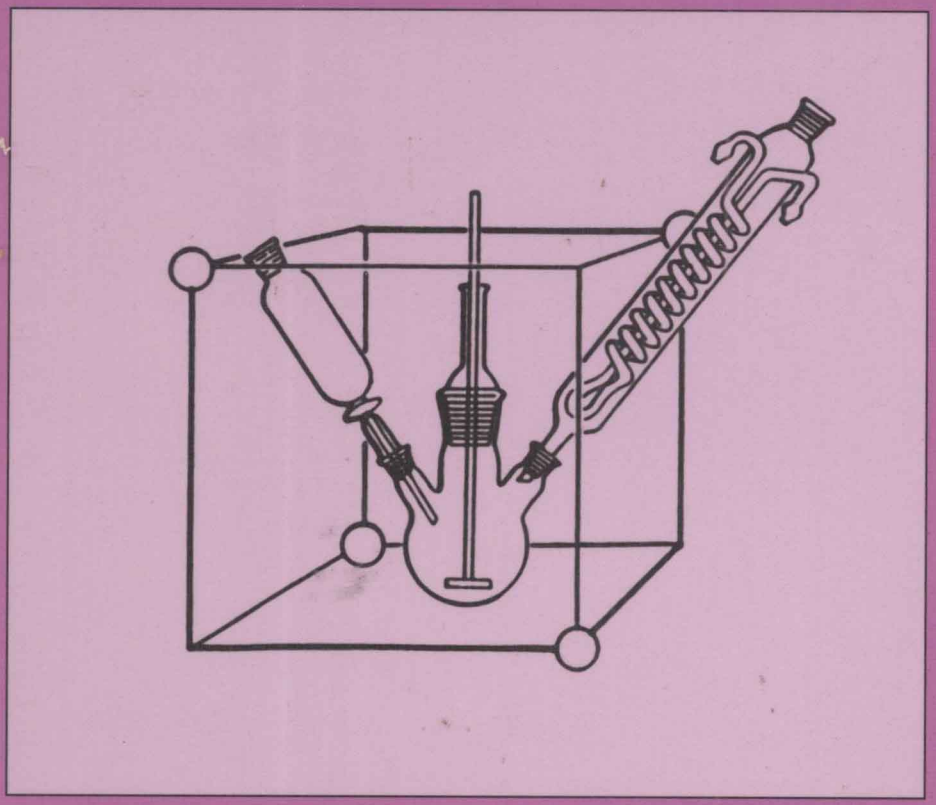
DATA HANDLING IN SCIENCE AND TECHNOLOGY

24

# Design and Optimization in Organic Synthesis
## Second revised and enlarged edition

ROLF CARLSON
JOHAN E. CARLSON



DATA HANDLING IN SCIENCE AND TECHNOLOGY — VOLUME 24
Advisory Editors: S. Rutan and B. Walczak

# Design and optimization in organic synthesis

## Second revised and enlarged edition

**ROLF CARLSON**
*Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway*

**JOHAN E. CARLSON**
*Luleå University of Technology, SE-976 34 Luleå, Sweden*

2005

ELSEVIER

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

**ELSEVIER B.V.**
**Radarweg 29**
**P.O. Box 211, 1000 AE Amsterdam**
**The Netherlands**

ELSEVIER Inc.
525 B Street, Suite 1900
San Diego, CA 92101-4495
USA

ELSEVIER Ltd.
The Boulevard, Langford Lane
Kidlington, Oxford OX5 1GB
UK

ELSEVIER Ltd.
84 Theobalds Road
London WC1X 8RR
UK



1st edition 2005

Library of Congress Cataloging in Publication Data
A catalog record is available from the Library of Congress.

British Library Cataloguing in Publication Data
A catalogue record is available from the British Library.

ISBN: 0-444-51527-5
ISSN: 0922-3497 (Series)

♾ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).
Printed in The Netherlands.

