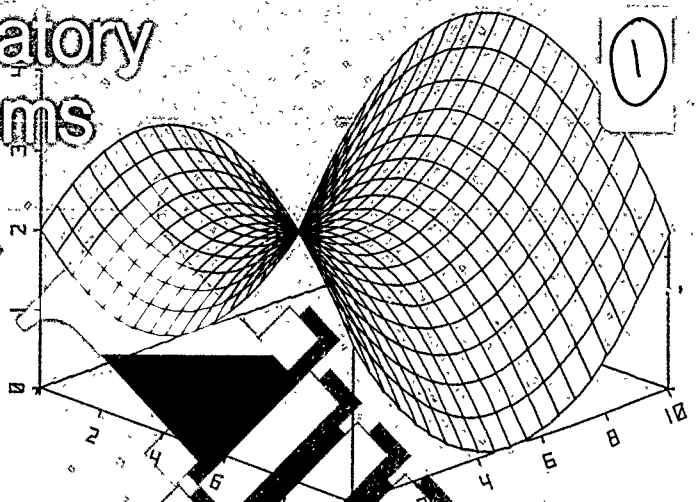
ALS 26665-1-MACF
ISSN 0169-7439
Vol. 10, No. 1-2
February 1991

Chemometrics and intelligent laboratory systems

Proceedings of the Mathematics in Chemistry Conference, College Station, Texas, U.S.A., 8-10 November 1989



Chemometrics and intelligent laboratory systems

An International Journal Sponsored by the Chemometrics Society

SCOPE OF THE JOURNAL

Chemometrics and Intelligent Laboratory Systems publishes articles about new developments on laboratory techniques in chemistry and related disciplines which are characterized by the application of statistical and computer methods.

Special attention is given to emerging new technologies and techniques for the building of intelligent laboratory systems, i.e. artificial intelligence and robotics.

The journal deals with the following topics:

chemometrics: The chemical discipline that uses mathematical and statistical methods to design or select optimal procedures and experiments, and to provide maximum chemical information by analyzing chemical data.

A non-exhaustive list of subjects is:
- statistical methods to evaluate performance characteristics of methods
- proficiency of laboratories and interlaboratory comparisons
- calibration models
- information theory
- correlation techniques and time series analysis
- optimization and experimental design
- regression
- transformation techniques
- deconvolution
- factor analysis
- pattern recognition and clustering
- artificial intelligence and expert systems
- process control
- graph theory
- operations research

computerized acquisition, processing and evaluation of data
- processing of instrumental data
- storage and retrieval systems
- computerized and automated analysis for industrial processes and quality control

robotics

developments in statistical theory and mathematics with application to chemistry

intelligent laboratory systems including self optimizing instruments, planned organic synthesis, data banks with interpretative facilities, and in general applications of expert systems and knowledge representation systems in analytical chemistry

application (case studies) of statistical and computational methods to chemical or related data obtained from natural (medical, geochemical, environmental, food science, pharmacological, toxicological, etc.) and industrial systems (including modelling of processes and quality control)

new software to implement the methods described above and problems associated with the use of software (validation of software for instance)

imaging techniques and graphical software applied in chemistry

PUBLICATION

Chemometrics and Intelligent Laboratory Systems has four volumes (Vols. 10-13) in 1991. The subscription price for 1991 is Dfl. 1232.00 (ca. US$ 733.35), including postage. Subscribers in the U.S.A., Canada, Japan, Australia, New Zealand, P.R. China, India, Israel, South Africa, Malaysia, Thailand, Singapore, South Korea, Taiwan, Pakistan, Hong Kong, Brazil, Argentina and Mexico receive their copies by air delivery. Back volumes (Vols. 1-9) are available at Dfl. 260.00 plus postage. Claims for missing issues will be honoured, free of charge, within three months after publication of the issue. Customers in the U.S.A. and Canada wishing additional bibliographic information on this and other Elsevier journals should contact Elsevier Science Publishing Company Inc., Journal Information Center, 655 Avenue of the Americas, New York, NY 10010. Tel. (212) 633-3750.

Subscription orders may be sent to ELSEVIER SCIENCE PUBLISHERS B.V., P.O. Box 211, 1000 AE Amsterdam, The Netherlands, tel (20)5803911, telex 18582 ESPA NL.

See inside back cover for General information.

MASTER COPY KEEP THIS COPY FOR REPRODUCTION PURPOSES

REPORT DOCUMENTATION PAGE

Form Approved
OMB No 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1991	Final 1 May 89 - 30 Apr 90

4. TITLE AND SUBTITLE
Mathematics in Chemistry Conference

5. FUNDING NUMBERS
DAAL03-89-G-0035

6. AUTHOR(S)
Clifford Spiegelman (principal investigator)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Texas A&M University
College Station, TX 77843

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P. O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARO 26665.1-MA-CF

11. SUPPLEMENTARY NOTES
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This was another important meeting where leading researchers in both the chemical and mathematical sciences exchanged ideas and discussed new results. There was ample time for participants to form new friendships and exchange ideas. One of the main benefits of these meetings is to get to meet and know colleagues from outside disciplines.

(23)

14. SUBJECT TERMS
Chemometrics, Intelligent Laboratory Systems, Chemical Modeling, Statistical Modeling, Statistics, Modeling Conference

15. NUMBER OF PAGES
270

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std 239-18
298-102

VOL. 10, NOS 1-2 CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS FEBRUARY 1991

Abstracts/Contents Lists published in Analytical Abstracts, ASCA, BioSciences Information Service, Chemical Abstracts, Chromatography Abstracts, Current Contents/Engineering, Technology & Applied Sciences, Current Contents/Physical, Chemical & Earth Sciences, Current Index to Statistics (core journal), Excerpta Medica, INSPEC, SCISEARCH, and Cambridge Scientific Abstracts

CONTENTS

MONITOR

Proceedings of the Mathematics in Chemistry Conference, College Station, Texas, U.S.A., 8–10 November 1989

CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS

An International Journal Sponsored by the Chemometrics Society

VOLUME 10, 1991

Available for $110.46 from Elsevier Science Publishers B.V., PO Box 211, 1000 AE Amsterdam, The Netherlands.

Per phonecon 5/13/91
JK

Chemometrics and intelligent laboratory systems

An International Journal Sponsored by the Chemometrics Society

Volume 10, 1991

ELSEVIER, AMSTERDAM—OXFORD—NEW YORK—TOKYO

NORTH-HOLLAND, AMSTERDAM

0169-7439/91/$03 50

This issue is printed on acid-free paper

Printed in The Netherlands

Monitor

Courses

COMETT 2 Project on Chemometrics and Qualimetrics

The EEC has awarded several projects to European chemometricians in its COMETT program. The objective of the COMETT program is to organize industry oriented training on a transnational level in advanced technological subjects. The program is open to all 12 EEC countries but also to the EFTA countries (Norway, Sweden, Finland, Iceland, Austria and Switzerland). Four types of projects were awarded, namely:

- Creation of a network for analyzing training needs, organizing exchanges, publicizing sources and learning material, etc. This network is called Eurochemometrics.
- Exchange of students and staff. Such an exchange must at the same time be transnational and involve both industry and university (for example a university in Belgium can send one of its students to an industry in Switzerland).
- Short course on method validation. This projected is coordinated by Dr. H. Smit (Universiteit van Amsterdam, Laboratorium voor Analytische Scheikunde, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands).
- Demonstration (pilot) project on a package of courses and training materials. (Chemometrics and qualimetrics for the chemical, pharmaceutical and agroalimentary industries).

Except for the course coordinated by Dr. Smit, the projects are coordinated by the author of this news item.

The most important project is, without doubt, the pilot project. It proposes courses on 4 levels:

- Introductory and integration courses. The introduction courses are 2 to 3 day general courses, meant for countries where chemometrics has progressed to a lesser extent. Integration courses are those which comprise chemometrics together with more familiar subjects. An example of such a course is that organized by Prof. Ducauze and Dr. Feinberg in Paris (in French, Institut National Agronomique, Laboratoire de Chimie Analytique, rue Claude Bernard 16, 75231 Paris Cedex 05, France). By teaching a course in which chemometrics is made available in the same program as instrumental laboratory methods, it aims at the integration of chemometrics in the more general knowledge of analytical chemical methodology.
- General long courses These courses are similar to those organized in the earlier, less ambitious COMETT 1 project. The course lasts about 5 days, is organized by different countries in turn and has many lecturers from industry and university. Such schools have been organized earlier in Aix en Provence, Gargnano, Tortosa, Bristol and Bruges. The next school will be organized in or around Nijmegen. (For details, write to Dr. L. Buydens, Laboratorium voor Analytische Chemie, Katholieke Universiteit Nijmegen, Toernooiveld, 6525 ED Nijmegen, The Netherlands.)
- Specialized short courses where subjects can be treated in greater depth. Many subjects are possible, but those that seem to be favoured are experimental design, multivariate calibration, method validation and quality assurance and expert systems. A list of courses available for in-house teaching will also be made available.
- European masters degree. The partners in the project will try to develop a degree, the aim of which is to train chemomet-

ricians with a sufficiently broad knowledge.

The Eurochemometrics consortium will also produce 'distance learning', courseware and teaching aids For instance, Olav Kvalheim will complete his SIRIUS ADVISER program with a videotape introduction course.

The total level of expenditure is about 2 5 million ECU (i e., about US$ 3.2 million) of which the EEC pays half. There are 70 partners (about 30 industries, 30 universities and 10 research institutes). The project is coordinated locally by 12 centres. One of these is devoted to distance learning (coordinator Dr. R. Brereton, University of Bristol, School of Chemistry, Cantock's Close, Bristol BS8 1TS, U.K.) and the other eleven to organizing courses and producing teaching aids and courseware. The list of these centres is given below, together with the name and address of the coordinator(s). Further information can be obtained from the author of this article or from the local centres.

Norway/Denmark. Coordinator: O. Kvalheim, University of Bergen, Department of Chemistry, Realfagbygget, Allegt. 41, 5000 Bergen, Norway

The Netherlands. Coordinator: L. Buydens, Katholieke Universiteit Nijmegen, Laboratorium voor Analytische Chemie, Toernooiveld, 6525 ED Nijmegen, The Netherlands

Sweden/Finland. Coordinator: P. Geladi, Umeå Universitet, Department of Organic Chemistry, 90187 Umeå, Sweden

Austria/Germany/Switzerland. Coordinator: W. Wegscheider, University of Technology Graz, Institut für Analytische Chemie, Tecknikerstrasse 4, 8010 Graz, Austria

France (North). Coordinator: M. Feinberg, I N A., Laboratoire de Chimie Analytique, 16 rue Claude Bernard, 75231 Paris Cedex 5, France

France (South). Coordinator: R. Phan-Tan-Luu, LPRAI Centre de St.-Jérome, Université d'Aix Marseille III, rue Henri Poincaré, 13397 Marseille Cedex 13, France

United Kingdom/Ireland. Coordinator: S J. Haswell, The University of Hull, School of Chemistry, Hull, HU6 7RX, U.K.

United Kingdom. R. Brereton, University of Bristol, School of Chemistry, Cantock's Close, Bristol BS8 1TS, U.K.

United Kingdom. S. Pringle, University of Bristol, Department for Continuing Education, Wills Memorial Building, Queen's Road, Bristol BS8 1HR, U.K.

Italy. Coordinator: M. Forina, Istituto di Analisi e Tecno, Farmaceut. ed Alimentari, Via Brigata Salerno, ponte, 16147 Genova, Italy

Spain/Portugal. Coordinator: F.X. Rius, Universitat de Barcelona, Depart. de Quimica, Pl. Imperial Tarraco 1, 43005 Tarragona, Spain

The first courses to be announced within the COMETT scheme are:

- Chemometrie und künstliche Intelligenz (8–12/4/91 — Ruhr-Universität Bochum). Information: W. Wegscheider
- Optimisation: stratégies et méthodes (13–15/3/91 — Paris). Information: M. Feinberg
- Qualité et validation des méthodes. La bonne pratique de laboratoire (9–11/10/91 — Paris). Information: M. Feinberg
- Echantillonnage et contrôle de qualité dans les industries agroalimentaires (10–12/4/91 — Paris). Information: C. Ducauze
- Information des laboratoires (27–29/11/91 — Paris). Information: M. Feinberg
- Multivariate optimization and experimental design (26–28/5/91). Information: O. Kvalheim
- 7th COMETT School on Chemometrics (date to be announced later — Nijmegen). Information: L. Buydens
- Etude dans un domaine experimental sans contrainte (18–22/3/91 — LPRAI Marseille). Information: R. Phan-Tan-Luu
- Sensibilisation et principes de base (15–19/4/91 — LPRAI Marseille). Information: R. Phan-Tan-Luu
- Formulation et mélanges (3–7/6/91 — LPRAI Marseille). Information: R. Phan-Tan-Luu
- Méthodes modernes d'élaboration de matrices d'epériences optimales (14–18/10/91 — LPRAI Marseille). Information: R. Phan-Tan-Luu
- Sensibilisation et principes de base (18–22/11/91 — LPRAI Marseille). Information: R. Phan-Tan-Luu
- Criblage et étude des facteurs (9–13/12/91 — LPRAI Marseille). Information: R. Phan-Tan-Luu

D.L. MASSART

News

Interlaboratory Testing Award Nominations

Nominations are now being accepted for the 1991 W.J. Youden Award in Interlaboratory Testing, sponsored by the American Statistical Association. Final date for receipt of nominations is April 1, 1991. The W.J. Youden Award in Interlaboratory Testing was established in 1985 to recognize publications that make outstanding contributions to the design and/or analysis of interlaboratory tests or describe ingenious applications to the planning and evaluation of data from interlaboratory tests. The award consists of US $1,000 and a suitable citation.

Eligible publications for the 1991 award must appear in professionally refereed journals or monograph series in 1989–1990. Nominations, along with 6 copies (in English) of the publication, should be sent to the Chair of the Award Committee, Paul vor Doehren, Searle, 4901 Searle Parkway, Skokie, IL 60077, U.S.A.

Book Review

Fourier Transforms in NMR, Optical, and Mass Spectrometry. A User's Handbook, by A.G. Marshall and F.R. Verdun

Elsevier, Amsterdam, 1989, xvi + 450 pages, price Dfl. 220.00, US$ 107.25 (hardcover), Dfl. 95.00, US$ 46.25 (paperback), ISBN 0-444-87360-0 (hardcover), 0-444-87412-7 (paperback)

Fourier transforms are becoming increasingly important for a range of spectroscopic techniques. Some of these techniques, such as NMR and infrared spectrometry, are now performed almost exclusively using FT instruments. The object of this book is to clarify the similarities and differences between the application of Fourier transforms to these different techniques. It provides, for the first time, a unified treatment of the mathematics of Fourier transforms and their application to the three most common forms of FT spectrometry. Despite the few limitations noted below, the aims of this book are achieved admirably.

The style of this book was obviously carefully thought out; the book is both easy to understand and very readable. The use of involved mathematics is avoided except where necessary, and extensive use of illustrations is made to clarify the most difficult points. Physical examples are also given frequently to show the relevance of particular theorems or concepts. A set of problems (with answers) is presented at the end of each chapter. These are particularly useful if the book is being used as a class text, but could also be valuable to readers who wish to consolidate their understanding of the material presented in each chapter. The only significant complaint about the style is that, because of the authors desire to keep the mathematics to a minimum, readers are regularly requested to verify a particular result for themselves. This is often justified, since most of this extra material would rarely be used and its inclusion would simply clutter the text. At other times, however, the added detail would be useful and the fact that readers are required to verify it for themselves could be irritating.

The book consists of ten chapters, the first six of which cover general material, and the last

four of which deal with specific types of Fourier transform spectrometry. Chapter 1 introduces spectral line shapes and explains the Fourier transform relationship between impulse response and continuous oscillation experiments. The origins of absorption mode and dispersion mode spectra are also covered.

Chapters 2 and 3 cover the mathematics of Fourier transforms of both continuous and discretely sampled waveforms. This includes topics such as dynamic range, aliasing, zero-filling, apodization, and phase correction

The stated purpose of Chapter 4 is to deal with experimental aspects that are common to all types of Fourier transform spectrometry. Although this is generally true, a significant amount of the material presented has little or no relevance to FT-optical spectrometry.

Chapter 5 deals with the different sources of noise that can occur in FT spectrometry, and which sources of noise lead to a multiplex advantage or disadvantage. The effects of signal averaging, dynamic range, and apodization on the signal-to-noise ratio are also discussed.

In Chapter 6 non-FT methods for converting data from the time to frequency domain are explained and compared with the FT method. It is worth noting that these initial, general chapters are written mainly in the language of FT-NMR or FT-mass spectrometry, which is not always the same as that of FT-optical spectrometry. Because of this, readers wishing to learn about FT-optical spectrometry (in particular, FT-IR) may find them somewhat confusing, and a rather large portion of the material irrelevant. For readers who are mainly interested in the areas of NMR and/or mass spectrometry, however, these initial chapters provide an excellent and comprehensive introduction to FT spectrometry.

Chapters 7, 8 and 9 deal with aspects of FT spectrometry that are unique to FT-mass spectrometry, FT-nuclear magnetic resonance spectrometry, and FT-optical spectrometry. Of these chapters, that on FT-optical spectrometry is by far the weakest. It is appreciably shorter than the other two chapters, and attempts to deal with FT-infrared, FT-ultraviolet/visible, FT-Raman and Hadamard transform-Raman spectrometries. Consequently, none of these techniques are covered in enough detail to give anything more than a very basic introduction.

Although Chapter 9 is rather poor, the two chapters on FT-NMR and FT-mass spectrometry give a good overview of the current state of the art, and enough information to give a solid grounding in the field of interest.

Chapter 10 provides a brief review of the application of FT methods to other forms of spectrometry. Finally, five appendices are included which give integrals and theorems for FT applications, a description and program listings in FORTRAN and BASIC for the fast Fourier transform algorithm, a comprehensive atlas of Fourier transform pairs, and other useful data. These appendices are a good addition, and mean that the book certainly qualifies as "a user's handbook".

This book is clearly aimed at students and scientists who need to learn about several types of FT spectrometry, and it is an excellent text for this purpose. It should prove to be particularly useful both as a teaching text and as a general reference for Fourier transform methods as they are applied to spectrometry.

For newcomers to the fields of FT-NMR or FT-mass spectrometry this is also an excellent introductory text, which puts the technique of interest into the context of other forms of FT spectrometry. Although those wishing to learn about FT-optical spectrometry may find this book to be rather confusing and the information in it somewhat limited, for those who already have a good grounding in these techniques considerable insight could be gained from the fresh look at old material.

Overall, this is a book to be recommended, and it should prove to be a valuable addition to many spectroscopists' bookshelves.

RICHARD S. JACKSON and
PETER R. GRIFFITHS
Department of Chemistry,
The University of Idaho,
Moscow, ID 83843, U.S.A.

Meeting Report

MADLUST 90, Chemometrics Towards 20000, Tromsø, Norway, 2–6 July 1990

MADLUST 90 was the third in a series of workshop seminars on chemometrics. The previous two, ASTMULD (1984) and MULDAST (1987) were very much local to Scandinavian chemometricians developing the theories and tools now widely accepted throughout the world. For MADLUST, the organisers (Kim Esbensen, Norway, Paul Geladi and Michael Sjöstrom, Sweden, and Pentti Minkkinen, Finland) took a worthwhile decision to broaden the focus of the meeting to include people from industry who apply chemometrics to their particular problems. The hope was, of course, that the two groups would spark ideas off each other. The hope was well realised. The meeting was organised around four main themes: Process Chemometrics, Statistics and Chemometrics, Chemometrics Towards 2000, and Image Analysis in Chemometrics. Each theme occupied a day and discussions on the theme were focused by presentations from a small group of speakers. This arrangement meant that plenty of time was available for discussion.

The Process Chemometrics session was perhaps the major innovation of the meeting. The presentations were by John MacGregor (MacMaster University, U.S.A.), Roy Tranter (Glaxo Manufacturing Services, U.K.), Randy Pell (University of Washington, U S A.) and a trio from the University of Washington, U.S.A., representing the Center for Process Analytical Chemistry (Jim Burger, Marybeth Seasholtz and Yondong Wang). The presentations are subsequent discussions highlighted three major areas, two of which are not normally considered by chemometricians. The interface between the process operator and chemometrics is very important and determines the acceptability of the method and, hence, its overall success. Part of the interface is the presentation of the results from the chemometrics and the concept of having a visible, variable-sized dustbin for all unexplained or unexpected effects proved to be novel and challenging to some. The third area — locally weighted models — has proved valuable but clearly needs more theoretical development to be generally applicable.

The session on Statistics and Chemometrics was more concerned with the theoretical development of chemometric and was presented by Tormod Naes (MATFORSTK, Norway), Age Smilde (University of Groningen, The Netherlands), Hans Berntsen (SINTEF, Norway) and Agnar Höskuldsson (DIA-M, Denmark). Four quite different subjects were discussed: local modelling, the analysis of three dimensional data arrays, the relation of the extended Kalman filter with bi-linear modelling and the optimisation of selecting t-vectors for inclusion in a PLS model. Each created considerable discussion and the first two, at least, showed how some of the problems highlighted in the first session could be resolved.

Chemometrics Towards 2000 allowed elements of art and culture to be introduced into chemometrics as well as consideration of some of the problems facing chemometrics. Erik Johansson (Hässle AB, Sweden), Willem Windig (Eastman Kodak, U.S.A.) and Harald Martens (Consensus Analysis A/S, Norway) raised the issues of the image of chemometrics in managers' minds. There is a need for simplicity of approach and the incorporation of techniques from outside chemometrics, if chemometrics is to survive and develop as a viable subject. These, and the major discussion session subtitled "The Chemometrics User Speaks Back" were a highlight of the week as they clarified a number of ideas that could increase the acceptability and usefulness of chemometrics in many areas, particularly in industry.

The final session, Image Analysis, was presented by Ewart Bengtson (Centre for Image Analysis, Sweden) and Hans Grahn (University of Umeå, Sweden). Here, the benefits of being able to extract from very large image data sets the parts of an image which are related to each other through chemical, physical or medical factors, were well described. As these techniques are essentially non-destructive as far as samples are concerned, they have potential in process analysis, thus bringing the meeting back to its starting point.

R.L. TRANTER
Glaxo Manufacturing Services Ltd., Barnard Castle, Co. Durham, U.K.

Meeting Announcement

2nd Scandinavian Symposium on Chemometrics Bergen, Norway, 28–31 May 1991

More than 60 contributions were received within the submission deadline (15 January) and we list a small selection below. The full program (second announcement) will be available by 15 February.

Special session: Relations between the latent-variable approach in chemometrics, biometrics, econometrics and psychometrics

P. Horst: *Sixty years with latent variables and still more to come*

J. Birks: *Reconstruction of past lake-water pH from biological data — applications of numerical calibrations to acid-rain research*

H.F.M. Boelens, B. van den Bogaert and H.C. Smit: *Determination of parameter values in a signal model using a matched linear system*

R. Carlson: *Synthetometrics. Recent developments*

L. Eriksson and M. Sjöström: *Rational ranking of chemicals according to environmental risk*

K. Esbensen and P. Geladi: *Multivariate image regression (MIR) — principal component regression for modeling and prediction*

K. Faber, L. Buydens and G. Kateman: *Determination of the number of significant factors in a data matrix*

P. Geladi: *A comparison of classification methods as applied to chemical multivariate image analysis*

M. Gerritsen, L. Buydens, B. Vandeginste and G. Kateman: *Quantitative multivariate analysis of HPLC-UV data by GRAM and ITTFA*

H. Grahn and J. Säf: *MRI, NMRS and MIR*

J. Havel, A. Hrdlicka, C. Moreno and M. Valiente: *Evaluation of ICP-AES multicomponent trace analysis data by PLS calibration*

S. de Jong: *Principal Covariates Regression*

J. Jonsson, M. Sandberg, S. Rännar, M. Sjöström and S. Wold: *Parametrization of nucleotides and the use of these characteristics in QSARs for regulatory DNA sequences*

E.J. Karjalainen and U.P. Karjalainen: *Simultaneous analysis of multiple GC runs and samples with alternating regression*

N. Kettaneh-Wold: *Mixture design and PLS modelling — some industrial applications*

O.M. Kvalheim, Yi-zeng Liang and T.V. Karstang: *A full rank solution to evolving factor analysis using selectivity and latent projections*

Y.-Z. Liang: *Qualitative and quantitative analysis of multicomponent data — methods for treating white, grey and black analytical systems*

R. Manne and T.V. Karstang: *Optimal scaling — a solution to the "size" problem in multivariate calibration*

H. Martens, B. Alsberg and E. Stark: *Multivariate preprocessing of NIR spectra by EMSC and SIS*

D.L. Massart, H. Keller and B. Bourguignon: *An operation research approach to multicriteria decision making*

P. Minkkinen: *Optimization of environmental emission measurement plans*

Å. Nordahl. *Computer controlled optimization of organic synthetic reactions*

R. Tranter: *Process monitoring and meaningful numbers*

B. Skagerberg: *Multivariate statistical process control (MSPC):*

A.K. Smilde, C.H.P. Bruins, P.M J. Coenegracht and D.A. Doornbos: *Combination of factorial design and three-way analysis to elucidate the influence of free silanol groups of the stationary phase on retention in RP-HPLC*

V.M. Taavitsainen: *Nonlinear multivariate data analysis*

N. Vogt: Quality by Design

W. Windig and C.E. Heckler: *Simple-to-use interactive self-modelling mixture analysis, new developments and applications in industry*

S. Wold: *Nonlinear PLS with splines*

R.-Q. Yu: *Chemometrics in China*

For registration (deadline 12 April) or information, contact: Laila Kyrkjebø or Olav M. Kvalheim, Department of Chemistry, University of Bergen, N-5007 Bergen, Norway. Fax: +47 5-329058

PROCEEDINGS OF THE MATHEMATICS IN CHEMISTRY CONFERENCE, COLLEGE STATION, TEXAS, U.S.A., 8–10 NOVEMBER 1989

CONTENTS

Proceedings of the Mathematics in Chemistry Conference, College Station, Texas, U.S.A., 8–10 November 1989

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 11–12
Elsevier Science Publishers B V., Amsterdam

Organizer's summary

This was another important meeting where leading researchers in both the chemical and mathematical sciences exchanged ideas and discussed new results. There was ample time for participants to form new friendships and exchange ideas. One of the main benefits of these meetings is to get to meet and know colleagues from outside disciplines. Participants enjoyed wine tasting at a local winery during the second night of the conference. During the first night the participants had a banquet dinner with after dinner speaker Dr. Herbert Hauptman co-winner of the 1985 Nobel Prize in chemistry. He at my request, gave a frank discussion of the difficulty of getting chemists to accept his and Karle's results. Part of these difficulties are presented in the written version of his talk. Dean Abe Clearfield of Texas A&M and Dr. E. Prince of the National Institute of Standards and Technology at my request have included in the proceedings their comments that followed Dr. Hauptman's talk.

As was the case at the 1985 Chemometrics Research Conference that I coorganized most invited talks had invited discussants. Chemists' talks were discussed by a mathematician and mathematicians' talks were discussed by a chemist. Some speakers were hard to classify as belonging to one field or the other. The main focus of the invited discussions was to explain and expand upon the main presentation to the broad audience.

The opening session was moderated by Lloyd Currie of NIST and the opening speaker was Leon Gleser whose talk demonstrated to the conference that measurement error models are often useful. The second talk was by Anne Thompson and she discussed chemical and statistical modeling to environmental science.

The second session dealt with making sense from multivariate data. Peter Jurs gave a survey of the use of clustering procedures in his laboratory.

The third session dealt with modeling in chemistry. Professor Steve Brown of the University of Delaware gave his change of time series procedures for calibration while Professor Don Watts of Queens University demonstrated how useful profile t and trace plots can be in obtaining interval estimates. The fourth session dealt with statistical mechanics issues during which the audience was treated to interesting fractal plots and interpretations. The Speakers were Fereydoon Family from Emory University, Dan ben-Avraham from Clarkson University, and David Weitz from Exxon Research Labs.

The sixth session gave an interesting description of how a graduate student in statistics working with a distinguished electrochemist can impact chemistry. This talk was given by Janet Osteryoung. The second talk at this session was also based upon joint work by an agricultural chemist and a statistician. They gave interesting case study examples of where PLS would and would not work. The third talk was an interesting statistical layout of receptor modeling given by Karen Bandeen-Roche.

In the next session Phil Hopke gave a tutorial on the use of receptor modeling and Ron Henry gave a lecture about the use of optimization methods in environmental modeling.

We had a dynamic session on structural modeling that included talks by Ted Prince of the NIST, Macolm Gerloch of Cambridge University, and Milan Randić of Drake University. Ted talked about the use of maximum entropy techniques to resolve structure. (Ted says that since the conference he and some colleagues have made important advances.) Macolm talked about ligand field theory and the electronic structure of inorganic complexes and Milan gave an interesting

talk on the use of graph theory as a companion procedure to more often used clustering techniques.

The final session was about multivariate analysis and design. It was enjoyed by all. Probably a humorous thing that many will remember for a long time is the 'honors' that Cris Nachheim tacked onto his name with the abstract such as FRS and ASPCA among others. Cris gave an interesting talk on experimental design and Pat Carey gave examples of successful application of PLS methods at Los Alamos.

C.H. SPIEGELMAN

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 13–18
Elsevier Science Publishers B.V., Amsterdam

History of X-ray crystallography

Herbert A. Hauptman
Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, NY 14203 (U.S.A.)

(Received 11 July 1990; accepted 19 July 1990)

Abstract

Hauptman, H.A., 1991. History of X-ray crystallography. *Chemometrics and Intelligent Laboratory Systems*, 10, 13–18

In this brief sketch of the history of X-ray crystallography I emphasize the important role played by the development of the direct methods which were devised to solve the central problem of X-ray crystallography, the so-called phase problem. I also stress the importance of cross disciplinary research, in particular the essential role which mathematics played in this development.

INTRODUCTION

In 1895 Wilhelm Röntgen discovered X-rays. With this discovery the stage was set for the creation of the modern science of X-ray crystallography.

In 1912 Paul Ewald was completing his doctoral dissertation concerned with the optical properties of a medium consisting of a regular arrangement of isotropic resonators. A crystalline solid which, on the sub-microscopic level, consists of a triply periodic, regular arrangement of atoms, or molecules, is therefore precisely the kind of medium with which Ewald was concerned. Since the smallest interatomic distances in a crystal are of the same order of magnitude as the wavelengths of X-rays, it occurred to Max von Laue, upon learning of Ewald's results, that a crystal might serve as a three-dimensional diffraction grating for X-rays. In order to test this hypothesis he prevailed upon the younger physicists Walter Friedrich and Paul Knipping to perform the necessary scattering experiment.

The scattering experiment indeed showed that when a beam of X-rays strikes a crystal, the crystal scatters the incident beam in many different directions and with different intensities. If these scattered X-rays strike a photographic plate they will blacken the plate at those points where the scattered rays strike the plate. In this way one obtains the so-called diffraction pattern. This experiment marked the birth of the science of X-ray crystallography and, because of its fundamental importance in determining crystal and molecular structures, must be regarded as a landmark event in twentieth century science. The major obstacle in the path leading from the observed diffraction pattern to the desired crystal structure is known as the phase problem, for reasons to be given shortly. I propose here to give a brief historical account of the methods devised to overcome this obstacle, the so-called direct methods of X-ray crystallography.

0169-7439/91/$03 50

THE DIFFRACTION PATTERN

It has already been remarked that a crystal may be regarded as a regular triply periodic arrangement of an array of atoms. One imagines three families of planes, the planes in each family being parallel to and equidistant from one another. In this way one obtains a tiling of the crystal space by means of congruent parallelepipeds each one of which is said to be a fundamental parallelepiped, or unit cell, of the crystal.

If each unit cell contains a molecule — a collection of atoms — in its interior, and if the atoms are arranged in precisely the same way in all the unit cells, then each unit cell and its contents are indistinguishable from every other unit cell and its contents.

There corresponds to each atom an electron density function; hence, by superposition of the individual atomic electron density functions, one obtains an overall electron density function $\rho(r)$, a nonnegative function of the position vector r which gives the number of electrons per unit volume at the position r. It is clear from the geometric construction that the electron density function in any unit cell is identical to that in every other unit cell. Hence $\rho(r)$ is a triply periodic function of position, and this property may be taken as the mathematical definition of a crystal.

If on the other hand we choose to regard a crystal as a triply periodic arrangement of an array of atoms, or molecules, then by a crystal structure we mean simply the arrangement and identities of the atoms in the unit cell and by a molecular structure the arrangement and identities of the atoms in the molecule.

It was recognized almost from the beginning that the diffraction pattern, that is the directions and intensities of the X-rays scattered by a crystal, is uniquely determined by the crystal structure; which is to say that if one knew the crystal structure — the arrangement of the atoms in the crystal — then one could calculate the diffraction pattern completely. It turns out that, conversely, diffraction patterns in general determine unique crystal and molecular structures, although this fact was not known until many years later. In short, the information content of a typical molecular structure coincides precisely with the information content of its diffraction pattern. It is a measure of the great advances made by the new science of X-ray crystallography that one nowadays can routinely transform the information content of a diffraction pattern into a molecular structure, at least for the so-called 'small' molecules, that is those consisting of some 150 or fewer non-hydrogen atoms.

THE PHASE PROBLEM

Since X-rays, like ordinary visible light, are electromagnetic waves, they have a phase as well as an intensity, just as any other wave disturbance. In order to work backwards, from diffraction patterns to crystal and molecular structures, it turns out to be necessary to measure not only the intensities of the X-rays scattered by the crystal but their phases as well. However, the phases cannot be measured in the ordinary kind of diffraction experiment; they appear to be irretrievably lost. Only the intensities can be directly measured. This then gives rise to the central problem of X-ray crystallography, the so-called phase problem, how to deduce the values of the phases of the X-rays scattered by a crystal when only their intensities are known. For some forty years after the landmark experiment of Friedrich and Knipping, all attempts to find a general method for going directly from the diffraction pattern, that is measured intensities alone, to the crystal structure, with or without the intervention of the phases — a method that would be useful for the complex structures of interest to chemists, biologists, and mineralogists — were defeated.

In fact, because the needed phase information was lost in the diffraction experiment, it was thought that one could use arbitrary values for the phases associated with the measured intensities of the scattered X-rays. In this way one obtains a myriad of different crystal structures, all consistent with the known intensities. It therefore came to be generally believed that a procedure for going directly from the measured intensities to crystal structures could not, even in principle, be

devised. By the same mode of thinking, the problem of deducing the values of the individual phases from the diffraction intensities, the so-called phase problem, was also thought to be unsolvable, even in principle. It wasn't until the early 1950s, through the exploitation of special properties of molecular structures and through a simple mathematical argument, that these erroneous conclusions were finally refuted.

Atomicity

The special property that all crystal and molecular structures possess may be summed up in one word: atomicity. Thus the electron density function $\rho(r)$ in a crystal takes on large positive values at the atomic position vectors and drops to small values between the atoms. If our goal is merely to determine the positions of the atoms — that is, the positions of the maxima of $\rho(r)$ — rather than the much more complicated electron density function associated with the distribution of atoms in the crystal, then our problem is greatly simplified, it turns out to be not only determinate but actually greatly overdetermined by the available X-ray diffraction intensities.

This is most easily seen by eliminating the lost phase information from the relationships between the diffraction pattern and the crystal structure. Doing this results in a system of equations relating the diffraction intensities alone with the atomic position vectors. Because the number of these relationships far exceeds (by a factor of ten or so) the number of unknown position vectors needed to define the crystal structure, our problem is greatly overdetermined. Thus it is clear that there exist relationships between the measured diffraction intensities and the lost phases that may be exploited. It follows that the phases of the scattered X-rays are also determined by their intensities. In short, the lost phase information is to be found among the available intensities, and the phase problem is therefore a solvable one, at least in principle. There remains the task of devising numerical algorithms leading from the abundance of experimentally measured diffraction intensities to the values of the individual phases. The techniques of X-ray crystallography that deduce the individual phases by exploiting relationships between measured diffraction intensities and phases are known as direct methods.

The argument just presented was in fact anticipated in 1927 by Heinrich Ott [1], who showed by algebraic analysis and applications that the method is capable of solving simple centrosymmetric structures, in which all phases must be either 0 or π. The method was further elaborated by Kedareswar Banerjee in 1933 [2] and Melvin Avrami in 1938 [3] but was clearly of only limited value in applications. While this early work of Ott, Banerjee and Avrami shed important light on the more general phase problem, it attracted little attention at the time and was not further developed, it appears now to be all but forgotten.

Solving the phase problem

My work on this problem started in 1948 about a year after I joined the Naval Research Laboratory in Washington, DC and commenced my collaboration with Jerome Karle. It had been some 35 years since Friedrich and Knipping had carried out their famous experiment, and by 1947 the phase problem, the central problem of X-ray crystallography, was still unsolved and generally regarded as unsolvable. The central importance of this problem and its strong mathematical component combined to provide a challenge that could not be denied.

Then too, there was a certain air of mystery surrounding the problem. On the one hand the simplicity and logic of the argument "proving" its unsolvability, even in principle, appeared to be overwhelming. On the other hand crystal and molecular structures were being solved, although the structures studied were almost always very simple ones involving a small number of atoms or larger structures containing one or a small number of heavy atoms, for which special techniques had been devised. It had not yet been generally understood that the implicit assumption of atomicity and the concomitant trial-and-error approach to most structure solutions had imposed a powerful restriction on the permitted values of the phases.

The first important contribution that Karle and I made was the recognition that it would be neces-

sary to exploit prior structural knowledge to transform the phase problem from an unsolvable one to one that was solvable, at least in principle. Our first step in this direction was to exploit the nonnegativity of the electron density function $\rho(r)$. Before our analysis was complete, however, David Harker and John Kasper published their famous paper [4] in which they derived inequalities in which the measured intensities restrict the possible values of the phases. This was a very mysterious paper, because nowhere in it does there appear any explicit mention of the basis for the inequality relations, and indeed the most important fact is conspicuous by its absence. It is simply that the electron density function is nonnegative everywhere. This fact is, however, implicit in Harker and Kasper's work. In very short order Karle and I completed our own analysis and derived the complete set of inequality relationships based on the nonnegativity of the electron density function [5]. It includes the Harker–Kasper inequalities as a special case, and many others besides. Although the complete set of inequalities greatly restricts the values of the phases, the relations appear to be too intractable to be useful in applications, except for the simplest structures, and their potential has never been fully exploited.

The recognition in 1950 and 1951 that molecules consist of atoms that to a good approximation may be regarded as points completely transformed the nature of the phase problem. While it meant accepting as fact that the observed diffraction intensities by themselves were indeed not sufficient to determine a unique electron density function, it also meant that they were more than sufficient, by far, to determine the atomic position vectors [6]. It meant as well that the phases corresponding to the point atom structure were greatly overdetermined by the available intensities. Finally, it meant that a formidable psychological barrier had been removed, because it now made sense to look for a solution to the phase problem, that is, for numerical algorithms leading from measured intensities to individual phases. In hindsight it is perfectly clear that owing to the great overabundance of diffraction data, a probabilistic approach is called for; some 40 years ago, however, this was not so apparent.

Before we could even get started, an unexpected complication arose. It turned out that because the values of the individual phases clearly depend not only on the crystal structure but also on the choice of origin, they are not uniquely determined by the crystal structure alone. It followed that the diffraction intensities alone do not determine unique values for the phases. The process leading from diffraction intensities to phases would have to include a recipe for specifying the origin. This required that we separate out two contributions to a phase, one due to the crystal structure alone and one due to the choice of origin. We clearly needed to study how a phase is transformed when the origin is shifted, a problem that was complicated by the fact that the permissible origins depend on the crystallographic elements of symmetry, which were usually known in advance.

The solution was made easier by the discovery that there are always certain linear combinations of the phases, the so-called structure invariants, that are uniquely determined by the crystal structure alone and are independent of the choice of origin. It is therefore only the values of the structure invariants that we can hope to estimate from the measured intensities. Once we have estimated a sufficient number of these we can then hope to evaluate the individual phases by a process that incorporates a recipe for specifying the origin.

What was clearly called for was the devising of a method for identifying the structure invariants, and then using these to come up with recipes for fixing the origin appropriate to the different elements of crystallographic symmetry that may be present. Once this was done there would remain the task of estimating the values of the structure invariants by means of their conditional probability distributions, assuming that an appropriately chosen set of diffraction intensities is known.

Probabilistic techniques

Beyond any doubt our most important contribution during the early 1950s was the introduction of probabilistic techniques — in particular, use of the joint probability distribution of several diffraction intensities and the corresponding phases

— as the central tool in the solution of the phase problem [7]. We assumed to begin with that all positions of the atoms in the unit cell of the crystal were equally likely, or, in the language of mathematical probability, that the atomic position vectors were random variables, uniformly and independently distributed. With this assumption the intensities and phases of the scattered X-rays, as functions of the atomic position vectors, are also random variables, and one can use the methods of modern mathematical probability theory to calculate the joint probability distribution of any collection of intensities and phases. A suitably chosen joint probability distribution leads directly to the conditional probability distribution of a specified structure invariant, assuming again an appropriately chosen set of diffraction intensities. The conditional distribution in turn leads to the structure invariant, an estimate of which is given, for example, by its most probable value. Once one has a sufficiently large number of sufficiently reliable estimates of structure invariants, one can use standard techniques to calculate the values of the individual phases, provided that the process incorporates a recipe for specifying the origin.

Although probabilistic methods played an essential role in the development of the direct method and provided it with its logical foundation, it must also be pointed out that non-probabilistic methods also played an important part. In this connection the early work of Sayre [8], Zachariasen [9], Cochran [10] and Woolfson [11] should be mentioned. In particular the well known Sayre equation, a relationship of fundamental importance among measured magnitudes and unknown phases, continues to be useful to the present day and lies at the heart of some of the more successful computer programs for solving crystal structures.

CONCLUDING REMARKS

I cannot conclude this brief account of the early history of the direct methods of X-ray crystallography without also describing the reception this work received at the hands of the crystallographic community. This was, simply, extreme skepticism, if not outright hostility. In hindsight I think this reaction was due, first, to the strong mathematical flavor of this early work, not well understood by most crystallographers, as well as the ingrained and almost universal belief that the phase problem was unsolvable in principle and that any claim to the contrary must therefore be flawed. This nearly universal skepticism and inability to understand the proposed solution no doubt explains why so few early attempts to apply the new methods were made. It wasn't until the 1960s, when easy to use computer programs became available, that widespread applications were made.

Today some 100 000 molecular structures are known, most determined by the direct methods, and about 5 000 new structures are added to the list every year. It is no exaggeration to say that modern structural chemistry owes its existence to this development.

Although no equations are shown in this article, it should be clear that the developments described here would not have been possible without strong dependence on mathematical techniques, in particular the modern theory of mathematical probability, and it is this interaction between mathematics and the phase problem of X-ray crystallography which I have tried to emphasize in this article. Work on the phase problem continues to this day and applications to structures of ever increasing complexity continue to be made. It still appears that progress is made only in proportion to our ability to bring more powerful mathematical techniques to bear on this fascinating problem.

ACKNOWLEDGEMENT

This work was supported by the National Science Foundation grant number CHE-8822296.

REFERENCES

1 H. Ott, Structure analysis, *Zeitschrift fur Kristallographie*, 66 (1927) 136–153.
2 K. Banerjee, Determination of the signs of the Fourier terms in complete crystal structure analysis, *Proceedings of*

the Royal Society of London, A; Mathematical and Physical Sciences, 141 (1933) 188–193.
3 M. Avrami, Direct determination of crystal structure from X-ray data, *Physical Review*, 54 (1938) 300–303.
4 D. Harker and J.S. Kasper, Phases of Fourier coefficients directly from crystal diffraction data, *Acta Crystallographica*, 1 (1948) 70–75.
5 J. Karle and H. Hauptman, The phases and magnitudes of the structure factors, *Acta Crystallographica*, 3 (1950) 181–187.
6 H. Hauptman and J. Karle, Relations among the crystal structure factors, *Physical Review*, 80 (1950) 244–248.
7 H. Hauptman and J. Karle, *Solution of the Phase Problem I. The Centrosymmetric Crystal*, American Crystallographic Association Monograph No. 3, Polycrystal Book Service, Dayton, OH, 1953.
8 D. Sayre, The squaring method: a new method for phase determination, *Acta Crystallographica*, 5 (1952) 60.
9 W.H.A. Zachariasen, A new analytical method for solving complex crystal structures, *Acta Crystallographica*, 5 (1952) 68.
10 W. Cochran, A relation between the signs of structure factors, *Acta Crystallographica*, 5 (1952) 65.
11 M. Woolfson, The statistical theory of sign relationships, *Acta Crystallographica*, 7 (1954) 61.

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 19
Elsevier Science Publishers B.V., Amsterdam

Comments on "History of X-ray crystallography" by Herbert A. Hauptman

E. Prince

National Institute of Standards and Technology, Gaithersburg, MD 20899 (U.S.A.)

My own career as a crystallographer corresponds very closely with the development of direct methods of phase determination. In fact my first exposure to crystallography was in the summer of 1949 when, freshly out of college, I had a temporary job in the laboratory of David Harker and John Kasper, who had recently completed the determination of the structure of decaborane, the first structure to be determined ab initio from diffraction data alone. I was an interested spectator during the early 1950s, when the work of Herbert Hauptman and Jerome Karle was the subject of sometimes bitter controversy, and I have a particularly vivid memory of an American Crystallographic Association meeting that was held at Harvard in the spring of 1954. (I can be absolutely positive about the date, because I was working at the time at Bell Labs, in New Jersey, while my fiancee was teaching in a school in the Boston suburbs. I had a strong incentive to get to that meeting.) The program at this meeting had a series of half a dozen paper whose titles were variations on the theme "Why the methods proposed by Hauptman and Karle won't work." These were followed by a paper by Clark, Evans and Christ, of the U.S. Geological Survey, entitled "The Structure of Colemanite, Solved Using the Methods of Hauptman and Karle." This paper did not completely silence the opposition (I remember also a rather sharp exchange between Jerry Karle and Michael Woolfson, who was later to become one of the leaders in the development of direct methods, at a meeting at Cornell in 1959), but acceptance of the ideas of direct methods had become quite general by the early 1960s.

0169-7439/91/$03.50

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 20
Elsevier Science Publishers B V., Amsterdam

Reminiscences

Abe Clearfield

Department of Chemistry, Texas A&M University, College Station, TX 77843 (U S.A.)

I remember well as a student, attending the first presentation to the crystallographic community, by Herb Hauptman and Jerry Karle, of their ideas on solving the phase problem. I believe it was an American Crystallographic Association Meeting at the University of Michigan. We were all assembled in a large auditorium and as Dr. Hauptman has stated, the presentation was quite mathematical. At the completion of the talks, there was a moment of stunned silence, then many hands shot up tc ask questions, I thought. Instead each of the then leading lights of crystallography felt obligated to reveal their own brilliance by putting these two young upstarts on their place. They began to criticize the methods and tried to point out the fallacy in the Hauptman–Karle approach. During this heated discussion, my major professor, Dr. Philip Vaughan leaned over and said to me "these guys really have something". Phil was only three years out of Cal Tech having worked with Linus Pauling and then worked as a postdoc with Eddie Hughes. Phil later went on to make his own modest contribution to 'Direct Methods' but then gave up what surely would have been a brilliant career to take over the family geology instruments business.

Much later, when Herb Hauptman came to the Medical Foundation of Buffalo, his initial experimental group included Bill Duax, who worked as a postdoc with me, and Dave Smith my first Ph.D. student. Later my second Ph.D. student, Bob Blessing, joined the group. These now senior level scientists, along with the other bright younger members of the group, have solved some exceedingly different problems in biological systems as part of the overall effort to apply 'Direct Methods' to crystallographic problems. The power of the method is still being developed and gives promise of revealing to us the intricate secrets of both the mineral and living worlds.

0169-7439/91/$03.50

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 21–43
Elsevier Science Publishers B V., Amsterdam

An introduction to receptor modeling

Philip K. Hopke
Department of Chemistry, Clarkson University, Potsdam, NY 13676 (U.S.A.)

(Received 8 November 1989; accepted 15 February 1990)

CONTENTS

Abstract

Hopke, P.K., 1991. An introduction to receptor modeling. *Chemometrics and Intelligent Laboratory Systems*, 10, 21–43.

A major problem facing air quality management personnel is the identification of sources of airborne particles and the quantitative apportionment of the aerosol mass to those sources. The ability to collect particle samples and analyze these samples for a suite of elements by such techniques as neutron activation analysis or X-ray fluorescence provides the data for the problem of resolving a series of complex mixtures into its components based on the profiles of the elements emitted by the various sources in the

0169-7439/91/$03.50

airshed. If all of the sources and their composition profiles are known, then the mass balance model becomes a multiple regression problem If a series of samples have been analyzed without substantial information being available on the sources, factor analysis methods can be employed In both situations, the analysis is complicated by higher levels of measurement error in these analyses than in typical spectrochemical problems. In addition, the source profiles can vary as the composition of input materials for the emission sources change in time Thus, there are limitations to the ability of statistical methods to resolve sources in real world problems The physical and statistical basis of these methods and their application to representative problems will be reviewed

1 INTRODUCTION

The advent of a U.S. national ambient air quality standard for total suspended particles (TSPs) in the early 1970s created the need to identify particle sources so that effective control strategies could be designed and implemented. The initial efforts at identification of particle sources focused on dispersion models of point sources and, in most cases, resulted in substantial reductions in TSP levels. However, as the increment of additional control needed to reach standard levels became smaller, the model uncertainties led to difficulties in identifying the actual sources of continuing problems. In addition, fugitive and other non-ducted emissions are generally not treated or are poorly handled in these models. Thus, additional methods were required to identify and quantitatively apportion particle mass to sources These new methods are called receptor models. In them, the measured properties of the collected ambient samples are used to infer the contributions of the sources to the ambient pollutant concentration These methods require that samples be obtained at locations of interest, receptor sites, and that the samples so collected be analyzed for the properties that are characteristic of the pollutant sources.

These requirements have arisen at a time when new analytical methods have been developed that permit multielemental analysis of large numbers of airborne particle samples or microscopic characterization of large numbers of individual particles. Thus, large data bases on the composition of airborne particles are available for use in these receptor models. Although much of the thrust of the model developments have been aimed at identification of sources of particle mass, they also can be used to elucidate the origins of the various measured species observed in the samples. It then becomes possible to quantitatively apportion the observed airborne concentrations such as airborne lead among the various source types.

The importance of receptor models as air quality management tools in the U.S. has recently been substantially increased by the promulgation of a new ambient air quality standard for particulate matter. This new standard requires all of the state and local air quality planning agencies to revise their plans for improving air quality and reducing the particulate level concentrations where they are expected to exceed the prescribed levels In the associated guidance documents provided by the U.S. Environmental Protection Agency [1], receptor models are explicitly approved for use in this planning process along with the traditional dispersion models. Thus, receptor models have now become an accepted part of the regulatory process for air quality management.

This paper will outline several of the applicable models, provide examples of their use in apportioning materials in a number of airsheds, and demonstrate how they can identify the influence of emissions on the overall airborne particle concentrations.

2 PRINCIPLE OF MASS CONSERVATION

All of the currently used receptor models are based on the assumption of mass conservation and the use of a mass balance analysis. For example, let us assume that the total airborne particulate lead concentration (ng/m^3) measured at a site can be considered to be the sum of contributions from independent source types such as motor vehicles, incinerators, smelters, etc.

$$Pb_T = Pb_{auto} + Pb_{incin} + Pb_{smelter} + \ldots \quad (1)$$

However, a motor vehicle burning leaded gasoline emits particles containing materials other than lead. Therefore, the atmospheric concentration of lead from automobiles in ng/cm^3, Pb_{auto}, can be considered to be the product of two cofactors: the gravimetric concentration (ng/mg) of lead in automotive particulate emissions, $a_{Pb\,auto}$, and the mass concentration (mg/m^3) of automotive particles in the atmosphere, f_{auto}.

$$Pb_{auto} = a_{Pb,auto} f_{auto} \tag{2}$$

The normal approach to obtaining a data set for receptor modeling is to determine a large number of chemical constituents such as elemental concentrations in a number of samples. The mass balance equation can thus be extended to account for all m elements in the n samples as contributions from p independent sources

$$x_{ij} = \sum_{k=1}^{p} a_{ik} f_{kj} \quad i = 1,m \quad j = 1,n \tag{3}$$

where x_{ij} is the ith elemental concentration measured in the jth sample, a_{ik} is the gravimetric concentration of the ith element in material from the kth source, and f_{kj} is the airborne mass concentration of material from the kth source contributing to the jth sample. There are several different approaches to receptor model analysis that have been successfully applied including chemical mass balance (CMB) and multivariate receptor models including principal components analysis and target transformation factor analysis (TTFA). These models can be applied to both particulate and gaseous species. The basis for each of these methods will be presented in subsequent sections of this paper along with examples of their application to the identification of pollution sources in the atmosphere.

3 CHEMICAL MASS BALANCE

3.1 Introduction

The chemical mass balance (CMB) sometimes called the chemical element balance solves eq. (3) directly for each sample by assuming that the number of sources and their compositions at the receptor site are known. This approach was first independently suggested by Winchester and Nifong [2] and by Miller et al. [3]. The composition of an ambient sample is then used in a multuple linear regression against source compositions to derive the mass contribution of each source to that particular sample. Miller et al. [3] modified eq. (3) to explicitly include changes in composition of the source material while in transit to the receptor

$$x_{ij} = \sum_{k=1}^{p} \alpha_{ik} a_{ik} f_{kj} \tag{4}$$

where α_{ik} is the coefficient of fractionation so that if a'_{ik} were the composition of the particles as emitted by the source, a_{ik} is the composition of the particles at the receptor site ($a_{ik} = \alpha_{ik} a'_{ik}$). In practice, it is generally impossible to determine the α_{ik} values and they are assumed to be unity ($a_{ik} = a'_{ik}$).

3.2 Previous applications

Early applications of this approach to urban aerosol mass apportionment included Pasadena, CA [4], Heidelberg, Germany [5], Ghent, Belgium [6], and Chicago, IL [7]. In all of these analyses, the quality of available source compositions severely limited the precision to which the ambient compositions could be reproduced.

Several major research efforts have subsequently resulted in substantially better source data. The source emission studies led to much improved resolution of the particle sources in Washington, DC [8,9]. In one of these studies, Kowalczyk et al. [8] used a weighted least-squares regression analysis to fit 6 sources with 8 elements for 10 ambient samples. In these analyses, the ambient elemental concentrations are weighted by the inverse of the square of the analytical uncertainty in that measurement.

Subsequently, Kowalczyk et al. [9] examined 130 samples using 7 sources with 28 elements included in the fit. Although 28 elements were used in the fitting process, the fit did not change appreciably with varying numbers of elements in-

cluded with the exception of some of the key tracer elements such as Na, Pb, and V. Cheng and Hopke [10] have recently examined these data using a variety of regression diagnostics. They found that these 'marker' elements can be clearly identified and their influence on the quality of the fit to the ambient data and the source mass contributions can be quantitatively estimated.

The elemental balance sheet allows the identification of the major sources of metals in the air. For example, vanadium and nickel primarily arise from oil-fired power plant emissions; 23 of 25 ng/m^3 for V and 4.0 of 17 ng/m^3 for Ni with most of the nickel unexplained. Subsequent studies have shown that Kowalczyk et al. [9] used an unusually low Ni/V ratio for the oil power plant profile which led to the underprediction of Ni. Zinc is mainly released by incinerator sources but also comes from motor vehicles (51 ng/m^3 from refuse incinerations and 7.3 ng/m^3 from motor vehicles). The reverse is true for lead with motor vehicles as the primary source and refuse incineration as a lesser but important source, 428 ng/m^3 from motor vehicles and 34 ng/m^3 from incineration. In this way sources of both particulate mass and specific elements can be identified.

Mayrsohn and Crabtree [11] presented the use of an iterative least-squares approach to apportion 6 sources of airborne hydrocarbon compounds. The sources were automotive exhaust, volatilization of gasoline and release of gasoline vapor, commercial natural gas, geological natural gas, and liquefied petroleum gas. They performed the least-squares fit to the hydrocarbon compound concentrations using gas chromatography to determine the concentrations of eight compounds. Their ordinary least-squares source reconciliation algorithm recognized that not all sources may contribute to every sample, and, if negative contributions were obtained, a different configuration of sources was employed with certain qualifying assumptions [12]. Each possible configuration with positive coefficients was considered and the one with the lowest standard error was chosen as the optimum solution. On the average, automotive exhaust was the source of almost 50% of observed hydrocarbons. Gasoline and its vapor contributed 30–30% by weight and the balance resulted from commercial and geological natural gas. Thus, automobiles and other highway-related sources were responsible for the majority of these hydrocarbons.

A similar study utilizing this mass balance approach for resolving hydrocarbon sources has been made. Nelson et al [13] have examined the atmospheric hydrocarbons in Sydney, Australia. They used a much more extensive hydrocarbon profiles for their sources and have obtained good agreement between the mass balance approach and a resolution based on an emission inventory. They also found that the major hydrocarbon sources were direct automobile exhaust ($36 \pm 4\%$) and evaporative emissions of gasoline ($32 \pm 4\%$) Thus, it was possible to identify the impact of highway emissions on gaseous as well as particulate pollutants.

In 1979, Watson [14] and Dunker [15] independently suggested a mathematical formalism called effective variance weighting that included the uncertainty in the measurement of the source composition profiles as well as the uncertainties in the ambient concentrations As part of this analysis, a method was also developed to permit the calculation of the uncertainties in the mass contributions. Effective-variance least squares has been incorporated into the standard personal computer software developed by the U.S. EPA for receptor modeling.

The most extensive use of effective-variance fitting has been made by Watson and co-workers [14,16] in their work on data from Portland, OR. Since that study, a number of other applications of this approach have been made including Medford, OR [17], Philadelphia, PA [18,19], and at a number of locations in the U.S. Environmental Protection Agency's Inhalable Particulate Network [20].

It must be made clear, however, that the CMB analysis works well in these examples because both the source and ambient samples were collected and analyzed during the same time period. A much less detailed resolution of lead sources was all that was possible in Kellogg, ID [21] when on-site samples could not be obtained. In an intercomparison study organized by the U.S. Environmental Protection Agency [22] to examine recep-

tor models, a set of ambient particulate elemental compositional data sets were analyzed by a number of investigators using similar CMB methods. The compositions of particles from sources in Houston, TX, were not available and were not measured during this program so that source composition profiles had to be obtained from literature sources. The lack of source data immediately raised problems in the use of the mass balance methods and comparison of results from different investigators [22]. It is not always certain exactly which sources should be included in the analysis. Although emission inventories may be available for the region, it may be that the measured source composition for a coal-fired power plant in Maryland burning eastern bituminous coal is not a particularly good representation for a lignite-burning plant in Texas.

An additional problem for receptor modeling is that the motor vehicle profile in the United States is undergoing rapid changes in lead and bromine concentrations with time as the new, catalyst-equipped cars, diesel cars and trucks replace the remaining leaded-fuel burning vehicles. An interesting solution to the problem of the changing lead concentration in motor vehicle emissions was recently provided by Dzubay et al. [19]. They obtained particle samples in the summer of 1982 in Philadelphia, PA and vicinity in the size ranges of < 2.5 μm and 2.5–10 μm using a dichotomous sampler. The samples were analyzed using ion chromatography for sulfate and nitrate, X-ray fluorescence (XRF) and instrumental neutron activation analysis (INAA) for elemental composition, and a thermo-optical method for organic and elemental carbon. Because there is also a non-ferrous metal smelter in the airshed, lead in the air comes from incinerators, the smelter, and tailpipe emissions. Using the other measured species in the data set, they derived the amounts of lead that could be attributed to all sources other than motor vehicles. They then used a second multiple regression analysis to relate the amount of unaccounted lead, total lead minus all sources other than vehicles, to the motor vehicle source and obtained a lead value of 6% lead in motor vehicle emissions. It appears that as long as sufficient leaded fuel is still in use, it will be possible to employ an approach such as this one to obtain the current fleet-weighted average. With leaded fuel having been phased out entirely, the lead and bromine are no longer useful tracers for motor vehicles [23]. A similar trend will now be starting in Europe as lead concentrations are reduced during the next few years

Since motor vehicles are an important source of particles, it is helpful to know that there may be other tracers appearing for automobiles. As part of the Philadelphia study discussed above, Olmez and Gordon [24] identified unusually high values of the rare earth elements lanthanum, cerium, and samarium arising from the catalysis support material from an oil refinery. It is likely that similar materials arise from the catalytic converters in automobiles and could serve as new markers for tailpipe emissions.

The results from Mayrsohn and Crabtree [11] and Nelson et al. [13] suggest that a mass balance is applicable for the gaseous aliphatic hydrocarbons. These species along with CO could possibly provide good tracers for particulate emissions from highways. Such a result is less likely to be obtained for more reactive species like olefins. There will be problems for semi-volatile species like polycyclic aromatic hydrocarbons (PAHs) because of the partitioning of the species between the gaseous and particulate phases. This problem has been recently reviewed by Pankow [25]. The sampling and analysis problems of reactive hydrocarbons and the modeling needed to account for their reactions in transit from source to receptor makes it very difficult to perform accurate receptor modeling and is an area of study that requires considerable additional effort.

There are alternative approaches to solving eq. (3). For example, it can be restated as a linear programming problem. Cheng and Hopke [26] have examined the use of the L_1 norm and linear programming approaches suggested by Hoagland [27], Henry et al. [28], and Henry [29]. Cheng and Hopke [26] found that a weighted, constrained L_1-norm approach was much more stable that either ordinary weighted least-squares or effective-variance weighted least-squares methods at least for the set of three data sets created for the EPA Receptor Model Intercomparison Workshop.

These data sets are described in detail by Currie et al. [30].

These same EPA data sets have also been reanalyzed using non-negative, weighted least-squares methods. In these studies, Wang and Hopke [31] concluded that these methods do provide valuable analysis of the rank of the source profile matrix and physically meaningful non-negative mass contributions. However, they suggest that the methods might lead to incorrect results if the proper source profiles are not used in the fitting process. Thus, there are statistical methods that are useful for extracting estimates of the mass contributions when both the source profiles and the ambient concentrations are known. However, it is often the case that the measured profiles are too similar to one another to be successfully resolved. Thus, other methods are needed to increase the amount of information available about the source and ambient particles.

This other method is computer-controlled scanning electron microscopy (CCSEM). The analysis of microscopic features of individual particles, such as their chemical composition, will provide much more information from each sample than can be obtained from bulk analysis. Therefore, the ability to perform microscopic analyses on a number of samples permits the use of CCSEM techniques in receptor models. CCSEM is an extension of individual particle characterization by optical microscopy and scanning electron microscopy (SEM). The microscope has long been employed to determine those characteristics or features that are too small to be detected by the naked eye. The use of optical microscopy in receptor models has been described by Crutcher [32]. Optical microscopic investigation of particle samples and its application to source apportionment have been illustrated in detail by Hopke [33]. The ability of the scanning electron microscope equipped with X-ray detection capabilities (SEM/XRF system) to provide size shape, and elemental constitution data extends the utility of microscopic examinations. For example, several studies have used the SEM in analysis of samples of coal-fired power plant ash [34,35] and volcanic ash [36]. However, these studies are limited in the number of particles detected, since SEM has the disadvantage of being time-consuming to examine particles manually

CCSEM can provide an important additional method in the area of receptor modeling. Casuccio et al. [37] and Hopke [33] have surveyed the initial applications of CCSEM in the particle elemental investigation and its ability of identifying particle sources in the receptor model studies. A number of previous studies have shown that CCSEM is capable of detecting the characteristics of individual particles [38,39]. The significant improvement of CCSEM is the coupling of a computer to control the SEM. Hence, three analytical tools are under computer control in the CCSEM: (1) the SEM, (2) the energy dispersive spectrometry X-ray analyzer, and (3) the digital scan generator for image processing [37]. CCSEM rapidly examines individual particles in samples and provides their elemental constitutions as well as their aerodynamic diameter and shape factors. Based on these characteristics of each particle, particles can be assigned to a number of well defined classes. These particle classes become the basis for characterizing sources so that accurate particle classification becomes a key step in using CCSEM data in receptor modeling.

The approach to the particle classification can be accomplished by agglomerative, hierarchical cluster analysis along with rule-building expert systems. The particles with similar composition are grouped by the cluster analysis. The sample-to-sample difference will be clearly distinguished by comparing cluster patterns of samples. Moreover, it is assumed that a source emits various types of particles. However, the mass fractions of particles in the various particle classes will be different from source to source and are the fingerprint for that source. The rule-building expert system can help automate the particle class assignments. This idea has been confirmed by the successful work on the samples collected in El Paso, TX [40] and particles from a coal fired power plant [41]. CCSEM analysis of individual particles can apportion the mass of particles to different sources in the airshed.

3.3 Illustrative example of CMB analysis

To illustrate the use of the CMB method, an example will be taken from the study of Glover et

al. [42] of the sources of airborne particulate matter in Granite City, IL. With the promulgation of the new National Ambient Air Quality Standards for Particulate Matter — 10 μm (PM_{10}) — it has been necessary to review the State Implementation Plan (SIP) in each state for those areas most likely to be out of compliance with the new standard. In Illinois, one such area is Granite City, an industrial city northeast of St. Louis, MO, that has a history of total suspended particulate and airborne lead non-attainment.

The locations of the major industries in Granite City and that of the ambient airborne particulate sampler are shown in Fig 1, the local industries include steel mills (American Steel and Granite City Steel), a secondary lead smelter (Terracorp),

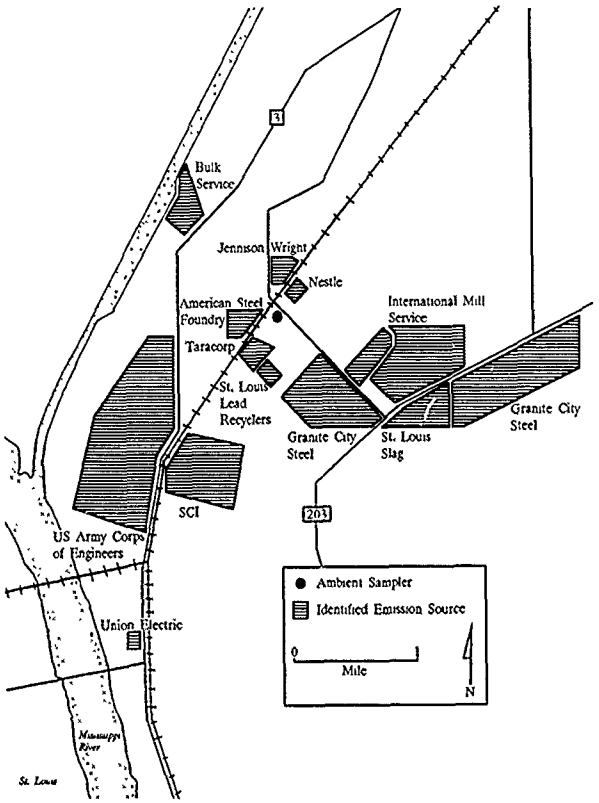


Fig. 1. Granite City local point sources.

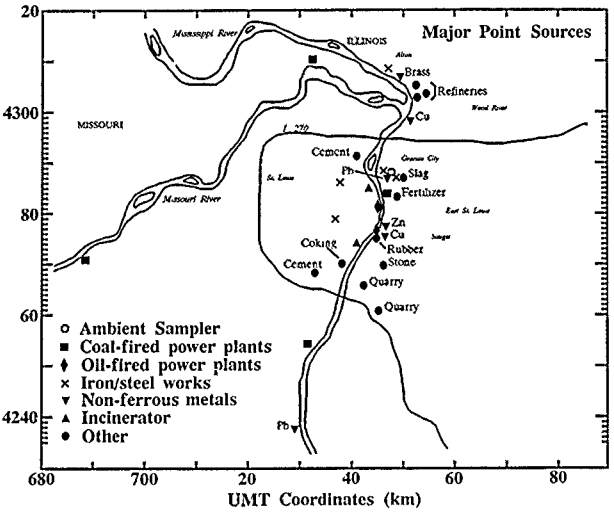


Fig 2. Point source locations in the Greater St. Louis, MO area.

an aluminum smelter (SCI) and a chemical plant (Jennison Wright). There is also a U.S. Army Corps of Engineers storage facility located at the edge of town. Fig. 2 shows the location of the major industries in the greater St. Louis Metropolitan area and their location relative to the ambient airborne particulate sampler.

As a part of the studies necessary to prepare an effective and efficient SIP, receptor modeling has been applied to elemental compositional data for 24 h airborne particle samples taken in Granite City by the Illinois State Water Survey using an automated dichotomous sampler. This sampler collects particles with an inlet that excludes large particles by having a 50% transmission efficiency for 10 μm particles. The particles that penetrate into the sampler are separated into two aerodynamic size fractions, < 2.5 μm (fine) and 2.5–10 μm (coarse). The particles are collected on Teflon filters which are then available for chemical analysis.

The particle samples were subjected to both XRF and INAA in order to provide the input data for receptor modeling, 48 sample pairs (fine and coarse) were thus analyzed for 33 elements. Each of these samples were then subjected to two CMB analyses. For the first analysis, the source profiles were taken from libraries available in the literature. To supplement the source profiles available in the literature, 12 dust samples were collected in and around Granite City, IL. These were aerosolized, sampled, and analyzed by XRF and INAA to provide site specific source profiles for the second CMB analysis.

In an attempt to account for more of the mass on each ambient filter, total carbon was measured seven times during the ambient sampling period. A Sierra PM_{10} sampler equipped with quartz fiber filters was collocated with the dichotomous sampler for this purpose. Each quartz filter was analyzed for total PM_{10} mass and total carbon mass. After the PM_{10} mass of each filter was

determined, the filter was treated with HCl to remove any carbonate. Each filter was then oxidized at 800°C, converting the elemental and organic carbon to CO_2. The amount of CO_2 released was measured with a Dohrmann carbon analyzer. A linear regression was used to relate the mass of total carbon to the total PM_{10} mass of each quartz filter. This regression is represented by

$$TC = 0.074 \times PM_{10} + 3.129 \quad (5)$$

where TC and PM_{10} are both measured in $\mu g/m^3$.

CCSEM [37] was used to partition the total carbon measurements between the fine and coarse fractions. The first and last quartz filters collected were analyzed by CCSEM. The number distribution, physical mass distribution, and aerodynamic mass distribution of the particles on each filter were determined along with an elemental analysis of the particles. The CCSEM measurements determined that the total carbon was apportioned between the fine and coarse fractions by

$$TC_{fine} = 0.919 \times TC \quad (6)$$

$$TC_{coarse} = 0.082 \times TC \quad (7)$$

The PM_{10} mass on each of the quartz filters was scaled to the PM_{10} mass collected on the Teflon disks. The mass of each pair of fine and coarse Teflon disks was added to find the total PM_{10} mass on the Teflon disks. TC_{fine} and TC_{coarse} for the Teflon disks were found by multiplying the scaling factor for each sample with eqs. (6) and (7), respectively.

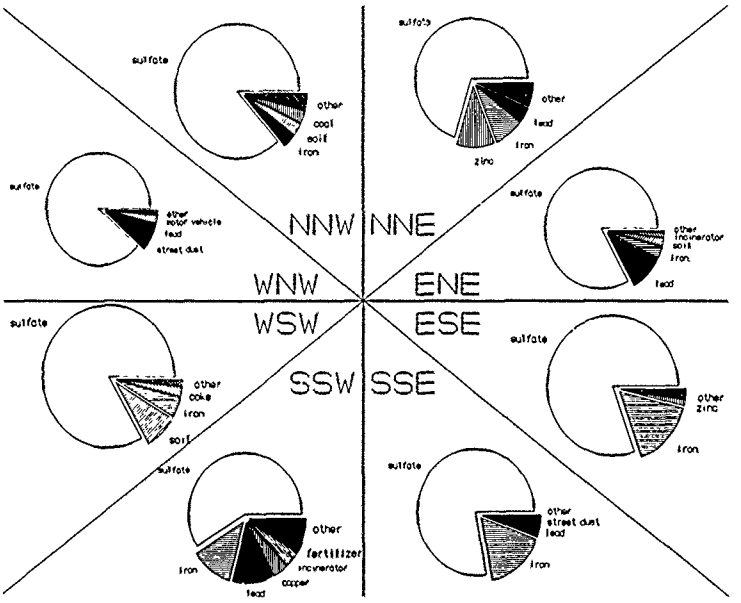


Fig. 3. Identified sources of fine fraction material.

3.3.1 Initial chemical mass balance

The initial CMB analysis identified several sources of particulate material in the Granite City area. Figs 3 and 4 show the identified source types for the fine and coarse fraction and the direction of each, relative to the sampler, based on the average wind direction during the time of sample collection Fig. 3 shows the regularity of the limestone and regional sulfate contributions to the fine fraction. Motor vehicle emissions were also observed to be coming from the highway to the north. Besides these fugitive and non-point sources, the local steel plants and lead smelters were observed to be major emission sources. Fugitive emissions from Granite City Steel appear as the urban dust coming from the southeast. The zinc source to the east is the galvanizing operations at Granite City Steel. This source is located to the west of the International Mill Service complex in Fig. 1. The coal-fired power plant identified to the east is Granite City Steels' coking operations while Taracorp's furnaces are the power plant identified to the southwest. Among the more distant source identified was a fertilizer plant located 5 km to the south of the sampler. The refinery complex 15 km to the north and the copper smelter 15 km to the south also appeared in the initial CMB analysis results. The coal-fired power plant that was identified to the north of the sampler is probably the facility located between the Mississippi and Missouri Rivers since there are no local sources with similar characteristics in that direction while the oil combustion source(s) to the southwest are the two oil-fired power plants

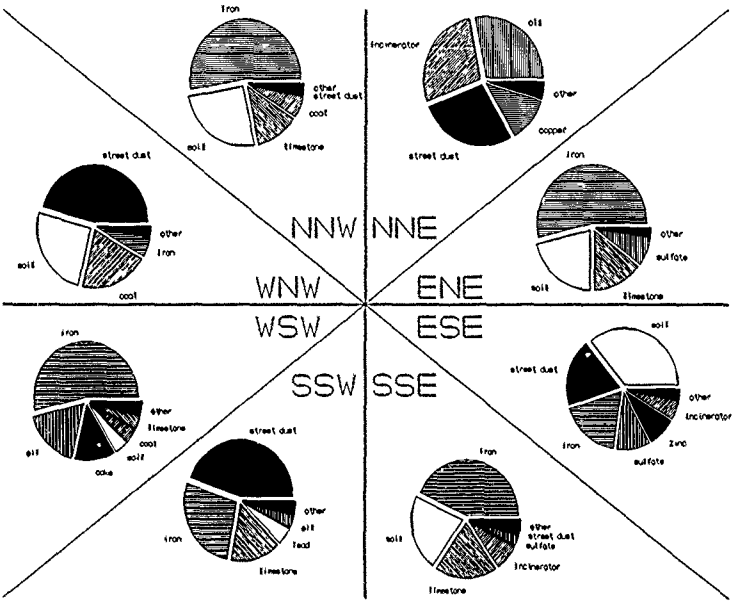


Fig. 4. Identified sources of coarse fraction material.

in that direction. The zinc smelter 12 km south of the sampler was expected to be a major source of fine zinc. However, the current study did not find appreciable amounts of zinc coming from the south.

Fig. 4 shows the predominance of the limestone and urban dust in the coarse fraction along with the local steel and lead sources. Besides the metal emissions from the steel plant, the coking operations at Granite City Steel appear as a combination of the sulfate emissions and coal-fired power plant profile. As it was found in the fine fraction, the galvanizing operation's zinc emissions and the lead smelter's combustion source appear to the east and southwest, respectively. The coke pile(s) identified to the west are at American Steel or Taracorp. American Steel is identified by the zinc source to the southwest and the coal sources to the west. The oil source to the northwest is the chemical treatment facility for railroad ties at Jennison Wright.

The initial CMB analysis results show that the composition of air pollution in the St Louis area has changed over the last ten years. Only one fifth of the fine profiles and one fourth of the coarse profiles used in the first CMB analysis were taken from the profiles derived from the 1975 to 1977 RAPS results These profiles accounted for 11 and

TABLE I

R^2 adjusted for degrees of freedom

Sample	Fine fraction values			Coarse fraction values		
	Initial	Final	Change	Initial	Final	Change
03/09/86	0 978	0 976	−0 002	0 812	0 811	−0 001
03/17/86	0 981	0 998	0 017	0.959	0 995	0 036
03/22/86	0 919	0 921	0 002	0.947	0 992	0 045
03/25/86	0 981	0 985	0 004	0 683	0 837	0.154
04/15/86	0.983	0 967	−0 016	0 683	0 837	0.154
04/18/86	0.933	0 993	0 060	0 810	0 970	0 160
04/21/86	0.949	0.956	0 007	0.941	0.950	0 009
05/23/86	0 891	0 926	0 035	0 878	0 983	0.105
05/23/86	0 947	0 960	0 013	0.797	0 982	0.185
05/25/86	0.957	0.964	0 007	0 862	0 857	−0 005
05/26/86	0.979	0 990	0 011	0 670	0 867	0.197
07/24/86	0.951	0.981	0 030	0.981	0 991	0 010
08/05/86	0 878	0.950	0 072	0 975	0.990	0 015
08/10/86	0 946	0.940	−0 006	0.968	0 990	0 022
10/18/86	0.991	0 958	−0 033	0 852	0.956	0.104
10/23/86	0 800	0 871	0 071	0 931	0 971	0 040
10/28/86	0 949	0 964	0 015	0 970	0 995	0 025
11/10/86	0.929	0 895	−0 034	0.947	0.994	0 047
11/11/86	0 965	0.967	0 002	0.972	0.969	−0 003
12/03/86	0 812	0 848	0 036	0.972	0 971	−0 001
12/07/86	0.802	0 843	0 041	0 805	0.969	0 164
01/29/87	0.985	0.965	−0 020	0 618	0 619	0 001
02/01/87	0 947	0.972	0 025	0.766	0 822	0 056
05/04/87	0.991	0 992	0 001	0 838	0.974	0.136
05/23/87	0.959	0.988	0 029	0 999	0.998	−0 001
05/25/87	0 852	0 857	0 005	0.742	0 827	0.085
06/06/87	0 979	0.990	0 011	0.911	0 955	0 044
06/12/87	0.985	0.983	−0 002	0.945	0.950	0 005
Average	0.936	0 950		0 875	0.935	
Avg. gain			0 023			0 073
Avg. loss			−0 016			−0 002

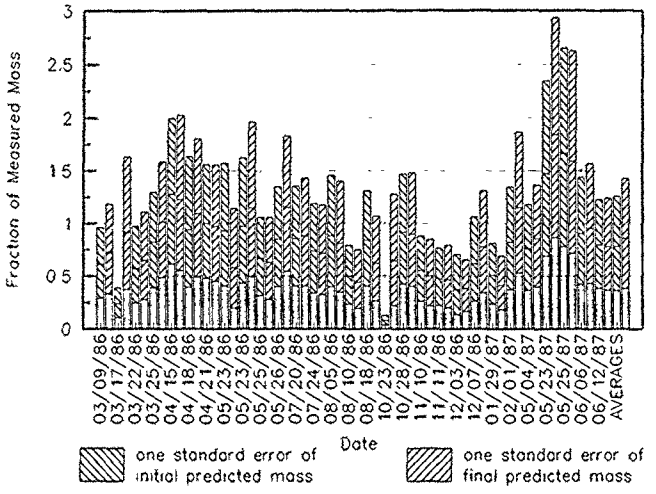


Fig. 5. Predicted mass fraction of selected fine fraction CMB Results, March 1986–June 1987, Granite City, IL

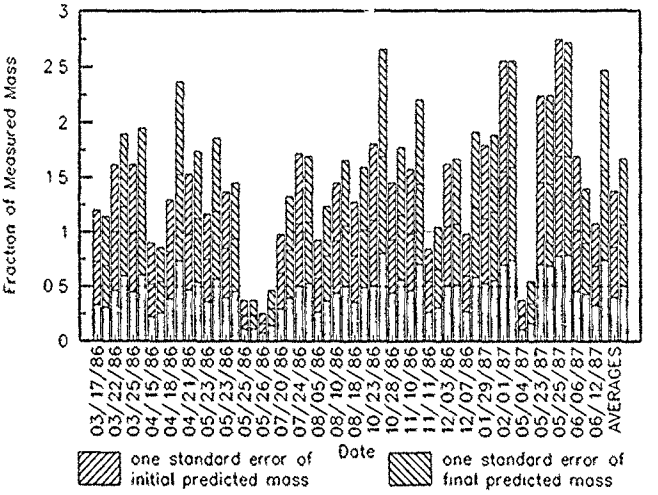


Fig. 6 Predicted mass fraction of selected coarse fraction CMB Results, March 1986–June 1987, Granite City, IL.

20% of all of the identified fine and coarse mass, respectively. The remaining profiles used in the current work were taken from more recent pollution source studies at various sites throughout the U.S.

3.3.2 Second chemical mass balance

By including the local dust samples among the potential source profiles in the second CMB analysis, a marked improvement in the quality of the predicted results was achieved The reanalysis did not change the types of sources identified by the CMB analysis. However, the apportionment between sources varied enough to cause the relative importance of sources to change. The improvement in the results can be seen in Table 1 where the average value of the adjusted R^2 increased for both fractions. (The adjustment in the R^2 values was made to account for the number of different sources that were identified for each sample.) This increase was especially apparent for the coarse fraction where the average negative change was less than one quarter of 1% while the average positive change was above 7%. Fig. 5 shows that the predicted mass of the fine fraction became closer to the observed mass with only a slight increase in error. (The error in the initial predicted results was influenced by the use of an artificial sulfur component, a source containing only sulfur, which caused the initial eror to be fairly low.) Fig. 6 shows that the predicted mass of the coarse fraction increased while the associated error decreased Fig. 7 shows that the predicted mass of the fine fraction fitting elements changed from an average over-prediction to an average under-prediction. Similar results were obtained for the coarse fraction samples Under-prediction is the more desirable error since during the fitting process, it is more difficult to explain mass that was not observed than to not explain all of the mass that

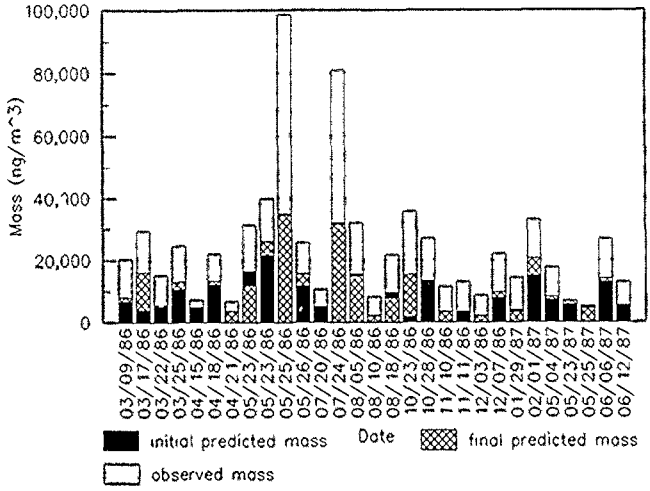


Fig. 7 Fine fraction ambient filter mass and mass of CMB fitting elements, March 1986–June 1987, Granite City, IL.

had been observed. There are always other unidentified sources that might explain the unaccounted for mass.

3.3.3 Total carbon results

In the CCSEM analyses, carbon was found to be a major component of the fine fraction. However, in the CMB analyses, carbon was never fit well. Lack of carbon information in many of the source profiles compounded the problem of having few ambient data.

3.3.4 Conclusions

By measuring ambient filters by both XRF and INAA, a relatively complete set of elemental measurements was obtained. The usefulness of these data was limited by the current unavailability of source profiles including these elements. The lack of data on carbon was a special problem in the present study since the limited ambient information did identify carbon as being an important part of the fine mass. The inclusion of site-specific profiles in a receptor-oriented source apportionment program improved the overall quality of the source apportionment results from those using only literature profiles. While not identifying new sources, the site-specific profiles significantly improved the R^2 of the coarse fretion. It also decrease the coarse fraction's predicted results error values. Considering that the initial fine fraction CMB required a unique sulfur factor to avhieve the best fit, the fine fraction results are also an indication that the better receptor modeling results are achieved by using site-specific profiles for fugitive emissions. The collection and analysis of site-specific fugitive dust profiles should be collected, if possible, during the course of future studies employing receptor models.

In many situations, locally measured source profiles are not available or there may have been significant changes in the particle producing activities in the airshed since the profiles were measured. Thus, it is helpful to have methods that can extract information from the ambient data alone as to the number, nature, and mass contributions of the particle sources in an area. These methods use multivariate statistical methods to obtain the receptor modeling information required.

4 MULTIVARIATE RECEPTOR MODELS

4.1 Introduction

Alternative approaches have been developed for identifying and quantitatively apportioning sources of airborne particles using multivariate statistical analysis. Eigenvector analysis has been the principal method that has been applied to airborne particle composition data. An eigenvector analysis tries to simplify the description of a system by determining the minimum number of new variables necessary to reproduce the measured attributes of the system. The mathematical basis of these methods has been described by Hopke [33].

Principal components and factor analysis are names given to several of the variety of forms of eigenvector analysis. It was originally developed and used in psychology to provide mathematical models of psychological theories of human ability and behavior [43]. However, eigenvector analysis has found wide application throughout the physical and life sciences. Unfortunately, a great deal of confusion exists in the literature in regard to the terminology of eigenvector analysis. Various changes in the way the method is applied has resulted in it being called factor analysis, principal components analysis, principal components factor analysis, empirical orthogonal function analysis, Karhunen–Loeve transform, etc., depending on the way the data are scaled before analysis or how the resulting vectors are treated after the eigenvector analysis is completed. All of the methods have the same basic objective; the compression of data into fewer dimensions and the identification of the structure of interrelationships that exist between the variables measured or the cases studied.

4.2 Mathematical procedures

The first step in the eigenvector analysis is the calculation of a dispersion matrix, the matrix that contains quantitative information on the relative variation of pairs of variables or pairs of samples (cases). There are two basic types of dispersion matrices. They are covariance matrices and corre-

lation matrices. For a correlation matrix, the data are scaled such that each variable or each case has an equal weight. The data are not scaled before calculating covariance In both instances, the data may be centered by subtracting a mean value before scaling and the calculation of the matrix elements. The choice of dispersion matrix depends on the nature of the data set to be analyzed. For many types of chemical spectroscopic data, the covariance matrix is the choise because each variable has the same measurement scale. For many geochemical problems, the difference in scale between major, minor, and trace components requires scaling to avoid domination of the analysis by the major components.

The dispersion matrix is then decomposed into a series of orthogonal vectors by the process outlined by Joreskog et al. [44] so that

$$U'DU = \Lambda \tag{8}$$

where U is the matrix of eigenvectors, U′ is its transpose, D is the dispersion matrix, and Λ is a diagonal matrix of eigenvalues where the trace of Λ is equal to the trace of D. If there were no errors in the data from which D is calculated, the number of non-zero eigenvalues would be the dimensionality of the problem called the rank of D. The rank for the original data matrix is the same as that for the dispersion matrix. However, experimental error generally results in a number of small but non-zero eigenvalues. The determination of the number of vectors containing significant information relative to those dominated by noise is often a difficult one. The lack of universally applicable criteria for determining the dimensionality of the data is a major problem in the application of factor analysis.

In the most commonly used approach to calculating the eigenvectors, the maximum amount of variance is packed into the first eigenvalue. The maximum possible amount of the remaining variance goes into the second and so forth. This compression of the information into a few components permits much of the variation in the data set to be displayed in a two- or three-dimensional plot. For many classification problems, the first few factors are able to reproduce most of the data structure and to remove some of the noise. The objects can then be plotted using the components axes and thus display the features of high-dimensional data in a few dimensions [45].

The compression of variance into the first factors will improve the ease with which the number of factors can be determined. However, their nature has now been mixed by the calculational method. Thus, once the number of factors has been determined, it is often useful to rotate the axes in order to provide a more interpretable structure.

The axis rotation can retain the orthogonality of the eigenvectors or cause them to be oblique. Depending on the initial data treatment, the axes rotations may be in the scaled and/or centered space or in the original variable scale space. The latter approach has proved quite useful in a number of chemical applications described by Malinowski and Howery [46] and in environmental systems as described by Hopke [33].

4.3 Previous applications

The first modeling applications of classical factor analysis were by Prinz and Stratmann [47] and Blifford and Meeker [48]. Prinz and Stratmann [47] examined both the aromatic hydrocarbon content of the air in 12 West German cities and data from Colucci and Begeman [49] on the air quality of Detroit. In both cases they found three factor solutions and used an orthogonal varimax rotation to give more readily interpretable results. Blifford and Meeker [48] used a principal component analysis with both varimax and a non-orthogonal rotation to examine particle composition data collected by the National Air Sampling Network (NASN) during 1957–1961 in 30 U.S. cities. They were generally not able to extract much interpretable information from their data. Since there are a very wide variety of particle sources among these 30 cities and only 13 elements were measured, it is not surprising that they were not able to provide much specificity to their factors.

The factor analysis approach was then reintroduced by Hopke et al. [50] and Gaarenstroom et al. [51] for their analysis of particle composition

data from Boston, MA and Tucson, AZ, respectively In the Boston data for 90 samples at a variety of sites, six common factors were identified that were interpreted as soil, sea salt, oil-fired power plants, motor vehicles, refuse incineration, and an unknown manganese–selenium source. The six factors accounted for about 78% of the system variance. There was also a high unique factor for bromine that was interpreted to be fresh automobile exhaust. Since lead was not determined, these motor vehicle-related factor loading assignments remain uncertain. Large unique factors for antimony and selenium were found. These factors represent emissions of species whose concentrations do not covary with other elements. Subsequent studies by Thurston and Spengler [52] where other elements including sulfur and lead were measured showed a similar result. They found that the selenium was strongly correlated with sulfur for the warm season (May 6 to November 5). This result is in agreement with the Whiteface Mountain results [53] and suggests that selenium is an indicator of long range transport of coal-fired power plant effluents to the northeastern U.S. They found lead to be strongly correlated with bromine and readily interpreted as motor vehicle emissions.

In the study of Tucson, AZ [51], whole filter data were analyzed separately at each site. They find factors that are identified as soil, automotive, several secondary aerosols such as $(NH_4)_2SO_4$, and several unknown factors. They also discovered a factor that represented the variation of elemental composition in their aliquots of their neutron activation standard containing Na, C, K, Fe, Zn, and Mg. This finding illustrates one of the important uses of factor analysis, screening the data for noisy variables or analytical artifacts.

One of the valuable uses of this type of analysis is in screening large data sets to identify errors [54]. With the use of atomic and nuclear methods to analyze environmental samples for a multitude of elements, very large data sets have been generated. Because of the ease in obtaining these results with computerized systems, the elemental data acquired are not always as thoroughly checked as they should be, leading to some, if not many, bad data points. It is advantageous to have an efficient and effective method to identify problems with a data set before it is used for further studies. Principal component factor analysis can provide useful insight into several possible problems that may exist in a data set including incorrect single values and some types of systematic errors.