DATA HANDLING IN SCIENCE AND TECHNOLOGY — VOLUME 24

# Design and optimization in organic synthesis

## Second revised and enlarged edition

DATA HANDLING IN SCIENCE AND TECHNOLOGY

**Advisory Editors:** S. Rutan and B. Walczak

---

Other volumes in this series:

| | |
|---|---|
| **Volume 1** | Microprocessor Programming and Applications for Scientists and Engineers, by R.R. Smardzewski |
| **Volume 2** | Chemometrics: A Textbook, by D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte and L. Kaufman |
| **Volume 3** | Experimental Design: A Chemometric Approach, by S.N. Deming and S.L. Morgan |
| **Volume 4** | Advanced Scientific Computing in BASIC with Applications in Chemistry, Biology and Pharmacology, by P. Valkó and S. Vajda |
| **Volume 5** | PCs for Chemists, edited by J. Zupan |
| **Volume 6** | Scientific Computing and Automation (Europe) 1990, *Proceedings of the Scientific Computing and Automation (Europe) Conference, 12–15 June, 1990, Maastricht, The Netherlands,* edited by E.J. Karjalainen |
| **Volume 7** | Receptor Modeling for Air Quality Management, edited by P.K. Hopke |
| **Volume 8** | Design and Optimization in Organic Synthesis, by R. Carlson |
| **Volume 9** | Multivariate Pattern Recognition in Chemometrics, illustrated by case studies, edited by R.G. Brereton |
| **Volume 10** | Sampling of Heterogeneous and Dynamic Material Systems: Theories of Heterogeneity, Sampling and Homogenizing, by P.M. Gy |
| **Volume 11** | Experimental Design: A Chemometric Approach (Second, Revised and Expanded Edition) by S.N. Deming and S.L. Morgan |
| **Volume 12** | Methods for Experimental Design: Principles and Applications for Physicists and Chemists, by J.L. Goupy |
| **Volume 13** | Intelligent Software for Chemical Analysis, edited by L.M.C. Buydens and P.J. Schoenmakers |
| **Volume 14** | The Data Analysis Handbook, by I.E. Frank and R. Todeschini |
| **Volume 15** | Adaption of Simulated Annealing to Chemical Optimization Problems, edited by J. Kalivas |
| **Volume 16** | Multivariate Analysis of Data in Sensory Science, edited by T. Næs and E. Risvik |
| **Volume 17** | Data Analysis for Hyphenated Techniques, by E.J. Karjalainen and U.P.Karjalainen |
| **Volume 18** | Signal Treatment and Signal Analysis in NMR, edited by D.N. Rutledge |
| **Volume 19** | Robustness of Analytical Chemical Methods and Pharmaceutical Technological Products, edited by M.W.B. Hendriks, J.H. de Boer, and A.K. Smilde |
| **Volume 20A** | Handbook of Chemometrics and Qualimetrics: Part A, by D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, and J. Smeyers-Verbeke |
| **Volume 20B** | Handbook of Chemometrics and Qualimetrics: Part B, by B.G.M. Vandeginste, D.L. Massart, L.M.C. Buydens, S. de Jong, P.J. Lewi, and J. Smeyers-Verbeke |
| **Volume 21** | Data Analysis and Signal Processing in Chromatography, by A. Felinger |
| **Volume 22** | Wavelets in Chemistry, edited by B. Walczak |
| **Volume 23** | Nature-inspired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks, edited by R. Leardi |

*To Inger...*

# FOREWORD BY TREVOR LAIRD

I first came across the work of Rolf Carlson in the early 1980's, when I was interested in statistical methods of optimisation as an important technique in process R&D. In fact when I wrote articles on chemical development issues [1,2] I often referred to Rolf's work, since it was published in journals which not everyone would have read (e.g. Acta Chem Scand) and I wished to spread the news that Rolf was a radical thinker with lots of new ideas.