Gatz [55] used a principal components analysis of aerosol composition and meterorological data for St. Louis, MO taken as part of project METROMEX [56,57]. Nearly 400 filters collected at 12 sites were analyzed for up to 20 elements by ion-induced XRF. Gatz [55] used additional parameters in his analysis including day of the week, mean wind speed, percent of time with the wind from NE, SE, SW, or NW quadrants or variable, ventilation rate, rain amount and duration. At several sites the inclusion of wind data permitted the extraction of additional factors that allowed identification of motor vehicle emissions in the presence of specific point sources of lead such as a secondary copper smelter. An important advantage of this form of factor analysis is the ability to include parameters such as wind speed and direction or particle size in the analysis.

In the early applications of factor analysis to particulate compositional data, it was generally easy to identify a fine particle mode lead–bromine factor that could be assigned as motor vehicle emissions. In many cases, a calcium factor sometimes associated with lead could be found in the coarse mode analysis and could be assigned as road dust. However, the problem of diminishing lead concentrations in gasoline discussed earlier for the CMB analysis also applies here. As the lead and related bromine concentrations diminish, the clearly distinguishable covariance of these two elements is disappearing. In a study of particle sources in southeast Chicago, IL based on samples from 1985 and 1986, much lower lead levels are observed and the lead–bromine correlation is quite weak [23]. Thus, the identification of highway emissions through factor analysis based on lead or lead and bromine is becoming more and more difficult and other analytic species are going to be needed in the future.

A problem that exists with these forms of factor analysis is that they do not permit quantitative source appointment of particle mass or of specific

elemental concentrations. In an effort to find an alternative method that would provide information on source contributions when only the ambient particulate analytical results are available, Hopke and co-workers [58–64] have developed target transformation factor analysis (TTFA) in which uncentered but standardized data are analyzed. In this analysis, resolution similar to that obtained from a CMB analysis can be obtained. However, a CMB analysis can be made on a single sample if the source data are known while TTFA requires a series of samples with varying impacts by the same sources, but does not require a priori knowledge of the source characteristics. The objectives of TTFA are (1) to determine the number of independent sources that contribute to the system, (2) to identify the elemental source profiles, and (3) to calculate the contribution of each source to each sample.

One of the first applications of TTFA was to the source identification of urban street dust [59]. A sample of street dust was physically fractionated by particle size, density, and magnetic susceptibility to produce 30 subsamples. Each subsample was analyzed by instrumental neutron activation analysis and atomic absorption spectroscopy to yield analytical results for 35 elements. The number of sources is determined by performing an eigenvalue analysis on the matrix of correlations between the samples. A target transformation determines the degree of overlap between an input source profile and one of the calculated factor axes. The input source profiles, called test vectors, are developed from existing knowledge of the emission profiles of various sources or by an iterative technique from simple test vectors [63]. The identified source profiles are then used in a simple weighted least-squares determination of the mass contributions of the sources [62].

In the analysis of the street dust, six sources were identified including soil, cement, tire wear, direct automobile exhaust, salt and iron particles. The lead concentration of the motor vehicle source was found to be 15% with a lead-to-bromine ratio of 0.39. This ratio is in good agreement with the values obtained by Dzubay et al. [65] for Los Angeles, CA freeways and in the range presented by Harrison and Sturges [66] in their extensive review of the literature. A comparison of the actual mass fractions with those calculated from the TTFA results shows that the TTFA provided a good reproduction of the mass distribution and source apportionments of the street dust that suggest that a substantial fraction of the urban roadway dust is anthropogenic in origin

One of the principal advantages of TTFA is that it can identify the source composition profiles as they exist at the receptor site. There can be changes in the composition of the particles in transit from the source to the receptor and approaches that provide these modified source profiles should improve the receptor model results. Chang et al. [63] have applied TTFA to an extensive set of data from St. Louis, MO to develop source composition profiles based on a subset selection process developed by Rheingrover and Gordon [67]. They select samples from a data base that were heavily influenced by major sources of each element. These samples were identified according to the following criteria:

1. Concentration of the element in question $X > X + Z_{cr}$ where X is the average concentration of that particular element for each station and size fraction (coarse or fine particle size fraction), Z_{cr} is typically set at about three for most elements, and is the standard deviation of the concentration of that element.
2. The standard deviation of the 6 or 12 h average wind directions for most samples, or minute averages for 2 h samples, taken during intensive periods is less than 20°.

Samples that are strongly affected by emissions from a source were identified through observation of clustering of mean wind directions for the sampling periods selected with angles pointing toward the source.

A number of studies of multivariate receptor models have used the data base from the Regional Air Pollution Study (RAPS) of St. Louis, MO. In the RAPS program, automated dichotomous samplers were operated over a 2 year period at 10 sites in the St. Louis metropolitan area. Fig. 2 shows the location of the 10 RAPS sampling stations. Ambient aerosol samples were collected in fine, < 2.4 μm, and coarse, 2.4–20 μm, fractions. Samples were analyzed at the Lawrence Berkeley

Laboratory for total mass by beta-gauge measurements and for 27 elements by XRF. The RAPS database contains results for almost 35000 samples.

Rheingrover and Gordon [67] screened the RAPS database according to the criteria stated above. With wind trajectory analysis, specific emission sources could be identified even in cases where the sources were located very close together [67] A compilation of the selected impacted samples was made so that TTFA could be employed to obtain elemental profiles for these sources at the various receptor sites.

Thus, TTFA may be very useful in determining the concentration of lead in motor vehicle emission as the mix of leaded fuel continues to change. Multivariate methods can thus provide considerable information regarding the sources of particles including highway emissions from only the ambient data matrix. The TTFA method represents a useful approach when source information for the area is lacking or suspect and if there is uncertainty as to the identification of all of the sources contributing to the measured concentrations at the receptor site.

Further efforts have recently been made by Henry and Kim [68] on extending eigenvector analysis methods. They have been examining ways to incorporate the explicit physical constraints that are inherent in the mixture resolution problem into the analysis. Through the use of linear programming methods, they are better able to define the feasible region in which the solution must lie. There exists a limited region in the solution space because the elements of the source profiles must all be greater than or equal to zero (non-negative source profiles) and the mass contributions of the identified sources must also be greater than or equal to zero. Although there has only been limited applications of this expanded method, it offers an important additional tool to apply to those systems where a priori source profile data are not available. These methods provide a useful parallel analysis with CMB to help insure that the profiles used are reasonable representations of the sources contributing to a given set of samples.

4.4 Illustrative example

4.4.1 Data description

In order to demonstrate the use of TTFA for the resolution of sources of urban aerosols, TTFA will be applied to a compositional data set obtained from aerosol samples collected during the RAPS program in St. Louis, MO [60]. The data for the samples collected during July and August 1976 from station 112 were selected for the TTFA process. Station 112 was located near Francis Field, the football stadium on the campus of Washington University, west of downtown St. Louis, MO.

During the 62 days of July and August, filters were changed at 12 h intervals, producing a total of 124 samples in each the fine and coarse fractions. Data were missing for 24 pairs of samples leaving a total of 100 pairs of coarse and fine fraction samples. Of the 27 elements determined for each sample, a majority of the determinations of 10 elements had values below the detection limits. Since a complete and accurate data set is required to perform a factor analysis, these 10 elements were eliminated from the analysis. For example, arsenic was excluded because almost all of the values were below the detection limits. Arsenic determinations by XRF are often unreliable because of an interference between the arsenic *K* X-ray and the lead *L* X-ray. A neutron activation analysis of these samples would produce better arsenic determinations. Reliable data for arsenic may be important to the differentiation of coal flyash and crustal material; two materials with very similar source profiles. The low percentage of measured elements can lead to distortions in the scaling factors produced by the multiple regression analysis. The remaining mass consists primarily of hydrogen, oxygen, nitrogen, and carbon. Although no measurements of carbon are included in the RAPS data, that portion of the sample mass must still be accounted for by the resolved sources. In order to produce the best possible source resolutions, it is vital to have accurate measurements of the mass of total suspended

particles (TSPs) as well as determinations for as many elements as possible.

The fine and the coarse samples were analyzed separately and only the fine-fraction results will be reported here. In this target transformation analysis, a set of potential source profiles was assembled from the literature to use as initial test vectors. In addition the set of unique vectors was also tested

4.4.2 Results

The eigenvector analysis provided the results presented in Table 2 Examination of the eigenvectors suggests the presence of 4 major sources, possibly 2 weak sources, and noise. To begin the analysis, a 4-vector solution was obtained. The iteratively refined source profiles are given in Table 3. The first 3 vectors can be easily identified as motor vehicles (Pb, Br), regional sulfate, and soil/flyash (Si, Al) based on their apparent elemental composition.

However, the fourth vector showed high K, Zn, Ba, and Sr was not initially obvious as to its origin. The resulting mass loadings were then calculated and the only significant values were for the sampling periods of noon to midnight on July 4 and midnight to noon on July 5. This was July 4, 1976 and there was a bicentennial fireworks display at this location. Thus, these two highly influenced samples change the whole analysis.

To illustrate this further, Table 4 gives the average values of the elemental composition of the fine fraction samples for the samples with and without the July 4 and 5 samples included It can be seen that these two samples from July 4 and 5 from 100 sample set have changed the average value of K by a factor of 2 and the average Sr by a

TABLE 2

Results of eigenvector analysis of July and August 1976 fine fraction data at Site 112 in St. Louis, MO

Factor	Eigenvalue	χ^2	Exner	Average % error
1	90.	210	0.324	204
2	5 0	156	0214	164
3	1.7	65	0.141	129
4	1.3	63	0064	93
5	0.16	55	0047	72
6	009	26	0034	68
7	003	24	0027	67
8	002	24	0021	58
9	002	15	0016	49

TABLE 3

Refined source profiles for the 4 source solution at RAPS Site 112, July–August 1976

Element	Motor vehicle	Sulfate	Flyash/ soil	Fireworks
Al	3.	0.9	62.	60
Si	0 0	2 8	140	0 0
S	0 0	232.	14.	26
Cl	5.2	1 6	0 31	19.
K	0.0	0 06	43.	580
Ca	12.	0 006	17.	0 27
Ti	2 8	1 8	2 3	0 0
Mn	1.5	0.1	0 8	3 6
Fe	5 8	3 8	38.	9.
Ni	0 2	0 06	0 05	0 3
Cu	1.9	0 2	0 03	4 6
Zn	9 8	1.4	0 0	24.
Se	0 1	0 1	0 0	0 01
Br	26.	0 0	2.7	2.
Sr	0 0	0 0	0 9	12
Ba	1.45	0.3	0 8	15.
Pb	105.	8.	3 8	0 0

TABLE 4

Comparison of data with and without samples from July 4 and 5, RAPS Station 112, July and August 1976 fine fraction

Element	Mean ± S D. (ng/m^3)	
	With	Without
Al	220 ± 30	200 ± 30
Si	440 ± 60	450 ± 60
S	4370 ±310	4360 ±320
Cl	90 ± 10	80 ± 9
K	320 ±130	150 ± 9
Ca	110 ± 10	110 ± 10
Ti	63 ± 13	64 ± 13
Mn	17 ± 3	17 ± 3
Fe	220 ± 20	220 ± 20
Ni	2.3 ± 0.2	2.3± 0 2
Cu	16 ± 3	15 ± 3
Zn	78 ± 8	75 ± 8
Se	2.7± 0 2	2.7± 0 2
Br	140 ± 9	130 ± 8
Sr	5 ± 4	1.1± 0.1
Ba	19 ± 5	15 ± 4
Pb	730 ± 50	720 ± 50

TABLE 5

Results of eigenvector analysis of July and August 1976 fine fraction data at Site 112 in St Louis, MO excluding July 4 and 5 data

Factor	Eigenvalue	χ^2	Exner	Average % error
1	87.	210	0.304	197
2	4 9	152	0.304	197
3	2.0	57	0 070	123
4	0 2	42	0 050	98
5	0.1	26	0 037	73
6	0.1	25	0 029	69
7	0 02	26	0 023	69
8	0 02	17	0 019	67
9	0.01	16	0 015	53

factor of 5. Thus, TTFA can find strong, unusual events in a large complex data set After dropping the samples from July 4 and 5, the analysis was repeated and the results are presented in Table 5 Now there are 3 strong factors, 2 weaker ones, and a continuum. Thus, a 5-factor solution was sought. These results are presented in Table 6.

The target transformation analysis for the fine fraction without July 4 and 5 indicated the presence of a motor vehicle source, a sulfate source, a soil or flyash source, a paint-pigment source, and a refuse source. The presence of the sulfate, paint-pigment, and refuse factors was determined by the uniqueness test for the elements sulfur, titanium, and zinc, respectively. In the paint-pigment factor, titanium was found to be associated with the elements sulfur, calcium, iron, and barium. This plant used iron titanate as its input material and the profile obtained in this analysis appears to be realistic. The zinc factor, associated with the elements chlorine, potassium, iron, and lead, is attributed to refuse-incinerator emissions. However, a high chlorine concentration is usually associated with particles from refuse incinerators [69,70]. This factor might also represent particles from zinc and/or lead smelters.

The results of this analysis provide quite reasonable fits to the elemental concentration and to the fine mass concentrations for this system. Thus, the TTFA provided a resolution of source types and concentrations that appear plausible although specific sources are not identified and quantitatively apportioned. From other studies with other data sets, it appears TTFA is typically able to identify 5 to 7 source types as long as they are reasonably distinct from one another.

TABLE 6

Refined source profiles (mg/g), RAPS Station 112, July and August 1976, fine fraction without July 4 and 5

Element	Vehicle	Motor sulfate	Soil/ flyash	Paint	Refuse
Al	5.	1.1	53.	0 0	0 0
Si	0 0	1.9	130.	0 0	7.
S	0 2	240.	19.	6.	0 0
Cl	2.4	1.1	0 0	4 6	22.
K	1.4	1.6	15.	5.7	48.
Ca	11.	0 0	16.	34.	1.2
Ti	0 0	0.7	2.5	110.	0 0
Mn	0 0	0 0	0.7	4.8	8 6
Fe	0 0	1.1	36.	90.	36.
Ni	0 08	0 04	0 042	0 011	0.7
Cu	0 6	0 01	0 0	0 0	8.7
Zn	0 8	0 0	0 0	3.7	65.
Se	0.1	0.1	0 001	0 2	0 2
Br	30.	0 3	2.5	0 0	0 05
Sr	0 09	0 01	0.15	0 1	0 001
Ba	0.7	0 035	0 07	28.	0 5
Pb	107.	6.5	5.	0 0	46.

5 SUMMARY

In this paper, several of the active areas of receptor modeling have been introduced. Their ability to determine the sources of particles in the air can be very useful in developing air quality management strategies and can potentially become enforcement tools as well. Since receptor models must of necessity be retrospective in nature, another important use can be in the calibration and testing of the prognostic dispersion models so that prediction of changes in air quality can serve as a more reliable basis for management decisions. The field of receptor modeling has grown and developed rapidly during the last several years and can be expected to continue to do so as methods are improved and new applications discovered.

ACKNOWLEDGEMENTS

Many of the studies reported here have been performed by students or post-doctoral associates in my group and their substantial contributions to the results presented here must be acknowledged. The work could not have been conducted without the support of the U.S. Department of Energy under Contract No. DE AC02-80EV10403, the U.S. Environmental Protection Agency under Grant No. R808229 and Cooperative Agreement R806236 and the National Science Foundation under Grant Nos. ATM 85-20533, ATM 88-10787, and ATM 89-96203.

REFERENCES

1 EPA, *PM10 SIP Development Guideline, Report No. EPA450/2-86-001*, U.S. Environmental Protection Agency, Research Triangle Park, NC, 1986.
2 J W. Winchester and G D. Nifong, Water pollution in Lake Michigan by trace elements from pollution aerosol fallout, *Water, Air, and Soil Pollution*, 1 (1971) 50–64.
3 M S Miller, S K. Friedlander and G M. Hidy, A chemical element balance for the Pasadena aerosol, *Journal of Colloid and Interface Science*, 39 (1972) 65–176.
4 S K. Friedlander, Chemical element balances and identification of air pollution sources, *Environmental Science Technology*, 7 (1973) 235–240.
5 J Bogen, Trace elements in atmospheric aerosol in the Heidelberg area measured by instrumental neutron activation analysis, *Atmospheric Environment*, 7 (1973) 1117–1125.
6 R Heindryckx and R. Dams, Continental, marine, and anthropogenic contributions to the inorganic composition of the aerosol of an industrial zone, *Journal of Radioanalytical Chemistry*, 19 (1974) 339–349.
7 D.F. Gatz, Relative contributions of different sources of urban aerosols. application of a new estimation method to multiple sites in Chicago, *Atmospheric Environment*, 9 (1975) 1–18.
8 G S. Kowalczyk, C E, Choquette and G.E. Gordon, Chemical element balances and identification of air pollution sources in Washington, DC, *Atmospheric Environment*, 12 (1978) 1143–1153.
9 G S. Kowalczyk, G E. Gordon and S W. Rheingrover, Identification of atmospheric particulate sources in Washington, DC using chemical element balances, *Environmental Science and Technology*, 16 (1982) 79–90.
10 M D. Cheng and P.K. Hopke, Identification of markers for chemical mass balance receptor model, *Atmospheric Environment*, 23 (1989) 1373–1389.
11 H. Mayrsohn and J.H. Crabtree, Source reconciliation of atmospheric hydrocarbons, *Atmospheric Environment*, 10 (1976) 137–143.
12 H Mayrsohn, J H Crabtree, M. Kuramoto, R.D. Sothern and S H. Mano, *Source Reconciliation of Atmospheric Hydrocarbons in the South Coast Air Basin, 1974, Report No DTS-76-2*, California Air Resources Board, El Monte, CA, 1975.
13 P F. Nelson, S M Quigley and M Y Smith, Sources of atmosphere hydrocarbons in Sydney. a quantitative determination using a source reconcilation technique, *Atmospheric Environment*, 17 (1983) 439–449.
14 J G Watson, *Chemical Element Balance Receptor Model Methodology for Assessing the Source of Fine and Total Suspended Particulate Matter in Portland, Oregon, Ph D Thesis*, Oregon Graduate Center, Beaverton, OR, 1979.
15 A M. Dunker, *A Method for Analyzing Data on the Elemental Composition of Aerosols, General Motors Research Laboratories Report MR-3074 ENV-67*, Warren, MI, 1979.
16 J.A. Cooper, J G. Watson and J.J. Huntzicker, The effective variance weighting for least squares calculations applied to the mass balance receptor model, *Atmospheric Environment*, 18 (1984) 1347–1355
17 J A Cooper, *Medford Aerosol Characterization Study. Application of Chemical Mass Balance to Identification of Major Aerosol Sources in the Medford Airshed, Interim Report to the Oregon Department of Environmental Quality*, Portland, OR, 1979.
18 J C. Chow, V. Shortell, J Collins, J G. Watson, T G. Pace and R. Burton, *A Neighborhood Scale Study of Inhalable and Fine Suspended Particulate Matter Source Contributions to an Industrial Area in Philadelphia, Paper No. 81-14.1*, Air Pollution Control Association, Pittsburgh, PA, 1981.
19 T.G. Dzubay, R.K. Stevens, G E. Gordon, I. Olmez, A E. Sheffield and W.J. Courtney, A composite receptor method applied to Philadelphia aerosol, *Environmental Science and Technology*, 22 (1988) 46–52
20 J C. Chow, J G. Watson and J.J Shah, *Source Contributions to Inhalable Particulate Matter in Major U.S. Cities, Paper No. 82-21.3*, Air Pollution Control Association, Pittsburgh, PA, 1982.
21 R.K. Stevens and T.G. Pace, Overview of the mathematical and empirical receptor models workshop (Quail Roost II), *Atmospheric Environment*, 18 (1984) 1499–1506.
22 T.G. Dzubay, R.K. Stevens, W.D. Balfour, H.J. Williamson, J A Cooper, J.E. Core, R.T DeCesar, E.R. Crutcher, S L. Dattner, B L. David, S L Heisler, J.J. Shah, P.K. Hopke and D L. Johnson, Interlaboratory comparison of receptor model results for Houston aerosol, *Atmospheric Environment*, 18 (1984) 1555–1566.
23 P.K. Hopke, W. Wlaschin, S. Landsberger, C. Sweet and S.J Vermette, in C.V. Mathai and D H. Stonefield (Editors), *PM-10. Implementation of Standards*, Air Pollution Control Association, Pittsburgh, PA, 1988, pp. 484–494.
24 I. Olmez and G.E. Gordon, Rare earths, atmospheric signatures for oil-fired power plants and refineries, *Science*, 229 (1985) 966–968.

25 J.F. Pankow, Review and comparative analysis of the theories on partitioning between the gas and aerosol particulate phases in the atmosphere, *Atmospheric Environment*, 21 (1987) 2275–2283.
26 M.D. Cheng and P.K. Hopke, Response to invited comments by L.J. Gleser on 'Investigation on the use of chemical mass balance receptor model numerical computations', *Chemometrics and Intelligent Laboratory Systems*, 1 (1986) 33–50.
27 E.S. Hougland, *Chemical Element Balance by Linear Programming, Paper No. 83-147*, Air Pollution Control Association, Pittsburgh, PA, 1983.
28 R.C. Henry, C.W. Lewis, P.K. Hopke and H.J. Williamson, Review of receptor model fundamentals, *Atmospheric Environment*, 18 (1984) 1507–1537.
29 R.C. Henry, *A Factor Model of Urban Aerosol Pollution. A New Method for Source Identification, Ph.D. Thesis*, Oregon Graduate Center, Beaverton, OR, 1978.
30 L.A. Currie, R.W. Gerlach, C.W. Lewis, W.D. Balfour, J.A. Cooper, S.L. Dattner, R.T. DeCesar, G.E. Gordon, S.L. Heisler, P.K. Hopke, J.J. Shah, G.D. Thurston and H.J. Williamson, Interlaboratory comparison of source apportionment procedures: results for simulated data sets, *Atmospheric Environment*, 18 (1984) 1517–1537.
31 D. Wang and P.K. Hopke, The use of constrained least-squares to solve the chemical mass balance problem, *Atmospheric Environment*, 23 (1989) 2143–2150.
32 E.R. Crutcher, Light microscopy as an analytical approach to receptor modeling, in S.L. Dattner and P.K. Hopke (Editors), *Receptor Models Applied to Contemporary Pollution Problems*, Air Pollution Control Association, Pittsburgh, PA, 1983, pp. 266–284.
33 P.K. Hopke, *Receptor Modeling in Environmental Chemistry*, Wiley, New York, 1985.
34 G. Fisher, D.P.Y. Chang and M. Brummer, Fly ash collected from electrostatic precipitators: microcrystalline structures and the mystery of the spheres, *Science*, 192 (1976) 553–555.
35 R.L. Carpenter, R.D. Clark and Y.F. Su, Fly ash from electrostatic precipitators: characterization of large spheres, *Journal of the Air Pollution Control Association*, 30 (1980) 679–681.
36 S.J. Rothenberg, P.B. Denee, C.R. Brundle, R.L. Carpenter, F.A. Seiler, A.F. Eidson, S.H. Weissman, C.D. Fleming and C.H. Hobbs, Surface and elemental properties of Mount St. Helens volcanic ash, *Aerosol Science and Technology*, 9 (1988) 263–269.
37 G.S. Casuccio, P.B. Janocko, R.J. Lee, J.F. Kelly, S.L. Dattner and J.S. Mgebroff, The use of computer controlled scanning electron microscopy in environmental studies, *Journal of the Air Pollution Control Association*, 33 (1983) 937–943.
38 P.C. Bernard and R.E. van Grieken, Classification of estuarine particles using automated electron microprobe analysis and multivariate techniques, *Environmental Science and Technology*, 20 (1986) 467–473.
39 D. Kim and P.K. Hopke, Classification of individual particles based on computer-controlled scanning electron microscopy data, *Aerosol Science and Technology*, 9 (1988) 133–151.
40 D. Kim and P.K. Hopke, Source apportionment of the El Paso aerosol by particle class balance analysis, *Aerosol Science and Technology*, 9 (1988) 221–235.
41 D.S. Kim, P.K. Hopke, G.S. Casuccio, R.J. Lee, S.E. Miller, G.M. Sverdrup and R.W. Garber, Comparison of particles taken from the ESP and plume of a coal-fired power plant with background aerosol particles, *Atmospheric Environment*, 23 (1989) 81–84.
42 D.M. Glover, P.K. Hopke, S.J. Vermette, S. Landsberger and D.R. D'Auben, Source apportionment from site specific source profiles, *Journal of Air and Waste Management Association*, submitted for publication.
43 H.H. Harman, *Modern Factor Analysis*, University of Chicago Press, Chicago, IL, 3rd ed., 1976.
44 K.G. Joreskog, J.E. Klovan and R.A. Reyment, *Geological Factor Analysis*, Elsevier, Amsterdam, 1976
45 R.E. Blackith and R.A. Reyment, *Multivariate Morphometrics*, Academic Press, London, 1971.
46 E.R. Malinowski and D.G. Howery, *Factor Analysis in Chemistry*, Wiley, New York, 1980.
47 B. Prinz and H. Stratmann, The possible use of factor analysis in investigating air quality, *Staub-Reinhalt Luft*, 28 (1968) 33–39.
48 I.H. Blifford and G.O. Meeker, A factor analysis model of large scale pollution, *Atmospheric Environment*, 1 (1967) 147–157.
49 J.M. Colucci and C.R. Begeman, The automotive contribution to air-borne polynuclear aromatic hydrocarbons in Detroit, *Journal of the Air Pollution Control Association*, 15 (1965) 113–122.
50 P.K. Hopke, E.S. Gladney, G.E. Gordon, W.H. Zoller and A.G. Jones, The use of multivariate analysis to identify sources of selected elements in the Boston urban aerosol, *Atmospheric Environment*, 10 (1976) 1015–1025.
51 P.D. Gaarenstrom, S.P. Perone and J.P. Moyers, Application of pattern recognition and factor analysis for characterization of atmospheric particulate composition in Southwest Desert atmosphere, *Environmental Science and Technology*, 11 (1977) 795–800.
52 G.D. Thurston and J.D. Spengler, A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston, *Atmospheric Environment*, 19 (1985) 9–26.
53 P.P. Parekh and L. Husain, Trace element concentrations in summer aerosols at rural sites in New York State and their possible sources, *Atmospheric Environment*, 15 (1981) 1717–1725.
54 B.A. Roscoe, P.K. Hopke, S.L. Dattner and J.M. Jenks, The use of principal components factor analysis to interpret particulate compositional data sets, *Journal of the Air Pollution Control Association*, 32 (1982) 637–642.
55 D.F. Gatz, Identification of aerosol sources in the St. Louis area using factor analysis, *Journal of Applied Meteorology*, 17 (1978) 600–608.

56 S.A. Changnon, R.A. Huff, P.T. Schickendenz and J L. Vogel, *Summary of METROMEX, Vol 1 Weather Anomalies and Impacts*, Illinois State Water Survey Bulletin 62, Urbana, IL, 1977.
57 B Ackerman, S A. Changnon, G. Dzurisin, D L. Gatz, R.C. Grosh, S D Hilberg, F A Huff, J.W Mansell, H.T. Ochs, M E. Peden, P.T. Schickedanz, R.G Semonin and J L. Vogel, *Summary of METROMEX, Vol. 2. Causes of Precipitation Anomalies*, Illinois State Water Survey Bulletin 63, Urbana, IL, 1978.
58 P.K. Hopke, R E. Lamb and D F S. Natusch, Multielemental characterization of urban roadway dust, *Environmental Science and Technology*, 14 (1980) 164–172.
59 D.J Alpert and P.K. Hopke, A quantitative determination of sources in the Boston urban aerosol, *Atmospheric Environment*, 14 (1980) 1137–1146
60 D.J. Alpert and P.K. Hopke, A determination of the sources of airborne particles collected during the regional air pollution study, *Atmospheric Environment*, 15 (1981) 675–687
61 K G. Severn, B A. Roscoe and P.K Hopke, The use of factor analysis in source determination of particulate emissions, *Particulate Science and Technology*, 1 (1983) 183–192.
62 P K. Hopke, Target transformation factor analysis as an aerosol mass apportionment method: a review and sensitivity analysis, *Atmospheric Environment*, 22 (1988) 1777–1792.
63 S.N. Chang, P K. Hopke, G E. Gordon and S.W. Rheingrover, Target transformation factor analysis of airborne particulate samples selected by wind-trajectory analysis, *Aerosol Science and Technology*, 8 (1988) 63–80
64 B A. Roscoe and P K. Hopke, Comparison of weighted and unweighted target transformation rotations in factor analysis, *Computers and Chemistry*, 5 (1981) 1–7.
65 T.G. Dzubay, R K. Stevens and L.W. Richards, Composition of aerosols over Los Angeles freeways, *Atmospheric Environment*, 13 (1979) 653–659.
66 R.M Harrison and W.T. Sturges, The measurement and interpretation of Br/Pb ratios in airborne particles, *Atmospheric Environment*, 17 (1983) 311–328.
67 S G. Rheingrover and G E. Gordon, Wind-trajectory method for determining compositions of particles from major air pollution sources, *Aerosol Science and Technology*, 8 (1988) 29–61.
68 R C. Henry and B M. Kim, Extension of self-modeling curve resolution to mixtures of more than three components. Part 1. Finding the basic feasible region, *Chemometrics and Intelligent Laboratory Systems*, 8 (1990) 205–216
69 R.R Greenberg, W.H. Zoller and G E. Gordon, Composition and size distribution of particles released in refuse incineration, *Environmental Science and Technology*, 12 (1978) 566–573.
70 R.R. Greenberg, G E. Gordon, W.H. Zoller, R.B. Jacko, D.W. Neuendorf and K.J. Yost, Composition of particles emitted from the Nicosia municipal incinerator, *Environmental Science and Technology*, 12 (1978) 1329–1332.

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 45–57
Elsevier Science Publishers B.V., Amsterdam

Measurement error models

Leon Jay Gleser
Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260 (U.S.A.)

(Received 8 November 1989, accepted 8 February 1990)

Abstract

Gleser, L.J., 1991. Measurement error models. *Chemometrics and Intelligent Laboratory Systems*, 10, 45–57.

An overview is given of linear measurement error models. Such models appear in many forms, including errors-in-variables regression and factor analysis, but are mathematically related to each other. Of particular interest to chemists are mass balance receptor models in which source profiles are estimated with error. A general model is given for errors in profiles, and the attention of chemists is directed toward recent advances in statistical model fitting and numerical analysis which may be of use in estimating source contributions.

1 INTRODUCTION

Measurement error models have been applied in virtually every area of science and technology. Perhaps most familiar to chemists are the models of factor analysis and errors-in-variables regression models, in which the predictors (independent variables) are observed subject to random errors of measurement.

Although measurement error models can be either linear or nonlinear, in the present paper attention is confined to linear measurement error models. Section 2 introduces such models, indicating the wide variety of mathematical forms in which these models can be stated. Some basic concepts, principles and terminology are introduced, with the goal of facilitating access by chemists to the broad statistical literature dealing with methods for fitting and analyzing measurement error models.

A brief survey of available statistical estimation methods, and related computer software, is given in Section 3. Particularly emphasized is an approach, called 'correction for attenuation' by psychometricians, which adjusts classical regression estimators (which ignore measurement errors in the predictor variables) for errors in the predictors. Besides permitting use of standard algorithms (both classical and more recent robust methods), this approach also has the merit of focusing the attention of users on ways to obtain and use available information about the sources and magnitudes of the measurement errors.

In environmental studies, chemists have used both factor analysis and errors-in-variables regression (which they call effective variance calculation) to identify source contributions to environmental pollution [1]. The statistical models used in these contexts stem from linear mass balance equations that relate the concentrations of certain 'aerosol properties' (e.g., chemical compounds) at a receptor to the total mass contributions from the

0169-7439/91/$03.50

sources. These applications will be used throughout the paper as concrete examples of linear measurement error models. In Section 4, some suggestions for possible improvements in the models and methods of statistical analysis used in this area will be presented.

2 LINEAR MEASUREMENT ERROR MODELS

Measurement error models have in common their attempt to describe situations in which the variables Y observed (denoted by capital letters) are of interest only because they reflect certain unobservable, or latent, variables y (denoted by corresponding lower case letters) that are measured by Y subject to random error. That is,

$$Y = y + e$$

where e is a random error of measurement having mean 0 and distribution functionally unrelated to the value of y. For the ith experimental unit or time period, we may have obtained measurements $Y_i^{(1)},\ldots,Y_i^{(m)}$ on m latent variables $y_i^{(1)},\ldots,y_i^{(m)}$, $i=1,\ldots,n$. Let $Y_i=(Y_i^{(1)},\ldots,Y_i^{(m)})'$ and $y_i=(y_i^{(1)},\ldots,y_i^{(m)})'$ be m-dimensional column vectors containing the observed and latent variables, respectively. Then

$$Y_i = y_i + e_i, \qquad i=1,\ldots,n \tag{1}$$

where the vectors e_i of measurement errors have mean vector 0 and distributions functionally unrelated to the values of the latent variables y_i. It is usually assumed that the error vectors e_i are independently distributed.

In a linear measurement error model, the elements of each latent vector y_i are assumed to satisfy a common set of linear relationships. Geometrically, this means that the y_is, represented as points in m-dimensional space, all lie in a hyperplane $\mathscr{H}$ of dimension r, $r<m$, passing through an origin α. The dimension r of $\mathscr{H}$ can be either known or unknown; in the latter case, r is a basic parameter of the model.

Three commonly used ways to restate the above geometric description of the model in an algebraic (parameterized) form are the following

$$y_i = \Lambda f_i + \alpha, \qquad i=1,\ldots,n \tag{2}$$

$$y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \end{pmatrix} = \begin{pmatrix} \mathrm{B} \\ \mathrm{I}_r \end{pmatrix} y_{i2} + \begin{pmatrix} a \\ 0 \end{pmatrix}, \qquad i=1,\ldots,n \tag{3}$$

$$\Delta y_i = \gamma, \qquad i=1,2,\ldots,n \tag{4}$$

In eqn. (3), I_r is the r-dimensional identity matrix.

The model (2) is the familiar model of factor analysis. The columns of the $m\times r$ factor loading matrix Λ are a basis for the hyperplane $\mathscr{H}$, while the factor score vectors f_i contain the coefficients representing each y_i as a linear combination of the basis elements (columns of Λ).

The model (3) is the model of errors-in-variables regression. Here, r elements of each y_i serve as predictor (independent) variables for the remaining $m-r$ variables. By renumbering components, we can allow the predictor variables chosen to form the r-dimensional subvector y_{i2} containing the last r elements of y_i. The slope matrix B, $(m-r)\times r$ and intercept vector a are basic parameters of the model.

Model (4) is a more symmetric way of writing a set of linear equations relating the elements of y_i, in that no distinction is made between independent and dependent variables (as was done in model (3)). This model is often referred to as an implicit linear functional relationship model. The coefficient matrix Δ, which is $(m-r)\times m$ and has full rank $m-r$, and the vector γ are basic parameters of the model.

Model (4) often results from consideration of families of simultaneous stochastic equations. In such models, observations X_{jt}, $j=1,\ldots,J$, are made at each of T time points t. It is assumed that these observations satisfy a set of linear equations

$$\sum_{j=1}^{J} a_{ij} X_{jt} = f_{it}, \qquad i=1,\ldots,I \tag{5}$$

where $(f_{1t},\ldots,f_{It})'=f_t$ are independent random vectors having mean vector 0 and a common distribution. The X_{jt} are quantities internal (endogenous) to a given system (in econometrics, an economic system), while the f_{it} represent random

influences external (exogenous) to the system that account for the linear combinations on the left side of eq (5) not being exactly equal to 0. If $J < I$, and the matrix $A = ((a_{ij}))$ has full rank J, we can renumber indices so that

$$A = (A_1, A_2), \qquad A_1: J \times J \text{ of rank } J$$

$$Z_t = (X_{1t}, \ldots, X_{Jt})', \qquad W_t = (X_{J+1,t}, \ldots, X_{It})'$$

and then (5) becomes

$$A_1 Z_t + A_2 W_t = f_t, \quad t = 1, \ldots, T$$

or

$$Z_t = -A_1^{-1} A_2 W_t + A_1^{-1} f_t$$

$$= \Pi W_t + f_t^*, \qquad t = 1, \ldots, T \tag{6}$$

Using classical multivariate linear regression methods, we can find an estimator $\hat{\Pi}$ of Π. To estimate the original matrix A of coefficients, it is necessary to impose restrictions. This is usually done by identifying certain of the a_{ij} as being equal to 0. Such restrictions on A_2 imply that certain elements of $A_1\Pi = A_2$ are zero. Since A_1 is unknown, this results in an implicit linear functional relationship model for Π. Here, the columns of $\hat{\Pi}$ become the observed Y_i and the columns of Π are the latent vectors y_i. Thus, the model (4) is applied to estimated regression slope matrices in a classical regression model. It is in this manner that measurement error models often appear in the econometrics literature. The potential application of similar stochastic equation models (5), and the resulting linear measurement error models (4), in chemistry and other physical sciences should be apparent. In these models some of the X_{jt} variables can be measurements of variables obtained at times prior to t (that is, lagged values), in which case (5) has the form of an ARIMA time series model. A thorough discussion of linear simultaneous stochastic equation models, and the related linear measurement error models, can be found in refs. 2–4.

In calibration models, estimated regression slopes can again serve as observed variables, with true slopes acting as latent variables. For examples, suppose that we fit a linear model

$$Z_i = \alpha + \beta W_i + e_i, \qquad i = 1, \ldots, k$$

and obtain the least squares estimators $\hat{\alpha}$, $\hat{\beta}$ of the intercept and slope. A new observation Z is obtained, and we wish to estimate the value of W that led to Z. Then

$$\begin{pmatrix} Z \\ \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \alpha + \beta W \\ \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \end{pmatrix}$$

has the form of a measurement error model with $Y_1 = (Z, \hat{\alpha}, \hat{\beta})'$, $y_1 = (\alpha + \beta W, \alpha, \beta)'$. One way to represent the linear restriction is in the form (4)

$$(1, 0, -W)\, y_1 = \alpha$$

Calibrations, and thus calibration models, are widely used in the physical sciences and engineering [5,6]. Although most calibrations involve estimation of a single predictor W from a single dependent variable Z (perhaps on many occasions), multivariate calibration models are also used [7,8] The calibration literature tends to emphasize methods based on classical linear (or approximately linear) multiple regression models, so that the connection to measurement error models is not widely known. Consequently, the calibration and measurement error model literatures have tended to develop in parallel.

It should be added that predictor variables in the physical sciences, and also the medical and behavioral sciences, are often measured indirectly through calibration. This is a source of measurement error in regression experiments that is frequently overlooked, at the cost of a possibly substantial bias in conclusions [9]. On the other hand, calibration experiments provide a useful way to assess measurement errors in predictors (Section 3).

In mass balance models, two distinct applications of linear measurement error models arise. First, we may have measurements C_{it} of concentrations of 'aerosol property' i at a receptor at time t for m properties ($i = 1, \ldots, m$) and T times ($t = 1, \ldots, T$). The true concentrations c_{it} are thought to result from the mass contributions s_{jt} of r sources, as represented by the linear mass

balance model:

$$c_{it} = \sum_{j=1}^{r} a_{ij} s_{jt}, \qquad i = 1,\ldots,m; \quad t = 1,\ldots,T \tag{7}$$

Letting

$$\mathbf{Y}_t = (C_{1t},\ldots,C_{mt})', \qquad \mathbf{y}_t = (c_{1t},\ldots,c_{mt})'$$
$$\mathbf{f}_t = (s_{1t},\ldots,s_{rt})', \qquad \Lambda = ((a_{ij}))$$

we have the factor analysis model

$$\mathbf{Y}_t = \mathbf{y}_t + \mathbf{e}_t, \quad \mathrm{E}(\mathbf{e}_t) = \mathbf{0}, \quad \mathbf{y}_t = \Lambda \mathbf{f}_t. \tag{8}$$

The intercept term α in model (3) does not appear here since it is usually assumed that all variables are centered at their sample means (The variables are also usually standardized by their standard deviations — a practice about which we will have more to say later.) In applications of this model, the coefficients a_{ij} of the mass balance equations and the number r of sources are usually assumed to be unknown.

A second application of linear measurement error models to mass balance problems occurs when we know the number r or sources, and also have unbiased measurements (or other similar prior information) for the coefficients a_{ij} in eq. (7). Here, only one measurement in time is usually taken, so that the 'aerosol properties' are treated as experimental units. That is, it is assumed that we observe

$$\mathbf{Y}_i = \begin{pmatrix} C_i \\ A_{i1} \\ \vdots \\ A_{ir} \end{pmatrix}, \qquad i = 1,\ldots,m$$

where

$$\mathbf{Y}_i = \begin{pmatrix} C_i \\ A_{i1} \\ \vdots \\ A_{ir} \end{pmatrix} = \begin{pmatrix} c_i \\ a_{i1} \\ \vdots \\ a_{ir} \end{pmatrix} + \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{i,r+1} \end{pmatrix} \tag{9}$$

$\mathrm{E}(e_{ij}) = 0$, $j = 1,\ldots,r+1$, and

$$c_i = \sum_{j=1}^{r} a_{ij} s_j = \mathbf{B} \begin{pmatrix} a_{i1} \\ \vdots \\ a_{ir} \end{pmatrix}, \qquad i = 1,\ldots,m,$$
$$\mathbf{B} = (s_1,\ldots,s_r) \tag{10}$$

This model has the errors-in-variables regression form (2) with the slope matrix B giving the mass contributions s_j, $j = 1,\ldots,r$, of the sources. Again, the intercept α in model (2) does not appear since all measured variables are centered at their sample means.

2.1 Model uniqueness

Although the idea of linear relationships among the latent variables is intuitively clear (with a concrete geometric interpretation), each of the models used to represent or parameterize the idea has elements of arbitrariness. First, note that the parameterizations in two of the models [(2) and [4)] that we have described are not uniquely defined. For example, in the factor analysis model (2), we can replace Λ by $\Lambda^* = \Lambda\mathrm{T}$ and $\mathbf{f}_i$ by $\mathbf{f}_i^* = \mathrm{T}^{-1}\mathbf{f}_i$ for any r-dimensional invertible matrix T without changing the validity of the model

$$\mathbf{y}_i = \Lambda \mathbf{f}_i + \alpha = \Lambda\mathrm{T}\mathrm{T}^{-1}\mathbf{f}_i + \alpha = \Lambda^*\mathbf{f}_i^* + \alpha$$

(Since the columns of Λ are a basis for the hyperplane H, and bases of vector spaces are not unique, this fact should not be surprising.) In the literature, this nonuniqueness problem is called factor indeterminacy. One can impose restrictions on Λ (and possibly other parameters of the model) to remove this indeterminacy, but such restrictions are exterior to the model (and data) and cannot be tested. Indeed, it is common for one set of restrictions to be imposed for computational convenience (usually to reduce the estimation problem to a type of principal components analysis), and then for investigators to search among the set of equivalent parameterizations of the fitted model for one which has meaning in the given context. (For example, the program VARIMAX searches to find permissible loadings λ_{ij} in Λ with maximum variability — either λ_{ij} is near 0 or very large.) The extra searching that such exploratory factor analysis methods do among equivalent parameterizations of the model (2) in the attempt to find a 'meaningful solution' is not accounted for by customary indices of accuracy (large-sample variances and covariances of the estimators). It is entirely possible for two investigators starting with the same data and the same initial solution for the

parameters to arrive at quite different 'meaningful solutions' (final fitted models). In confirmatory factor analysis, on the other hand, a set of restrictions is imposed a'priori (usually based on previous experience with the variables being studied), regardless of computational convenience, and then such a model is fitted, and also tested against other less restrictive models (particularly models allowing a larger number of factors).

Similar comments about indeterminacy apply to model (4). Here, the coefficient matrix Δ and the vector γ can be replaced by $A\Delta$ and $A\gamma$, for any $(m-r)$-dimensional invertible matrix A, without affecting the validity of the equation defining the model. Again, restrictions needed to identify the parameters cannot be tested by the given data.

By imposing suitable extra-model restrictions, one can reduce both model (2) and model (4) to the errors-in-variables regression form (3). (This is intuitively clear from the fact that all three models describe the same geometric assumption that the latent vectors y_i lie in the hyperplane $\mathscr{H}$.) Verification of this assertion can be found in refs. 3 and 10. However, even model (3) requires prior separation of the elements of y_i into a vector of predictor (independent) variables y_{i2} and a vector of dependent variables y_{i1}. In the factor analysis model (2), this also means that the factors f_i are identified with certain of the components of y_i. Where there is a natural such separation of variables (such as in the second mass balance model (9) and (10) above), it is then reasonable to prefer the model (3), since the parameters B and a are uniquely defined by the model. However, in other contexts, this violation of the symmetry of the relationships among the variables causes experimenters some concern. For example, if it were actually the case that $y_i = (y_i^{(1)}, y_i^{(2)}, y_i^{(3)})'$, $r = 2$, and $y_i^{(2)} = 5y_i^{(3)}$, and we chose $y_{i1} = (y_i^{(1)})$, $y_{i2} = (y_i^{(2)}, y_i^{(3)})'$ in model (3), we would not be able to recover the linear relationship among the elements of y_i. Nevertheless, it is always true that model (3) for some choice of y_{i1}, y_{i2} yields one of the permissible (equivalent) solutions (fitted models) for models (2) and (4).

Observing the arbitrariness involved in parameterizing the models (even model (3)), and in contrast the uniqueness of the hyperplane $\mathscr{H}$ which geometrically describes the linear relationships among the elements of the latent vectors y_i, it is natural to try to parameterize $\mathscr{H}$ directly. One way to do this is by the angles θ_j, $j = 1, \ldots, m-1$, between the hyperplane $\mathscr{H}$ and any $m-1$ of the m axes in m-dimensional space. This approach is mentioned in ref. 11, where it is applied in the case $m = 2$, $r = 1$ (a linear relationship between two latent variables). However, generalizations to general m, general r, appear to be computationally and analytically difficult. Further, the angles θ_j are not in themselves usually of intrinsic interest.