I was therefore delighted when the first edition of "Design and Optimization in Organic Synthesis" was published in 1992. It was always on my list of recommended books for development chemists. In the 1990's, the popularity of using factorial designs and other statistical methods to optimise organic chemical reactions and processes increased, and became "standard practice" in many organisations. These methods are now used routinely in most pharmaceutical companies for optimisation, fine tuning and in validation programmes.

I first had the opportunity to meet Rolf when he visited the UK to give a talk at the annual SCI Process Development Symposium; I witnessed for the first time his unique lecturing style. Later on I visited Norway and met him several times and observed that he is always pushing into new territory. His publications in the 1980's on principal component analysis were ahead of their time and have only recently been taken up in industry, where they have proved of value in many areas.

It was during a visit to Norway that Rolf asked me if I would write the foreword for the new edition; I had no hesitation in accepting. His work has had a great impact on industrial chemistry; it is regrettable that the methodologies he has helped to develop are not taught more widely in universities especially outside Scandinavia!

It is 12 years since the first edition of "Design and Optimization in Organic Synthesis" and there have probably been more publications on statistical methods of optimisation applied to organic chemistry in the last decade than all previous decades put together. The journal "Organic Process Research and Development", has provided a forum for papers discussing optimisation using statistical methods.

The revised updated edition of "Design and Optimisation in Organic Synthesis" is most welcome, and timely and I hope it will be even more successful than the first edition. It is appropriate that his son has been involved in the revision and updating. The new edition should help to continue spreading the word about the value of "design of experiments" in organic chemistry.

Trevor Laird
Scientific Update and Editor,
Organic Process R&D journal

## References:

[1] Laird, T, The neglected science of chemical development, Chemistry in Britain, 1208 (1989)

[2] Laird, T, Development and scale up of processes for the manufacture of new pharmaceuticals, in Comprehensive Medicinal Chemistry, P G Sammes and J B Taylor (eds), Pergamon Press, Vol. 1, p. 321

# PREFACE TO THE SECOND EDITION

## Objectives

The first edition was intended as an introductory textbook on how to use statistically designed experiments and multivariate methods in experimental synthetic organic chemistry.

This is also the objective of the second edition. We have maintained the didactic style used in the first edition in the second edition. When the book was first published twelve years ago, statistically designed experiments in synthetic chemistry were not very well known in the academic world and it was felt that an introductory textbook for synthetic chemists was needed. Unfortunately, the situation in academia is much the same today [1]. In process organic chemistry, however, the situation is different. Over the past decade, there has been an increasing use of statistically designed experiments and multivariate modelling for the development of new synthetic procedures and for optimising reaction conditions. It was felt that an upgraded text was needed, both as a means for evangelisation in academia and as a textbook for chemists engaged in process development.

After some discussion, it was decided that the new edition should be written by two authors: Rolf Carlson and Johan E. Carlson. We, the authors, wish to explain the reason for this: Rolf Carlson has a Ph.D. degree in organic chemistry. He is professor of organic chemistry (organic synthesis) at the University of Tromsø, Tromsø, Norway. Johan E. Carlson has an M.Sc. degree in computer science and engineering and his Ph.D. degree in electrical engineering. He is assistant professor at Luleå University of Technology, Luleå, Sweden. His research is in the field of statistical signal processing and data analysis for ultrasonic measurement systems. We are family, father and son, and our vivid discussions on multivariate problems at the dinner table have brought us to the conclusion that our competencies and experimental backgrounds are complementary and that this might cross-fertilise a new edition of the book.

## New in the second edition

The new edition of the book has been revised and enlarged. We have been able to keep the length of the book similar to the first edition. This is due to: (a) A stricter typography, where the text has been produced by LATEX, with the typefaces Computer Modern and Helvetica. (b) Data tables

of descriptors for different classes of compounds have been removed from the book and placed as Excel and text files on the CD that accompanies the book.

The contents of the old chapters have been upgraded and explanatory comments and new examples have been added. New sections have also been included. A section on the use of Ishikawa diagrams for identifying variables in a process has been added to Chapter 4. A section on links between a theoretical model and the experimental response surface has been added to Chapter 12. Three new chapters have been added: Chapter 13 which describes how kinetic parameters can be estimated by sequential response surface modelling. Chapter 19 which presents an new strategy for designing combinatorial libraries. Chapter 20 which summarises some techniques that can be used for optimisation when there are several response variables.

A CD accompanies the book and includes programs and data files.

- Programs: MATLAB routines for: Canonical analysis, extraction of kinetic data from response surfaces, and PLS modelling.
- Data files: Excel and text files with descriptor data for: Solvents, aldehydes, amines, ketones, and Lewis acids.

All models in the book have been recalculated to detect errors in the first edition. Formulae, schemes and figures have been redrawn. Formulae were drawn using ChemDraw, graphs and plots have been made with graphic routines of MATLAB, more "artistic" drawings were produced using CorelDraw.

The examples given in the book have largely been taken from our own experimental studies. This is not made to highlight our own results, there is another reason. With our own results we know all the details, the past history, the observations made, which considerations were made in the designs, as well as the analytical procedure. In examples from the literature, such details are sometimes obscured.

Tromsø, September 2004
Rolf Carlson and Johan E. Carlson

# References

[1] T. Laird, "Design of experiments (DoE)", Org. Proc. Res. Dev., Vol. 6, No. 4, p. 337, 2002.

# PREFACE TO THE FIRST EDITION

## Short background to why this book was written

Organic synthesis is an important area in chemical research. New synthetic methods are invented at an ever increasing rate which open up new ways to produce interesting chemical compounds. These factors are of tremendous importance for the practical use of synthetic methods both for academic research and for industrial applications.

All synthetic methods have emerged either as the result of an unforeseen observation or as the result of innovative thought. However, it is rare that the first experiments along a new train of thought give satisfactory results. Much tedious work is often required before a new idea can become established as a synthetic method. It is necessary to explore the reaction conditions to determine how they should be adjusted to obtain optimal results. It is also interesting to find out whether the reaction can be used as a general method for a number of similar substrates. To this end, it is necessary to determine the scope and limitations of the reaction. This in turn calls for more experimentation.

Chemical phenomena are rarely the result of single causes. Instead, a number of factors are likely to be involved and, unfortunately, their influence will depend on still other factors. In order to be able to take such interactions into account, it is necessary to use multivariate methods which allow all pertinent factors to be considered simultaneously, both for designing experiments and for analyzing the result. Knowledge of multivariate methods is not, however, widely spread in the community of synthesis chemists. Therefore, many new methods are still being investigated through poorly designed experiments and hence, new procedures are not properly optimized. Still, the most common method to carry out "systematic studies" is to consider "one factor at a time", although such an approach was shown by R.A. Fisher to be inappropriate over 60 years ago [1], when several factors are to be considered.

I believe that the reason why organic synthesis chemists do not apply statistical principles in their experiments is that they do not know how to use such methods. In general, they do not bother to read text-books on statistics because such books rarely describe how statistics may be relevant to their chemical problems. My personal experience may illustrate this: When I started my chemical career some 20 years ago I was asked by my professor to study and optimize a

chemical reaction. I asked how to proceed and he answered me "You must do it by trial and error". My immediate reaction to this was that there must be a more systematic way. A friend suggested that I should read a book on "Biological Analysis of Variance".[2] However, I was unable to translate Latin squares, crop variation, and effects of fertilizers into my world of three-necked flasks, dropping funnels and mechanical agitation. The pioneering work by George E.P. Box and coworkers published in the fifties [3] would have answered my questions at that time. Unfortunately, their work was obviously unknown to my professor and to the academic circles working with synthetic chemistry. The situation is probably the same today. I am therefore convinced that there is a need for a book which spans a bridge between practical organic synthesis and statistics. Such a book must describe chemical problems as they are seen by the synthesis chemist and must introduce statistical tools so that the results obtained are chemically relevant to the chemist.