2.2 *Identifiability*

Apart from questions of uniqueness of parameterization, there is also the problem of identifying the linear relationships from data. This is caused by the fact that we do not directly observe the latent variables y_i, but instead observe $Y_i = y_i + e_i$. Linear associations (covariance) among the elements of the error vectors e_i can thus be mistaken for (confounded with) linear relationships among the elements of y_i, since both types of association can result in covariation between elements of the observed Y_i. Consequently, assumptions about the form of the joint distribution of the elements of the error vectors, e_i is required in order to identify the linear relationships of interest (among the elements of the latent vectors y_i).

Because normal distributions are determined by their mean vectors and covariance matrices, this problem of identifiability always arises for normally distributed Y_is. Interestingly, only normal distributions suffer from this problem, since information about latent linear relationships can otherwise be obtained from higher moments or cumulants of the distribution [12,13]. Thus, normal distributions in measurement error models play the unusual role of the most 'nonrobust' or 'worst-case' distribution (in contrast to their 'best-case' role in most other types of inference). Because use of sample higher moments or cumulants in estimation is computationally cumbersome, adds a large component of variability to estimates, and also requires knowledge of which

moments or cumulants to use, the problem of nonidentifiability in normal distributional cases (because it reflects on any procedure based on sample mean vectors and covariance matrices) is also relevant even to situations where we are certain the data are not normally distributed.

Basically, in normal distributional cases, linear relationships among the elements of y_i cannot be identified (consistently estimated) without knowledge about the error covariance matrices

$$\Sigma_i = \mathrm{Cov}(e_i)$$

of the error vectors e_i. This knowledge can either come from parametric assumptions about the Σ_i, or from independent estimates of these matrices obtained from other experiments (calibration data) or replications of Y_is for fixed y_is — that is,

$$Y_{ij} = y_i + e_{ij}, \qquad j = 1, \ldots, J_i$$

For factor analysis models, the classical assumption made is that the Σ_is are all equal to the same diagonal matrix. This diagonality assumption is usually justified by the belief that choice of a large enough value of r (the number of factors) removes all common sources of variation from the errors.

For errors-in-variables regression models, a wide variety of assumptions about the Σ_is have been used, and software packages exist to fit many of these models [14]. The sensitivity of the resulting estimates to the assumptions used is still an open question, although some information is available for the simple case $r = 1$. Common to all of these assumptions is the basic requirement that the regression slopes of the elements of e_{i1} on the elements e_{i2} are known [15]. Here, e_{i1} contains the errors in the observations Y_{i1} of the latent dependent vectors y_{i1}, and e_{2i} contains errors in the observations Y_{2i} of the latent independent vector y_{i2}. This requirement is clearly essential, since otherwise such regression slopes will be confounded with the matrix B in model (3). In most applications of errors-in-variables regression models, the measurements of Y_{i1} and of Y_{2i} are made separately, and it is reasonable to assume that the regression slopes of the e_{i1} on the e_{i2} are zero.

2.3 Structural and functional models

An important distinction that is made in the statistical literature on measurement error models is between models in which the latent vectors y_i are treated as unknown constants (functional models), and models in which the y_i are assumed to be independent random vectors (structural models). In the former case, the y_i are themselves parameters of the model. The fact that the number of such parameters increases as the sample size n increases causes major problems for statistical theory. For example, maximum likelihood estimators for the parameters of functional measurement error models need not exist [16,17]; or if they exist, need not be consistent. No completely satisfactory large sample optimality theory exists for functional measurement error models.

In contrast, structural measurement error models are typically parameterized by a finite number of parameters. Consequently, classical statistical theory (e.g., the theory of maximum likelihood estimation and likelihood ratio tests) can be applied. Even so, some problems remain: complicated finite sample distributions, nonexistence of all moments of the maximum likelihood estimator, etc. For example, in a strict mathematical sense, finite-length $1 - \alpha$ confidence intervals for the parameters of linear measurement error models ((2), (3) or (4); structural or functional cases) do not exist [18]. Commonly used confidence intervals (e.g., large-sample intervals) have arbitrarily small coverage probability when the measurement error variances are very large relative to the spread of the true latent variables. (See ref. 18a for exact results in the case $r = 1$ of model (3).) Fortunately, this theoretical result has minimal importance in most physical science applications because practitioners usually have some idea of the magnitudes of the measurement errors (and error variances) in their experiments. If not, some useful checks to verify that large-sample confidence intervals have desired coverage probability are available (see ref. 19, pp. 1134–1135). Alternatively, the Creasy–Fieller method of constructing $1 - \alpha$ confidence regions [20,21] can be

used, although such regions will not always be intervals.

The distinction between functional and structural measurement error models is similar to the distinction between fixed (designed) factors and random factors in the analysis of variance. In most contexts where factor analysis models are used, investigators are willing to assume that the factors f_i (and thus the latent vectors y_i) are random — for example, in the first mass balance model discussed above, the factors f_i represent mass contributions from the sources and might reasonably be assumed to vary randomly over time. On the other hand, one would be less certain that the proportions of mass $a_{i,j}$ from the r sources would vary randomly across 'aerosol properties' i in the second mass balance model. Such latent variables seem to be fixed characteristics of the 'aerosol properties'. Consequently, this second mass balance model appears to be a functional measurement error model.

Nevertheless, arguments given in ref. 22 show that for every functional model one can construct a similarly parameterized structural model. Using this structural model, one can more easily determine restrictions insuring identifiability for the key parameters of both models (structural and functional). Further, the maximum likelihood solution for the structural model (which is the best asymptotic normal estimator of the parameters in that model) is typically also the best asymptotic normal estimator of the corresponding parameters in the functional model [22,23]. Consequently, even when one believes that one has a functional linear measurement error model, it is worth while starting one's statistical analysis by studying identifiability and choice of estimators for the corresponding structural model. An additional advantage of adopting structural model assumptions is that natural estimators (predictors) of the latent variables y_i based on the observed values Y_i can be defined. These are the conditional expected values $E[y_i \mid Y_i]$.

3 ESTIMATION AND SOFTWARE

A structural linear measurement error model yields a covariance structure model for the observations Y_i. Since the latent variables y_i are random, and the model (1) assumes that the (conditional) distribution of e_i does not depend on y_i, it follows that e_i and y_i are statistically independent. Thus,

$$\mathrm{Cov}(Y_i) = \mathrm{Cov}(y_i) + \mathrm{Cov}(e_i)$$

where the assumption that y_i varies in an r-dimensional subspace $\mathscr{H}$ of m-dimensional space implies that $\mathrm{Cov}(y_i)$ is singular of rank r. For example, in the factor analysis model (2),

$$\mathrm{Cov}(Y_i) = \Lambda\psi\Lambda' + \mathbf{D}_\theta \tag{11}$$

where ψ is the (common) covariance matrix of the factor vectors f_i (which are random because y_i is random) and $\mathbf{D}_\theta$ = diagonal $(\theta_1, \theta_2, \ldots, \theta_m)$ is the common covariance matrix of the error vectors e_i.

A very popular general computer program for fitting multivariate covariance structure models of reduced rank is the program LISREL VI [24]. This program also provides estimated large-sample variances and covariances for the resulting estimators, and tests of fit for models of various ranks r. There is a substantial literature dealing with special problems connected with this software (and method of estimation). Many of the relevant papers appear in the journal *Psychometrika*, although some significant papers in this area also have appeared in such journals as *Biometrika*, *South African Statistical Journal*, and the *Annals of Mathematical Statistics*. Although LISREL VI assumes that the data vectors Y_i are normally distributed, the large-sample properties of the estimators hold for certain nonnormal distributions, and methods exist for adjusting the estimators and tests for elliptical distributions with heavier tails than the normal [25].

LISREL VI is available as part of the SPSS statistical software system, or as an independent program. A similar program, ISU FACTOR [26], can be used with the SAS statistical software system.

One common misconception that users of factor analysis computer software programs have is that the sample correlations of the Y_i can be used in place of the sample covariances without effecting the estimates (particularly in large samples). This is incorrect [3,16]. Although use of sample

correlations removes the problem of choice of scale for the data, the model actually fitted is not the same as that assumed for the linear measurement error model. Adjusting the estimates to the correct scale does not correct for the difference in models. Consequently, the typical use of factor analysis in the chemical literature for the first mass balance model discussed in Section 2 does not necessarily find the mass contributions of the sources as specified in the original model.

Although the errors-in-variables regression model (3) in the structural case can be fitted using LISREL or ISU FACTOR, alternative computer software exists to fit this model directly. First, there exists a substantial numerical analysis literature on total least squares [27,28] dealing with fitting functional and structural errors-in-variables regression models under various assumptions on the error covariance matrices $\Sigma_i = \mathrm{Cov}(e_i)$. These approaches make use of generalized singular value decompositions of the data matrices $\mathbf{Y} = (Y_1, \ldots, Y_n)$ in place of the principal-component type analyses of the sample covariance matrix of the Y_is used by LISREL VI and other covariance structure model software. This yields greater numerical stability and reduced computational complexity (and time). However, the range of models that can be treated by the new total least squares methods is somewhat limited. Such programs also do not provide large-sample measures of accuracy for the estimators, or tests of goodness-of-fit.

Alternatively, the computer program SUPER CARP [14] can handle a wide variety of linear errors-in-variables regression models of both functional and structural type, including models in which the error covariance matrices are heterogeneous ($\mathrm{Cov}(e_i) = \Sigma_i$, with the Σ_i possibly unequal). This program also has the advantages of producing estimated large sample variances for the estimators, providing some diagnostics for goodness-of-fit of the models, and also tests of fit. The estimators produced by SUPER CARP incorporate methods suggested in Ref. 14 that produce solutions that have better performance in samples of moderate size than the maximum likelihood algorithms (which tend to occasionally produce extremely large estimates, or not to converge at all).

A third very useful program is ORDPACK [29,30]. Although this program is somewhat limited in the types of linear measurement error models that it can handle, it has the advantage of also being able to fit nonlinear measurement error models of the functional type. It also incorporates up-to-date numerical analytical optimization methods.

Mention should also be made of robust fitting methods for functional and structural measurement error models. These methods, which use either Huber's approach [31] to robust estimation or Hampel's outlier-resistant theory [32] based on measures of influence of extreme observations, are still under development, but offer the promise of less sensitivity to outliers and other deviant measurements. Some recent references which discuss robust approaches are refs. 33–36. In the chemical mass balance literature, a pioneering effort in this direction is presented in ref. 37.

My own recent research on estimation methods for fitting errors-in-variables regression models has concentrated on a type of 'correction-for-attenuation' approach long used by psychometricians [14,15,38]. To discuss this approach it is convenient to switch to a less subscripted notation for model (3). Let Y_i be the measurements on the latent dependent variables y_i and let X_i be the measurements on the latent predictor variables x_i. Thus

$$y_i = \mathbf{B}x_i + a, \quad Y_i = y_i + e_i, \quad X_i = x_i + f_i \qquad (12)$$

where x_i, e_i, f_i are independent of each other, $E(e_i) = 0$, $E(f_i) = 0$. (Note that the structural form of the model is being assumed, however, recall that good estimators for the structural model are also good estimators for the corresponding functional model.)

Assume that

$$E(x_i) = \mu, \quad \mathrm{Cov}(x_i) = \Sigma_x, \quad \mathrm{Cov}(f_i) = \Sigma_f,$$

$$i = 1, 2, \ldots, n$$

and let

$$\Xi = \Sigma_x(\Sigma_x + \Sigma_f)^{-1} \tag{13}$$

Then if the X_i are normally distributed,

$$E[x_i \mid X_i] = \Xi X_i + (I_r - \Xi)\mu, \qquad i = 1, \ldots, n \tag{14}$$

Even if the X_i are not normally distributed, the right-hand side of (14) is the best linear predictor of x_i given X_i in the sense of minimizing the expected squared-error loss of prediction. The matrix Ξ in (13) is called the reliability matrix of the measurements X_i of the latent predictor variables x_i. If Ξ is known (or can be consistently estimated), substituting $\Xi X_i + (I_r - \Xi)\overline{X}$ for x_i in (12) yields a classical regression model

$$Y_i = B\Xi(X_i - \overline{X}) + (a + B\overline{X}) + e_i^* \tag{15}$$

where

$$e_i^* = B(x_i - \Xi X_i - (I_r - \Xi)\overline{X}) + e_i$$

is uncorrelated with $X_i - \overline{X}$. This model can now be fit by classical least squares or robust regression methods [31] to yield estimates $\hat{\Gamma}$, $\hat{\xi}$ of $\Gamma = B\Xi$ and $\xi = a + B\overline{X}$. Since Ξ is known (or we have a consistent estimator $\hat{\Xi}$ of Ξ), the equations

$$\hat{\Gamma} = \hat{B}\Xi, \qquad \hat{\xi} = \hat{a} + \hat{B}\overline{X}$$

can be solved for $\hat{B}$ and $\hat{a}$. The resulting estimators $\hat{B}$, $\hat{a}$ are then consistent estimators of B and a. In the normal-X_i, normal-Y_i case, $\hat{B}$ and $\hat{a}$ are best asymptotically normal estimators when $\hat{\Gamma}$ and $\hat{\xi}$ are fit by least squares (or maximum likelihood) from (15), and Ξ is either known or the maximum likelihood estimator $\hat{\Xi}$ of Ξ based on the data $X_1, \ldots, X_n$ is substituted for Ξ [15]. Standard confidence regions for the elements of Γ and ξ can easily be converted to confidence regions (and intervals) for the elements of B and a. The method also can be extended to nonlinear errors-in-variables regression models [38].

The main advantage of this approach is apparent. One can use existing statistical software (and confidence region procedures) for classical regression to estimate the parameters. However, there is a price to pay — one must know or estimate Ξ. As noted by Gleser [15,38], this requires either replications on the X_i for each distinct x_i, or the use of independent calibration data for the X_i-measurements. The latter approach is familiar to chemists — for example, one can observe the X_is obtained for known values of the x_i in laboratory experiments. Just how one estimates Ξ depends upon the context — what one is willing to assume about the relationship of the calibration experiments to the experimental context in which the measurements Y_i, X_i are obtained. Although extra information is required for this approach, there is a welcome bonus, in that from Ξ one can determine the accuracy of estimation that can be expected in estimating B and a (and can also spot such potential problems as multicollinearity in the latent variables x_i). Constraints of space do not allow further detail, so individuals interested in this approach should consult Gleser [15,38].

4 LINEAR MASS BALANCE MODELS

As in most other real applications, mass balance models present the statistician with a choice between the desire to reflect all sources of variation and the need for parametric simplicity. For example, both of the mass balance models discussed in Section 2 assumed that the mass fractions (source compositions) a_{ij} do not vary over time. As Cheng and Hopke [37, p.49] note, this is not realistic for all sources j. Consequently, any measurements A_{ij} of the mass fractions a_{ij} taken at a particular time t_0 may not be valid for other times $t \neq t_0$.

Cheng and Hopke [37, p.49] also point out that mass balance models are probably never exactly correct, since some mass may be lost due to chemical reaction along the path taken by particles to the receptor, while on the other hand there may be contributions of mass from sources not accounted for in the model.

A model which reflects the abovementioned sources of variation is the following:

$$C_{it} = c_{it} + \omega_{it}$$

$$c_{it} = \sum_{j=1}^{r} a_{ij}(t)s_{jt} + \epsilon_i(t) \qquad (16)$$

$$A_{ij} = a_{ij}(0) + e_{ij}, \quad 1 \le i \le m, \quad 1 \le j \le r, \quad 1 \le t \le T$$

Here, C_{it} is the measurement at the receptor of the mass of property i at time t, c_{it} is the true mass of property i at time t, ω_{it} reflects errors of measurement in C_{it}, and $\epsilon_i(t)$ is the error in the mass balance equations due to unidentified sources and mass lost to chemical reactions in transit. Also, $a_{ij}(t)$ is the mass fraction of property i from source j at time t. This mass fraction is measured by A_{ij} at time $t = 0$, with error e_{ij}.

In this model, all quantities are assumed to be random. (If $a_{ij}(t)$, $1 \le i \le m$, stays constant over times t for any source j, we will simply assume that the variances of $a_{ij}(t)$, $1 \le i \le m$, $0 \le t \le T$, are zero.) Realistically, those quantities indexed by the time index t should have a time series correlation structure. However, as a strong simplifying assumption, we may assume that the times t at which observations are taken are sufficiently spread out that such correlations are negligible, yet that the underlying process is also sufficiently stable that we can assume that the joint distributions of time-indexed quantities are identical at each time point t. Consequently, we assume that the random matrices $((a_{ij}(t)))$ are independently and identically distributed (i.i.d.) with unknown mean matrix $\Lambda = ((\lambda_{ij}))$. Similarity, we assume that the vectors $s(t) = (s_{1t}, \ldots, s_{rt})'$, $t = 1, \ldots, T$, of mass contributions from sources $1, \ldots, r$ are i.i.d., that the vectors $\epsilon(t) = (\epsilon_1(t), \ldots, \epsilon_m(t))'$, $t = 1, \ldots, T$, are i.i.d. with mean vector 0, and that the vectors $\omega(t) = (\omega_{1t}, \ldots, \omega_{mt})'$, $t = 1, \ldots, T$, of measurement errors in the C_{it} are i.i.d. with mean vector 0.

Let

$$u(t) = \left(\left(a_{ij}(t) - \lambda_{ij}\right)\right) = \left(\left(u_{ij}(t)\right)\right) \qquad (17)$$

The $u_{ij}(t)$ are the values of the random mass fractions $a_{ij}(t)$ centered at their means λ_{ij}. Consequently, the $u(t)$, $0 \le t \le T$ are i.i.d random matrices with $E[u(t)] = 0$.

We assume that the $u(t)$, $0 \le t \le T$, the $s(t)$, $1 \le t \le T$, and the measurement errors $\omega(t)$ and e_{ij}, $1 \le t \le m$, $1 \le j \le r$, are mutually statistically independent. This assumption is reasonable since the variations of the mass fractions and mass contributions are likely to be unrelated to each other, or to errors made in measurement.

Substitution of (17) into (16) yields the following model for the observed quantities C_{it}, A_{ij}:

$$C_{it} = \sum_{j=1}^{r} \lambda_{ij}s_{jt} + \left[\omega_{it} + \epsilon_{it} + \sum_{j=1}^{r} u_{ij}(t)s_{jt}\right] = \sum_{j=1}^{r} \lambda_{ij}s_{jt} + g_{it},$$

$$A_{ij} = \lambda_{ij} + u_{ij}(0) + e_{ij} = \lambda_{ij} + f_{ij}, \quad 1 \le i \le m, \quad 1 \le j \le r, \quad t = 1, \ldots, T \qquad (18)$$

If we let

$$C(t) = (C_{1t}, \ldots, C_{mt})', \qquad g(t) = (g_{1t}, \ldots, g_{mt})'$$

$$A = \left(\left(a_{ij}\right)\right), \qquad f = \left(\left(f_{ij}\right)\right)$$

we can write (18) in vector–matrix form as

$$C(t) = \Lambda s(t) + g(t), \qquad t = 1, \ldots, T$$
$$A = \Lambda + f \qquad (19)$$

The model (19) has the form of a factor analysis model, but with the important addition of an unbiased and independent estimator A of the factor loading matrix Λ. Such a model has not previously been considered in the literature. However, it should be noted that the error term $g(t)$ in (19) does not meet the requirements for classical factor analysis. To see this, note from (18) that $g(t)$ is a function of the vector $s(t)$ of mass contributions from the sources, and also of the equation error vector $\epsilon(t) = (\epsilon_{1t}, \ldots, \epsilon_{mt})$. Since $\epsilon(t)$ reflects both loss of mass due to chemical reaction (which may be related to the total mass released into the environment) and also other unidentified sources of mass (which may be correlated with mass produced by identified sources), any assumption that $\epsilon(t)$ and $s(t)$ are independent could be erroneous. For this reason, and the previously mentioned fact that $g(t)$ is a function of $s(t)$, the usual assump-

tion made in classical factor analysis that $s(t)$ and $g(t)$ are independent seems to be excessively strong. Fortunately, the large-sample properties of classical factor analysis estimates continue to hold under weaker assumptions concerning the joint distribution of $s(t)$ and $g(t)$—see ref. 39. Nevertheless, if we obtain estimates of Λ and the $s(t)$ using classical assumptions, it will be necessary to check that these (or similar) assumptions hold. Such verification will have to be reserved for later research

In the light of model (19), both of the models for linear mass balance mentioned in Section 2 have serious deficiencies. The first (factor analysis) model discussed lacks the identifiability properties of model (19), treats mass fractions $a_{ij}(t)$ as constant over time, and ignores the prior knowledge of estimates A of the factor loading (mass fraction) matrix Λ for known sources. However, this model does share with model (19) the flexibility of allowing an unknown number of sources additional to those explicitly modeled, and in modeling the variation of the source mass contribution vector $s(t)$ over time.

The second linear mass balance model discussed in Section 2 has identifiable parameters and incorporates estimates of Λ. Unfortunately, this model is static (ignores variation in the $a_{ij}(t)$ and s_{jt} over time), and requires prior knowledge of the number r of sources. It also makes the very strong extra distributional assumption that $(C_{it}, A_{it1}, \ldots, A_{itr})'$ are independent, $t = 1, \ldots, m$.

Neither of the two models discussed allows for random errors ϵ_{it} in the mass balance equation due to loss of mass in transit by chemical reaction.

Due to lack of space, it is only possible to sketch an approach to estimation of the parameters in model (19). Any such approach will require us to model the common distributions of the error vectors $g(t)$ and error matrix f, particularly the covariance matrices of their elements.

My own favored mode of approach would be Bayesian (or empirical Bayesian) based on recent work of Press and Shigemasu [40]. These authors provide an approximate (in large samples — here, large T) Bayesian approach to factor analysis using normality assumptions for the $g(t)$ and conjugate priors for the parameters (Λ, the common mean vector and covariance matrix of the $s(t)$, the common covariance matrix of the $g(t)$). Using the prior for Λ, and the data $A = (A_{ij})$, one can update the prior to form a posterior for Λ given A. This posterior distribution can then play the role of the prior distribution of Λ in Press and Shigemasu's Bayesian analysis. Note that this is an appropriate way to use the information conveyed by the measurements A_{ij}, since these are often not really measurements but instead may be partly obtained from subjective judgments of the experimenters. Press and Shigemasu's analysis [40] yields posterior modal estimators of Λ and posterior mode 'predictors' for the source contribution vectors $s(t)$, as well as posterior credible regions (Bayesian confidence regions) for these quantities and tests of fit for the model (particularly for the number of sources r). As already noted, it will be necessary to check whether the large-T properties claimed for these procedures continue to hold under the violations of classical factor analysis assumptions which we have noted in the model (19).

5 CONCLUDING REMARKS

A subject as vast and varied as that of linear measurement error models cannot possibly be covered in a single survey paper. For this reason, the comprehensive surveys in the books of Fuller [14] and Kendall and Stuart [41] are highly recommended. The present paper has highlighted common models, themes and problems in the measurement error literature in the hope that this brief introduction will help chemists gain access to that literature for use in their own research. The modeling and treatment of measurement (and equation) errors is a fundamental problem in the statistical analysis of physical data which must be properly addressed if conclusions reached by scientists are to be valid. Although the problems that arise are analytical difficult, they are unavoidable. Fortunately, some of the best minds in science have addressed these problems over the last fifty years, and there are many useful methods available to practitioners. In the context of linear mass balance models, the strengths and weaknesses of two of

these approaches have been mentioned and a new model incorporating their strengths (in a modeling sense) has been proposed. It is hoped that further research on this and similar models will yield improvements on methods currently used to analyze data based on linear mass balance models.

ACKNOWLEDGEMENT

Research for this paper was supported by National Science Foundation Grant DMS-8901922.

REFERENCES

1 R C. Henry, C.W. Lewis, P K. Hopke and H.J. Williamson, Review of receptor model fundamentals, *Atmospheric Environment*, 18 (1984) 1507–1515.
2 T.W. Anderson, Estimating linear restrictions on regression coefficients for multivariate normal distributions, *Annals of Mathematical Statistics*, 22 (1951) 327–351.
3 T.W. Anderson, Estimating linear statistical relationships, *Annals of Statistics*, 12 (1984) 1–45.
4 T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, 2nd ed., 1987.
5 R Carroll, J. Sacks and C. Spiegelman, A quick and easy multiple use calibration curve procedure, *Technometrics*, 30 (1988) 137–142.
6 G. Knafl, J. Sacks and C. Spiegelman, Calibrating for differences, in L.J. Gleser, M.D. Perlman, S.J. Press and A R. Sampson (Editors), *Contributions to Probability and Statistics. Essays in Honor of Ingram Olkin*, Springer-Verlag, New York, 1989, pp. 335–348.
7 P.J. Brown, Multivariate calibration (with discussion), *Journal of the Royal Statistical Society, Series B*, 44 (1982) 287–321.
8 R. Sundberg and P.J. Brown, Multivariate calibration with more variables than observations, *Technometrics*, 31 (1989) 365–372.
9 J. Buonaccorsi, Errors-in-variables with systematic biases, *Technical Report No. 31*, Societal Institute of the Mathematical Sciences, Department of Biostatistics, Harvard School of Public Health, Boston, MA, 1988.
10 L.J. Gleser, T.W. Anderson's contributions to the study of linear statistical relationship models, in G.P.M. Styan (Editor), *The Collected Papers of T.W. Anderson: 1943–1985*, Wiley, New York, 1990, pp. 1607–1613.
11 T.W. Anderson, Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics, *Journal of the Royal Statistical Society, Series B*, 38 (1976) 1–20.
12 O. Reiersol, Identifiability of a linear relation between variables which are subject to error, *Econometrica*, 18 (1950) 375–389.
13 R.C. Geary, Determinations of linear relations between systematic parts of variables with errors of observation the variances of which are unknown, *Econometrica*, 17 (1949) 30–58
14 W A. Fuller, *Measurement Error Models*, Wiley, New York, 1987.
15 L.J. Gleser, The importance of assessing measurement reliability in multivariate regression, *Technical Report No. 88-19*, Department of Statistics, Purdue University, 1988
16 T.W. Anderson and H. Rubin, Statistical inference in factor analysis, in J. Neyman (Editor), *Proceedings of the Third Berkeley Symposium*, Vol. V, University of California Press, Berkeley, CA, 1956, pp. 111–150
17 L.J. Gleser, Estimation in a multivariate "errors-in-variables" regression model, large sample results, *Annals of Statistics*, 9 (1981) 24–44.
18 L.J. Gleser and J.T. Hwang, The nonexistence of $100(1-\alpha)\%$ confidence sets of finite expected diameter in errors-in-variables and related models, *Annals of Statistics*, 15 (1987) 1351–1362.
18a L.J. Gleser, Confidence intervals for the slope in a linear errors-in-variables regression model, in A K. Gupta (Editor), *Advances in Multivariate Statistical Analysis*, D. Reidel, Dordrecht, 1987, pp. 85–109.
19 N A. Hasabelnaby, J H. Ware and W.A. Fuller, Rejoinder to comments by Leon Jay Gleser, *Statistics in Medicine*, 8 (1989) 1133–1135.
20 M A. Creasy, Confidence limits for the gradient in the linear functional relationship, *Journal of the Royal Statistical Society, Series B*, 18 (1956) 65–69.
21 E.C. Fieller, Some problems in interval estimation, *Journal of the Royal Statistical Society, Series B*, 16 (1954) 175–185.
22 L.J. Gleser, Functional, structural and ultrastructural errors-in-variables models, *Proceedings of the Business and Economics Section*, American Statistical Association, Washington, DC, 1983, pp. 57–66.
23 L.J. Gleser, A note on G.R. Dolby's ultrastructural model, *Biometrika*, 72 (1985) 117–124.
24 K.G. Jöreskog and D. Sörbum, *LISREL VI; Analysis of Linear Structural Relationships by the Method of Maximum Likelihood*, Scientific Software, Chicago, IL, 1984.
25 A. Shapiro and M.W. Browne, Analysis of covariance structures under elliptical distributions, *Journal of the American Statistical Association*, 82 (1987) 1092–1097
26 S G. Pantula, *ISU FACTOR*, Department of Statistics, North Carolina State University, 1983
27 G H. Golub and C.F. Van Loan, An analysis of the total least squares problem, *SIAM Journal of Numerical Analysis*, 17 (1980) 883–893.
28 S. Van Huffel and J. Vandewalle, Analysis and properties of the generalized total least squares problem $AX \approx B$ when some or all columns in A are subject to error, *Matrix Analysis and Applications*, 10 (1989) 294–315.
29 P.T. Boggs, R.H. Byrd, J.R. Donaldson and R.B Schnabel, *Reference Guide for ORDPACK Software for Orthogonal Distance Regression* (Version 1.3, February 4, 1987), National Institute of Standards and Technology, Gaithersburg, MD, 1987.

30 P.T. Boggs, C H Spiegelman, J R. Donaldson and R B Schnabel, A computational examination of orthogonal distance regression, *Journal of Econometrics*, 38 (1988) 169–201.
31 P.J Huber, *Robust Statistics*, Wiley, New York, 1981
32 F R Hampel, The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, 62 (1974) 1179–1186.
33 R.J Carroll and P P Gallo, Some aspects of robustness in functional errors-in-variables regression models, *Communications in Statistics, Part A*, 11 (1982) 2573–2585.
34 R H Zamar, Robust estimation in the errors in variables model, *Technical Report No. 60*, Department of Statistics, University of British Columbia, 1987.
35 R.H. Zamar, Bounded influence estimation in the errors-in-variables model, *Contemporary Mathematics*, in press
36 J Van Ness and C.L. Cheng, Bounded influence errors-in-variables estimation, *Contemporary Mathematics*, in press.
37 M D Cheng and P K. Hopke, Investigation on the use of chemical mass balance receptor model: numerical computations (with discussion), *Chemometrics and Intelligent Laboratory Systems*, 1 (1986) 33–50
38 L.J Gleser, Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models, *Contemporary Mathematics*, in press
39 T.W. Anderson and Y. Amemiya, The asymptotic normal distribution of estimators in factor analysis under general conditions, *Annals of Statistics*, 16 (1988) 759–771.
40 S.J. Press and K Shigemasu, Bayesian inference in factor analysis, in L.J. Gleser, M D. Perlman, S.J Press and A R Sampson (Editors), *Contributions to Probability and Statistics Essays in Honor of Ingram Olkin*, Springer-Verlag, New York, 1989, pp. 271–287.
41 M G. Kendalll and A Stuart, *The Advanced Theory of Statistics*, Vol. 2, Hafner, New York, 4th ed., 1979

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 59–67
Elsevier Science Publishers B.V., Amsterdam

Metrological measurement accuracy: Discussion of "Measurement error models" by Leon Jay Gleser

L.A. Currie

National Institute of Standards and Technology, Gaithersburg, MD 20899 (U.S.A.)

INTRODUCTION

Professor Gleser has provided an exquisite overview and integration of the error structure and statistical modeling that may be employed to characterize the results of modern, multivariable chemical metrology. His demonstration of the equivalence of three representations of the linear, multivariate statistical relationship — as factor analysis (Gleser's eq. (2)), errors-in-variables regression (eq. (3)), and implicit functional (eq. (4)) models — is especially satisfying, in that it makes plain the fact that we may approach linear models in chemistry from apparently different, yet intrinsically equivalent perspectives. His 'new' model (eq. (19)) for treating the inevitable nonlinearities or unsatisfied assumptions in real chemical experiments should prove particularly interesting to those involved in difficult environmental and field studies. Finally, the essential difference between structural and functional models reveals a basic dichotomy: that in the physical sciences we generally find causal (functional) relationships, often involving fixed latent variables, yet the statistical estimation procedures that we must use are 'satisfactory' (in terms of existence and consistency) for the multivariate structural models. Resolution is promised, however, through the asymptotic behavior of the estimators.

The relevance of Professor Gleser's essay to chemical metrology follows from the facts that all of the chemical variables that we measure are subject to error, and at a rapidly increasing pace our measurement systems are producing high dimensional data. Except for defined standards, the 'error-free' independent variables of classical univariate chemistry are, in fact, simply unattainable asymptotes covered by the more general linear measurement error models. Furthermore, the division into dependent and independent classes becomes increasingly problematic as the number of variables increases.

In Gleser's paper we have been given a fundamental overview of statistical issues and statistical references. In keeping with the spirit of chemometrics, I shall attempt to complement that with some chemical approaches, assumptions, and references.

THE METROLOGICAL CONTEXT

As noted above, effectively *all* of our metrological parameters must be viewed as estimates, complete with error (generally random and systematic). Certain characteristics of metrology in the physical sciences, however, have important implications for the measurement error models discussed by Gleser. The most important of these are: (1) theoretical and/or controlled, laboratory-based estimates for the error-covariance matrix; and (2) multiple levels of measurement, where estimated quantities (latent variables) may be more and more remote from the directly observed sig-

nals of chemical sensors. Point-1, already alluded to by Gleser, means that in many cases the variances and correlations may be precisely estimated from physical theory or presumed distribution functions (e.g., Poisson), or they may be derived from extensive, controlled laboratory evaluations. That is, we may often supply the covariance matrix at the outset, rather than estimating it with the linear model that we are fitting. The second point is illustrated by the following metrological level-diagram:

Level	Variable	Realization
1	y: instrumental signal	direct observation (sensor response)
2	x: species-x concentration	calibration, deconvolution of y
3	Θ: source strength/system property	calibration, deconvolution of x

The essential point is that the 'measured quantities' appearing as parameters in the linear measurement error models may themselves be the product of modeling. As we move from level-1 toward level-3, the measurements become more and more indirect. For example, we can never directly observe the concentration (x) of a chemical substance; we must compute it from a calibration model and the response of a chemical sensor. Similarly, we cannot directly observe the strength of a pollutant source (Θ) at a receptor site; we must compute it from the computed chemical concentration vector or matrix (x) obtained at that site.

The importance of the multiple levels of metrology to the application of measurement error models is that the associated deconvolution modeling of signals and concentrations can lead to model error (missing components, systematic model/parameter error, ...) as well as correlated estimates. Further comment on this matter will be given under the subheadings of factor analysis and measurement refinement.

FACTOR ANALYSIS

Factor analysis (FA) is employed in the physical sciences in at least three different ways. As with cluster analysis, it can serve as a very useful exploratory tool, particularly in its graphical mode, to make inferences (or conjectures) concerning concealed relationships in multivariable chemical systems [1] Using principal components projections, one can obtain rather efficient vizualization of high dimensional space, and draw inferences concerning clusters and/or classes of objects, lower dimensional (lines, planes) mixture relations among end member classes, important non-linearities, and possible outliers and/or 'unusual' samples. Beyond pure visualization, one may seek to simplify the representation by removing factors (components) that appear to derive largely from noise, or perform some simple rotations to inspire chemical insight. These applications of FA can be extremely powerful when linked with the well-trained eye or the inspired scientific mind. They are replete with pitfalls, if employed as automatic routines.

A second application of factor analysis is to provide an empirical, linear approximation of the multivariate structure of a chemical class. Such 'class modeling', based on the first few principal components of a class of 'similar' chemical members, commonly known as 'soft modeling', has become one of the major descriptive and discriminating tools for chemical classification and pattern recognition studies [2,3].

The third role for factor analysis is for linear functional modeling. Casual use is ruled out in this case. Assumptions and parameterization must be recognized — viz., we are explicitly treating the model

$$y = xA + e \tag{1}$$

where y is the matrix ($t \times i$) of responses for a given set of variables; x is the matrix ($t \times j$) of pure component concentrations; A is a design or chemical profile matrix ($j \times i$), reflecting normalized responses or spectra of pure components; and e is the measurement error matrix ($t \times i$). (Eq. (1) is the transpose of Gleser's FA equation; it follows the convention of putting 'objects' or samples by the rows of y [4].) The fundamental *chemical* factor analytic issue is that eq. (1) represents a linear functional relationship, it is *not* an eigenvector equation. In other words, the factor score

matrix x has meaning in terms of chemical components, having chemically characteristic spectra (or fingerprints or profiles) represented by the loading matrix A. Thus, although FA should lead to the proper estimate for the number of linearly independent (estimable) chemical components, ad hoc manipulations such as VARIMAX cannot in general be expected to produce chemically correct loadings. (For one thing, chemical profiles are rarely orthogonal.)

It should be noted that eq. (1) is employed broadly, not only in the area of environmental source apportionment ('mass balance' as used by Gleser), but also in the chemical laboratory, where the xs represent concentrations of chemical components of the system being analyzed. These two types of application reflect levels-3 and -2 respectively of the chemical metrology level structure presented earlier. In both cases, residual variance may be employed to estimate measurement error, or to test presumed measurement error.

Several issues related to the validity and application of eq (1) deserve exposure. *First*, is the number of linearly independent components, r. Unfortunately, r is rarely known, except in the case of single components or fully isolated components (as in high resolution spectrometry or chromatography). One of the most important functions of FA, therefore, is to make possible an estimate of r, given an appropriate data matrix. A number of magic rules exist to produce such estimates. One of the more reliable approaches appears to be an F-test, as outlined by Malinowski [5], subject to the constraints that the errors be homogeneous (constant variance over all factors) and uncorrelated. Starting with the least significant principal component, error eigenvalues are tested sequentially for statistical significance. A *second* issue, also treated in ref. 5, relates to the testing of possible target vectors (columns of A matrix) for significance, given the 'abstract factor space' deriving from principal component analysis (PCA). Malinowski observes that this procedure "brings target factor analysis from the quagmire of heuristic reasoning to the realm of statistical inference."

Target factor analysis [6] is one of the approaches for deriving chemically meaningful factors for use with eq. (1). It speaks to the *second* issue, namely model uniqueness, using Gleser's terminology. Among other recommended approaches, perhaps the most famous is that of 'Self modeling curve resolution', invented by Lawton and Sylvestre [7] This technique was developed for two-component systems, and it works well if the samples reasonably span the factor space The extreme samples set inner limits for the unknown spectra or profiles, and non-negativity constraints set outer limits. If unique variables exist for each of the chemical components, then spectra rather than spectral bands may be estimated. Other workers later extended the Lawton and Sylvestre approach to three [8] or more components [9]. Uncertainties for estimated end member (isolated) spectra have been derived by the error propagation technique of Roscoe and Hopke [10,11]. Other means for deriving chemical factors take into account clustering of loadings using the variance diagram technique [12], incorporate physicochemical modeling [13], and compare derived FA spectral windows with spectrochemical data bases [14]. For an excellent review of the several approaches to 'mixture (factor) analysis' see Gemperline [15,16].

The question of finding mutually exclusive, factor-specific (unique) variables is closely related to the 'MLR(T)' technique. Here, one designs the measurement process to contain as many unique tracers as possible. Multiple Linear Regression on the Tracer species then produces spectral or profile estimates for the corresponding sources. This has been especially useful in sorting out the information contained in environmental (mass balance) data matrices [17–19].

The *third* issue. Almost without exception, experts with chemical factor analysis (as embodied in eq. (1)) recommend *avoiding* standardization of the data matrix prior to factor analysis. This is in keeping with the assumption of error homogeneity, and Gleser's comment (Section 3) regarding misuse of the sample correlation matrix. On the other hand, if variables are measured on quite different scales, or exhibit quite different measurement errors, then initial 'scaling' (standardization) is recommended [20]. That means use of a correlation matrix. Quoting Mellinger [21], "the covari-

ance–variance matrix may be used... only when the variables have essentially equal variances." An interesting discussion of the four alternatives — centering or not, scaling or not — is given by Malinowski and Howery [6] A device to use standard FA software (which centers data) for FA about the origin, for environmental source apportionment, was developed by Thurston and Spengler using a fictional null vector [22]

Not unrelated to the question of scaling is the *fourth* issue, data matrix weighting. As noted by Gleser in his discussion of identifiability, classical FA models treat the error-covariance matrices as though they were equal to the same diagonal matrix, independent of sample. This assumption generally does not hold in chemical applications, for several reasons. The primary reason is that chemical measurement error usually increases with increasing concentration; and the concentration of a given element (chemical variable) may vary widely depending on both the relative and absolute amounts of the predominant components in a given sample. A log transform might help, when the relative standard deviation is fixed. A weighted FA solution to the problem has been offered by Cochran and Horne [23], where the variance for data matrix element y_{it} is treated as a product function characteristic of row-t and column-i. These authors demonstrated that classical PCA, which ignores this row–column dependence of the variance, leads to incorrect results.

The *fifth* issue relates more specifically to identifiability — i.e., the confounding of covariance among chemical components, with that associated with their measurement errors. The problem derives from the fact that chemical concentrations (level-2 in the metrological level diagram) are often estimated from a least squares fit to overlapping signals from level-1. This happens for example in the deconvolution of a gamma ray multi[illegible]t, and in corrections for mutual interference in optical or X-ray spectrometry. Thus the error-covariance matrix for the response data matrix used in FA is not necessarily diagonal. Perhaps methods exist for treating known off-diagonal elements in FA, but untreated, they will confound the component estimates. Further comments on this issue will be given in the section on measurement refinement.

Sixth, and last, is the matter of random sampling. In Section 4 of his paper, Gleser observes that all of the quantities in the linear mass balance models, though containing a time index, are assumed to be random, that time series correlation should be made negligible by the sampling strategy. This may be possible in a number of instances, but in many chemical experiments time (and space) variations of chemical component intensities are turned into an advantage. One illustration is found in chromatography, and the relatively new technique of evolutionary factor analysis [24]. Here, cyclic appearances and disappearances of components in time-partitioned data matrices are detected as periodically changing numbers (r) of chemically significant principal components. The time sequence of changes in the number of significant components serves as the first step in identification of species that have different chromatographic elution times. Clearly, analogous temporal phenomena are associated with the transport of atmospheric species; so evolutionary factor analysis could become a very important part of linear mass balance modeling.

ERRORS-IN-VARIABLES REGRESSION

Gleser's 'new' model (his eq. (19)) serves as an excellent conjunction linking the discussion of FA and errors-in-variables regression (EVAR), for it promises incorporation of the best features of each, while compensating for some common deficiencies. Of special interest is the utilization of both the full sample data matrix and prior estimates of the factor loading matrix (chemical spectra or profiles). Classical FA ignores this prior information, while classical EVAR treats data from only one sample at a time. At the Quail Roost-II Workshop on Receptor Modeling via Chemical Mass Balance and Factor Analysis Models, some creative attempts were made to incorporate these two types of information, but no generally satisfactory solution was put forth [25]. Later analyses, based on the same data sets, showed further creative approaches, such as linear programming (LP) and partial least squares (PLS) [26–28]. The PLS solution, in fact was a two-block factor analytic technique that related the principal eigenvectors

of the source profile matrix to those of the sample data matrix — i.e., as with Gleser's new model, it utilized all of the samples together with prior estimates of the source profiles Comments on other advantages of Gleser's new model will appear in the next section.

Returning to 'single sample' EVAR, it is noteworthy that the maximum likelihood estimation (MLE) for two-variable chemical problems has long been recognized as important. MLE has been employed especially in intercalibrations involving two measured variables and in intercomparisons involving two laboratories. Both biased and unbiased methods for incorporating the concomitant 'errors-in-x' are found in the chemical literature [29,30]. Multivariate manifestations are found in the areas of multicomponent gamma ray spectrometry and multicomponent source apportionment (chemical mass balance modeling) [31,32].

Because of the importance of this topic in modern environmental and analytical chemistry, Beebe and Currie undertook an empirical evaluation of popular algorithms/software for treating the problem [33]. Specifically, the methods mentioned in Gleser's paper, effective variance weighted least squares (EVWLS), orthogonal distance regression (ODR) [34] and the MLE (structural model) method of Fuller [35], were tested with bi- and trivariate data sets having known structure. Details will be found in ref. 33, but two of the essential conclusions were that ODR was relatively less precise, but unbiased, while EVWLS gave accurate precision estimates, and was as precise as MLE, but biased. This was surprising, because the formulation of EVWLS in ref. 32 seemed equivalent to MLE. On further reading, however, one finds an approximation that makes its implementation equivalent to iteratively weighted least squares (IWLS) which is known to produce biased estimates [29]. This is a rather serious discovery, for EVWLS is the currently accepted method for chemical mass balance (regression) calculations.

Gleser's proposals for correcting for attenuation (bias) are especially welcome, given the foregoing observation. The reliability matrix (his eq. (13)) and the expanded regression model error (below eq. (15)) hold the key. This very facile solution to an important class of chemical problems is all the more practicable, because it can be applied using standard linear regression software. The 'price' we must pay, estimation of the reliability matrix, is not unreasonable. As Gleser shows in ref. 36, the covariances comprising the reliability matrix come directly from: (a) the set of observed variable values (Σ_x), and (b) the difference ($\Sigma_x - \Sigma_f$) where Σ_f represents the covariance matrix of measurement errors These latter are the same errors (variances) we now employ in EVWLS and IWLS; they may be estimated through replication or 'theory'.

MEASUREMENT REFINEMENT

In the last parts of this discussion I should like to comment on aspects on the problem where the chemist can make his most important contributions, given the insights concerning measurement error models provided by the mathematician-statistician This represents the synergism which is the true benefit of cross-disciplinary research. By refining the measurement process, the chemist can reduce or eliminate errors associated with multicollinearity, identifiability, and certainly model uniqueness. By model refinement, using known physicochemical relationships, otherwise erroneous, linear model assumptions may be averted.

Perhaps the most obvious measurement refinement relates to the relative magnitudes of the measurement errors across species and/or samples. (Reducing the absolute magnitudes of the measurement errors, of course, always helps; this should be done to the extent feasible.) Planning measurements to control the relative magnitudes of measurement errors is interesting because it can influence multicollinearity. For example, the matrix to be inverted in weighted regression analysis is $A'WA$, where A is the design matrix and W is the diagonal matrix of weights (inverse variances). Altering the relative weights thus alters the 'condition' of this critical matrix of linear regression. In fact, an optimum may be achieved by maximizing the determinant of this matrix, the Fisher information [18]. Chemical insight is related to this

issue in two ways: deciding which variables are most important for increased weight (depends on the mix of likely source components), and deciding how to accomplish the measurement task. When weights depend on signal magnitude, as they often do in chemical measurements, then iteration is necessary to take into account the y dependence. The basic question is one of iterative, intelligent design of the chemical measurement process.

Closely related is the issue of chemical interference and the corresponding off-diagonal elements of the sample covariance matrix. This is a very real issue for overlapping spectra or chromatographic peaks in laboratory analysis, and it has important consequences for environmental mass balance studies where level-2 metrological data (estimated chemical concentrations) are employed in level-3 models. Covariance among concentration estimates must be avoided for classical FA; quoting Anderson: "an essential assumption is that the [error] covariance matrix is diagonal" [37]. To achieve this costs money. To illustrate the point, in air particulate receptor modeling it is common to measure a host of element concentrations using X-ray fluorescence analysis (XRF). The method is inexpensive (ca. $40/sample) but insensitive for certain elements (e.g., those with low atomic number, such as carbon, boron), and exhibits interferences for others (e.g., lead L-X rays interfere with arsenic K-X rays). Correction for interference, often done by regression techniques, necessarily induces covariance between the estimated (corrected) concentrations. A more expensive technique (by a factor of three to five), neutron activation analysis (NAA), will often overcome both limitations, though special interferences (dependent on nuclear properties) may occur here. Unique tracer techniques generally cost even more, but they may *eliminate* collinearity among certain sources; and often the specialized, single species measurement process has no interspecies interference. The price is higher. A case in point is ^{14}C, which we measure to unambiguously resolve fossil from biospheric carbon sources (cost: ca. two to five times that of NAA).

Use of ^{14}C illustrates measurement refinement by paying attention to the chemical question concerning *what* to measure. By employing a tracer of this sort that is both unique and absolute, one can accomplish other ends. Namely, inexpensive (XRF) unique tracers (mineral-corrected potassium, lead) that are not absolute can be calibrated, thus achieving reliability for a given airshed, but at reduced cost [18,38]. Reliable (orthogonal) tracers can also be added to the design of the overall experiment. An example is a recent EPA sponsored study of carbonaceous aerosol sources in Roanoke, VA, U.S.A. Here, ^{14}C was employed in the validation/calibration mode discussed above; this step resolved wood-burning carbon from fossil carbon in the atmosphere. As a second step, stable rare earth isotopes were purposely added to label fuel oil in the area. Their signatures provided added 'orthogonal' resolution of this component of the atmospheric soot from the fossil component from motor vehicles [39]. A statement by Rao marvelously supports the philosophy of such approaches to measurement refinement in quite another field: "Possibly what is wrong with the economists is that they are not trying to refine their measurements or trying to measure new variables which cause economic change. That is far more important that dabbling with whatever data are available and trying to make predictions based on them" [40].

MODEL REFINEMENT

Not far removed is the subject of model refinement. Gleser's proposed model (eq. (19)) speaks to this. As recognized also by Cheng and Hopke [26], *real* receptor models are not linear. There are selective changes in particle composition during transport, including physical effects (agglomeration, settling) and chemical effects (reaction). I believe that the most effective way to account for such nonlinearities is to employ carefully constructed physicochemical models of the respective processes. The alternative, which will not be further discussed here, is to use chemical knowledge and data to select those species that are 'chemically robust' — i.e., conservative (linear) tracers that resist change, isotopes and nonreactive gases being classic examples. Physicochemical modeling for source apportionment has been dubbed 'hybrid modeling'. Examples are seen in the use of

reaction rate constants to help model the gas-to-particle conversion of sulfur dioxide to sulfate [41] and the selective oxidation of polycyclic aromatic hydrocarbons during atmospheric transport [42]. An interesting statistical challenge, for better representing 'real' behavior, would be to describe individual source profiles (columns of the A matrix, eq (1)) as empirical principal component class models [2] to serve as prior information for source apportionment by FA and EVAR.

Model refinement can be considered in a larger, more generic sense. Realizing that our models are imperfect 'cartoons' or caricatures of reality generally emphasizing (distorting?) particular perspectives or parameters, it is meaningful to consider classes of models, having varying degrees of refinement (and corresponding increases in cost). In atmospheric chemistry, for example, we may look beyond the relatively simple hybrid models mentioned above, to two and three dimensional (spatial) models of the temporal processes taking place Such 'full dynamic modeling' relies heavily on highest quality numerical methods, plus statistics, but it must be fundamentally based on sound, detailed physical and chemical analysis of the system. Pertinent illustrations of such model classes are given in Table 1, for atmospheric chemistry together with two other fields of endeavor. This viewpoint was presented for atmospheric modeling in ref. 43, it was inspired by Hofstadter [44].

Considerable insight into the relation between model realism and viewpoint, and metrological accuracy, can be gained by examining the evolution of oceanographic models Like atmospheric models, they have been designed to describe the state of the fluid system, including concentrations and transport of chemical constituents. In both areas of environmental science, the simplest models frequently serve quite well for estimation and prediction of a limited set of parameters. In oceanography, one of the driving forces has been the need to understand the effect of anthropogenic carbon dioxide perturbations on the atmosphere–ocean system — a central problem for forecasts of global warming. The earliest models simply treated the spatially averaged atmosphere and world oceans as two or three reservoirs [45,46]. Far more realistic is the box-diffusion model for vertical transport in the ocean, which treats the upper layer as well mixed and describes the ocean below the thermocline as an infinite set of boxes — i.e., as a diffusive medium [47]. This model, which still describes a fictitious 'average' ocean, has been compared with more realistic representations of the ocean which take into account horizontal transport as well as upwelling of deep ocean water in the equatorial zone and downwelling in the temperate and polar zones. It was found that the box-diffusion model "gives an excellent representation of atmospheric CO_2 and $^{14}CO_2$ interactions on time scales up to several tens of years" and hence near-term effects of fossil fuel combustion on global climate [48]. Expanding the temporal scale (to glacial times) and the number of chemical variables observed required a considerably more complex (realistic?) model. 'Pandora's Box' [49] *.

TABLE 1

Model refinement

Music [44]	Atmospheric science	Oceanography
Muzak	Linear models [19] (conservative tracer)	2-box [45,46] (above/below thermocline)
Jazz	Hybrid [26] (SO_4 [41], PAH [42])	Box-diffusion [47] (surface, deep ocean)
Classical music	1, 2, 3D dynamic, reacting system ↕ reality	'Pandora' [49] (multicompartment/ flows)

* Another, cogent illustration of environmental model complexity and relevance has just come to my attention, from the field of ground water hydrology. As with the several imperfect views of the ocean (and the classic multiple perspectives of the elephant), the particular perspective of reality embodied in the hydrological model (or 'cartoon') determined its predictive validity. In this case, a construct was created to describe the behavior of ground water in fractured zones, and it was parameterized with the most accessible observable, the fluctuating ground water level. Once calibrated, the model did well at predicting ground water levels; but when a new need arose, forecasting transport of pollutants, it failed completely (E.A. Prych, personal communication, 1990).

Thus, enormous advances in geochemical sampling, multivariable chemical measurement, and computational power make possible model refinements that approach reality. The challenge to the chemist and statistician, however, is to define just what level of complexity is appropriate — i.e., to provide guidance as to the nature and magnitude of errors in measurements and in models that are actually relevant.

REFERENCES

1 R.R. Meglen, Chemometrics: its role in chemistry and measurement sciences, *Chemometrics and Intelligent Laboratory Systems*, 3 (1988) 17–29.

2 S. Wold and M. Sjöström, SIMCA: A method for analyzing chemical data in terms of similarity and analogy, in B.R. Kowalski (Editor), *Chemometrics: Theory and Application*, ACS Symposium Series 52, American Chemical Society, Washington, DC, 1977, p. 243.

3 K.J. Schostack and E.R. Malinowski, Preferred set selection by iterative key set factor analysis, *Chemometrics and Intelligent Laboratory Systems*, 6 (1989) 21–29.

4 S. Wold, C. Albano, W.J. Dunn III, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg and M. Sjöström, Multivariate data analysis in chemistry, in B.R. Kowalski (Editor) *Chemometrics: Mathematics and Statistics in Chemistry*, D. Reidel, Dordrecht, 1984, pp. 17–96.

5 E.R. Malinowski, Statistical *F*-tests for abstract factor analysis and target testing, *Journal of Chemometrics*, 3 (1988) 49–60.

6 E.R. Malinowski and D.G. Howery, *Factor Analysis in Chemistry*, Wiley, New York, 1980.

7 W.H. Lawton and E.A. Sylvestre, Self modeling curve resolution, *Technometrics*, 13 (1971) 617–633.

8 O.S. Borgen and B.R. Kowalski, An extension of the multivariate component-resolution method to three components, *Analytica Chimica Acta*, 174 (1985) 1–26.

9 R.C. Henry, Self-modeling curve resolution and other linear constraints on factor analysis of airborne particle composition data, *Mathematics in Chemistry Conference, College Station, 8–10 November 1989*.

10 B.A. Rosece and P.K. Hopke, Error estimates for factor loadings and scores obtained with target transformation factor analysis, *Analytica Chimica Acta*, 132 (1981) 89–97.

11 E.R. Malinowski, Obtaining the key set of typical vectors by factor analysis and subsequent isolation of component spectra, *Analytica Chimica Acta*, 134 (1982) 129–137.

12 W. Windig, W.H. McClennen and H.L.C. Meuzelaar, Determination of fractional concentrations and exact component spectra by factor analysis of pyrolysis mass spectra of mixtures, *Chemometrics and Intelligent Laboratory Systems*, 1 (1987) 151–165.

13 E.A. Sylvestre, W.H. Lawton and M.S. Maggio, Curve resolution using a postulated chemical reaction, *Technometrics*, 16 (1974) 353–368.

14 B.G.M. Vandeginste, Application of chemometrics, *Pure and Applied Chemistry*, 55 (1983) 2007–2016.

15 P.J. Gemperline, Mixture analysis using factor analysis. I. Calibration and quantitation, *Journal of Chemometrics*, 3 (1989) 549–568.

16 J.C. Hamilton and P.J. Gemperline, Mixture analysis using factor analysis. II: Self-modeling curve resolution, *Journal of Chemometrics*, 4 (1990) 1–13.

17 C.W. Lewis, R.E. Baumgardner, R.K. Stevens, L.D. Claxton and J. Lewtas, The contribution of woodsmoke and motor vehicle emissions to ambient aerosol mutagenicity, *Environmental Science and Technology*, 22 (1988) 968–971.

18 L.A. Currie, K.R. Beebe and G.A. Klouda, What should we measure? Aerosol data: past and future, *Proceedings of the 1988 EPA/APCA International Symposium on Measurement of Toxic and Related Air Pollutants*, Air Pollution Control Association, Pittsburgh, PA, 1988, pp. 853–863.

19 R.C. Henry, C.W. Lewis, P.K. Hopke and H.J. Williamson, Review of receptor model fundamentals, *Atmospheric Environment*, 18 (1984) 1507–1515.

20 S. Wold, Principal component analysis, *Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 37–52.

21 M. Mellinger, Multivariate data analysis: its methods, *Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 29–36

22 G.D. Thurston and J.D. Spengler, A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston, *Atmospheric Environment*, 19 (1985) 9–25.

23 R.N. Cochran and F.H. Horne, Statistically weighted principal component analysis of rapid scanning wavelength kinetics experiments, *Analytical Chemistry*, 49 (1977) 846–853.

24 K. Schostack, P. Paruch, S. Patel and E.R. Malinowski, Evolutionary factor analysis, *Journal of Research of the National Bureau of Standards*, 93 (May/June 1988) 256–257.

25 L.A. Curne, R.W. Gerlach, C.W. Lewis, W.D. Balfour, J.A. Cooper, S.L. Dattner, R.T. DeCesar, G.E. Gordon, S.L. Heisler, R.K. Hopke, J.J. Shah, G.D. Thurston and H.J. Williamson, Interlaboratory comparison of source apportionment procedures: results for simulated data sets, *Atmospheric Environment*, 18 (1984) 1517–1537.

26 M.D. Cheng and P.K. Hopke, Investigation on the use of chemical mass balance receptor model: numerical computations, *Chemometrics and Intelligent Laboratory Systems*, 1 (1986) 33–50; L.J. Gleser, Invited comments.

27 I.E. Frank and B.R. Kowalski, Statistical receptor models solved by partial least squares, in J.J. Breen and P.E. Robinson (Editors), *Environmental Applications of Chemometrics*, ACS Symposium Series 292, American Chemical Society, Washington, DC, 1985, pp. 271–279.

28 L.A. Currie, The limitations of models and measurements as revealed through chemometric intercomparison, *NBS*

Journal of Research, 90 (1986) 409; L.J. Gleser, Invited comments.
29 B.D. Ripley and M. Thompson, Regression techniques for the detection of analytical bias, *Analyst (London)*, 112 (1987) 377–383.
30 D. York, Least squares fitting of a straight line, *Canadian Journal of Physics*, 44 (1966) 1079–1086.
31 R.M. Parr and H.F. Lucas, Jr., A rigorous least squares analysis of complex gamma-ray spectra with partial compensation for instrumental instability, *IEEE Transactions on Nuclear Sciences*, NS-11 (3) (1964) 349.
32 J.G. Watson, J.A. Cooper and J.J. Huntzicker, The effective variance weighting for least squares calculations applied to the mass balance receptor model, *Atmospheric Environment*, 18 (1984) 1347–1355.
33 K.R. Beebe and L.A. Currie, A comparison of errors in variables regression procedures, in preparation.
34 P.T. Boggs and J.R. Donaldson, Orthogonal distance regression, *NIST Internal Report 89 4197* (1989). Proceedings of the American Mathematical Society Joint Summer Research Conference on Statistical Analysis of Measurement Error Models and Their Applications, Humboldt State University, CA, June 1989.
35 W.A. Fuller, *Measurement Error Models*, Wiley, New York, 1987.
36 L.J. Gleser, Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models, *Contemporary Mathematics*, (1989) in press.
37 T.W. Anderson, Estimating linear statistical relationships, *Annals of Statistics*, 12 (1984) 1–45.
38 R.K. Stevens, C.W. Lewis, T.G. Dzubay, R.E. Baumgardner, R.B. Zweidinger, V.R. Highsmith, L.T. Cupitt, J. Lewtas, L.D. Claxton, L.A. Currie, G.A. Klouda and B. Zak, Mutagenic atmospheric aerosol sources apportioned by receptor modeling, in W.L. Zielinski, Jr. and W.D. Dorko (Editors), *Monitoring Methods for Toxics in the Atmosphere*, ASTM, Philadelphia, PA, 1990, pp. 187–196.
39 W.R. Kelly, personal communication, 1989.
40 M.H. DeGroot, A conversation with C.R. Rao, *Statistical Science*, 2 (1987) 53–67.
41 C.W. Lewis and R.K. Stevens, Hybrid receptor model for secondary sulfate from an SO_2 point source, *Atmospheric Environment*, 19 (1985) 917–924.
42 S.K. Friedlander, New developments in receptor modeling theory, in E.S. Macias and P.K. Hopke (Editors), *Atmospheric Aerosol. Source/Air Quality Relationships*, ACS Symposium Series 167, American Chemical Society, Washington, DC, 1981, Ch. 1.
43 L.A. Currie, Chemometrics and standards, *Journal of Research of the National Bureau of Standards*, 93 (May/June 1988) 193–205.
44 D.R. Hofstadter, Flexible concepts and creative analogies: a computer model, lecture given at NASA—Goddard, May 1986.
45 R. Revelle and H.E. Suess, Carbon dioxide exchange between atmosphere and ocean, and the question of an increase of atmospheric CO_2 during the past decades, *Tellus*, 9 (1957) 18–27.
46 B. Bolin and E. Eriksson, Changes in the carbon dioxide content of the atmosphere and sea due to fossil fuel combustion, in *The Atmosphere and the Sea in Motion, Rossby Memorial Volume*, Rockefeller Institute Press, New York, 1959, pp. 130–143.
47 H. Oeschger, The contribution of radioactive and chemical data to the understanding of the environmental system, in L.A. Currie (Editor) *Nuclear and Chemical Dating Techniques: Interpreting the Environmental Record*, ACS Symposium Series 176, American Chemical Society, Washington, DC, 1982, pp. 5–42.
48 W.S. Broecker, T.-H. Peng and R. Engh, Modeling the carbon system, *Radiocarbon*, 22 (1980) 565–598.
49 W.S. Broecker and T.-H. Peng, Carbon cycle 1985, glacial to interglacial changes in the operation of the global carbon cycle, *Radiocarbon*, 28 (1986) 309–327.

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 69–79
Elsevier Science Publishers B.V., Amsterdam

How chemical kinetics uncertainties affect concentrations computed in an atmospheric photochemical model

Anne M. Thompson * and Richard W. Stewart
NASA / Goddard Space Flight Center, Laboratory for Atmospheres, Greenbelt, MD 20771 (U S A.)

(Received 7 November 1989; accepted 20 July 1990)

Abstract

Thompson, A M and Stewart, R W., 1991. How chemical kinetics uncertainties affect concentrations computed in an atmospheric photochemical model. *Chemometrics and Intelligent Laboratory Systems*, 10: 69–79.

Tropospheric photochemical models are used increasingly as predictive tools to assess the chemical response of the lower atmosphere to changes in physical and chemical conditions which influence trace species distributions. Among the many uncertainties in the modeling process are imprecisions in reaction rate data used in formulating model continuity equations. In this paper we evaluate the propagation of these kinetics uncertainties to computed species distributions in a photochemical model.

A one-dimensional kinetics-diffusion model having 72 reactions among 24 species is used. Non-chemical sources and initial background concentrations are chosen to be representative of clean continental mid-latitude air. Chemical reaction rate data are mostly those of the NASA Kinetics Evaluation Panel No. 8 (1987) and include imprecisions in photolysis rates and binary and ternary reactions. A Monte Carlo technique is used to estimate uncertainties in computed concentrations due to the given rate uncertainties.

We compute uncertainties in odd hydrogen species (the radicals OH and HO_2) and in hydrogen peroxide ranging from 22–41%. Uncertainties for O_3 and CO are, respectively, 17% and 30%. Odd nitrogen uncertainties range from 18% for NO to 72% for N_2O_5. The smallest uncertainty is that for nitric acid at 6%, but this is neglecting uncertainties in physical sources and sinks, such as precipitation scavenging. The uncertainty in OH (31%) is important when using the model to predict tropospheric oxidant levels because OH determines the lifetime of numerous naturally and anthropogenically emitted trace gases.

INTRODUCTION

One-dimensional photochemical models are used to simulate vertical profiles of trace gas distributions (O_3, NO_x, CO, OH, H_2O_2) in the atmosphere. We have used a model of the troposphere to predict changes in atmospheric composition, primarily levels of the oxidants O_3, OH, and H_2O_2, as emissions of NO, CO, and CH_4 change over the next several decades [1–3]. We also use the model to interpret trace gas measurements in selected field experiments, calculating ozone production in convective situations [4].

In both predictive and interpretive modes, the photochemical model gives results that are uncertain at least to the degree that key photochemical reaction rates are uncertain and mechanistic pathways for some reactions are not known in detail. We have evaluated some of these effects and report on an investigation of uncertainties in calculated trace gas concentrations due to the imprecision of photochemical reaction rates.

A Monte Carlo method is used to specify sets of photochemical reaction rates, with means and uncertainties given from a standard tabulation of kinetics and absorption spectra. The overall uncertainty or likely range of concentrations for a given species is determined by hundreds of runs in which each rate coefficient is selected randomly and a steady-state solution is computed for all species. The Monte Carlo rate kinetics study is carried out for one type of background chemistry, simulating a northern mid-latitude continental environment. Each solution describes a unique atmospheric composition and when these are averaged together, the mean is taken as representative of this type of chemical regime. In a related study [5] we report on how species uncertainties vary with mean composition when other chemical environments are simulated.

METHOD

Photochemical model

A one-dimensional photochemical-kinetics model solves the continuity equation for the concentration of the ith species, c_i, as a function of time, t:

$$\frac{\partial}{\partial z}\left[K(z,t)N(z)\frac{\partial \chi_i}{\partial z}(z,t)\right] + P_i(z,t) - L_i(z,t) = \frac{\partial c_i}{\partial t}(z,t) \quad (1)$$

where z = altitude (cm, in our model); $K(z)$ is an eddy diffusion coefficient (in $cm^2\ s^{-1}$, assumed to be time-independent); $N(z)$ is molecular density (cm^{-3}); $\chi_i(z, t)$ is mixing ratio or mole fraction of species i. $P_i(z, t)$ and $L_i(z, t)$ are photochemical production and loss terms, respectively, for species i. Photochemical reactions making up production and loss include photodissociation or thermal dissociation reactions, in which the species i is a fragment formed by a unimolecular process; bimolecular reactions between two free radicals or a free radical and a nonradical species; and three-body processes in which combination of two radicals in concert with an energetically stabilizing third body leads to formation of a nonradical molecule.

Our photochemical model spans 0–15 km (the latter taken as mean height of the tropopause)

TABLE 1

Trace gases and boundary conditions in photochemical model

Species	Upper boundary (15 km)
O_3	influx, $5 \times 10^{10}\ cm^{-2}\ s^{-1}$
$O(^3P)$	influx, $4 \times 10^3\ cm^{-2}\ s^{-1}$
CH_3, CH_3O, CH_3O_2, CH_3OOH, $C_2H_5O_2$, H_2O_2, C_2H_5OOH, CH_3CO_3, H, OH, HO_2	photochemical equilibrium
$NO_y(NO + NO_2 + NO_3 + HNO_3 + HNO_4 + 2N_2O_5)$	influx, $2.5 \times 10^8\ cm^{-2}\ s^{-1}$
H_2CO, PAN, CH_3CHO	zero flux
CO, C_2H_6	troposphere-to-stratosphere transfer
	Lower boundary (0 km)
O_3	deposition
$O(^3P)$	deposition
CH_3, CH_3O, CH_3O_2, $C_2H_5O_2$, CH_3CO_3, H, OH, HO_2	photochemical equilibrium
NO	flux
NO_2	deposition
NO_3, N_2O_5	deposition
PAN *	deposition
H_2CO, CH_3OOH, CH_3CHO, C_2H_5OOH *	deposition
H_2O_2, HNO_3, HNO_4 *	deposition
C_2H_6	fixed, 1.5 ppbv
CO	flux

* These species also rained out with first-order removal below 6 km.

TABLE 2

Photolysis reactions

Model reaction	Photodissociation	% Standard dev., NASA/JPL [8]	% Standard dev *, Monte Carlo
J_1	$O_3 + h\nu \rightarrow O_2 + O$	10	9.7
J_2	$O_3 + h\nu \rightarrow O_2 + O(^1D)$	40	40
J_3	$NO_2 + h\nu \rightarrow NO + O$	30	30
J_4	$HNO_3 + h\nu \rightarrow OH + NO_2$	30	29
J_5	$H_2O_2 + h\nu \rightarrow OH + OH$	40	43
J_6	$NO_3 + h\nu \rightarrow NO + O_2$	100	83
J_7	$NO_3 + h\nu \rightarrow NO_2 + O$	100	81
J_8	$H_2CO + h\nu \rightarrow H + HCO$	40	38
J_9	$CH_3OOH + h\nu \rightarrow OH + CH_3O$	40	37
J_{10}	$HNO_4 + h\nu \rightarrow HO_2 + NO_2$	100	89
J_{11}	$CH_3CHO + h\nu \rightarrow CH_3 + HCO$	40 **	37
J_{12}	$N_2O_5 + h\nu \rightarrow NO_2 + NO_3$	100	82
J_{13}	$H_2CO + h\nu \rightarrow H_2 + CO$	40	35
J_{14}	$C_2H_5OOH + h\nu \rightarrow C_2H_5O + OH$	40 ***	37
J_{15}	$PAN + h\nu \rightarrow CH_3CO_3 + NO_2$	30 §	27

* From 800 model runs.

** Specified uncertainty assumed in analogy with H_2CO.

*** Specified uncertainty assumed in analogy with CH_3OOH.

§ Specified uncertainty assumed in analogy with HNO_3.

with 24 grid points [1,6]. Spacing is at 1-km intervals between 1 and 15 km and on a refined grid below 1 km to give better simulations of gradients in the boundary layer. Several types of boundary conditions are specified, depending on the species: photochemical equilibrium, flux, fixed mixing ratio, or removal at surface or tropopause with a specified transfer velocity. We calculate vertical profiles of 24 trace species, a standard complement of odd oxygen (O_3, $O(^3P)$); odd hydrogen (H, OH, HO_2), odd nitrogen (NO, NO_2, NO_3, N_2O_5, HNO_3, $HNO_4 = HO_2NO_2$), hydrocarbons derived from oxidation of CH_4 (CH_3, CH_3O_2, H_2CO, CH_3OOH, CO) and C_2H_6 and its oxidation products, including peroxy acetyl nitrate ($C_2H_5O_2$, C_2H_5OOH, CH_3CHO, CH_3CO_3, PAN). A list of species and boundary conditions is given in Table 1. The set of chemical reactions used in the model appears in Tables 2 and 3.

Eq. (1) is solved by finite differencing after converting to a set of nonlinear algebraic expressions of form:

$$\frac{d\bar{X}}{dt} = \bar{f}(\bar{X}, t)$$

$$\bar{X} = X_1^1, X_2^1, X_3^1, \ldots, X_{ns}^1, X_1^2, X_2^2, \ldots, X_{ns}^2, \ldots, X_1^{np}, X_2^{np}, X_3^{np}, \ldots, X_{ns}^{np} \quad (2)$$

where $X_i^j = i$th species mixing ratio at altitude grid point j; $\bar{f}$ = forcing function which is a sum of flux divergence, and rates of chemical reaction; ns = total number of chemical species; np = total number of spatial grid points. The mixing ratios are obtained from integration of (2).

In performing sensitivity calculations, as for example in simulating perturbed emissions or varying reaction rate coefficients, a steady-state version of the model is used. This means simultaneous solution of eqs. (2) where $d\bar{X}/dt = 0$ and diurnally averaged reaction rates and species concentrations are computed according to the method of Turco and Whitten [7]. Diurnally averaged rate coefficients and photolysis rates are used in the steady-state version and the desired means are approximated:

$$\overline{k_{ij}X_iX_j} = (DF)_{ij}k_{ij}\bar{X}_i\bar{X}_j \quad (3)$$

The reaction or loss term in eq. 3 is the product of diurnally averaged species mixing ratios $\bar{X}_i$ and $\bar{X}_j$ and the diurnally averaged rate coefficient is

$$\bar{k}_{ij} = (DF)_{ij}k_{ij} \quad (4)$$

where $(DF)_{ij}$ is a diurnal averaging factor and k_{ij}

TABLE 3

Photochemical reactions, rates, and uncertainties

Model reaction number	Bimolecular reaction	Rate and uncertainty factors		
		A-Factor	$E/R \pm \Delta E/R$	$f(298)$
2	$O + O_2 \rightarrow 2O_2$	8.0×10^{-12}	2060 ± 250	1.15
4	$O(^1D) + N_2 \rightarrow O + N_2$	1.8×10^{-11}	$-(110 \pm 100)$	1.2
5	$O(^1D) + O_2 \rightarrow O + O_2$	3.2×10^{-11}	$-(70 \pm 100)$	1.2
6	$NO + O_3 \rightarrow NO_2 + O_2$	2.0×10^{-12}	1400 ± 200	1.2
7	$NO_2 + O \rightarrow NO + O_2$	6.5×10^{-12}	$-(120 \pm 120)$	1.1
8	$NO_2 + O_3 \rightarrow NO_3 + O_2$	1.4×10^{-13}	2500 ± 140	1.15
9	$NO + NO_3 \rightarrow NO_2 + NO_2$	1.7×10^{-11}	$-(150 \pm 100)$	1.3
12	$N_2O_5 + N_2 \rightarrow NO_2 + NO_3 + N_2$	5.7×10^{14}	10600	
13	$O(^1D) + H_2O \rightarrow OH + OH$	2.2×10^{-10}	0 ± 100	1.2
14	$O(^1D) + CH_4 \rightarrow OH + CH_3$	1.4×10^{-10}	0 ± 100	1.2
15	$O(^1D) + CH_4 \rightarrow H_2 + H_2CO$	1.4×10^{-11}	0 ± 100	1.2
16	$O(^1D) + H_2 \rightarrow OH + H$	1.0×10^{-10}	0 ± 100	1.2
17	$H + O_3 \rightarrow OH + O_2$	1.4×10^{-10}	470 ± 200	1.25
19	$OH + O_3 \rightarrow HO_2 + O_2$	1.6×10^{-12}	940 ± 300	1.3
20	$HO_2 + O_3 \rightarrow OH + 2O_2$	1.1×10^{-14}	$500 + 500/-100$	1.3
21	$OH + O \rightarrow H + O_2$	2.2×10^{-11}	$-(120 \pm 100)$	1.2
22	$HO_2 + O \rightarrow OH + O_2$	3.0×10^{-11}	$-(200 \pm 100)$	1.2
23	$H_2O_2 + O \rightarrow OH + HO_2$	1.4×10^{-12}	2000 ± 1000	2.0
24	$OH + CH_4 \rightarrow CH_3 + H_2O$	2.3×10^{-12}	1700 ± 200	1.2
25	$HO_2 + NO \rightarrow OH + NO_2$	3.7×10^{-12}	$-(240 \pm 80)$	1.2
26	$OH + CO \rightarrow CO_2 + H$	$1.5 \times 10^{-13} \times (1 + 0.6p)$	0 ± 300	1.3
27	$OH + H_2 \rightarrow H_2O + H$	5.5×10^{-12}	2000 ± 400	1.2
29	$OH + HNO_3 \rightarrow H_2O + NO_3$	**		1.3
30	$OH + H_2O_2 \rightarrow H_2O + HO_2$	3.3×10^{-12}	$200 + 100/-300$	1.3
31	$OH + HO_2 \rightarrow H_2O + O_2$	4.6×10^{-11}	$-(230 \pm 200)$	1.3
32	$OH + OH \rightarrow H_2O + O$	4.2×10^{-12}	240 ± 240	1.4
33	$OH + H_2CO \rightarrow H_2O + HCO$	1.0×10^{-11}	0 ± 200	1.25
34	$HO_2 + HO_2 \rightarrow H_2O_2 + O_2$	2.3×10^{-13}	$-(600 \pm 200)$	1.3
36	$H + HO_2 \rightarrow H_2 + O_2$	7.3×10^{-12}	0 ± 200	1.3
37	$H + HO_2 \rightarrow H_2O + O$	3.2×10^{-12}	0 ± 200	1.3
38	$H + HO_2 \rightarrow OH + OH$	7.0×10^{-11}	0 ± 200	1.3
39	$H_2CO + O \rightarrow OH + HCO$	3.4×10^{-11}	$+1600 \pm 250$	1.25
41	$CH_3O_2 + NO \rightarrow CH_3O + NO_2$	4.2×10^{-12}	$-(180 \pm 180)$	1.2
42	$CH_3O_2 + HO_2 \rightarrow CH_3OOH + O_2$	1.7×10^{-13}	$-(1000 \pm 500)$	1.3
43	$CH_3OOH + OH \rightarrow CH_3O_2 + H_2O$	1.0×10^{-11}	0 ± 200	2.0
44	$CH_3O + O_2 \rightarrow H_2CO + HO_2$	3.9×10^{-14}	900 ± 300	1.5
45	$H_2 + O \rightarrow OH + H$	8.8×10^{-12}	4200	
47	$HNO_4 + M \rightarrow HO_2 + NO_2 + M$	1.0×10^{14}	10350	
48	$HNO_4 + OH \rightarrow H_2O + O_2 + NO_2$	1.3×10^{-12}	$-(380 + 270/-500)$	1.5
49	$HCO + O_2 \rightarrow CO + HO_2$	3.5×10^{-12}	$-(140 \pm 140)$	1.3
50	$C_2H_4 + OH \rightarrow C_2H_5 + H_2O$	1.1×10^{-11}	1100 ± 200	1.2
51	$C_2H_5O_2 + NO \rightarrow C_2H_5O + NO_2$	4.2×10^{-12}	$-(180 \pm 180)$	1.2
52	$C_2H_5O + O_2 \rightarrow CH_3CHO + HO_2$	1.2×10^{-13}	1350 ± 300	1.5
53	$C_2H_5O_2 + HO_2 \rightarrow C_2H_5OOH + O_2$	6.5×10^{-13}	$-(650 \pm 200)$	1.3
54	$CH_3CHO + OH \rightarrow CH_3CO_3 + H_2O$	6.0×10^{-12}	$-(250 \pm 200)$	1.4
55	$CH_3CO_3 + NO \rightarrow CH_3 + CO_2 + NO_2$	2.4×10^{-12}		
57	$PAN + M \rightarrow CH_3CO_3 + NO_2 + M$	6.3×10^{-2}	12785	1.5

TABLE 3 (continued)

Model Reaction Number	Three-body reaction	Rate * k_0^{300}	n	k_∞^{300}	m
1	$O + O_2 + M \rightarrow O_3 + M$	$(6.0 \pm 0.5) \times 10^{-34}$	2.3 ± 0.5		
3	$O + O + M \rightarrow O_2 + M$	4.27×10^{-28}			
10	$NO + O + M \rightarrow NO_2 + M$	$(9.0 \pm 2.0) \times 10^{-32}$	1.5 ± 0.3	$(3.0 \pm 1.0) \times 10^{-11}$	0 ± 1
11	$NO_2 + NO_3 + M \rightarrow N_2O_5 + M$	$(2.2 \pm 0.5) \times 10^{-30}$	4.3 ± 1.3	$(1.5 \pm 0.8) \times 10^{-12}$	0.5 ± 0.5
18	$H + O_2 + M \rightarrow HO_2 + M$	$(5.7 \pm 0.5) \times 10^{-32}$	1.6 ± 0.5	$(7.5 \pm 4.0) \times 10^{-11}$	0 ± 1
28	$OH + NO_2 + M \rightarrow HNO_3 + M$	$(2.6 \pm 0.3) \times 10^{-30}$	3.2 ± 0.7	$(2.4 \pm 1.2) \times 10^{-11}$	1.3 ± 1.3
35	$OH + OH + M \rightarrow H_2O_2 + M$	$(6.9 \pm 3.0) \times 10^{-31}$	$0.8 + 2.0/-0.8$	$(1.0 \pm 0.5) \times 10^{-11}$	1.0 ± 1.0
40	$CH_3 + O_2 + M \rightarrow CH_3O_2 + M$	$(4.5 \pm 1.5) \times 10^{-31}$	2.0 ± 1.0	$(1.8 \pm 0.2) \times 10^{-12}$	1.7 ± 1.7
46	$HO_2 + NO_2 + M \rightarrow HNO_4 + M$	$(1.8 \pm 0.3) \times 10^{-31}$	3.2 ± 0.4	$(4.7 \pm 1.0) \times 10^{-12}$	1.4 ± 1.4
56	$CH_3CO_3 + NO_2 + M \rightarrow PAN + M$	4×10^{-29} ***			

* $k(z) = \dfrac{k_0(T)[M]}{1 + k_0(T)[M]/k_\infty(T)} 0.6^{\{1 + [\log_{10}(k_0(T)[M]/k_\infty(T))]^2\}^{-1}}$, $k_0(T) = k_0^{300}(T/300)^{-n}$ and $k_\infty(T) = k_\infty^{300}(T/300)^{-m}$.

** Expression for this reaction is sum of three terms given in ref. 8.

*** Use overall $f(298) = 1.5$.

is a bimolecular rate coefficient between species i and j. The factors are determined from eq. 3 by running the time-dependent model to equilibrium, i.e. to periodic 24-hour behavior, and evaluating all the averages in (3). All the species concentrations illustrated in this study are diurnally averaged mixing ratios, $\bar{\chi}_i$.

Looking at eq. 3 it is clear that the diurnal factor $(DF)_{ij}$ depends on equilibrium concentrations of species, i.e. composition, and that as the calculated equilibrium composition changes in response to a different set of rate coefficients, the factors also change. Thus, in performing the Monte Carlo study, a time-dependent run must be carried out to obtain factors self-consistent with the diurnally averaged $\bar{\chi}_i$ from steady-state calculation. The initial 'perturbed' set of rates coefficients is always run with the time-dependent model and the diurnally averaged rates are supplied to the steady-state model for final calculation of the diurnally-averaged (or steady-state) concentrations.

The expression 'unperturbed' chemistry refers to the atmospheric composition as simulated by the model with the standard set of 72 reaction rate coefficients at mean values (Tables 2 and 3). Atmospheric measurements are used in specification of mixing ratios or flux values for NO and CO and for O_3 deposition velocity. The 'unperturbed' chemical profiles simulate 'Clean Continental' northern mid-latitude regions: $O_3 = 44$ ppbv, CO = 135 ppbv, $NO_x = 0.20$ ppbv, with $CH_4 = 1.70$ ppmv at the surface. Vertical profiles of O_3, CO, NO_x, and HNO_3 appear in Fig. 1.

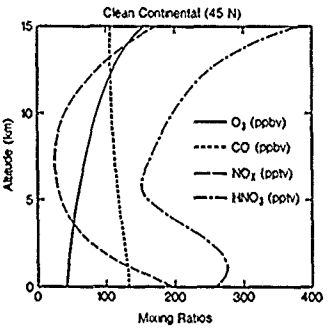


Fig. 1. Vertical profiles of O_3, CO, HNO_3, and NO_x typical of the relatively clean continental mid-latitude troposphere. Concentrations are given in mixing ratio by volume (mole fraction).

Monte Carlo calculations

Method

We vary the 72-reaction set of rate coefficients for each model run as follows A given set of perturbed rate coefficients is generated from a random number generator and each perturbed run is made with a different set of 72 reactions. The set of reaction rates is based on uncertainties in chemical rates as described in the JPL/NASA Panel 8 Evaluation [8] to derive corresponding uncertainties in the species concentrations.

The perturbed reactions are used in the time-dependent version of the model, which is integrated for two days to produce diurnally averaged rates (4) and mixing ratios. These mixing ratios are not 'converged' to equilibrium in that the 24-hour cycle of each species is not periodic. It would take many days of integration to achieve this because several constituents (e g. O_3, CO, and PAN) have photochemical lifetimes over a week. This is not computationally practical because each day of integration takes several minutes on the VAX 11/780 and attached processor.

We have compared diurnally averaged rates computed after two and ten day time dependent model runs. The maximum difference as a percentage of the imprecision occurs for the photolysis of N_2O_5 (rate J_{12} in Table 2) and is 1.6%. Only one other rate (rate 9 in Table 3) has a percentage difference as great as 1%. We do not expect the variances in species concentration computed over a set of model runs to be sensitive to small errors in averaged rates for each individual model, and this approximate averaging should be adequate.

Assignment of rate coefficient uncertainties

Most of the 72 reactions used in the photochemical model have an associated uncertainty given by the NASA panel evaluation [8]. As noted in this report, the assigned uncertainties are subjective judgments of the panel and are not based on rigorous statistical analysis because there have been an insufficient number of laboratory investigations.

We have assumed that the uncertain parameters entering into reaction rate calculations have simple probability density functions, Gaussian or lognormal, depending on whether the parameter is intrinsically positive or not. At the beginning of a model run, values are selected from these distributions for each parameter entering into the reaction rate. Each run gives different values for the concentrations corresponding to the randomly selected rates for that run. After a sufficiently large number of trials (runs), histograms showing the percentage deviation of each species concentration from its mean over all runs are obtained numerically for each species. We show results after 800 runs. The computed means and variances for each species are nearly constant as runs are added at this point. The maximum difference in the ratio of the standard deviation in the mean from the 700 run results is for C_2H_5OOH which changed by 1.6% after 800 runs. Similar calculations for stratospheric chemistry carried out by Stolarski and coworkers show that convergence to 1–2% is obtained after ~ 1000 runs [9,10].

Most of the reactions used in the photochemical model fall into one of three categories, photolysis, bimolecular, and termolecular, as noted in the discussion following eq. (1). The uncertainties in reaction rates are stated differently for each category in ref. 8 which requires some difference in the treatment for each one.

Uncertainties in photolysis rates used in the calculations are given as an overall fractional uncertainty in the rate, rather than as measurement uncertainties in the various fluxes, cross sections, and quantum yields which determine these rates [8]. The photodissociation reactions are given in Table 2. We have assumed a lognormal distribution for the photolysis rates with a standard deviation corresponding to the stated fractional uncertainty for each.

Most second order rates are obtained from the product of a rate coefficient and an exponential factor containing the activation energy. The general expression for binary rates is

$$k(T) = A \exp(-E/RT) \qquad (5)$$

where $k(T)$ is the overall reaction rate is the rate coefficient multiplying the exponential factor, E is the activation energy, R the gas constant, and T the temperature. JPL/NASA [8] give uncertainties in activation energy, ΔE, as well as an uncer-

tainty, $f(298)$, in the overall rate at 298 K. The overall uncertainty at other temperatures is calculated from the expression

$$f(T) = f(298) \exp|\Delta E/R(1/T - 1/298)| \quad (6)$$

For purposes of generating perturbed binary rates for a Monte Carlo series of model runs we assume that the overall uncertainty given in the JPL/NASA Panel 8 Tabulation is given by an uncertainty in the rate coefficient A in eq. 5 with A being lognormally distributed. The temperature dependent factor in eq. 5 is always evaluated at the standard value of the activation energy. The JPL/NASA [8] convention is followed in Table 3, which means that $f(298) = 1.2$ signifies a $1-$ sigma uncertainty of 20%. Note that the column labeled $f(298)$ in Table 3 is the overall uncertainty and is not necessarily identical to that which would be computed using the stated uncertainty in activation energy. Temperatures in the 1-dimensional model decrease with altitude and we have chosen to evaluate the binary rate uncertainties in eq. 6 at the surface temperature of 288 K. This is a conservative assumption in that it gives smaller rate uncertainties in model mixing ratios, but it is reasonably good for evaluating uncertainties in the boundary layer in which we are primarily interested.

The general expression used to evaluate termolecular rates is more complicated (Table 3). The general form of a termolecular reaction is A + B + M → AB + M where M is a quenching third body. Low pressure, k_0, and high pressure, k_∞, limiting rates are given in the form

$$k_0(T) = k_0^{300}(T/300)^{-n},$$
$$k_\infty(T) = k_\infty^{300}(T/300)^{-m} \quad (7)$$

and these are combined in a rate expression applicable to general conditions of atmospheric temperature and pressure by

$$k(z) = \frac{k_0(T)[\mathrm{M}]}{1 + k_0(T)[\mathrm{M}]/k_\infty(T)} \times 0.6^{\{1+[\log_{10}(k_0(T)[\mathrm{M}]/k_\infty(T))]^2\}^{-1}} \quad (8)$$

The factor [M] in eq. (8) is the concentration of third bodies involved in the termolecular reactions, specified by the model as the sum of O_2 + N_2. Uncertainties are given for the coefficients k_0^{300}, and k_∞^{300} and for the exponents n and m in the temperature dependent factors. Since k_0^{300} and k_∞^{300} must be positive they are assumed lognormally distributed, but the exponents n and m may be assumed normally distributed. The overall rate is thus a function of four random variables and the nature of its distribution does not follow immediately from the assumptions, as does that of the binary rate, though it is clearly always positive.

RESULTS AND DISCUSSION

Reaction rate uncertainties

Variability in some of the reaction rates important in the odd hydrogen balance of the troposphere is shown in Fig. 2. Fig. 2a shows the distribution in the rate of photolysis of ozone to produce $O(^1D)$ which initiates most tropospheric photochemistry:

$$O_3 + h\nu(\lambda = 295\text{–}310\ \mathrm{nm}) \rightarrow O(^1D) + O_2$$

The stated uncertainty in this rate is 40% and a lognormal distribution is assumed for photolysis reactions. Here apparent lognormality and an uncertainty close to the one given in JPL/NASA [8] are recovered from the numerical results. Fig. 2b shows the distribution of the $O(^1D) + H_2O$ reaction which is the primary source of tropospheric OH.

Fig. 2c shows the distribution of the termolecular rate for the reaction OH + OH + M → H_2O_2 + M forming hydrogen peroxide. Although the distribution appears somewhat skewed towards positive values we cannot characterize it as lognormal since, as noted previously, it results from a relatively complicated relationship among four random variables. Indeed, the termolecular distributions we have examined appear to be more symmetric about their means than would be the case if strictly lognormal.

Computed constituent uncertainties

Fig. 3 shows the calculated variability in odd hydrogen species, OH and HO_2, and in hydrogen

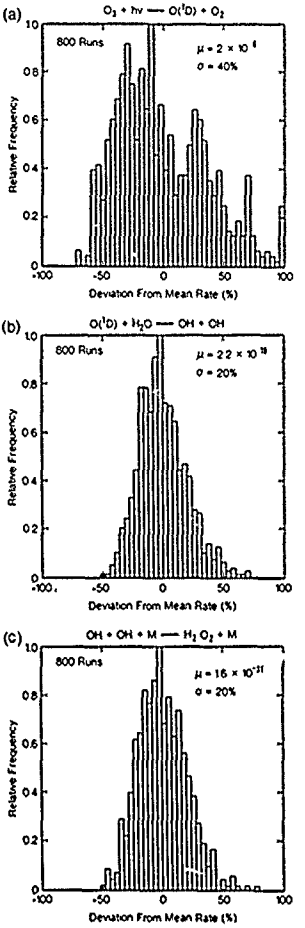


Fig. 2. Histograms showing distributions of several reactions calculated at 0 km (288 K) affecting the concentration of OH. Statistics are based on 800 model runs. (a) O_3 UV photolysis in s^{-1}; b) the major formation reaction for OH with mean rate 2.2×10^{-10} cm^3 s^{-1}; (c) ternary rate for OH and OH recombination with mean rate $= 1.6 \times 10^{-31}$ cm^6 s^{-1}.