The present book is a humble attempt to use these principles as guidelines. Professional statisticians may complain that this is yet another introductory text on statistics, written by a non-statistician. This is true, my devotion and professional training are in the field of synthetic chemistry. However, in exploring this field I felt a need for statistical principles to guide my way. This book is an attempt to transmit my personal experiences to my fellow colleagues. If anyone among them should respond to the message and be stimulated to learn more about statistics, I would feel that my mission as an evangelist had been successful.

The book is intended as an introductory text-book on multivariate methods in experimental organic synthetic chemistry. I have tried to describe how various statistical tools can be applied to common problems encountered when a chemical reaction is elaborated into a synthetic method. The methods are illustrated throughout by examples from organic synthesis. Many of the examples have been taken from my own experiments, not with a view to supporting our own results, but because the reasoning behind the experiment is known in detail. In the examples furnished by others, it is sometimes possible to trace the logic behind the experimental set-up, but quite often certain details remain obscured.

Statistical principles will be presented in the context of chemical examples to show how statistical inferences can be linked to chemical consequences. No previous knowledge of statistics is required.

Some of the reasoning in the book uses matrix formalism. This is for the sake of convenience, since some quantitative relations are more easily expressed in matrix language than otherwise. Readers without any previous experience of matrix calculus may skim those paragraphs in which matrix calculus is used without losing too much of the essential message. As matrix formalism is used in many contexts in physical and theoretical chemistry, it is my hope that this will not be a major obstacle to the reader. For those unfamiliar with matrix calculation, a short Appendix has been included at the end of this book.

Umeå, June 1991
Rolf Carlson

# References

[1] R. A. Fisher and W. A. MacKenzie, J. Agr. Sci. 13 (1923) 311.

[2] G. Bonnier and O. Tedin, Biologisk Variationsanalys Svenska Bokförlaget (Bonniers), Stockholm 1940.

[3a] G. E. P. Box and K. B. Wilson, J. Roy. Stat. Soc. Ser. B 13 (1951) 1;

[3b] G. E. P. Box, Biometrics 10 (1954) 15;

[3c] G. E. P. Box and P. V. Youle, ibid. 11 (1955) 287.

## Acknowledgements

First, we wish to thank Dr. *Trevor Laird* for taking the time to read the manuscript and for writing the Foreword.

Many readers of the first edition have given their comments and criticism and we are grateful for this. Special thanks are due to *S. Stanley Young, Michael R. Emptage,* and *Anna-Lena Sunesson* for their very thorough, detailed, encouraging, and critical comments.

We thank past end present co-workers for their contributions to the examples given. Without their experimental effort and enthusiasm, writing this book would not have been possible.

We thank *Alemayehu Mekonnen* for his careful reading and helpful comments on parts of the manuscript of the present edition.

We thank Senior Lecturer *Elinor Ytterstad*and Professor *Anders Grennberg* for their critical reading and their helpful comments on statistical and mathematical details in the manuscript of the present edition.

We thank Professor *Roger Phan-Tan-Luu* for the possibility to use his library for collecting useful references.

We thank the *Journal of Chemometrics*, *New Journal of Chemistry* and *Acta Chemical Scandinavica* for their kind permission to reproduce details from published works as illustrations in this book. References to these works are given in the text.

We thank Professor *Christian Reichardt*, and *Wiley-VCH Verlag, Weinheim*, for their kind permission to reproduce data of solvent properties.

One of the authors, R. C., thanks the *Norwegian Research Council* for their support in the research field described in the book.

We have dedicated this book *To Inger*. She is the wife of one author and the mother of the other author. Her patience and support in the process of writing this book have been invaluable, and we express our sincerest gratitude to her.

Tromsø, November 2004
Rolf Carlson and Johan E. Carlson

# CONTENTS