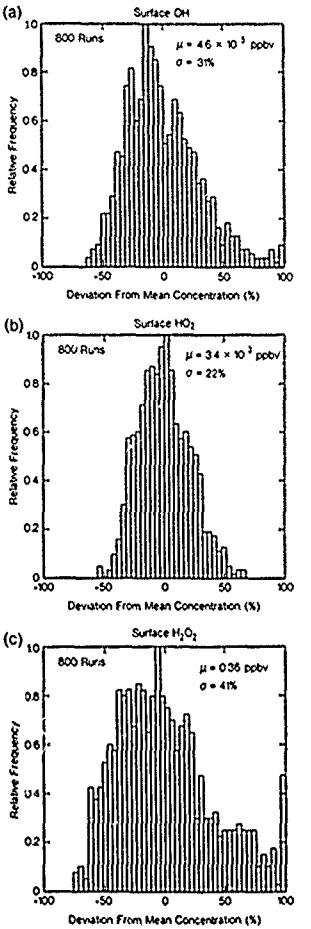


Fig. 3. Histograms of odd hydrogen species distributions at 0 km after 800 model runs.

peroxide, H_2O_2. All deviations greater than 100% above the mean are placed in the rightmost vertical bar on these histograms plots. We expect the variance of OH and HO_2 to be relatively large since they participate in more reactions than any other species. Hydrogen peroxide is readily absorbed in cloud droplets and may be an important component in the liquid phase production of sulfate and consequent decrease in droplet pH [11,12]. We note that wet removal of H_2O_2 is included in our model continuity equations for H_2O_2 as a first order rate coefficient but this rate is not varied. We have previously explored the possibility of increases in future peroxide levels resulting from projected changes in methane and CO emissions and from possible climate changes [3,13]. We estimate global change for H_2O_2 responding to continuing 0.5–1%/yr CO and CH_4 increases to be about 20% over the next fifty years [13]. The present study would imply that this change is smaller than the model's precision for computing H_2O_2 under a given set of conditions. Fortunately, we can make atmospheric measurements of key species (O_3, CO) to better precision than we compute in the Monte Carlo study and this suggests that we can improve on the calculated uncertainties for all species by constraining the model with observations [5].

Fig. 4 shows the calculated variability in members of the odd nitrogen family, nitric oxide (NO), nitrogen dioxide (NO_2), and nitric acid (HNO_3). The uncertainty in HNO_3 is one of the smallest occurring in our calculations. As for H_2O_2, uncertainties in HNO_3 due to rainout are not included in this study. These are likely to substantially increase the HNO_3 variance [14].

Fig. 5 shows the calculated variability of ozone (O_3) and carbon monoxide (CO). These species are less reactive than free radicals, peroxides or acids. We expect smaller variability for O_3 due to rate uncertainties because external sources of O_3 as well as chemical reactions, are important to its atmospheric distribution. A fixed flux into the troposphere is assumed for ozone at the tropopause. The uncertainty at the surface is 17%. The uncertainty for CO is higher, ~31%, even though an upflux of CO is very important to boundary layer CO. The reason is a fractional

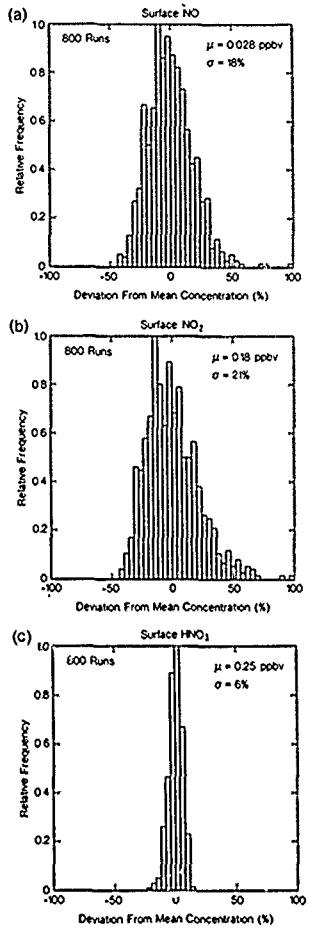


Fig. 4. Histograms of odd nitrogen species distributions at 0 km after 800 model runs.

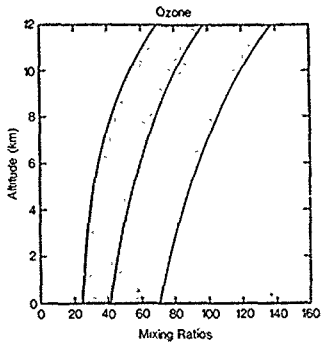


Fig. 5. Vertical profiles (0–12 km) based on 800 simulations for (a) O_3 and (b) CO. The mean is indicated with the central vertical profile, shading indicates 1 – sigma deviation from the mean.

uncertainty in the rate of the reaction, OH + CO, (30%), which is the major CO sink. When model runs are performed without varying this reaction rate, the variation in CO is less than 20%.

SUMMARY

Estimated imprecisons in chemical reactions rates important in tropospheric photochemistry have been used to estimate the resulting uncertainty in model calculated trace species distributions. A Monte Carlo approach is used with tabulated kinetics imprecisions specified for 72 reactions. The tabulated imprecisions are reproduced closely by the model after several hundred model runs and the propagated uncertainty in 24 trace constituents is calculated. Uncertainties for ozone and carbon monoxide are 17% and 31%, respectively. For CO this is 2–3 times greater than the imprecision which typically affects CO measurements in the atmosphere.

Odd nitrogen uncertainties are ~20% for NO and NO_2 and only 6% for HNO_3 because imprecision in precipitation scavenging, an important loss for nitric acid, has not been included in the study. Hydroxyl radical (OH) has a computed uncertainty of 31%, which somewhat limits the model assessment capability for precise evaluation of oxidant changes.

In a related study [5] we report on correlation analysis between rates and species to identify those reactions which contribute most to the variance of selected species. This also helps in developing in-situ measurement strategies to reduce the overall computational variance found in the present study and in identifying the photochemical processes at which further laboratory investigation might be most effectively directed.

ACKNOWLEDGEMENTS

Thanks are due to an anonymous reviewer for comments and to M.A. Huntley for programming assistance. This work was supported by the U.S. EPA (Interagency Agreement DW 80933962-0) and the NASA Tropospheric Chemistry Program.

REFERENCES

1 A.M. Thompson and R.J. Cicerone, Possible perturbations to atmospheric CO, CH_4, and OH, *Journal of Geophysical Research*, 91 (1986) 10853–10864.

2 A.M. Thompson, M.A. Huntley and R.W. Stewart, Perturbations to trophosphenc oxidants: 1. Model calculations of ozone and OH in chemically coherent regions, *Journal of Geophysical Research*, 95 (1990) 9829–9844.
3 A.M. Thompson, M A. Owens and R.W. Stewart, Sensitivity of troposphenc hydrogen peroxide to global chemical and climate change, *Geophysical Research Letters*, 16 (1989) 53–56
4 A M Thompson and R.W. Stewart, How chemical kinetics uncertainties affect concentrations computed in an atmospheric chemical model, *Journal of Geophysical Research*, (1990) submitted for publication.
5 K.E. Pickering, A.M. Thompson, R R. Dickerson, W.T. Luke, D.P. McNamara, P.R. Zimmerman and J.P. Greenberg, Model calculations of troposphenc ozone production potential following observed convective events, *Journal of Geophysical Research*, 95 (1990) 14049–14062.
6 A M. Thompson and R.J. Cicerone, Clouds and wet removal as causes of variability in the trace-gas composition of the marine troposphere, *Journal of Geophysical Research*, 87 (1982) 8811–8826.
7 R.P. Turco and R.C. Whitten, A note on the diurnal averaging of aeronomical models, *Journal of Atmospheric and Terrestrial Physics*, 40 (1978) 13–20.
8 JPL/NASA, *Chemical kinetics and photochemical data for use in stratospheric modeling, NASA Kinetics Panel Evaluation, 8 Pub 87-41*, JPL, Pasadena, CA, 1987.
9 R S. Stolarski, D.M. Butler and R.D. Rundel, Uncertainty propagation in a stratospheric model 2. Monte Carlo analysis of imprecisions due to reaction rates, *Journal of Geophysical Research*, 83 (1978) 3074–3078.
10 R.S Stolarski and A.R. Douglass, Sensitivity of an atmospheric photochemistry model to chlorine perturbations including consideration of uncertainty propagation, *Journal of Geophysical Research*, 91 (1986) 7853–7864.
11 S A. Penkett, B M.R. Jones, K.A. Brice and A.E.J. Eggleton, The importance of atmospheric ozone and hydrogen peroxide in oxidizing sulphur dioxide in cloud and rainwater, *Atmospheric Environment*, 13 (1979) 123–137.
12 W L. Chameides and D D. Davis, The free-radical chemistry of cloud droplets and its impact upon the composition of rain, *Journal of Geophysical Research*, 87 (1982) 4862–4878.
13 A.M. Thompson, M.A. Huntley and R.W. Stewart, Perturbations to troposphenc oxidants. 2. Model calculations of hydrogen peroxide in chemically coherent regions, *Atmospheric Environment*, (1990) in press.
14 R.W. Stewart, A.M. Thompson, M A. Owens and J.A. Herwehe, Comparison of parameterized nitric acid rainout rates using a coupled stochastic–photochemical tropospheric model, *Journal of Geophysical Research*, 94 (1989) 5219–5226.

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 81–83
Elsevier Science Publishers B.V., Amsterdam

Analysis of chemical structure–biological activity relationships using clustering methods

Peter C. Jurs * and Richard G. Lawson
152 Davey Laboratory – Chemistry, The Pennsylvania State University, University Park, PA 16802 (U.S.A.)

(Received 8 November 1989; accepted 8 December 1989)

Abstract

Jurs, P.C. and Lawson, R.G., 1991. Analysis of chemical structure-biological activity relationships using clustering methods. *Chemometrics and Intelligent Laboratory Systems*, 10, 81–83.

The importance of calculating clustering tendency of a data set as part of a complete methodology is described. A new method for evaluating the clustering tendency is illustrated with artificially clustered, random, and actual chemical data sets. This new index is shown to be more useful than the original one.

Cluster analysis is a useful and increasingly popular method for exploring data represented in high-dimensional spaces. Questions that can be approached using cluster analysis arise in pharmaceutical and agricultural chemistry in the context of structure-activity relationships. For example, a common exploratory approach to SAR is to retrieve those compounds which have a particular structural fragment from a large data base of compounds. Then it is of interest to seek subsets of compounds with structural similarities, that is, clusters. Other examples come from toxicology, where it is of interest to examine sets of compounds for structural similarities so that these similarities can be related to toxicity. A third example involves the examination of a number of possible conformations for complex structures as provided by a molecular mechanics routine to see if they fall in natural subgroupings.

The exploration of multivariate data via clustering involves many steps: data collection, initial screening of the variables, exploration of clustering tendency, application of clustering strategies, and validation and interpretation of the results. Often the entire process is iterative. Once a data set has been selected for analysis, the examination of clustering tendency prior to the development of clusters is important because it allows the experimenter to be sure that the clustering exercise has a chance of finding real clusters. Most algorithms designed to find clusters will find some regardless of the structure of the data. This work focusses on the evaluation of clustering tendency via Hopkins statistic and a recently proposed variation of it.

0169-7439/91/$03.50

Hopkins statistic has been shown previously to be a very good method for assessing clustering tendency [1].

Hopkins statistic [2,3] is intended to assess whether or not a given data set differs from a set of uniform random numbers. The statistic is calculated with the following equation.

$$H = \frac{\sum U_i}{\sum U_i + \sum W_i} \qquad \begin{matrix} U_i\text{: random to real} \\ W_i\text{: real to real} \end{matrix}$$

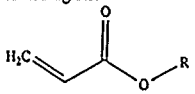
Each U_i value is the distance from a randomly selected position within the sampling window to the nearest data point, and each W_i value is the distance from a randomly selected data point to its nearest neighbor data point. The sums are over the number of sampling points, which is usually selected to be 5% to 10% of the number of points in the data set. The U_i positions (the sampling points) are chosen from a uniform distribution within the sampling window. H has values near 1/2 for unclustered data, that is, data with a uniform distribution. H has values greater than 1/2 for clustered data, and 1.0 is the upper limit for extremely clustered data. For reasonable assumptions, H has a beta distribution, so the probability for rejection of the null hypothesis (no clustering) can be quantitatively stated. For example, for 15 sampling points and a value $H = 0.65$, the probability of rejection of the null hypothesis is 0.90.

The ordinary Hopkins statistic has several shortcomings. One is its sensitivity to the size of the sampling window and hence to outliers. Another is that the criterion of comparison to a uniform distribution is weak since almost any measured or calculated data will be more clustered than the uniform distribution.

We have investigated [4] a modified form of the Hopkins statistic, H', designed to overcome these shortcomings. Instead of choosing the sampling points from a uniform distribution, we choose them from the actual univariate distributions of the data under investigation. This allows us to investigate whether the clustering tendency observed for the data set is due to the multivariate nature of the observations or due only to the univariate distributions of the variables.

Tests of this modified Hopkins statistic with two-dimensional and ten-dimensional artificial data sets designed to be extremely clustered, and with an eight-dimensional chemical data set, show the modified statistic to be more conservative in its estimation of clustering than the original Hopkins statistic. The modified statistic also is not sensitive to outliers.

The chemical example used for testing the modified Hopkins statistic consists of 143 acrylate compounds with the general structure shown. This set of data was analyzed in the context of a structure–toxicity relationship investigation [5]. Each of the 143 acrylates was represented by a set of eight calculated structural descriptors which were chosen to best represent the structures. A principal components plot of the data shows no apparent clustering. However, the data do show substantial clustering tendency with the original Hopkins Statistic: $H = 0.82$. When the original Hopkins statistic was calculated for scrambled data, $H = 0.77$. This shows that there is substantial clustering tendency due to the univariate distribution of the eight structural descriptors. The value for the modified Hopkins statistic was $H' = 0.65$. This shows that the multivariate data contain more information than merely their univariate distributions. This data set was analyzed for clustering using the well-known K-means and Isodata clustering method, and five stable clusters were found. These five clusters made good sense when the structures of the compounds in each class were considered by knowledgeable chemists and toxicologists.



The modified Hopkins statistic can also be used for feature selection, that is, for selection of these variables which support clustering in a data set. Preliminary studies have shown that the use of partial sums of U_i and W_i can be used effectivly for deletion of the least useful variables thereby focussing on those variables that best support clustering.

REFERENCES

1 G. Zeng and R C. Dubes, A comparison of tests for randomness, *Pattern Recognition*, 18 (1985) 191–198
2 B. Hopkins, A new method for determining the type of distribution of plant individuals, *Annals of Botany*, 18 (1954) 213–227.
3 A. Jain and R Dubes, *Algorithms For Clustering Data*, Prentice Hall, Englewood Cliffs, NJ, 1988, pp. 136–137.
4 R G. Lawson and P.C. Jurs, New index for clustering tendency and its application to chemical problems, *Journal of Chemical Information and Computer Science*, 30 (1990) 36–41.
5 R G. Lawson and P C. Jurs, Cluster analysis of acrylates to guide sampling for toxicity testing, *Journal of Chemical Information and Computer Science*, 30 (1990) 137–144

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 85–86
Elsevier Science Publishers B V., Amsterdam

Comments on "Analysis of chemical structure–biological activity relationships using clustering methods" by Peter C. Jurs and Richard G. Lawson

Leon Jay Gleser

Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260 (U.S.A.)

Cluster analysis shares with other scaling methods (such as principal components, factor analysis) the ideal that there is an underlying structure which influences observed variables, but which is *not* entirely revealed by these variables. This underlying structure, as compared to the observed variables themselves, may also be more highly predictive of other phenomena. For example, structural similarities of sets of compounds may reflect underlying chemical structures that are related to the biological toxicity of these compounds. Thus, clustering is seen as a valid alternative to regression analysis as a way of predicting these other phenomena (e.g., toxicity). Once clusters have been identified, analysis of variance can be used to demonstrate the predictive ability of the clusters. A similar approach has been used in educational research to find predictors of improvement in mathematics achievement of junior high school students [1].

Although the modified Hopkins statistic discussed by Jurs and Lawson [2] can be useful in determining whether multivariate data reflect underlying clusters, it has some disadvantages. One disadvantage is that this statistic depends heavily on the scales of the variables measured. Simply changing the scale of measurement on any single variable measured will change the value of H. More generally, the value of H is affected by the standard deviations of the variables being considered. (This disadvantage of H is shared by principal component analysis, where scale changes can influence principal values and principal vectors in complex ways.) Lawson and Jurs [3] are aware of this problem, and standardize their variables before clustering. However, if sample standard deviations are used for standardization, rather than the actual population standard deviations, the distribution of the Hopkins statistics is no longer necessarily a beta distribution (even if reasonably large samples are used to estimate standard deviations).

A second possible disadvantage of the modified Hopkins statistic is that it concentrates on clustering as a multivariate phenomenon (i.e., due to dependence of the variables). This excludes from consideration clusters that can form in the multivariate space because the individual variables themselves show clustering (multimodality) in their marginal distributions, while yet being independent. Since the ideal in scaling is a latent structure which relates the observed values, and this latent structure is of primary interest, this may not be a serious defect. However, it does raise the conceptual question of what constitutes a cluster.

It is to Jurs and Lawson's credit that they have eliminated the major disadvantage of the original Hopkins statistic–namely, the insistence on assuming that unclustered variables were independent and *uniformly* distributed. Few variables en-

0169-7439/91/$03.50

countered in nature have uniform distributions. Although it is possible to transform marginal distributions so that they are uniform (by the probability integral transformation performed variable by variable), such transformations destroy comparability of distances to nearest neighbors It is these distances between data points that most intuitively convey the notion of 'clustering'. (If all that is meant by clustering were lack of independence, then tests of independence based on either the Kolmogorov–Smirnov distance between multivariate distributions or Pearson chi-squared tests of independence based on grouped data in contingency tables could be used. The distances utilized by these tests have little resemblance to Euclidean distances between data points.)

Besides use of the Hopkins statistic, there are other ways that the 'reality' of observed clusters can be demonstrated. Using more than one distinct method for searching for clusters (e.g., *K*-means and Isodata, as used by Jurs and Lawson [2] in their chemical data) is one good method. If different search methods arrive at similar (numbers of) clusters, one can be less worried that the clusters are artifacts of a particular search method. Additionally, one can hold back a randomly selected subset of variables in an initial clustering search, and then see if adding these variables changes the conclusions. (This approach assumes that no small subset of variables by itself defines the true underlying clusters.) Instead of withholding variables, one can randomly divide data points (cross-validation) into two or more groups and see if similar clusters arise in such data sets. This approach is associated with a formal statistical theory that is currently discussed under the terminology 'bootstrap analysis' [4,5]. There is also a resemblance between the subsampling of data used in the Hopkins test and the resampling methods used in bootstrap analysis. Finally, as demonstrated by Jurs and Lawson, one can see if the clusters found make sense in the light of existing chemical (and biological, in the case of toxicity) knowledge. If the clusters successfully predict other phenomena (e.g. toxicity), this is further evidence that such clusters are not artifacts of the data.

As Jurs and Lawson so clearly show, cluster analysis has the potential to yield important insight and direction in the study of classes of chemical compounds.

REFERENCES

1 J.E. Lockley, A comparative study of cluster analysis and MANCOVA in the analysis of mathematics achievement data, in L.J. Gleser, M.D. Perlman, S.J. Press and A.R. Sampson (Editors), *Contributions to Probability and Statistics. Essays in Honor of Ingram Olkin*, Springer-Verlag, New York, 1989, pp. 241–270.

2 P.C. Jurs and R.G. Lawson, Analysis of chemical structure-biological activity relationships using clustering methods, *Chemometrics and Intelligent Laboratory Systems*, 10 (1991) 81–83.

3 R.G. Lawson and P.C. Jurs, New index of clustering tendency and its application to chemical problems, *Journal of Chemical Information and Computer Science*, 30 (1990) 36–41.

4 B. Efron, The jackknife, the bootstrap, and other resampling plans, *Society for Industrial and Applied Mathematics, CBMS – National Science Foundation Monographs*, 1982, No. 38.

5 B. Efron and G. Gong, A leisurely look at the bootstrap, the jackknife, and cross-validation, *American Statistician*, 37 (1983) 36–48.

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 87–105
Elsevier Science Publishers B.V., Amsterdam

Rapid parameter estimation with incomplete chemical calibration models

Steven D. Brown
Department of Chemistry, University of Delaware, Newark, DE 19716 (U.S.A.)

(Received 8 November 1989; accepted 8 February 1990)

Abstract

Brown, S.D., 1991. Rapid parameter with incomplete chemical calibration models. *Chemometrics and Intelligent Laboratory Systems*, 10: 87–105.

The use of calibration models to predict analyte concentrations in samples showing responses from poorly calibrated components, or samples showing drift in the instrumental response function, is seldom successful. These incomplete calibration models cannot account for variations not encountered in the calibration step. Simple modifications are possible which remedy this difficulty for classical least squares (CLS) regression. By using sequential regression for the prediction step, extensions are possible which lessen errors due to overfitting, and permit prediction of well-modelled components in the presence of unmodelled components. Implementation of the sequential regression is conveniently done through use of the Kalman filter. Use of filter models for dynamics and measurement also permits correction of drift of various types. The use of CLS calibration with Kalman filter prediction is presented and tested with simulated spectroscopic data. Comparisons are made to other calibration and prediction methods.

INTRODUCTION

Care in the calibration step is very important for a successful multicomponent analysis. During the initial phase of a calibration, when standard mixtures of analytes are measured, effort must be made to calibrate over the widest possible range of instrumental conditions, analyte concentrations, and potential interferences. From these calibration data, a calibration model is generated which explains as much as possible of the variations seen during the calibration step. The model is used to predict analyte concentrations from further measurements made on predictors. Care in collecting calibration data and generating a calibration model is repaid in the range over which the calibration remains valid during prediction.

Even with great care in calibration, there is still the likelihood of instrumental drift with time, and the chance that small changes in the nature of the sample may appear in the form of unexpected (and uncalibrated) components. Drift and unmodelled responses present two significant challenges to calibration schemes. Both can be regarded as unmodelled components in the calibration, but the effects of these unmodelled components are

0169-7439/91/$03.50

seen during prediction. The development of calibration methods that are more robust to the effects of instrumental drift and unmodelled components would greatly extend the useful range of many calibration schemes.

Considerable research has been directed at methods for improving the modelling of the calibration step. Methods based on regression of data onto factor models or on the relation of latent variables have been developed to improve the calibration process by lessening the effects of noise in the calibration model [1,2] These methods have shown success in generating very reliable models for the calibration step, but they are less successful at predicting concentrations for multicomponent samples, especially those that are observed under conditions far removed from the conditions of calibration. Other methods, for example those based on rank annihilation, might be more suited to treatment of chemical measurement of samples containing well-modelled components coexisting with unknown contaminants [3]. Because these methods presume identical spectral or temporal behavior for any well-modelled components, so that second-order or higher data can be rank annihilated, they are more suited to arrays to bilinear spectra than to time-varying calibration systems, which may contain time jitter from run to run [4]. That jitter makes registration of the bilinear arrays uncertain, and it causes difficulties in the rank reduction process. Drift in the instrumental response is also problematic to rank reduction methods because of the lack of reproducibility of the time varying responses of standards and samples.

The prediction step can be considered a time-series process, and it seems reasonable to apply methods intended for time series analysis in attempting to create calibration models which are more robust to errors in the prediction step. Since the time-series involved are multivariate, given the multicomponent chemical models and the multicomponent responses observed, a multivariate approach is appropriate.

One multivariate, time-based approach that might be examined is the Kalman filter. Although many of its time-series properties have not been used to full advantage in applications in analytical chemistry, this algorithm has been extensively used for analysis of multicomponent data [5]. Previous work from this laboratory [6,7] has demonstrated that modified Kalman filter methods may be advantageous for multicomponent analysis in the presence of unanticipated and unmodelled responses in a multicomponent signal. Some work on drift compensation of univariate systems [8] has also appeared.

This paper demonstrates that one form of calibration, classical least squares (CLS) calibration, is directly compatible with ordinary Kalman filtering, either in vector or in scalar (sequential regression) form. Additions to the CLS calibration model which account for random drift and for unmodelled responses are presented and discussed. All methods are tested with simulated spectroscopic data.

THEORY

Classical least squares calibration

For analysis of a set of compounds contained in a mixture, any of the standard methods of multicomponent calibration can be used. CLS calibration, sometimes called *K*-matrix calibration, is convenient for use here because of its assumption of the least-squares causal model relating the measured response $\mathbf{A}_s$ of standards to their known concentrations $\mathbf{C}_s$

$$\mathbf{A}_s = \mathbf{C}_s\mathbf{K} + \mathbf{e} \tag{1}$$

where the $n \times p$ matrix K relates the m spectra collected over p sensor channels to the $m \times n$ concentrations in $\mathbf{C}_s$. From the calibration step, where both $\mathbf{A}_s$ and $\mathbf{C}_s$ are known, matrix K is easily obtained from

$$\mathbf{K} = \left(\mathbf{C}_s^T\mathbf{C}_s\right)^{-1}\mathbf{C}_s^T\mathbf{A}_s \tag{2}$$

The columns of the 'K-matrix', K, are estimates of the pure-component spectra of species involved in the calibration. Once the calibration is completed, the matrix K can also be used to estimate the concentrations of analytes $\mathbf{C}_u$ in unknown samples, since

$$\mathbf{C}_u = \mathbf{A}_u\mathbf{K}^T\left(\mathbf{K}\mathbf{K}^T\right)^{-1} \tag{3}$$

The accuracy of the estimates C_u obtained from the prediction step depends on the adequacy of the calibration model K, and the presence of additional, unexpected components altering the multicomponent response A_u. These can be additive, as might be the case when additional constituents are present, and these constituents responses contribute to the multicomponent signal. They also might be multiplicative, as would be the case when linear or proportional drift caused a change in the instrumental response expected for a given concentration of analytes.

While calibration based on classical least-squares is well-understood, since it is one form of ordinary multiple linear regression, it may not always be the best method for calibration. Some of the undesirable features of a calibration based on CLS regression include possible overfitting of data to the calibration models, where parts of the unknown response are fitted to noise in the calibration models [1].

Sequential regression for prediction

One way to alter CLS calibration is to perform the regression of unknown response onto models sequentially, rather than in a single step. Sequential regression of data onto the classical causal model of equation 1 is well-established [9–11]. The algorithm is given by three equations, one for the update of the regression parameters (here, the unknown concentrations), one for the update of the covariance of the estimates, and one for the correction of the current estimates C_u and P to account for the information contained in new data. If the regression parameters are contained in the $n \times 1$ vector C_u with covariance P, the recursion relations, expressed for the ith channel of a p-channel spectrum, are

$$C_u(k) = C_u(k-1) + L(k)\left[A_u(k) - C_u^T(k)K(k)\right] \quad (4)$$

$$L(k) = \frac{P(k-1)K(k)}{1/a(k) + K^T(k)P(k-1)K(k)} \quad (5)$$

$$P(k) = P(k-1) - \frac{P(k-1)K^T(k)P(k-1)}{1/a(k) + K^T(k)P(k-1)K(k)} \quad (6)$$

In these equations, $a(k)$ is the weight given to observation $A_u(k)$, and $L(k)$ is the correction factor used to update $C_u(k)$ and $P(k)$. Careful choice of appropriate values for $a(k)$ will reduce the problem of overfitting mentioned above. The calibration matrix $K(k)$ can be calculated directly from eq. (2) above, or it also can be obtained by application of sequential regression of the spectra obtained during calibration runs onto the standard concentrations, using a regression approach analogous to that in eqs. (4)–(6).

Sequential regression requires initial guesses $C_u(0)$ and the covariance matrix P(0), a measure of the uncertainty of the initial guess $C_u(0)$. The covariance matrix has units of concentration squared, and its diagonal elements are the variance associated with each element of the concentration vector C_u.

With correct regression models, the sequential estimates $C_u(k)$ quickly become independent of the initial guess $C_u(0)$, provided that a 'reasonable' value is selected for P(0). Values of about 1–100 times $C_u(0)$ work well for the diagonal values of P(0); the off-diagonal elements may be set to zero. Larger values of P(0) typically aid in getting rapid convergence. When P(0) is selected too small, biased results for C_u will result from the sequential regression [9].

While sequential regression may not always be as computationally efficient as ordinary regression, it sometimes can be more computationally efficient, depending on the number of parameters to be fitted, the dimension of the measurement, and the weighting factors. Cases where sequential regression has a computational advantage over ordinary regression arise where the few parameters are to be fitted to a high-dimension measurement, and where weighting data are available for use in the fitting; this situation is common in the analysis of multicomponent data in analytical chemistry. Sequential regression also offers other advantages. Two of these advantages are the elimination of the need for matrix inversion, and the possibility of using prior information on the values and/or distribution of C_u and P. A third advantage is the ease with which the regression problem can be recast into forms suited to analysis by regression methods based on loss functions other than simple least squares.

Within a Bayesian framework, for example, C_u can be considered a random parameter vector with some prior distribution, and the set of observations should be correlated with C_u. The posterior probability density function for C_u is desired at some point k, that is $p(C_u|A_u)$. The estimate $\hat{C}_u$ can be obtained from the distribution; a common approach is to use the value for which the distribution attains a maximum — the maximum a posterior (MAP) estimate. For a symmetric distribution, the MAP estimate coincides with the mean of the distribution, and it is also the value that minimizes the parameter error variance $E[(C_u - \hat{C}_u)(C_u - \hat{C}_u)^T]$. The problem is to determine the evolution of the density function (or its mean) with added data. In general, solution of this problem is not possible, but if measurement noise e is taken as Gaussian, an exact solution is possible. Under these constraints, it is found that optimal weighting of observations is given by the relation

$$1/a(k) = E\left[(e(k) - \bar{e}(k))^T(e(k) - \bar{e}(k))\right] \tag{7}$$

Given this definition of the weighting, the sequential regression can also be cast into a form amenable to use with the scalar form of the Kalman filter, with system dynamics model

$$X(k+1) = F(k)X(k) + w(k) \tag{8}$$

and a measurement model

$$z(k) = H^T(k)X(k) + v(k) \tag{9}$$

where, for simple K-matrix prediction, the filter state X is the vector C_u, the filter measurement matrix H is the calibration matrix K, the filter measurement z is the spectral datum A_u, and the filter noise parameter v describes the calibration measurement error e. If the filter dynamics matrix is set to identity for this time-dependent problem, and the filter systems noise w is taken as zero, eqs. (4)–(6) may be seen to be identical with the update equations from the scalar Kalman algorithm (eqs. (A3)–(A5) in the Appendix), where the vector quantity $L(k)$ is the Kalman gain. The filter time projection eqs. (A1) for the state, and (A2) for the state covariance, are identities in this example, because the filter model for systems dynamics (eq. (8)) is an identity in this analysis, and because $Q(k) = E[w(k)w^T(k)] = 0$ and $R(k) = E[v(k)v^T(k)] = 1/a$ when the system noise is zero and the measurement noise is defined by eq. (7) above. Use of Kalman filter methods therefore offers a general, flexible framework for classical least squares calibration and prediction, since classical least squares can be taken as a subset of the more general filtering approach, with identity systems dynamics and uniform weighting.

Modelling drift in CLS prediction

The systems dynamics matrix $F(k)$ of the Kalman filter need not be identity, however A model for drift can be used to describe filter state dynamics, thus extending the CLS calibration model to track drifting multicomponent systems. Random and linear drift models are believed to describe many chemical systems [8]. A drifting parameter X is generally described by a linear equation

$$X(t) = X(t-1) + d(t) \tag{10}$$

where $d(t)$ is the drift. Random drift occurs when $d(t)$ is a random parameter, while linear drift results when $d(t)$ varies systematically with time. If the state is defined as $X(t) = [C_u(t), d(t)]$, this systems model leads to a simple systems dynamics model, namely

$$\begin{bmatrix} C_u(t) \\ d(t) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} C_u(t-1) \\ d(t-1) \end{bmatrix} + w(t) \tag{11}$$

with the measurement matrix as the time-independent quantity

$$H = \begin{bmatrix} K \\ 0 \end{bmatrix} \tag{12}$$

This dynamic model is observable if matrix $[HF^TH(F^T)^2H\ldots(F^T)^{n-1}]$ is of rank n for the n-dimensional state vector X [11]. In this instance, this matrix is of full rank if K is of full rank, and if duplicate measurements are made on each sample, so that drift variables in d can be characterized.

Other forms of instrumental drift are just as easily modelled. With proportional drift, the response at some time t might be related to the

response at an earlier time $t-1$, by $z(t) = 1/\gamma(t) \cdot z(t-1)$ where $\gamma(t)$ is a time-dependent, random parameter. This leads to a systems dynamics equation that is now a function of time, t

$$C_u(t+1) = \gamma(t)C_u(t) + w(t) \tag{13}$$

Now, the set of parameters C_u and the random parameter γ must both be estimated to obtain C_u in the presence of random drift. Define the system state as $X(t) = [C_u(t), \gamma(t-1)]$. Then the filter models are

$$\begin{bmatrix} C_u(t) \\ \gamma(t-1) \end{bmatrix} = \begin{bmatrix} \gamma(t-1) & 0 \\ 0 & \end{bmatrix} \begin{bmatrix} C_u \\ \gamma(t-1) \end{bmatrix} + \begin{bmatrix} \gamma \\ 0 \end{bmatrix} w(t) \tag{14}$$

to account for drift over time between spectral measurements, and

$$z(t) = KC_u + v \tag{15}$$

to describe the calibration during the prediction of this particular spectral measurement. As discussed above, the prediction step may be solved directly, with a matrix inversion, to obtain state estimates C_u, or it may be broken down to a series of scalar relations defining the sequential regression of z onto K

$$z(k) = K(k)C_u + v(k) \tag{16}$$

Such decomposition of the measurement vector z into a sequence of scalar measurements $z(k)$ is common in the engineering literature [10,12]. For the filter models described by eqs. (13) and (16), the index t describes time between spectral measurements, while index k describes scalar components of the measurement. The state X will be both time- and wavelength-dependent, but since only state estimates are the end of the update process are of interest, and not the evolution of states during the sequence of scalar updates, states are given in terms of time for this model. State projection occurs between measurement of full spectra, while state update occurs for each spectral channel.

In this treatment, it is assumed that spectral measurement is fast, and that drift during collection of a spectrum is negligible. If so, the systems dynamics matrix can be expressed as the time- and state-dependent quantity

$$f(X,t) = \begin{bmatrix} \gamma(t)C_u(t) & 0 \\ 0 & 1 \end{bmatrix} \tag{17}$$

and the measurement matrix is defined as the time-independent quantity defined in eq. (12) above. This filter model is nonlinear, since the system dynamics depends upon the present value of the filter state. The states of this model may be estimated by use of the extended Kalman filter [5,9–11]. In essence, the extended filter provides a way to linearize the systems dynamics matrix f about the current state estimates, so that

$$F(\hat{X}) = \frac{\partial f(X,t)}{\partial X}\Big|_{X=\hat{X}} = \begin{bmatrix} \gamma(t-1) & \hat{C}_u(t) \\ 0 & 1 \end{bmatrix} \tag{18}$$

where $\hat{C}_u(t)$ and $\hat{\gamma}(t-1)$ are the current state estimates in the extended Kalman filter. It is possible to perform sequential regression over the spectral data to obtain estimates of states C_u for a given spectrum and time, then proceed though the extended Kalman filter to provide predictions of drift between spectral measurements, as described above. If a good estimate of the system noise $Q(t)$ is available, accurate estimation of the true concentrations and the apparent drift in concentration should be possible using these simple modifications to CLS prediction.

Examination of the equations for the Kalman filter (eqs. (A1)–(A7)) demonstrates that the equations for updating state estimates are decoupled from those used to project states ahead in time. There is no reason why other regression-based prediction methods which employ externally-supplied initial guesses cannot be used in conjunction with the projection equations used in the Kalman filter. In this way, other calibration methods might be extended to account for drift between samples, or for other time-dependent effects.

Compensating for unmodelled responses in prediction

If the measurement model is in error, ordinary regression of data onto the spectral models will

produce inaccurate estimates of concentration. With any recursive algorithm, there is also the possibility of skipping the processing of data that is corrupted by the existence of poorly modelled signals. This feature can be used to avoid regions of data for which models are in error, provided some means of evaluating the model quality can be found.

Adaptive filtering for estimation of noise processes

Several indicators exist for model quality. The most reliable are based on the filter innovations, a measure of how well the filter model can predict new data. For scalar Kalman filtering, the filter innovations are defined as

$$\nu(k) = z(k) - \mathbf{H}^T(k)\mathbf{X} \qquad (19)$$

where $\mathbf{X}(k|k-1)$ is the projected state at point k, based on information up through point $k-1$. One possible way to evaluate innovation quality is to compare the observed innovations sequence $\nu(k)$ with that expected from the filter theory. With a correct filter model, the filter innovations are given by $\mathbf{H}(k)\mathbf{P}(k)\mathbf{H}^T(k)$, assuming no correlation of state and measurement noise. This quantity accounts for the presence of error in $z(k)$ which is not part of the filter model $\mathbf{H}(k)$. With a correct model, the error in $z(k)$ is random, and its variance is $\mathbf{R} = \mathbf{E}[v(k)v^T(k)]$. According to theory, for a correct filter model, with Gaussian noise on the measured data, the filter innovations will also be Gaussian. In addition, the innovations will have a mean value of 0 and a standard deviation of $\sqrt{(R)}$. When the observed innovations deviate significantly from theory, model error must be present [11,13].

The actual error being evaluated in any comparison of observed and theoretical innovations is error in modelling R, and not H, however. In the theory of the Kalman filter, it is assumed that, in addition to being Gaussian noise processes, with covariances R and Q, the noise sequences $v(k)$ and $w(t)$ have zero means. Any error in modelling the measurement matrix H will be indicated by a nonzero mean for $v(k)$, while errors in modelling F will appear as a nonzero mean for $w(t)$. An adaptive filter tests the modelling of the filter noise variances. If the additional assumption is made that R and Q are well-modelled, however, any modelling error detected may then be assigned to non-zero noise means For an adaptive filter based on matching of theoretical and experimental innovations, the error is attributed to deviations in the presumed mean of v. This model error can be 'covered up' by artificially increasing the measurement variance $R(k)$, which effectively down-weights the parts of the spectral data that are not well-modelled. Any regression done with incomplete models, however, is suboptimal, and the results obtained from adaptive filtering are not always minimum variance estimates. Operation of this filter requires averaging of a set of innovations prior to comparison with theory, for better statistical properties [6]. The lag introduced by the averaging process makes the filter slow to converge to good estimates of states. Estimates obtained from the covariance-matching adaptive filter are very dependent on the initial guesses used to begin filtering, and simplex optimization has been needed to locate the best filtering results, as well as to automate this adaptive filter [7]. Because of these undesirable features, the covariance-matching adaptive filter was not used here

Adaptive filtering by innovations correlation matching

Another check on model quality can be done by investigation of the autocorrelation of the innovations. Matching observed innovations autocorrelation over a part of the innovations sequence to that expected from filter theory permits estimation of the noise variances required for the filter model. For correlation matching, the autocorrelation function σ is calculated for the innovations over some window of autocorrelation lags. Then, the experimental autocorrelation ϕ is related to the theoretical autocorrelation Φ by the equation

$$\phi(k,l) = \Phi(k,l)\alpha + \eta(k,l) \qquad (20)$$

for datum k and lag l, where $\eta(k,l)$ is a zero-mean, white noise term, and α is the fitted parameter, taken here as independent of k. The noise

variances are expressed as linear functions of α, so that

$$\mathbf{R}(k) = \sum_{i=1}^{N} R_i \alpha_i \tag{21}$$

and

$$\mathbf{Q}(k) = \sum_{i=1}^{N} Q_i \alpha_i \tag{22}$$

The parameter α is obtained from the observed innovations autocorrelation by the sequential regression

$$\alpha(k) = \alpha(k-1) + \Theta(k)\Phi^T(k)\mathbf{W}^{-1}(k) \times [\phi(k) - \Phi(k)\alpha(k)] \tag{23}$$

where the covariance parameter Θ is propagated by

$$\Theta(k) = \Theta(k-1) - \Theta(k-1)\Phi^T(k)[\mathbf{W}(k) + \Phi(k)\Theta(k-1)\Phi^T(k)]^{-1}\Phi(k) \times \Theta(k-1) \tag{24}$$

and where W is a weight matrix determined by the autocorrelation at lag 0 [14].

While the computationally simpler matching of theoretical and experimental innovations can also be used to estimate noise variances, the results of Monte Carlo studies show that noise estimates from these adaptive filters tend to be biased [15]. Further, with matching of observed and theoretical innovations autocorrelation, it is possible to estimate both noise variances (R and Q) at once, and these estimates are not strongly affected by measurement model error in the data [14]. Once these quantities have been estimated, subsequent estimation of noise means (deviations of $E(w)$ and $E(v)$ from zero) can be performed. For this reason, innovations correlation was used to obtain estimates of R and Q for filter studies throughout this work.

Adaptive filtering by estimating innovations variance

A third approach to adaptive filteri..ɓ makes use of the available error information carried in the filter innovations and state covariance matrix P. For a chemical system with no systems dynamics, the variance in the innovations is expected to be a function of the measured variables z, the states X and the measurement model H according to the relation

$$\sigma^2_{\nu(k)} = \sigma^2_{z(k)}\left[\frac{\partial \nu(k)}{\partial z(k)}\right]^2 + \sigma^2_{\mathbf{H}(k)}\left[\frac{\partial \nu(k)}{\partial \mathbf{H}(k)}\right]^2 + \sigma^2_{\mathbf{X}(k)}\left[\frac{\partial \nu(k)}{\partial \mathbf{X}(k)}\right]^2 \tag{25}$$

which yields upon substitution the relation

$$\sigma_{\nu(k)} = \left[\mathbf{R}(k) + \mathbf{X}(k|k-1)\mathbf{Q}\mathbf{X}^T(k|k-1) + \mathbf{H}(k)\mathbf{P}(k|k-1)\mathbf{H}^T(k)\right]^{1/2} \tag{26}$$

This equation reflects the fact that the innovation uncertainty σ_ν must remain large when states are not well known, but must be decrease to the limit of measurements noise when states are well known. Since this relation is based on the knowledge of R, it presumes accurate estimates of noise variances, but is permits rapid rejection of incorrectly modelled data if knowledge of noise variances is available. Data may be filtered normally, or rejected, based on comparison of the innovations $\nu(k)$ and the value of $\sigma_\nu(k)$: innovations falling within $\pm 3\sigma_\nu$ may be considered 'within those expected for a correct model', but those innovations falling outside of this range are clear indicators of error in the model.

In this connection, it should be noted that the standardized innovations $n(k)$

$$n(k) = \nu(k)\sigma_\nu(k)^{-1} \tag{27}$$

may also be defined. The squared, standardized innovations observed for filtering p measurements with an n-dimensional state model distribute as chi-square, with $p - n$ degrees of freedom [10]. With this relation, a simple test can be used to evaluate model quality. A threshold can be set, so that innovation values falling within the threshold are filtered normally, while those falling outside the threshold are ignored in the filtering, and affect neither the filter states or covariances. For example, innovations well below the threshold

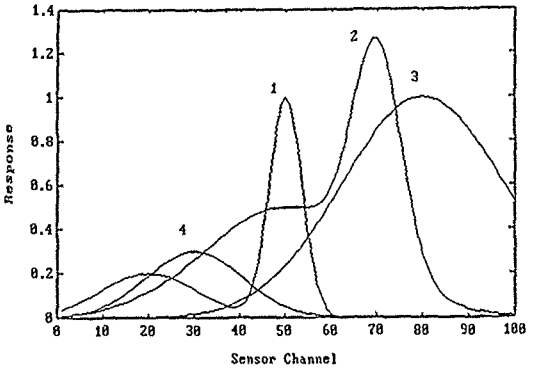


Fig 1 Simulated spectral responses for components present in multicomponent data used for calibration and prediction studies Numbers 1–4 refer to the response function created for the three calibrated components and the unmodelled component.

$3|\sigma_r|$ have a fairly high probability of occurring by chance, while those with values greater than $3|\sigma_r|$ are not likely. In practice, though, an asymmetric threshold on innovations is desirable. Large positive innovations imply model error (for chemical responses with positive peaks), while large negative innovations might be expected as state estimates are refined. However, several consecutive, large, negative innovations may indicate that state estimates may be affected by the model error. In this situation, it is necessary to alter the covariance matrix $P(k|k-1)$, both to increase the uncertainty in state estimates and to increase σ_r. Measurements following this change are processed as before. For work reported here, the absolute threshold was set to $3|\sigma_r|$, and two consecutive measurements producing innovations below $3|\sigma_r|$ caused reset of the diagonal elements of the covariance matrix for all state components contributing more than 5% of the predicted measurement. This selective reset was done to avoid altering state estimates that were not likely to have been influenced by the model error. Calculation of the innovations threshold is fast, and the filtering is set to that most of the data processed are well-modelled. For these reasons, rapid convergence of filter estimates is usually observed, even though the filter is not strictly optimal, because the filter model is incomplete. External optimization methods and extensive iteration are not needed when this adaptive filter is used to correct filter models. When consecutive negative innovations are encountered, however, at least one more iteration should be performed to insure satisfactory estimates of states and covariances.

IMPLEMENTATION

Programs for CLS calibration, partial least-squares calibration, principal components calibration, and Kalman filtering were all developed in the MATLAB programming environment. Kalman filtering programs included the linear drift filter, the proportional drift filter based on an extended Kalman filter, the second-order adaptive filter based on covariance matching, and the innovations variance-based adaptive filter for detection of model errors. In all Kalman filters, the Kalman algorithm (eqs. (A1)–(A5)) was used. The MATLAB environment was run on an Apple MacIntosh SE equipped with 68020 processor, 8 Mbytes of memory, hard disk and a 68882 numeric coprocessor. No effort was made to optimize any of these programs for execution speed.

Data for evaluation of these filter calibration

methods were generated using simulated visible spectra. Validation and training sets were made by randomly generating sets of component concentrations, then by multiplying the true spectra by the concentrations, and finally by adding noise. Noise added to spectra was drawn from Gaussian or uniform distributions; typically, the noise level was such that the peak S/N ratio was 20.1. Data sets containing drift were generated by calculating concentrations using the drift models defined by eqs. (10) and (13). The drift-corrupted concentrations were multiplied by the true spectra, and noise was added to produce sets of noisy, drift-corrupted spectra.

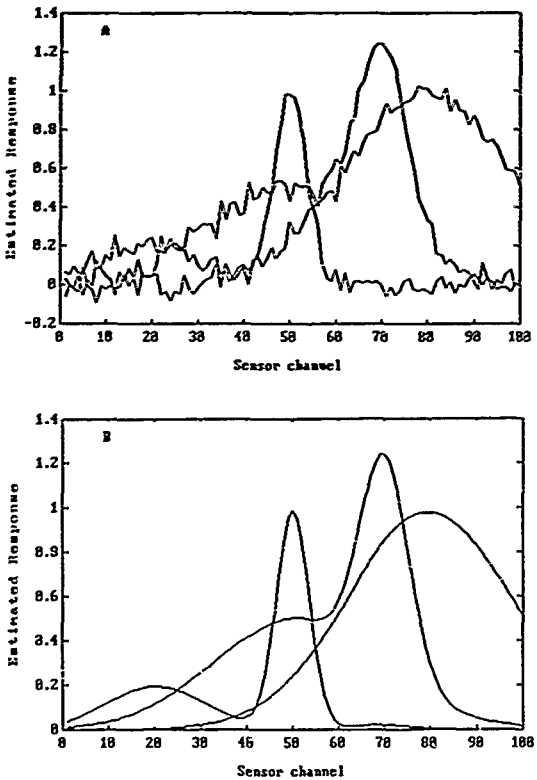


Fig. 2. Estimated spectra from CLS and sequential regression, A. Columns of the K matrix, from CLS regression as applied to 20 member training set, with random, Gaussian noise added to obtain a maximum S/N of 180. B. Columns of the A matrix, from sequential regression of the same 20 member calibration set as above.

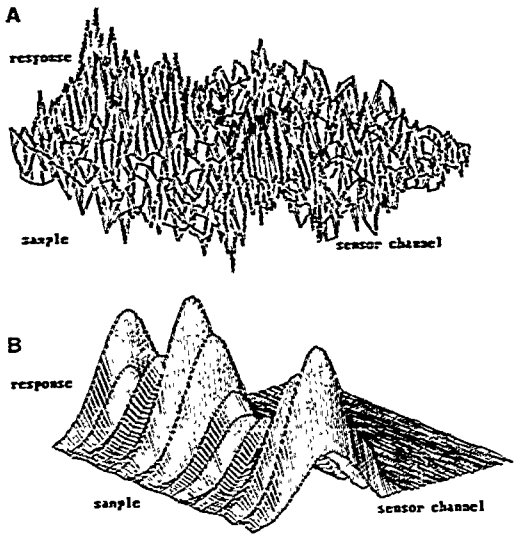


Fig. 3. Innovations from adaptive filtering of incompletely modelled data. A. Innovations from filtering of 25-member validation set with a maximum S_f/λ of 50, and unmodelled component as given in Fig. 1. A sequential regression calibration was used to generate the adaptive filter model, and estimated values for R (2.5×10^{-1}) and Q (1.0×10^{-10}) were used in the filtering. B. Innovations from filtering of 25-member validation set with maximum S_f/λ of 2000, and unmodelled component as given in Fig. 1. Filtering was done as in (A), but with estimated values for R (1.1×10^{-6}) and Q (1.0×10^{-12}) used for filtering.

TABLE 1

Estimation of component concentrations in absence of model error and drift

Method	Prediction set S/N (max.)	Calibration model	R	Q	PRESS *
KF	39	S.R. **	2.6×10^{-3}	0 0	0.0907
CLS	39	CLS ***	0	0	0.804
KF	39	CLS	2.6×10^{-3}	$* 0 \times 10^{-5}$	0.822
CLS	39	S.R.	0	0	0.00
KF	18	S R.	1.06×10^{-2}	0 0	0.524
CLS	18	CLS	0	0	1.246
KF	18	CLS	1.06×10^{-2}	1.0×10^{-5}	1.244
CLS	18	S.R.	0	0	0 524

* PRESS. The sum of squared error for the predicted spectrum as compared to the true, noise-free spectrum, summed over all calibration components for the 20 members of the prediction set The same prediction set was used, with different amounts of added noise, in each case, and each method was applied to each set, so that direct comparison is possible.

** S R., model from sequential regression of absorbance onto standard concentrations, using estimated measurement error variance of 2.6×10^{-3}.

*** CLS: model from classical least squares regression of calibration data, without weighting.

RESULTS AND DISCUSSION

A set of calibration standards was prepared by simulation. The component spectra were as shown in Fig 1. After calibration had been accomplished, prediction was attempted on simulated validation sets, for which the true values of component concentrations, noise variance, and drift were known.

Equivalence of Kalman filter and CLS prediction

To demonstrate the essential equivalence of CLS prediction and sequential regression, the two methods were compared for a well-behaved set of data. Fig. 2 shows plots of the columns of matrix K obtained from sequential regression of calibration spectra onto the standard concentrations, and as estimated from CLS calibration applied to the same training set data. The noise estimate used in the sequential regression was 1.0×10^{-4}, close to the true noise variance contained in the calibration data. Improvement in the estimation obtained through use of the sequential regression is apparent. Overfitting of the calibration data has been diminished substantially by use of the sequential regression and the agreement of the columns of K estimated from sequential regression with the true spectra is excellent.

Table 1 shows results from a time-independent, scalar Kalman filter and CLS prediction applied to a typical sets of validation data, with noise taken from a uniform distribution. In this study, the spectral models were generated two ways: one set was generated by CLS calibration; these contained noise, as demonstrated in Fig. 2. The other set was generated by sequential regression of calibration data; these models were virtually noise-free. Both models were used for Kalman filtering and for CLS prediction. The results from the time-independent Kalman filter were identical to those obtained from CLS prediction when the CLS calibration model was used. However, if the calibration model noise was treated as a form of system dynamics noise, and a suitable value was used for Q (see eq. (A7) in the Appendix) in the Kalman filter, improved estimates resulted. In fact, these estimates tracked estimates obtained from filtering with noise-free calibration models.

TABLE 2

Estimation of component concentrations in presence of unmodelled response

Method	Prediction set S/N (max.)	Calibration model	R	Q	PRESS *
CLS	39	CLS	–	–	3.64
CLS	1944	CLS	–	–	3.86
CLS	1944	S.R.	–	–	2.67
AKF **	39	S.R.	2.6×10^{-3}	1.0×10^{-12}	1.00
AKF	1944	S.R.	2.6×10^{-3}	1.0×10^{-12}	0.13
AKF	39	CLS	2.6×10^{-3}	1.0×10^{-5}	2.21
AKF	1944	CLS	9.9×10^{-7}	1.0×10^{-5}	0.37
PLS ***	39	PLS §	–	–	3.20
PLS	1944	PLS	–	–	2.17
PCR §§	39	PCR §§§	–	–	3.25
PCR	1944	PCR	–	–	2.26

* Modified to reflect unmodelled component presence. PRESS was calculated for the accuracy of prediction of all modelled components.

** AKF. Innovations variance-based adaptive filtering, using innovations limit of $3\sigma_e(k)$, and reset for covariance upon restart of 3×10^{-2}.

*** PLS: Partial least squares prediction, using the algorithm given in Geladi and Kowalski [2].

§ PLS. Partial least-squares calibration, with the model defined by cross-validation. A five-factor model was used in these fits.

§§ PCR: Principal components regression, using the algorithm given in Geladi and Kowalski [2].

§§§ PCR. Principal components calibration, with the model defined by the first three eigenvectors of the calibration data scatter matrix.

Calibration model noise is "explained" by treating it as a form of dynamic noise in the sequential regression, and increasing Q to a realistic value prevents the overfitting of the noisy model to data. In general, Kalman filtering of the prediction sets produced results that were superior to those obtained from CLS prediction, unless CLS prediction was carried out with noise-free calibration models.

Estimation in the presence of unmodelled components

When an extra, non-calibrated component is added to the set of species producing the set of calibration responses, the prediction error increases. Fig. 2 shows the response of four components, the calibration model included concentrations and responses on the last three, since component 4 was absent from the calibration. As indicated in Table 2, the presence of this unanticipated and uncalibrated response produces a significant decrease in the accuracy of estimation of the well-modelled components. Components 1 and 2, whose responses show significant overlap with the unmodelled response of component 1, carry the largest error in estimation. Component 3, with a response that is somewhat separated from the unmodelled component, still carries some error in estimation. Further, the error in estimation is present despite the calibration method employed. Even methods based on regression of data onto factor-based calibration models are unable to compensate for the unmodelled component in the prediction step. The innovations variance-based adaptive filter, on the other hand, successfully compensates for most of the effects of the unmodelled component, and shows significantly less error in the estimated concentrations of components 1 and 2, and slightly less error, on average, in the estimation of component 3.

The results observed for partial least squares (PLS) and principal components regression (PCR) calibration were strikingly different than those observed for fitting by other means. PRESS values obtained from fitting the calibration model to the validation data set were almost independent of the noise contained in the validation data, and they were considerably better than all but those produced by adaptive filtering. The results can be explained by noting that PLS and PCR methods rely on a factor model for the multicomponent system. This factor model is produced during the calibration, and it is set up to remove most of the noise present in the calibration. The three PCR factors (or five PLS factors) fit validation data with or without noise equally well: the dominant error is model error from the uncalibrated component. In this case, the model error is relatively small, and good estimates resulted from factor-based calibration. Residuals from fitting of the validation data showed overfitting of components near the uncalibrated component, just as in the CLS fitting, however.

Sensitivity of innovations variance adaptive filter to noise estimates

As would be expected from its derivation, the adaptive filter based on innovations variance is somewhat sensitive to values used for system and measurement noise. To obtain the results summarized in the table above, values of 1×10^{-9} and 2.6×10^{-3} were used for Q and R, respectively. These values, obtained from the innovations correlation adaptive filter as discussed above, very slightly overestimated the actual noise contributions, which were 0 and 2.5×10^{-3} for system and measurement noise variances. For state values near unity, the third term in eq. (26) will dominate at the start, when the state covariances are large, but the first term will quickly become the dominant term as state covariance decreases. With state values near unity, the second term will, in general, always be small unless system noise is sizable, this term probably could be neglected to decrease computational overhead, if desired. After the first 20 points, and on subsequent filter passes with better state estimates and decreased covariances, estimates of measurement noise variance will dominate the calculation of filter innovations variance, and will therefore set the region where acceptable innovations will be found. Significant over- or underestimation of the measurement noise variance may be expected to significantly affect the performance of the adaptive filter. To test the

TABLE 3

Sensitivity of innovations variance adaptive filter to noise variance estimates

Method	Prediction set S/N (max.)	Calibration model	R	Q	PRESS *
AKF	1944 **	S.R.	1.0×10^{-3}	1.0×10^{-10}	6.04
AKF	1944 **	S.R.	2.5×10^{-3}	1.0×10^{-10}	1.04
AKF	1944 **	S.R.	5.0×10^{-3}	1.0×10^{-10}	1.74
AKF	1944 **	S.R.	1.0×10^{-3}	1.0×10^{-9}	1.81
AKF	1944 **	S.R.	2.5×10^{-3}	1.0×10^{-9}	1.06
AKF	1944 **	S.R.	5.0×10^{-3}	1.0×10^{-9}	1.74
AKF	1944 **	S.R.	1.0×10^{-3}	1.0×10^{-6}	4.05
AKF	1944 **	S.R.	2.5×10^{-3}	1.0×10^{-5}	1.19
AKF	1944 **	S.R.	5.0×10^{-3}	1.0×10^{-5}	1.87
AKF	1944 ***	S.R.	2.5×10^{-3}	1.0×10^{-9}	1.06
AKF	1944 ***	S.R.	5.0×10^{-3}	1.0×10^{-9}	1.75
AKF	1944 ***	S.R.	1.0×10^{-3}	1.0×10^{-5}	1.51

* Modified PRESS. See Table 2 for details. True noise variances for data were $R = 2.6\times10^{-3}$, and $Q = 1.0\times10^{-9}$

** Noise was generated from uniform distribution for this data set.

*** Noise was generated from normal distribution for this data set.

sensitivity of the filtering to errors in the estimation of the measurement and system noise variances, two data sets were prepared for examination, one with noise taken from a uniform distribution, and one with noise taken from a normal distribution. Filtering was done on these data, using the same filter model, and the same initial guesses. Only the guesses for the system and measurement noise was changed from run to run. Table 3 presents the results of that study. Fig. 3 shows the innovations after adaptation of the filter model for two noise levels. The unmodelled fourth component is visible, even when noise in the measurement approaches the size of the unmodelled component (case A) When noise levels are small (case B), good correction of the models for the unmodelled component is observed from the flat innovations over the data set.

TABLE 4

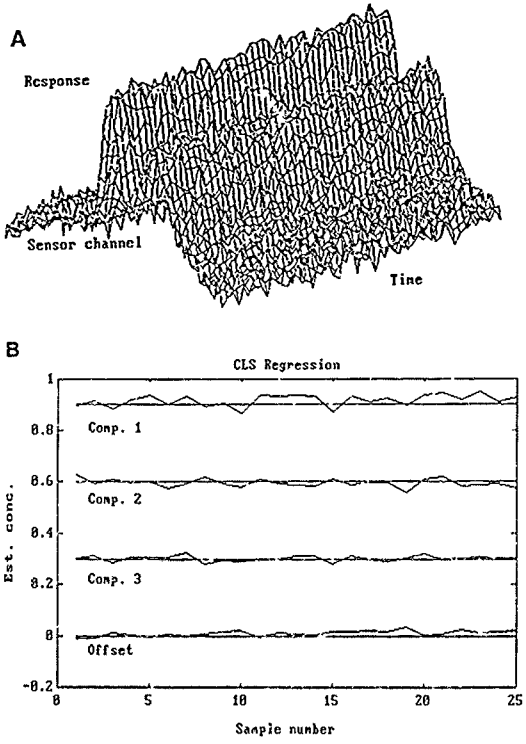
Estimation of noise variance in data

Model type *	Init R	Est R	Init Q	Est Q	True R	True Q
N	1.0×10^{-10}	4.1×10^{-4}	0	1.1×10^{-12}	4.0×10^{-4}	0
N	1.0×10^{-8}	4.1×10^{-4}	0	1.3×10^{-12}	4.0×10^{-4}	0
N	1.0×10^{-4}	4.1×10^{-4}	0	1.3×10^{-12}	4.0×10^{-4}	0
N	1.0×10^{-2}	4.3×10^{-4}	0	1.1×10^{-12}	4.0×10^{-4}	0
D	3.1×10^{-3}	4.1×10^{-4}	0	1.1×10^{-12}	4.0×10^{-4}	0
D	1.0×10^{-6}	4.1×10^{-6}	1.0×10^{-12}	4.4×10^{-8}	4.0×10^{-6}	4.0×10^{-8}
N	1.0×10^{-6}	4.1×10^{-6}	1.0×10^{-6}	5.1×10^{-8}	4.0×10^{-6}	4.0×10^{-8}
D	1.0×10^{-6}	4.1×10^{-6}	1.0	3.4×10^{-4}	4.0×10^{-6}	4.0×10^{-8}
A	1.0×10^{-6}	4.4×10^{-6}	1.0×10^{-8}	4.4×10^{-8}	4.0×10^{-6}	4.0×10^{-8}

* N Complete model, with initial state guess of 0 and covariance guess of I No drift was present D Incomplete model with random drift between measured spectra Initial guesses of state and covariance were as in case N above A Incomplete model, with random drift and unmodelled components present The initial state guess was within 10 × of true state value, and covariance guess of I

Estimation of noise variance parameters

To test the accuracy of estimation of noise variances by the innovations correlation adaptive filter, this filter was applied to several spectra with different noise structures. As above, noise from uniform and normal distributions was used, and the adaptive filter was applied to the data. For this study, a complete spectral model was used, and no adaptation of noise means was necessary. Initial guesses for the noise variances were typically near zero, and guesses for the covariance of the noise estimates were taken as 1. Table 4 summarizes the results from this study. In general, better estimation of noise variance was seen when initial guesses of noise variance were lower than the true values. These overly optimistic guesses converged quickly to accurate noise estimates. When noise estimates were tried that overestimated the noise contributions, convergence was slower, and the final estimates were not as accurate. For this system, estimates of the measurement variance R were found to be more accurate than those for the systems noise Q, but both



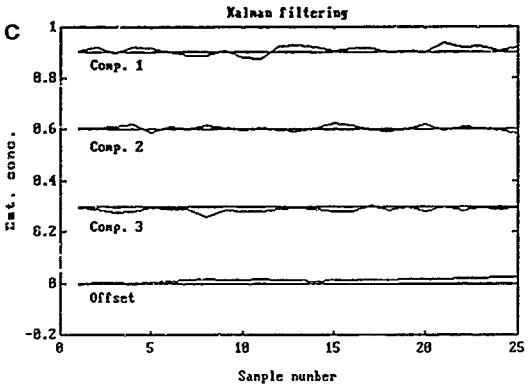


Fig. 4. A. Simulated data showing linear drift in response and offset for Drift3 data set. B. Estimated concentrations for data showing linear drift from CLS regression with offset term. C. Estimated concentrations for data showing linear drift from Kalman filter with drift model.

TABLE 5

Estimation of component concentrations from multicomponent data corrupted by linear drift

Method	Data set *	Prediction set S/N (max.)	Calibration model	Rel. error (%)			PRESS
				Comp. 1	Comp. 2	Comp. 3	
CLS	Drift1	39	CLS **	1.86	0.28	−1.61	28.77
KF	Drift1	39	SR ***	1.85	0.28	−1.61	28.65
CLS	Drift1	39	SR **	0.06	2.80	−0.33	29.83
KF	Drift2	39	SR ***	1.07	−0.05	−2.14	0.09
CLS	Drift2	39	SR **	1.11	−0.09	−2.12	0.25
KF	Drift3	39	SR ***	0.86	0.75	−4.07	0.57
CLS	Drift3	39	SR **	1.67	−0.06	−1.94	0.60
KF	Drift3 §	39	SR ***	0.96	0.57	−0.63	0.18

* Data sets had linear drift in all parameters, including the instrumental response functions for each component and in the offset term. Drift1 noise variances were 1.0×10^{-8} for drift in all instrumental response parameters, and 1.0×10^{-4} in the offset term. Drift2 noise variances were 1.0×10^{-8} in all instrumental response parameters and offset terms, in which the exception of the response function for component 1, which had a drift variance of 1.0×10^{-7}. Drift3 had noise variances for all instrumental response variables of 1.0×10^{-7}, and drift in the offset term with variance 1.0×10^{-8}. All data sets had added measurement noise, with a variance of 2.5×10^{-3}.

** Regression model augmented to include an offset term.

*** Filter model augmented to include an offset term and drift parameters in responses and offset. Duplicate measurements made at each point.

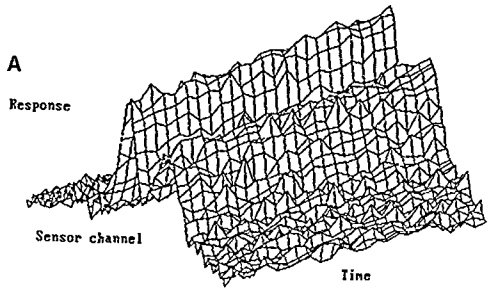
§ Four replicates were used in the measurement step for this run.

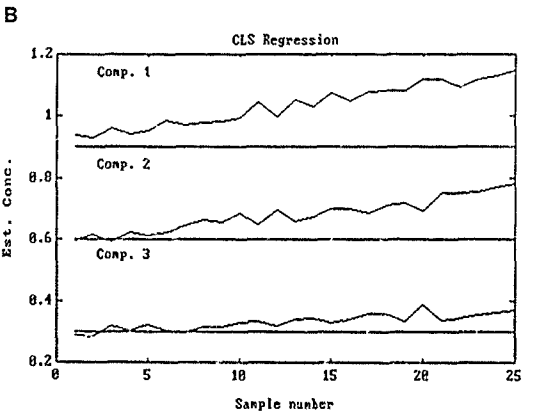
estimated values were close enough to the true noise variances to be useful in filtering data. PRESS values were generally lower for data with normally-distributed noise than for data with uniform noise, as might be expected from the derivation of the adaptive filter.

Correction of drift in multicomponent prediction

Correction for linear drift was briefly studied for multicomponent prediction. For the three component chemical system used in these studies, the state vector used for filtering was $X = [C_1\ C_2\ C_3\ b\ d_1\ d_2\ d_3\ d_b]$, where b is the offset term in the filter measurement model, and d_i describes drift in state i. Definition of the state in this way meant that eight parameters were fitted to the multicomponent data. Duplicate data were used to fit to ensure filter observability, as discussed above. Data with linear drift in response and offset were generated to test this filter. A typical data set is shown in Fig. 4A.

Results from this study are summarized in Table 5. Filtering results are not significantly better than those obtained from CLS regression with an offset term in the calibration model, but filtering





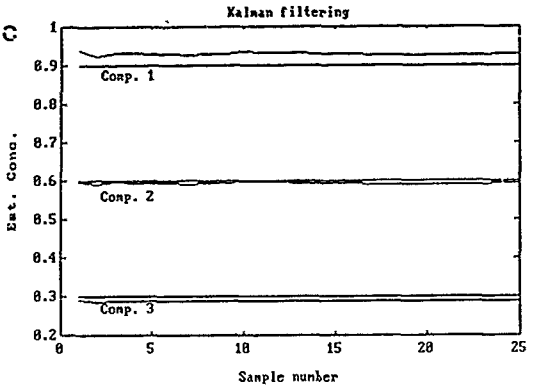


Fig 5 A Simulated data showing proportional drift in response and offset for Drift4 data set B Estimated concentrations for data showing proportional drift from CLS regression with offset term C Estimated concentrations for data showing proportional drift from extended Kalman filter with drift model

typically produced lower PRESS values than CLS regression. Fig. 4B and C clarify the advantage gained by the use of the much more complex filter model over the simpler CLS model. significantly reduced fluctuations in the estimation of concentrations from a drift-corrupted prediction set. It is clear from the figure that, when properly modelled, linear drift can be effectively removed from multicomponent systems Part of the failure to achieve better drift correction with the filter lies in the weak observability of the filter drift model With single measurements made on a drifting sys-

TABLE 6

Estimation of component concentrations from multicomponent data corrupted by proportional drift

Method	Data set *	Prediction set S/N (max)	Calibration model	Rel. error (%)			PRESS
				Comp 1	Comp. 2	Comp 3	
EKF **	Drift4	39	S.R. ***	3 15	−1 07	−3 80	0 09
CLS	Drift4	39	S R.	14 8	13 3	10 57	18 35
CLS	Drift4	39	S R ¶	14 5	13 1	9 78	18 39
EKF	Drift4	39	S R. §§	3 80	−0 34	−3 10	0 09
EKF	Drift5	39	S R §§	1.39	0 31	0 51	0 20
CLS	Drift5	39	S R. ¶	3 79	2 76	3.19	1.16
EKF	Drift5	39	S R. ***	1.12	0 03	0 24	0 19

* Data sets had proportional drift in the response parameters for all components, along with random drift in the offset Drift4 had proportional drift of 0 9905, and random drift variance of 1.0×10^{-8} Drift5 had proportional drift of 1 0016, and random drift of 1.0×10^{-8}. Both sets had measurements noise with variance 2.5×10^{-3}

** Extended Kalman filter, with filter models as defined by eqs. (14) and (15). Filter state included component concentrations and proportional drift parameter.

*** Using correct concentrations as the initial guess, with guessed proportional drift of 1

¶ Using augmented regression model, with offset term.

§§ Using zero as initial guess of component concentrations, and guessed proportional drift of 1.

tem, the filter drift model is not observable, and estimates of component concentrations fluctuate wildly Use of duplicate measurements insures that the filter model is observable. It is only weakly observable, however, and the estimated component concentrations are not as stable as in other filtering applications reported here. Collecting even more replicates improves filter results, since this has the effect of making the filter model more observable. The size of the drift also has an effect on the precision of the concentration estimates. As drift magnitude increases, Q also must increase, and as indicated in eq (A5), the covariance in the final filter states must also increase Thus, even though linear drift can be corrected and its effects removed with sufficient replicates, the precision of the predictions is degraded.

The correction of proportional drift by use of an extended Kalman filter was also investigated. The same spectral models and calibration was used as in the other studies reported above, and proportional drift was introduced into the set of spectral data to be used for prediction. A typical data set is shown in Fig. 5A. The extended Kalman filter was applied to data with proportional drift, using the filter model described in eqs. (15) and (17) above. Linearization of the system dynamics was done as in eq. (18). As before, estimates of noise variance were obtained from the innovations correlation adaptive filter. Table 6 shows results of fitting data with proportional drift.

Correction of proportional drift is better than correction of linear drift, but unlike the results obtained from the linear drift study, the initial guess for the states used in filtering is important in convergence of drift estimates. Even with poor initial guesses, though, very good compensation of drift occurs as is apparent from the error in the estimated results and the PRESS values. As with linear drift, correction of proportional drift results in degraded precision in the estimated component concentrations.

CONCLUSIONS

This work has demonstrated the ease with which a CLS calibration may be accomplished by Kalman filtering. This approach to CLS calibration and prediction provides for improved calibration models and improved prediction accuracy in noisy data. Another benefit is the ability to reject unmodelled component responses in the prediction of analyte concentrations in unknown mixtures. The correction of drift in the response of the chemical system during the prediction step is also possible, provided that a suitable model for the drift process can be generated. All these corrections represent relatively simple additions to the calibration model The classical least squares calibration model is especially convenient because the Kalman filter has been derived for use with a causal measurement model Given the general definition of the filter state, and the possibility of extending sequential regression to the inverse model, however, there is no difficulty in extending the time-series concepts of Kalman filtering to other calibration models, at least on an ad hoc basis

All filtering routines presented here are relatively fast. These could be realized in a suitable real-time language, if desired, for on-line use.

ACKNOWLEDGEMENT

This work was supported by grant DE-FG02-86ER13542 from the Division of Chemical Sciences, Office of Basic Energy Sciences, U S Department of Energy.

APPENDIX

Scalar Kalman algorithm

Propagation of filter states in time

$$\mathbf{X}(k) = \mathbf{F}(k)\mathbf{X}(k-1) \tag{A1}$$

Propagation of state covariance in time

$$\mathbf{P}(k|k-1) = \mathbf{F}(K)\mathbf{P}(k-1|k-1)\mathbf{F}^T(k) + \mathbf{Q}(k) \tag{A2}$$

Kalman gain

$$\mathbf{K}(k) = \mathbf{P}(k|k-1)\mathbf{H}(k)\left[\mathbf{H}^T(k)\mathbf{P}(k|k-1) \times \mathbf{H}(k) + R(k)\right]^{-1} \tag{A3}$$

State update

$$\mathbf{X}(k|k) = \mathbf{X}(k|k-1) + \mathbf{K}(k)\left[z(k) - \mathbf{H}^T(k) \times \mathbf{X}(k|k-1)\right] \quad \text{(A4)}$$

Covariance update

$$\mathbf{P}(k|k) = \mathbf{P}(k|k-1) - \mathbf{K}(k)\mathbf{H}^T(k)\mathbf{P}(k|k-1) \quad \text{(A5)}$$

Initial guesses

$$\mathbf{X}(0|0) = \mathbf{X}_0 \mathbf{P}(0|0) = \mathbf{P}_0 \quad \text{(A6)}$$

Noise assumptions

$$E\left[v(k)v^T(k)\right] = R(k)$$
$$E\left[w(k)w^T(k)\right] = \mathbf{Q}(k)$$
$$E\left[w(k)v^T(j)\right] = 0 \quad \text{for all} \quad j, k \quad \text{(A7)}$$

REFERENCES

1 K R Beebe and B R Kowalski, An introduction to multivariate calibration and analysis, *Analytical Chemistry*, 57 (1988) 1007A–1017A

2 P Geladi and B R Kowalski, Partial least-squares regression A tutorial, *Analytica Chimica Acta*, 185 (1986) 1–17

3 C-N Ho, G D Christian and E.R Davidson, Application of the method of rank annihilation to quantitative analyses of multicomponent fluorescence data from the video fluorometer, *Analytical Chemistry*, 50 (1978) 1108–1113

4 B E Wilson and B R Kowalski, Quantitative analysis in the presence of spectral interferents using second-order non-bilinear data, *Analytical Chemistry*, 61 (1989) 2277–2284

5 S D. Brown, The Kalman filter in analytical chemistry, *Analytica Chimica Acta*, 181 (1986) 1–26

6 S C. Rutan and S D Brown, Model error compensation in multicomponent analysis using adaptive Kalman filtering, *Analytica Chimica Acta*, 160 (1984) 99–119.

7 S C. Rutan and S.D Brown, Simplex-optimized adaptive Kalman filtering, *Analytica Chimica Acta*, 167 (1985) 39–50

8 P.C. Thijssen, S M. Wolfrum, G. Kateman and H C. Smit, A Kalman filter for calibration, evaluation of unknown samples and quality control in drifting systems, *Analytica Chimica Acta*, 156 (1984) 87–101

9 J Ljung and T Soderstrom, *Theory and Practice f Recursive System Identification*, MIT Press, Cambridge, MA, 1985

10 B D O Anderson and J B Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979

11 A Gelb (Editor), *Applied Optimal Estimation*, MIT Press, Cambridge, MA, 1974

12 G J Bierman, A comparison of discrete linear filtering algorithms, *IEEE Transactions on Aerospace and Electronic Systems*, AES-9 (1973) 28–37

13 A H Jazwinski, *Stochastic Process and Filtering Theory*, Academic Press, New York, 1970

14 P R Belanger, Estimation of noise covariance matrices for a linear time-varying stochastic process, *Automatica*, 10 (1974) 267–275

15 C Neethling and P Young, Comments on 'Identification of optimum filter steady-state gain for systems with unknown noise covariances', *IEEE Transactions on Automation and Control*, AC-19 (1974) 623–625

16 N R Draper and H Smith, *Applied Regression Analysis*, Wiley, New York, 2nd ed., 1981

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 107–116
Elsevier Science Publishers B.V., Amsterdam

Model building in chemistry using profile t and trace plots

Douglas M. Bates *
Department of Statistics, University of Wisconsin, Madison, WI 53706 (U S A)

Donald G. Watts
Department of Mathematics and Statistics, Queen's University, Kingston, Ontario K7L 3N6 (Canada)

(Received 1 April 1990, accepted 12 July 1990)

Abstract

Bates D M and Watts D G , 1991 Model building in chemistry using profile t and trace plots *Chemometrics and Intelligent Laboratory Systems*, 10 107–116

The aim of model building is to determine the correct model, which means that the equation describing the phenomenon under study includes all the important factors in the correct form, and excludes unimportant factors Practically, of course, we can only use the data at hand to fit a model which is 'adequate' In linear and nonlinear regression, a model which is inadequate because an important factor is not included, or because a factor is incorporated in a wrong form, can often be detected by examining plots of the residuals And in linear regression, models which include too many factors or too many parameters can often be detected by examining the parameter correlation matrix, or the parameter estimates and their standard errors For nonlinear models, however, such linear approximation summaries are not reliable To aid in the development of nonlinear models, we recommend using profile likelihood plots The plots are simple to generate and appear to be especially useful in detecting models which could be simplified by removing factors or by equating parameters In this paper we use data sets from chemical engineering to illustrate the value of profile t and profile trace plots in model building

INTRODUCTION

Linear regression

Consider a set of data consisting of values of a set of factors, x_{np}, $n = 1, 2, \ldots, N$, $p = 1, 2, \ldots, P$, and the corresponding values of a response, y_n, which are well described by a model of the form

$$Y_n = \beta_1 x_{n1} + \beta_2 x_{n2} + \ldots + \beta_P x_{nP} + Z_n \qquad (1)$$

where Y_n is a random variable corresponding to the observation for the nth case, and Z_n is a random variable corresponding to the 'noise' infecting that case. The noise for each case is assumed to be normally distributed with mean 0 and variance σ^2, and independent from case to case The model for all N cases can be written in matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z} \qquad (2)$$

0169-7439/91/$03.50

where Y is the $N \times 1$ vector of random variables representing the responses, X is the $N \times P$ derivative matrix, β is the $P \times 1$ vector of unknown parameter values, and Z is the vector of random variables representing the noise. The quantity $X\beta$ is called the expected value of Y and the model is termed linear because the derivative of the expected value with respect to any parameter does not depend on any of the parameters [1].

For a linear model of the form (2) with normally distributed noise, classical statistical analysis (see, for example ref. 2) establishes that the 'best' estimate of β, given data y, can be formally written as

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{3}$$

where $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_P)^T$, is the least squares estimate. Furthermore the associated estimator can be shown to have the properties that it is normally distributed with expected value β and variance-covariance matrix $(X^T X)^{-1}\sigma^2$. The pth parameter thus has estimated standard error

$$se(\hat{\beta}_p) = s\sqrt{\left\{(X^T X)^{-1}\right\}_{pp}} \tag{4}$$

where $s^2 = S(\hat{\beta})/(N - P)$ is the variance estimate given by the minimum sum of squares divided by its 'degrees of freedom', $N - P$. A $1 - \alpha$ confidence interval for that parameter is given by

$$-t(N - P, \alpha/2) \leqslant \delta(\beta_p) \leqslant t(N - P; \alpha/2) \tag{5}$$

where

$$\delta(\beta_p) = \frac{\beta_p - \hat{\beta}_p}{se(\hat{\beta}_p)} \tag{6}$$

is the studentized parameter and $t(N - P; \alpha/2)$ is the value which isolates an area $\alpha/2$ under the right tail of the Student's t distribution with $N - P$ degrees of freedom. A $(1 - \alpha)$ joint parameter inference region for the parameters is given by

$$(\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta}) \leqslant P s^2 F(P, N - P; \alpha) \tag{7}$$

where $F(P, N - P, \alpha)$ is the value which isolates an area α under the right tail of Fisher's F distribution with P and $N - P$ degrees of freedom

TABLE 1

Second-order rate constants for reactions with CN^- at 25°C

$x = \ln(k_x/k_H)$	$x = \sigma_x$
−0.215	−0.251
0.00	0.000
0.28	0.063
0.35	0.110
0.69	0.193
0.83	0.210
1.64	0.388

Michael addition

Gross and Hoz [3] obtained data on the relative reaction rate of the addition of CN^- to a series of 1,1-diaryl-2-nitroethylenes for which the linear model

$$Y_n = \beta_1 + \beta_2 x_n + Z_n \tag{8}$$

is appropriate. In eq. (8), Y_n relates to the natural logarithm of the relative rate constant, $\ln(k_x/k_H)$, and x to the substituent constant, σ_x. The data are listed in Table 1. The row 0, 0 corresponds to hydrogen.

For these data, the least squares estimates are $\hat{\beta} = (-0.031, 4.13)^T$ with parameter standard errors 0.036, 0.19 respectively. The variance estimate is $s^2 = 0.00474$ with five degrees of freedom, and we have

$$X^T X = \begin{pmatrix} 7.00000 & 0.91800 \\ 0.91800 & 0.25252 \end{pmatrix}$$

$$(X^T X)^{-1} = \begin{pmatrix} 0.27302 & -0.99254 \\ -0.99254 & 7.56837 \end{pmatrix}$$

Joint confidence regions are ellipses, as shown in Fig. 1.

Nonlinear regression

Now consider a model of the form

$$Y_n = f(\theta, x_n) + Z_n \tag{9}$$

where f is a nonlinear expectation function, x_n is a vector of regressor variables for the nth case, and $\theta = (\theta_1, \ldots, \theta_P)^T$ is a $P \times 1$ vector of unknown parameters. (We use θ to emphasize the difference from linear models. As before, the dis-

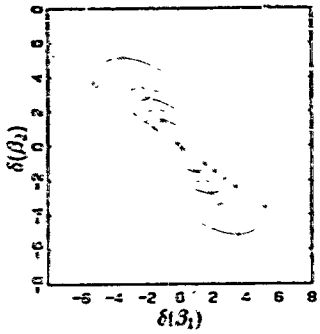


Fig. 1. Joint 60, 80, 90, 95, and 99% confidence regions for the parameters β_1, β_2 for the CN^- Michael addition data. Also shown is the parameter trace $\hat{\beta}_2(\beta_1)$ which intersects the ellipses at the vertical tangents, and the parameter trace $\hat{\beta}_1(\beta_2)$ which intersects the ellipses at the horizontal tangents.

turbances Z_u are assumed to be normal $(0, \sigma^2)$ and independent. A model $f(\theta, x_u)$ is nonlinear if at least one derivative of the expectation function with respect to at least one of the parameters involves at least one of the parameters [1].) For example, the Michaelis–Menten model for enzyme kinetics, $f = \theta_1 x/(\theta_2 + x)$ is deemed nonlinear because the derivative $\partial f/\partial\theta_1 = x/(\theta_2 + x)$ involves θ_2.

Unlike the linear model, eqn. (1), there are no analytic results for the estimates and their distributions. Indeed, there is not even an explicit solution for the least squares estimates; to find the least squares estimates, we must resort to search or iterative techniques. Properties of the estimates are usually assumed to be well represented by linear approximations evaluated at the least squares estimates $\hat{\theta}$, for example, the linear approximation variance–covariance matrix is taken to be $(\mathbf{V}^T\mathbf{V})^{-1}s^2$, where $s^2 = S(\hat{\theta})/(N-P)$ is the variance estimate, $\mathbf{V} = \partial\eta/\partial\theta^T$ is the derivative matrix evaluated at $\hat{\theta}$, and $\eta(\theta) = (f(\theta, x_1), \ldots, f(\theta, x_u))^T$ is the vector of function values evaluated at the design points.

The linear approximation standard error for the parameter θ_p is, by analogy with eq. (4),

$$se(\hat{\theta}_p) = s\sqrt{\{(\mathbf{V}^T\mathbf{V})^{-1}\}_{pp}} \quad (10)$$

and a linear approximation $(1-\alpha)$ marginal confidence interval is, by analogy with eq. (5),

$$-t(N-P;\alpha/2) \leq \delta(\theta_p) \leq t(N-P,\alpha/2) \quad (11)$$

where the studentized parameter is defined by analogy with eq. (6),

$$\delta(\theta_p) = \frac{\theta_p - \hat{\theta}_p}{se(\hat{\theta}_p)} \quad (12)$$

Finally, a linear approximation $(1-\alpha)$ joint parameter inference region for the parameters is taken to be

$$(\theta-\hat{\theta})^T\mathbf{V}^T\mathbf{V}(\theta-\hat{\theta}) \leq Ps^2F(P, N-P;\alpha) \quad (13)$$

Unfortunately, linear approximation inference regions are not trustworthy [1].

Profile t plots

Because the sampling theory approach is not adequate for nonlinear regression, it is necessary to use a more general approach based on the likelihood function. Fortunately, for noise which is independently normally distributed with constant variance, the likelihood function depends primarily on the sum of squares function

$$S(\theta) = (y-\eta(\theta))^T(y-\eta(\theta)) \quad (14)$$

and so drawing inferences about the parameters reduces to summarizing the sum of squares function efficiently and meaningfully.

For linear models of the form of eq. (2), the sum of squares function is quadratic in β and so contours of constant likelihood, which correspond to contours of constant relative plausibility of parameter values, are concentric ellipsoids. For example, the elliptical confidence regions in Fig. 1 are also sum of squares contours. To summarize the likelihood function for a linear model then, we only need to specify the (common) center of the ellipsoids $(\hat{\beta})$, and their size and orientation. This can be done mathematically for any number of

parameters, and essentially reduces to the summary eq. (7), but visualizing the joint region in more than three dimensions is very difficult.

For nonlinear models, the sum of squares surface is not quadratic, and so the problem becomes one of interpreting or visualizing a complicated surface in multiple dimensions. To do this, we focus on the characteristics of the sum of squares function when viewed in one or two dimensions.

A useful view in one dimension is given by its "shadow" in that dimension: the profile sum of squares function. For a model of the form (2) or (9), the profile sum of squares function for the pth parameter can be written

$$\tilde{S}(\theta_p) = \min_{\theta_{-p}} S\left[\left(\theta_p, \theta_{-p}^T\right)^T\right] = S\left(\left(\theta_p, \tilde{\theta}_{-p}^T\right)^T\right) \tag{15}$$

where the trace vector

$$\tilde{\theta}_{-p} = \left(\tilde{\theta}_1, \ldots, \tilde{\theta}_{p-1}, \tilde{\theta}_{p+1}, \ldots, \tilde{\theta}_p\right)^T \tag{16}$$

is the least squares estimate of θ_{-p} conditional on the profile parameter θ_p. The notation $(\theta_p, \tilde{\theta}_{-p}^T)$ denotes the vector with elements

$$\left(\tilde{\theta}_1, \ldots, \tilde{\theta}_{p-1}, \theta_p, \tilde{\theta}_{p+1}, \ldots, \tilde{\theta}_p\right)$$

For a linear model, the profile sum of squares function is a parabola, and can be written in terms of the studentized parameter as

$$\tilde{S}(\beta_p) = S(\hat{\beta}) + s^2\delta(\beta_p)^2 \tag{17}$$

By rearranging this equation, we can write

$$\delta(\beta_p) = sign\left(\beta_p - \hat{\beta}_p\right)\sqrt{\tilde{S}(\beta_p) - S(\hat{\beta})}\,/s \equiv \tau(\beta_p) \tag{18}$$

where $\tau(\beta_p)$ is the profile t function. That is, for a linear model, the profile t function is identically equal to the studentized parameter function.

For a nonlinear model, the profile t function is defined as

$$\tau(\theta_p) = sign\left(\theta_p - \hat{\theta}_p\right)\sqrt{\tilde{S}(\theta_p) - S(\hat{\theta})}\,/s$$

and is, in general, not equal to the studentized parameter function. The profile t function is similar to the ξ statistic used by Bliss and James [4].

The profile t function is valuable because plots of the profile t function versus the studentized profile parameter provide useful information on the nonlinearity of an estimation situation. This is because, for a linear model, a plot of $\tau(\beta_p)$ versus the studentized profile parameter $\delta(\beta_p)$ is a straight line through the origin with unit slope. Departures of the profile t plot of $\tau(\theta_p)$ versus $\delta(\theta_p)$ from the 45 degree line reveal nonlinearity in the parameter, and determining where $\tau(\theta_p)$ intersects the horizontal line at height $\pm t(N-P, \alpha/2)$ determines an accurate nominal $(1-\alpha)$ likelihood interval for θ_p.

Profile traces

Additional important information can be obtained from pairwise plots of the components of the trace vector versus the profile parameter. That is, we overlay plots of $\tilde{\theta}_q(\theta_p)$ versus θ_p and $\tilde{\theta}_p(\theta_q)$ versus θ_q.

For a linear model, a plot of the trace $\tilde{\theta}_q(\beta_p)$ versus β_p will be a straight line through the origin with slope given by the correlation between the parameters (derived from the appropriate element of the matrix $(\mathbf{X}^T\mathbf{X})^{-1}$). Furthermore, the traces will intersect the parameter joint confidence ellipses at points of horizontal or vertical tangency of the ellipses. (See Fig. 1 or a plot of the profile traces for the CN^- Michael addition data.)

For a nonlinear model, the traces will be curved, but will still intersect the parameter joint likelihood regions at points of vertical and horizontal tangency. This information, together with information from the profile t plots, can be used to obtain accurate sketches of the joint regions, as described in ref. 1, Appendix 6. The traces and sketches reveal useful information about the interdependence of the parameter estimates caused by the form of the model for the expectation function and by the experimental design used in the investigation. Such information can provide valuable insights for inference, for model building, and for design, as demonstrated in the next section.

CODIMER HYDROGENATION

Tschernitz et al. [5] obtained and analyzed data on the vapor phase hydrogenation of mixed isooctenes over a solid supported nickel catalyst in a study to determine the most plausible mechanism for a reaction. The data consisted of the average reaction temperature T (Kelvin), the average partial pressures of hydrogen (x_1), of codimer (x_2), and of hydrogenated codimer (x_3), (atmospheres), and the reaction rate (lb/(h) (lb catalyst)).

Eighteen mechanisms were postulated for the reaction, and the most plausible one is found to be that in which the reaction between molecularly adsorbed hydrogen and adsorbed codimer is controlled by the surface reaction, so the reaction rate is

$$r = \frac{\theta_1\theta_2\theta_3 x_1 x_2}{(1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3)^2} \qquad (19)$$

(ref. 5, model d). The parameters θ_2, θ_3, and θ_4 represent adsorption equilibrium constants and θ_1 is the product of the adsorption velocity constant of hydrogen and codimer molecules $\times sL$, where sL represents the 'activity' of the catalyst. The parameter θ_1 also has the interpretation of the proportion of the surface area of the catalyst which is covered by the reactants.

It is assumed that each of the constants can be expressed as a function of temperature by means of an Arrhenius relation,

$$\theta_i = \exp(\alpha_i/R + \beta_i/RT)$$

where $R = 1.986$ is the gas law constant, α is the effective entropy change, and β is the negative effective enthalpy change under the assumption that the catalyst activity remains unaltered with change in temperature.

The linear summary statistics for model d, using the data at all temperatures and the Arrhenius form for the velocity and equilibrium constants, are shown in Table 2. To improve the behavior of the estimates, we scaled and centered the data using $x_0 = 1000(1/T - 1/548)$, and, to avoid confusion, define $\phi_i = \alpha_i + \beta_i/548$ and $\gamma_i = \beta_i/1000$. Inspection of the parameter estimates and their standard errors in Table 2 suggests that ϕ_1 and γ_1 could be zero, that ϕ_4 and ϕ_2 could be equal, and that γ_3 and γ_2 could be equal. However, we must be careful about incorporating model reductions which involve the ϕs since they depend on the arbitrary centering temperature T_0 and the associated γ.

To demonstrate the kinds of information which is available from profiling, we present selected profile trace plots in Fig. 2 and discuss various aspects of the plots. Superimposed on the trace plots are sketches (dashed and solid closed curves) of the joint likelihood regions. The horizontal and vertical tangents of contours which are incompletely determined are shown by short bars on the traces. Also shown is the straight line (solid or dashed) corresponding to equal values of the parameters with the × indicating the point at which both parameters equal zero

From Fig. 2a it can be seen that ϕ_1 could be zero, since the point corresponding to $\phi_1 = 0$ lies

TABLE 2

Parameter summary for codimer hydrogenation data, model d

Parameter	Est	se	Correlation						
			ϕ_1	ϕ_2	ϕ_3	ϕ_4	γ_1	γ_2	γ_3
ϕ_1	−0.16	0.35							
ϕ_2	−2.78	0.37	−0.95						
ϕ_3	−1.08	0.31	−0.88	0.81					
ϕ_4	−2.97	0.87	−0.19	0.30	0.48				
γ_1	−2.66	1.79	−0.41	0.35	0.38	0.10			
γ_2	6.38	1.84	0.19	−0.14	−0.09	0.06	−0.91		
γ_3	5.41	1.22	0.14	0.02	−0.09	0.15	−0.72	0.73	
γ_4	17.75	2.48	−0.03	−0.03	−0.18	−0.83	0.11	−0.10	−0.02

well within the vertical 'shadow' of the joint likelihood region and has a studentized value near zero. However, $\phi_2 = 0$ is not plausible, nor is $\phi_2 = \phi_1$. Similarly, from Fig. 2e, $\gamma_1 = 0$ is eminently plausible, as is $\gamma_1 = \phi_4$. The latter is meaningless, however, since the two parameters are of different

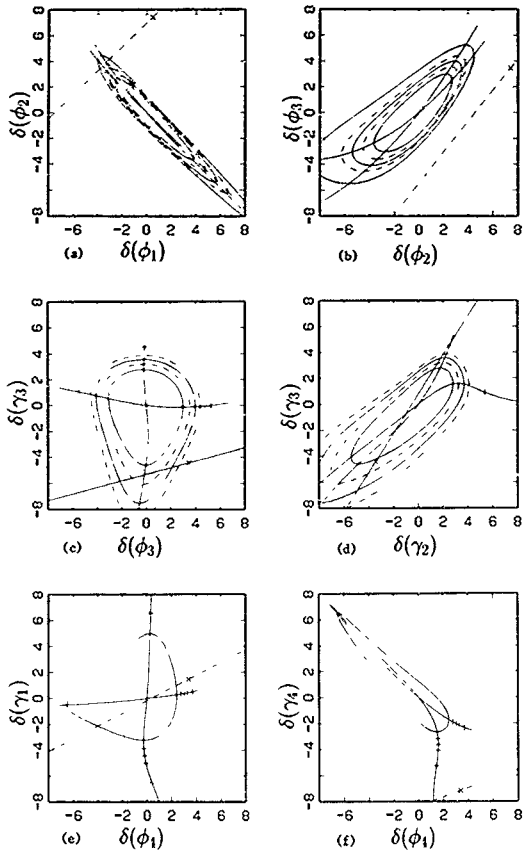


Fig. 2 Selected profile traces for codimer hydrogenation, model d (a) ϕ_2 vs ϕ_1, (b) ϕ_3 vs. ϕ_2, (c) $\gamma_3$vs ϕ_3, (d) γ_3 vs γ_2, (e) γ_1 vs ϕ_4, (f) γ_4 vs ϕ_4 The solid and dashed closed curves denote the 60, 80, 90, 95, and 99% joint likelihood boundaries The solid or dashed straight line is the line of equality of two parameters, and the × indicates the point corresponding to θ_p, $\theta_q = 0, 0$ Short vertical and horizontal bars on the traces show the boundaries of contours which are not completely determined

types. Comparing Fig. 2b and d, it can be seen that $\phi_3 = \phi_2$ is not plausible, but that $\gamma_3 = \gamma_2$ is highly plausible.

The high parameter correlations between the parameters manifest themselves as long ridges in the 8-dimensional inference region, as illustrated in Figs 2a and f, where only the 60% contour is complete. High parameter correlations often indicate overparametrization, but it appears that overparametrization can also manifest as large joint likelihood regions, especially unclosed ones, as shown in Fig. 2e, where there is negligible correlation, but the joint region is very large. This indicates a subspace in which the sum of squares surface is very flat, which could be due to overparametrization

Because the ϕs depend on the arbitrary centering temperature T_0, we first considered model reductions involving only the γs We refitted the model, first holding γ_1 at zero, and still found that $\phi_1 = 0$ was plausible, as was $\gamma_3 = \gamma_2$ Setting $\gamma_3 = \gamma_2$ and $\gamma_1 = 0$ gave the results shown in Table 3.

The residual sum of squares went from $2\,2456 \times 10^{-4}$ with 32 degrees of freedom to $2\,3543 \times 10^{-4}$ with 34 degrees of freedom, so there is not a statistically significant extra sum of squares At this point we noted that two response values gave rise to large studentized residuals and so these rows were deleted and the model refitted. The main effect of this was to reduce the residual sum of squares by about 30% and to reduce the parameter standard errors by about 15%.

Because $\gamma_1 = 0$, it is legitimate to follow through and set $\phi_1 = 0$ as well The results from this model, using the edited data, are shown in Table 4 The

TABLE 3

Parameter summary for codimer hydrogenation data, model d with $\gamma_1 = 0$ and $\gamma_3 = \gamma_2$

Parameter	Estimate	se	Correlation				
			ϕ_1	ϕ_2	ϕ_3	ϕ_4	γ_2
ϕ_1	−0 27	0 32					
ϕ_2	−2 71	0 36	−0 96				
ϕ_3	−0 96	0 29	−0 86	0 82			
ϕ_4	−2 89	0 85	−0 19	0 30	0 51		
γ_2	3 77	0 62	−0 36	0 43	0 57	0 42	
γ_4	17 78	2 40	0 00	−0 07	−0 22	−0 85	0 08

TABLE 4

Parameter summary for edited codimer hydrogenation data, model d with $\phi_1 = 0$, $\gamma_1 = 0$, and $\gamma_3 = \gamma_2$

Parameter	Estimate	se	99% Likelihood		Correlation			
			lower	upper	ϕ_2	ϕ_3	ϕ_4	γ_2
ϕ_2	−3 03	0 09	−3 28	−2 77				
ϕ_3	−1 13	0 13	−1 48	−0 70	−0 12			
ϕ_4	−3 14	0 79	−6 90	−1 24	0 38	0 67		
γ_2	3 63	0 49	2 25	5 12	0 28	0 54	0 38	
γ_4	17 51	2 32	11 96	28 39	−0 20	−0 45	−0 87	0 07

profile trace plots (selected examples of which are shown in Fig 3) still show considerable nonlinearity in the model–data set–parametrization combination. Parameters ϕ_4 and γ_4 are the worst affected, both individually and jointly, as shown by the strongly curved profile t plots The asymptotic behavior in the profile t plots causes the joint likelihood regions to be open at levels above 90% Although the line $\phi_2 = \phi_4$ passes through the center of the joint likelihood region, Fig 3a, it makes no sense to equate these parameters because they depend on the centering temperature and the γ parameters We conclude, therefore, that the simplest form of model d has been obtained

It is useful to note that of the ten trace pair plots only three (ϕ_3 vs ϕ_2, γ_2 vs ϕ_2, and γ_2 vs ϕ_3) gave closed contours at the 95 and 99% levels Since the model has been pared to a sensible minimum number of parameters, this suggest that improvement in the behavior of the likelihood surface could only be achieved by incorporating more data. From the remaining seven trace plots, and from the profile t plots, fig 3e and f, it is clear that the open contours are due to lack of information on ϕ_4 and γ_4. (In Fig 3e the profile t approaches an asymptote as ϕ_4 reduces, and in Fig 3f, the profile t approaches an asymptote as γ_4 increases. In Figs 3a, b and c, the contours are open because of ϕ_4, and in Figs 3b and d, because of γ_4. The parameters ϕ_2, ϕ_3, and γ_2 are all well behaved in these plots and in the other trace pair plots.) A future design should therefore be constructed so as to provide more information on ϕ_4 and γ_4 both individually and jointly, possibly using

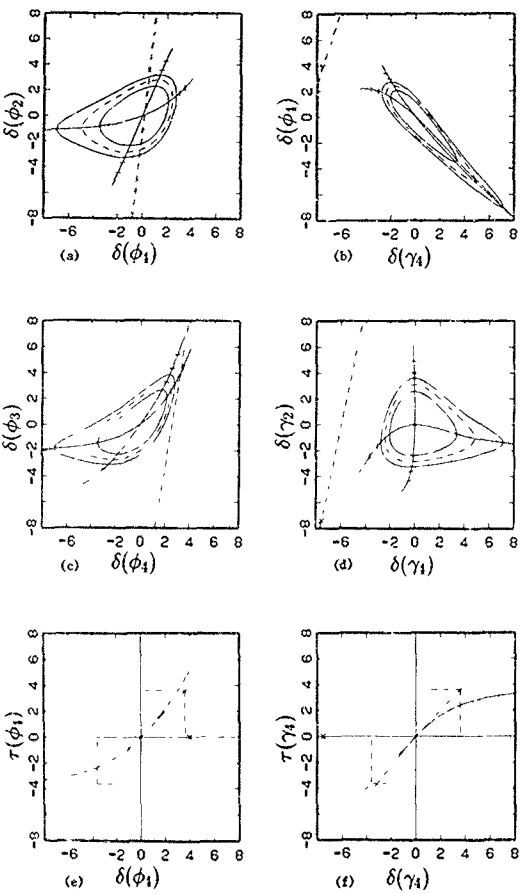


Fig 3 Selected profile t and trace plots for codimer hydrogenation, edited data, model d with $\phi 1 = 0$, $\gamma_1 = 0$, and $\gamma_3 = \gamma_2$ (a) ϕ_2 vs ϕ_4, (b) ϕ_4 vs γ_4, (c) γ_3 vs ϕ_4, (d) γ_2 vs ϕ_4, (e) profile t for ϕ_4, (f) profile t for γ_4 In (a)-(d), the solid and dashed closed curves denote the 60, 80, and 90% joint likelihood boundaries and the short bars on the traces indicate horizontal and vertical tangents of the 95 and 99% contours which are incompletely determined The dashed straight line is the line of equality of two parameters, and the ×, when present, indicates the point corresponding to θ_p, $\theta_q = 0, 0$ In (e) and (f), the solid line is the profile t function and the dashed line is the linear reference Dotted lines show nominal 60, 80, 90, 95, and 99% likelihood intervals. The × is the point corresponding to $\theta_p = 0$

subset designs as proposed by Box [6] and Hill and Hunter [7].

DISCUSSION

The profile plot approach to summarizing the inferential results of a statistical analysis has much to recommend it. The computations for the profile t and profile trace plots are very efficient because we start from excellent estimates based on the previous calculation, and because the problem is of reduced dimension $(P-1)$. Also, at each value of the profile parameter we simultaneously generate the profile t value and the converged values of the trace vector, which provides the data to make the profile pair plots. And for all the calculations, only minor modifications to standard software are necessary.

Profile plots provide important detailed information about the estimation situation. In addition to providing accurate marginal likelihood regions for each parameter, the profile t plots reveal how nonlinear each parameter is. Similarly, the profile trace plots and the associated likelihood contour sketches provide useful information on the pairwise behavior of the parameters. Superimposing the line of equality on the trace plots is a simple but extremely effective aid to model building. Perhaps more importantly, however, the plots collective provide insights into the experimental situation, so that steps can be taken to obtain more informative data [8].

Ratkowsky [9] has suggested rewriting rational model functions, such as in the codimer model, by factoring the numerator parameters into the denominator term. For example model d would become

$$r = \frac{x_1 x_2}{(\beta_1 + \beta_2 x_1 + \beta_3 x_2 + \beta_4 x_3)^2}$$

where $\beta_1 = 1/\sqrt{\theta_1 \theta_2 \theta_3}$, $\beta_2 = \sqrt{\theta_2/\theta_1 \theta_3}$, and so on. Profile plots for the βs are much better behaved than those for the θs, producing almost perfectly straight profile t plots and traces. One consequence of this is that marginal and joint linear approximation regions and summaries for the β parameters, are extremely accurate.

This illustrates a situation where linear approximation inferences for one set of parameters for a nonlinear regression model are much more accurate than for another set of parameters. However, the ease with which profile t and profile trace plots can be produced renders reparametrization considerably less important, since accurate marginal and joint likelihood regions can be obtained directly for the original parameters, which are usually more meaningful to the researcher.

For univariate reparametrization, say $\phi_p = g(\theta_p)$, the profile t plot and associated profile traces for ϕ_p can be obtained directly from the profile t plot and associated profile traces for θ_p: there is no need to reparametrize the model function or do any reestimation. This, of course, is a consequence of invariance of the likelihood function.

Profiling provides extremely valuable information for experimental design, as demonstrated in the codimer hydrogenation example. There it was clearly evident from the profile t and trace plots that further data was required to provide better information about ϕ_4 and γ_4. No such indication was evident from the linear summary statistics.

Finally, profiling can be applied to very general situations, including multiresponse estimation, as we have shown, and both univariate and multivariate time series analysis. The univariate situation has been discussed by Lam and Watts [10]. One can also use profiling to determine likelihood intervals for fitted values of the model function, by reparametrizing the model so that a new parameter, say ϕ_1, is equal to the fitted value at a specified design point.

REFERENCES

1 D.M. Bates and D.G. Watts, *Nonlinear Regression Analysis and Its Applications*, Wiley, New York, 1988

2 N.R. Draper and H. Smith, *Applied Regression Analysis*, Wiley, New York, 2nd ed., 1981

3 Z. Gross and S. Hoz, Radical-anionic nature of the transition state in the Michael addition reaction, *Journal of the American Chemical Society*, 110 (1988) 7489–7493

4 C I Bliss and A.T. James, Fitting the rectangular hyperbola, *Biometrics*, 22 (1966) 573–602
5 J L. Tschernitz, S. Bornstein, R.B. Beckman and O A Hougen, Determination of the kinetics mechanism of a catalytic reaction, *Transactions of the American Institute of Chemical Engineers*, 42 (1946) 883–905
6 M.J Box, An experimental design criterion for precise parameter estimation of a subset of the parameters in a nonlinear model, *Biometrika*, 58 (1971) 149–153
7 W J Hill and W.G. Hunter, Design of experiments for subsets of the parameters, *Technometrics*, 16 (1974) 425–434.
8 G.E.P Box and H.L Lucas, Design of experiments in nonlinear situations, *Biometrika*, 46 (1959) 77–90
9 D.A Ratkowsky, A statistically suitable general formulation for modelling catalytic chemical reactions, *Chemical Engineering Science*, 40 (9) (1985) 1623–1628
10 R.L.H Lam and D G Watts, Profile summaries for ARIMA time series model parameters, *Journal of Time Series Analysis*, in press.

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 117–122
Elsevier Science Publishers B V., Amsterdam

Diffusion in disordered media

Daniel ben-Avraham
Physics Department, Clarkson University, Potsdam, NY 13676 (U S.A)

(Received 8 November 1989, accepted 23 February 1990)

Abstract

ben-Avraham, D., 1991 Diffusion in disordered media *Chemometrics and Intelligent Laboratory Systems*, 10 117–122

Diffusion in disordered media is anomalous in that it does not follow the regular Fickian law of diffusion in homogeneous systems This has important implications for the physics of transport phenomena in disordered media Fractals and scaling theory have been particularly illuminating in this area of research An elementary exposition of anomalous diffusion in disordered media and its physical consequences, based on the concept of fractals, are presented

INTRODUCTION

Diffusion is among the most common phenomena in nature. One would find it relatively easy to provide with several examples of systems where diffusion plays a decisive role, in most areas of scientific research. In homogeneous, ordered media diffusion obeys Fick's law,

$$\langle R^2 \rangle \propto t \tag{1}$$

i.e., the mean square displacement of a diffusing particle increases proportionally to the time. This basic result is universal in that it applies whether diffusion takes place in one, two, or any dimension of regular Euclidean space [1]. We have become so much accustomed with this universality that the realization that Fick's law is violated for diffusion in disordered media came as a big surprise. In nonhomogeneous, disordered systems the diffusion law becomes anomalous [2,3],

$$\langle R^2 \rangle \propto t^{2/d_w} \tag{2}$$

with $d_w > 2$ This slowing down of the transport is caused by the delay of the diffusing particles in the dangling ends, bottlenecks and backbends existing in the disordered structure.

The concepts of fractals and fractal dimensionality have helped us understand better than ever before the physics of disordered systems such as porous earth, powders, amorphous materials, and aggregates. In this brief overview, we explain these concepts and how fractals are used to model disordered systems. We then show how diffusion is anomalous in disordered media and point at some of the physical consequences of this remarkable irregularity.

FRACTALS AND DISORDERED MEDIA

We begin with the definitions of the most basic properties of fractals [4]. Fractals are mathematical objects with a Hausdorff–Besicovitch dimension that is not an integer. They are most easily

0169-7439/91/$03.50

constructed in a recursive way. Thus, for example, the Koch curve (Fig. 1) is constructed by starting with a unit segment. The middle third section of this segment is erased and replaced by two other segments of equal length 1/3. Next, the same procedure is repeated for each of the four resulting segments (of length 1/3). This process is iterated ad infinitum. The limiting curve is of infinite length yet it is confined to a finite region of the plane. The best way to characterize it is by using its Hausdorff–Besicovitch or fractal dimension, d_f. In a Koch curve magnified by a factor of three there fit exactly four of the original curves. Therefore its fractal dimension is given by $3^{d_f} = 4$, or $d_f = \ln 4/\ln 3 = 1.262$. The fractal dimension is a generalization of the integer dimensions that we associate with regular objects of classical Euclidean geometry.

An important property of fractals which renders them particularly useful for the modelling of disordered media is their self-similarity. This can be seen by examining the Koch curve or the Koch snowflake, as it is frequently called. One can see a central object reminiscent of a snowman. To the right and to the left of this central snowman, there are two other snowmen, each being an exact reproduction only smaller by a factor of 1/3. Each of the smaller snowmen has in turn two still smaller copies of themselves to their right and left, etc.

In recent years, it has become clear that many disordered systems are best characterized by a symmetry of invariance under dilatation [5]. This fundamental symmetry is essentially the same as the self-similarity of fractals, only that disordered systems occurring in nature exhibit this self-similarity only in a statistical sense. For these objects a fractal dimension d_f is still easily defined by the scaling of their mass M with their linear size L

$$M \propto L^{d_f} \tag{3}$$

The Koch curve can serve as a model for a linear polymer chain. Likewise, the Sierpinski sponge of Fig. 2 is an obvious model for porous media. It is constructed by subdividing a cube into $3 \times 3 \times 3 = 27$ smaller cubes, and eliminating the central small cube and its six nearest neighbors. Each of the remaining 20 cubes is processed in the same way and the whole procedure is iterated indefinitely. Notice that the volume of the sponge is zero, while its surface area is infinite. This agrees intuitively with the fact that its fractal dimension $d_f = \ln 20/\ln 3 = 2.727$ lies in between $d = 2$ and $d = 3$. Fractals have been used to model an immense variety of disordered systems. Nature

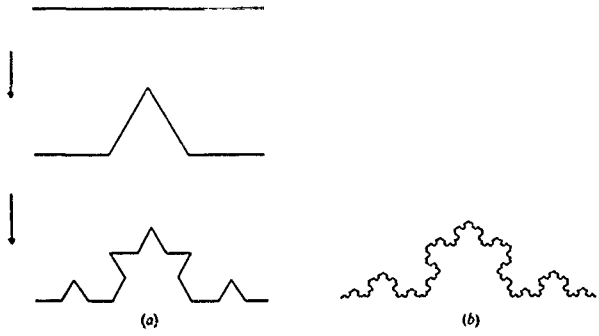


Fig. 1. Koch curve. (a) the iterative process by which it is constructed, (b) self-similarity – the central snowman is surroundered by two exact copies of itself.

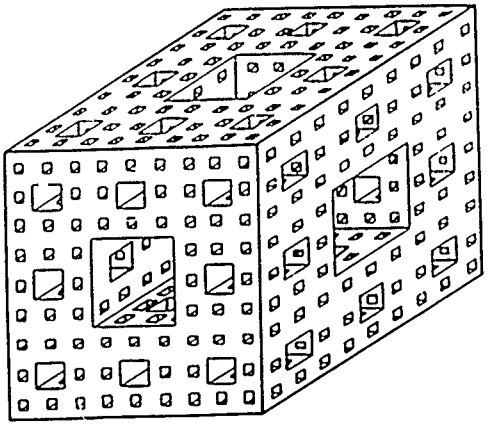


Fig 2 The Sierpinski sponge

abounds with examples of self-similar objects. This has been made clear by several excellent books published in recent years which helped popularize fractals.

ANOMALOUS DIFFUSION

It is convenient to refer to a simple random walk as a model for diffusion [1]. In a simple discrete random walk the walker advances one step in a unit time. Each step is taken with equal probabilities to any of the nearest neighbors of the present site. Denote the steps of such a walker by $u_1, u_2, \ldots, u_t$. Then, the mean square displacement at time t, $\langle R^2(t)\rangle$, is given by

$$\langle R^2(t)\rangle = \left\langle \left(\sum_{i=1}^{t} u_i \right)^2 \right\rangle = t + 2 \sum_{i>j}^{t} \langle u_i \cdot u_j \rangle \qquad (4)$$

For regular lattices the correlations $\langle u_i \cdot u_j \rangle$ are all zero. Thus, in homogeneous systems one has the usual result for normal diffusion that $\langle R^2(t)\rangle = t$. Disordered systems are characterized by irregular lattices. The nearest neighbors of a site are not distributed symmetrically and the correlations $\langle u_i \cdot u_j \rangle$ are *not* zero. This may lead to anomalous diffusion.

Interestingly, a random walk itself is a statistically self-similar object. To see this, consider the random walk as it looks when one regards n consecutive steps as one single 'superstep'. Each of the supersteps is a random jump r on the lattice. The random supersteps are distributed according to some probability $P_n(r)$ with a finite moment $\langle r^2\rangle = n$. In the limit $n \gg 1$, $P_n(r)$ tends to a Gaussian distribution. This is a simple result of the central limit theorem. It is evident that statistically the same random walk results for different values of n. The only difference between walks with $n = n_1$ and $n = n_2$ is that in the first case a step is performed every n_1 time units and every n_2 time units in the latter. Also, the average length of a step is $n_1^{1/2}$ and $n_2^{1/2}$ respectively, for the different walks. This means that if we scale time as $t \to \lambda t$ and length as $r \to \lambda^{1/2} r$ then two walks with $n_1 = \lambda n_2$ are exactly equivalent under this scaling. Hence, the simple random walk is statistically self-similar. In fact it is a statistical fractal. Upon dilation of space by a factor of $\lambda^{1/2}$,

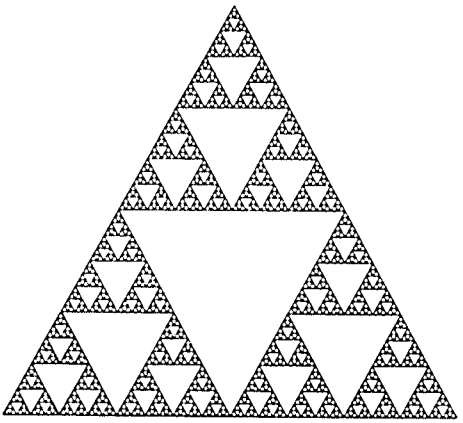


Fig 3 A Sierpinski gasket drawn to the sixth generation

the number of steps (or 'mass' of the walk) increases by a factor of λ. Therefore the fractal dimension of a random walk is $d_w = \ln \lambda / \ln \lambda^{1/2} = 2$. It is interesting that random walks performed on disordered, but statistically self-similar structures are still self-similar themselves, exactly as for regular lattices. The important difference is that the usual diffusion exponent, $d_w = 2$, is no longer equal to 2. Diffusion is anomalous.

We will now illustrate anomalous diffusion by considering a random walk on the Sierpinski gasket of Fig. 3. The Sierpinski gasket is perhaps the most widely used fractal lattice for theoretical applications. This is because of the fact that it is a finitely ramified fractal, i.e., one needs cut only a finite number of bonds to isolate a subset of the gasket. This property facilitates the exact analysis of various physical models, including the random walk problem.

At each step the walker chooses randomly to move to one of the four nearest neighbors on the gasket. As stated above, we expect the walk to be statistically self-similar. The mean square displacement would grow as $\langle R^2 \rangle \propto t^{2/d_w}$, where d_w is the anomalous diffusion exponent. Note that d_w is in fact the fractal dimensionality of the path of the random walker on the lattice. In Fig. 4 we show a plot of ln (t) against ln $\sqrt{\langle R^2 \rangle}$ as obtained from an exact enumeration of all possible walks. The slope of the resulting curve is $d_w = 2.32 \pm$

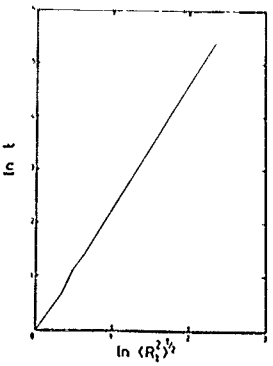


Fig. 4 Plot of $\ln(t)$ as a function of $\ln\sqrt{\langle R^2(t) \rangle}$ on a Sierpinski gasket, using exact enumeration of the walks. The slope is $d_w = 2.32 \pm 0.01$

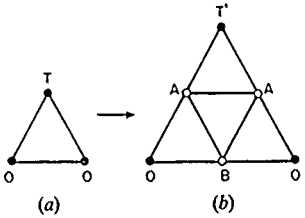


Fig 5 Rescaling of first passage time for exiting the gasket The walker enters the gasket at the top vertex and (a) takes a time T to exit through the lower O-vertices. (b) The rescaled gasket, $T \to T'$ and A and B are exit times from the internal (decimated) vertices to the lower O-vertices

0 01. This shows clearly the anomaly of diffusion on fractal lattices.

One can exploit the finite ramification of the Sierpinski gasket to obtain an exact value of the exponent d_w in the following way Consider the mean first passage time T to traverse a gasket unit from one of its vertices to either of the remaining two vertices O (Fig 5a) One can then calculate the corresponding mean first passage time T' for exiting a rescaled gasket unit by a factor of 2 (Fig. 5b) This is done by making use of the Markov property of the random walk. Thus, T' equals the time T to exit the first gasket unit, plus A, the mean first passage time to leave the rescaled unit from then on. Using the same reasoning for the times A and B (the mean exit times starting from the decimated internal vertices), one has

$$\begin{aligned} T' &= T + A \\ 4A &= 4T + A + B + T' \\ 4B &= 4T + 2A \end{aligned} \tag{5}$$

The solution is $T' = 5T$ (and $A = 4T$, $B = 3T$), which is the rescaling of time for a diffusion process on the gasket upon the rescaling of length by a factor of 2. Hence, $d_w = \ln 5/\ln 2 = 2\,322$ Notice the agreement with the result obtained from exact enumeration. This anomalous diffusion is characteristic of all fractal lattices, as well as of statistically self-similar objects such as percolation cluster and aggregates [6].

ANOMALOUS TRANSPORT PHYSICS

Diffusion is closely related to transport physics. Anomalous diffusion results in anomalous transport physics An excellent example is the relation between diffusion and conductivity of a medium. In homogeneous systems it is given by the Einstein relation

$$\sigma_{dc} = \frac{e^2 n}{k_B T} D \tag{6}$$

where σ_{dc} is the d.c. conductivity, n is the carrier density and D is the diffusion constant

$$D = \langle R^2 \rangle / t \quad t \gg 1 \tag{7}$$

The carrier density n is proportional to the mass density of the bulk. For fractal substrata, this scales as $n \propto R^{d_f - d}$. The conductivity exponent $\tilde{\mu}$ is defined by its scaling with the linear size R, $\sigma_{dc} \propto R^{-\tilde{\mu}}$ From eqs. (2) and (7), $D \propto t^{1/d_w - 1}$ and using it in the Einstein relation of eq (6) we get $t \propto R^{2 - d + \tilde{\mu} + d_f}$ Comparing this to eq (2) we obtain the relation

$$d_w = 2 - d + d_f + \tilde{\mu} \tag{8}$$

This is to be compared to the classical conductivity exponent $\tilde{\mu} = 0$ of homogeneous media (for which $d_f = d$ and $d_w = 2$), showing the anomalous conductivity that results because of anomalous diffusion.

A more fundamental consequence of anomalous diffusion arises when one looks at the density of states in a disordered substrate. The density of states is relevant for any physical phenomenon which is described by an equation of motion that contains the operator ∇^2 This includes, for example, electromagnetic, elastic, and quantum phenomena. The density of states $\rho(\epsilon)$ is related to diffusion through $P(0,t)$, the probability of a walker to be back at the origin at time t:

$$P(0,t) = \int_0^\infty \rho(\epsilon) \exp(-\epsilon t)\, d\epsilon \tag{9}$$

By the time t, a random walker has visited the sites within a volume $R(t)^{d_f} \propto t^{d_f/d_w}$. Therefore the probability of returning to the origin scales as $1/R^{d_f} \propto t^{-d_f/d_w}$. Using this result in eq (9), one finds

$$\rho(\epsilon) \propto \epsilon^{d_f/d_w - 1} = \epsilon^{d_s/2 - 1} \tag{10}$$

where d_s is the spectral [7], or fracton [8] dimensionality for the density of states. This is similar to the usual expression for homogeneous media, $\rho(\epsilon) \propto \epsilon^{d/2-1}$, except that d is replaced by the anomalous $d_s = 2d_f/d_w$.

As a final example of the physical consequences of anomalous diffusion we would like to mention diffusion–reaction systems in contrived geometries. It is well known that the reaction rate in diffusion-limited reactions is proportional to the volume covered by a diffusing reactant particle per unit time (this is known as the Wiener sausage problem). Clearly, this is critically affected by the irregularities of diffusion when the substrate is a statistical fractal. This intriguing topic is discussed in detail in the paper by Kopelman et al. [9].

SUMMARY

We have presented an elementary discussion of the basic properties of fractals and how fractals are useful for the modelling of disordered media. Diffusion in disordered media was shown to be anomalous in that rather than following Fick's law $\langle R^2 \rangle \propto t$, it obeys $\langle R^2 \rangle \propto t^{2/d_w}$, where d_w is the anomalous diffusion exponent and is dependent upon the specific characteristics of the substrate in question. We then discussed some of the dramatic consequences of anomalous diffusion, as manifested in bulk conductivity, the density of states, and reaction rates in diffusion–reaction systems. The interested reader is referred to more complete reviews and to the specialized literature of the field.

REFERENCES

1 For a review of regular diffusion see G.H. Weiss and R.J. Rubin, Random walk theory and application, *Advances in Chemical Physics*, 52 (1983) 363–505.

2 S. Havlin and D. ben-Avraham, Diffusion in disordered media, *Advances in Physics*, 36 (1987) 695–798. This and ref. 3 are complimentary review papers. Refer to the references in these reviews.

3 J.W. Haus and K.W. Kehr, Diffusion in regular and disordered lattices, *Physics Reports*, 150 (1987) 263–416.

4 See, for example, B.B. Mandelbrot, *Fractals: Form, Chance and Dimension*, Freeman, San Francisco, CA, 1977; *The Fractal Geometry of Nature*, Freeman, San Francisco, CA, 1982.

5 L. Pietronero and E. Tosatti (Editors), *Fractals in Physics*, North Holland, Amsterdam, 1986.

6 F. Family, *Chemometrics and Intelligent Laboratory Systems*, in press.

7 R. Rammal and G. Toulousse, *Journal de Physique, Lettres (Orsay, France)*, 44 (1983) L13

8 S Alexander and R. Orbach, *Journal de Physique, Lettres (Orsay, France)*, 43 (1982) L625

9 R. Kopelman, L.W. Anacker, E. Clement, L. Li and L. Sander, Low dimensional reaction kinetics and self-organization, *Chemometrics and Intelligent Laboratory Systems*, 10 (1991) 127–132.

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 123–126
Elsevier Science Publishers B.V., Amsterdam

Discussion of "Diffusion in disordered media" by Daniel ben-Avraham

George H. Weiss
National Institutes of Health, Bethesda, MD 20892 (U.S.A.)

Professor ben-Avraham, in his lucid article, has indicated some of the simplest characterizations of transport in a disordered medium. What makes the general analysis of such problems so difficult is that the characteristic function cannot easily be used to generate explicit representations of the solution to problems in which the medium is not translationally invariant. Nevertheless, because phenomena related to disordered media arise naturally in a variety of scientific fields the general area of diffusion in such media has become one of central interest in contemporary chemistry, mathematics, and physics. A sampling of some of the many applications of the theory is to be found in the a review by Alexander et al. [1], a proceedings of a meeting edited by Klafter et al. [2]. Excellent more comprehensive reviews of the subject have been given by Haus and Kehr [3], and by Havlin and ben-Avraham [4].

Since one cannot, in general, find solutions to the equations describing transport in a disordered medium, how does one go about calculating some of the properties of anomalous diffusion? Naturally, in a field which has been so widely studied, a great many theoretical techniques have been tried, most of which lead to approximations to a solution rather than explicit solutions. While a precise definition of the term "explicit solution" may contain some ambiguity, the only nontrivial model of a disordered medium for which all of the interesting transport properties are basically known is one originally suggested by Sinai [5]. The exact solution is due to Kesten [6]. Sinai's model is that of a random walk on a one-dimensional lattice in which, on a given step, the random walker can move from site i to $i+1$ with probability p_i or to $i-1$ with probability $1-p_i$. The p_i are assumed to be independent, identically distributed random variables which satisfy the conditions

$$E\left\{\ln\frac{p_i}{1-p_i}\right\}=0, \qquad E\left\{\ln^2\frac{p_i}{1-p_i}\right\}=\sigma^2<\infty \tag{1}$$

Let X_n be the location of the random walker at step n. Kesten shows that the random variable $\sigma^2 X_n/\ln^2 n$ converges in distribution, and finds an explicit representation of the distribution as an infinite series. Unlike the examples cited by ben-Avraham, the mean-squared displacement of the random walk satisfies

$$\frac{E\{X_n^2\}}{\ln^2 n}\to\text{constant} \tag{2}$$

as $n\to\infty$. There are many obvious generalizations of this model for which one would want to see a solution, but for which there are no exact results available either in the literature of mathematics or physics. For example, it would be most desirable to extend these results for the Sinai model to analogous random walks in higher dimensions, in addition to removing the restrictions on possible steps of the random walks that they be to nearest neighbors only. Another useful generalization is that of obtaining an exact solution of the corresponding first passage problems for such random walks in the presence of absorbing boundaries.

Related material by Solomon for the one-dimensional case has been presented in the mathematical literature [7], and a more heuristic approach has been taken in the physics literature to find the asymptotic survival probability for a Sinai random walk on a finite line bounded by traps at either end [8]. Clearly, it would be most useful to have further models for transport in a random environment that can be solved exactly, if only because most analyses of such problems that have appeared are approximate and one always likes to have a benchmark for purposes of comparison.

In the absence of general methods for solving problems of transport in disordered media investigators have resorted to a large number of both approximate (which start from a rigorous formulation of the dynamics) and heuristic techniques which enable one to understand the dynamics of such processes. We will mention just two of these because of their popularity, although not necessarily their accuracy, in any given problem. The first rather general method goes under the heading of the effective medium approximation, although there are many variants in the literature. To see the basic ideas behind this method in the context of a grossly simplified model, let us consider a lattice random walk on a line in which the random walker moves in one direction only, which we choose to be the positive x direction. Let k_i be the rate constant for the random walker to move from i to $i+1$, and assume that the random walker is initially at $i=0$. We will assume that the k_i are identically distributed independent random variables. Let $p_n(t)$ be the probability that the random walker is at n at time t. These probabilities satisfy the equations

$$\begin{aligned}\dot{p}_0(t) &= -k_0 p_0(t)\\ \dot{p}_n(t) &= k_{n-1}p_{n-1}(t) - k_n p_n(t), \quad n = 1,2,3,\ldots\end{aligned} \tag{3}$$

While these equations are readily solved exactly, I will use them to illustrate the basic ideas behind the effective medium approximation as well as a number of related techniques which have been used in solid state physics. In the context of the present problem, one assumes that there are a set of probabilities, $\{q_n(t)\}$, which approximate to the solution to eq. (3). These are taken to be the solution to the coupled set of equations

$$\begin{aligned}\dot{q}_0(t) &= -\int_0^t K(t-\tau)q_0(\tau)\,d\tau\\ \dot{q}_n(t) &= \int_0^t K(t-\tau)\left[q_{n-1}(\tau) - q_n(\tau)\right]d\tau,\\ n &= 1,2,3,\ldots\end{aligned} \tag{4}$$

Thus, the Markovian equations in eq. (3) are to be replaced by the coupled set of non-Markovian equations in eq. (4) in terms of an as yet undetermined kernel, $K(t)$. What we observe in the formulation of eq. (4) is that the approximating random walk takes place on a line whose properties are translationally invariant. The crucial step in the effective medium approximation is a technique for calculating the kernel $K(t)$ in terms of properties of the k_i.

A formal solution to eq. (4) is readily found. Introduce the Laplace transforms $\hat{q}_n(s)$ and $\hat{K}(s)$ by

$$\begin{aligned}\hat{q}_n(s) &= \int_0^\infty e^{-st} q_n(t)\,dt\\ \hat{K}(s) &= \int_0^\infty e^{-st} K(t)\,dt\end{aligned} \tag{5}$$

One readily verifies that the Laplace transform of the solution to the system of equations in eq. (4) is

$$\hat{q}_n(s) = \hat{K}^n(s)/\left[s+\hat{K}(s)\right]^{n+1} \tag{6}$$

In order to find an expression for $\hat{K}(s)$ we replace the original formulation given in eq (3) by a model in which only a single rate is random (it doesn't matter which one) while the remainder of the medium is regarded as having the properties of the effective medium defined in eq (4). Let k_j be the single random rate constant, and let $p_{n,e}(t)$ be the probability in this modified model, the random walker is at n at time t. The $p_{n,e}(t)$ satisfy the set of equation in eq. (4) with the exception of the indices j and $j+1$ for which the equations become

$$\begin{aligned}\dot{p}_{j,e}(t) &= \int_0^t K(t-\tau)p_{j-1,e}(\tau)\,d\tau - k_j p_{j,e}(t)\\ \dot{p}_{j+1,e}(t) &= k_j p_{j,e}(t) - \int_0^t K(t-\tau)p_{j+1,e}(\tau)\,d\tau\end{aligned} \tag{7}$$

The Laplace transform of the kernel $K(t)$ is then found from the solution to the transform of the set of equations in eq. (7) by requiring that the expectation of the solution to the modified system be equal to the solution for the state probabilities in the effective medium, i.e.:

$$q_n(t) = E\{p_{e,n}(t)\} \tag{8}$$

A solution for the Laplace transforms $\hat{p}_{j,e}(s)$ and $\hat{p}_{j+1,e}(s)$ is readily calculated from the combination of eqs. (4) and (7) to be

$$\hat{p}_{j,e}(s) = \frac{\{\hat{K}(s)\}^j}{(s+\hat{K}(s))^j(s+k_j)}$$
$$\hat{p}_{j+1,e}(s) = \frac{k_j\hat{K}^j(s)}{(s+k_j)[s+\hat{K}(s)]^{j+1}} \tag{9}$$

On making use of the requirement in eq. (8) we find that $\hat{K}(s)$ is the solution to

$$\frac{1}{s+\hat{K}(s)} = E\left\{\frac{1}{s+k}\right\} \tag{10}$$

where we have omitted the subscript on k because of our assumption that the random rate constants are identically distributed. It is easy to confirm that the $\hat{q}_n(s)$ can be expressed as

$$\hat{q}_n(s) = E^n\left(\frac{k}{s+k}\right)E\left(\frac{1}{s+k}\right) \tag{11}$$

which implies that the crucial quantity for our model is the expectation $E[k/(s+k)]$ or, equivalently, $E[1/(s+k)]$

In the present completely trivial model it is possible to show that eq. (11) is equivalent to the result found by taking the expectation of the exact solution of eq. (3). This solution is

$$\hat{p}_n(s) = \frac{k_0k_1k_2\cdots k_{n-1}}{(s+k_0)(s+k_1)\cdots(s+k_n)} \tag{12}$$

The identification of exact and approximate solutions is not readily demonstrated for more general models, and in fact the solution to the analogue of eq (10) generally requires the solution to a transcendental equation [3] What this means in practice is that one is practically limited to the calculation of properties in the limit $s \to 0$, or equivalently, in the limit $t \to \infty$ A discussion of the errors incurred in the use of the effective medium approximation in the context of a simple one-dimensional example is contained in the review by Haus and Kehr [3]. One of the attractive features of the effective medium approximation is that it is no harder to treat problems in three dimensions than it is for one-dimensional problems and the accuracy of the approximation generally increases as the number of dimensions increases. This is not true for a number of other techniques that have been applied to this general class of problems (e.g., the renormalization group approach suggested in ref. 9 which is restricted to one dimension only).

Finally, we mention a complete phenomenological approach that has been successfully applied to problems of the transport of carriers in amorphous semiconductors [10,11], as well as to models for chromatographic kinetics [12] In the first of these applications the transport is generally non-diffusive, while in the second it may or may not be diffusive. The model on which the analyses are based is known as the continuous-time random walk (CTRW) in the literature of physics and physical chemistry [13,14] This class of models is based on the simplest picture of a random walk in which the displacement on a given step and the time between successive steps of the walk are both assumed to be identically distributed independent random variables The space and time variables are often assumed to be uncorrelated so that the probability (or probability density) for the displacement r, that follows an interstep time t can be written in factorized form as

$$f(r,t) = p(r)\psi(t) \tag{13}$$

Only in the case in which $\psi(t) = k\exp(-kt)$ is the resulting process Markoffian However, it is known that provided that the first moment of $\psi(t)$ is finite and the variance-covariance matrix for the displacement consists of finite elements, the asymptotic properties of the random walk in an infinite medium will be those calculated by means of the central limit theorem, which is equivalent to ordinary diffusion [14].

The principal idea put suggested by Scher and Lax is that a detailed description of randomness in a medium can be replaced by the randomness

inherent in the pausing time density $\psi(t)$ appearing in eq. (13). This *ansatz* cannot be justified in any detailed way although it has been justified in the calculation of one quantity of physical interest by Klafter and Silbey [15]), but the consequences of the theoretical development have been shown in a number of studies, to yield results in good agreement with experimental data [16]. The key assumption in many of these calculation is that $\psi(t)$ behaves asymptotically as

$$\psi(t) \sim T^{\alpha}/t^{\alpha+1} \tag{14}$$

where $0 < \alpha \leqslant 1$, so that the first moment of the interstep time is infinite. Symmetric CTRWs which have the property that single-step displacement probabilities have a finite expectation as well as the properties in eqs. (13) and (14) can be shown to have the asymptotic property

$$E(r^2) \sim (t/T)^{\alpha} \tag{15}$$

where $E(r^2)$ is the expected value of the mean-squared displacement. When eq. (14) is valid the asymptotic form of the probability density of the displacement at time t will also differ from the Gaussian form that holds in ordinary diffusion.

Some of the properties and many of the applications of CTRWs satisfying eq. (14) have recently been reviewed by Shlesinger [17] and some of the mathematical properties of CTRWs based on the assumption of eq. (14) are given in ref. 18. It must be said that, because arguments based on CTRW models can only be characterized as having a hand-waving character, the applicability of such models to transport in disordered media can only be ascertained on a case-by-case basis, and it is by no means clear that CTRW approximations always lead to useful results.

I have presented only a small sample of an enormous number of different approaches taken in the literature of physics and chemistry to the general problem of transport in a disordered medium. The development of techniques for the solution, complete or partial, of these problems is a particularly active area of research in physical chemistry, physics, and in probability theory. Some of the basic phenomena that warrant investigation have been very ably outlined in the article by Professor ben-Avraham.

REFERENCES

1 S Alexander, J. Bernasconi, W R. Schneider and R. Orbach, Excitation dynamics in random one-dimensional systems, *Reviews in Modern Physics*, 53 (1981) 175–198
2 J. Klafter, R.J Rubin and M F Shlesinger (Editors), *Transport and Relaxation in Random Materials*, World Scientific Publishers, Singapore, 1986
3 J W. Haus and K.W Kehr, Diffusion in regular and disordered lattices, *Physics Reports*, 150 (1987) 263–406
4 S. Havlin and D ben-Avraham, Diffusion in disordered media, *Advances in Physics* 36 (1987) 695–798
5 Ya G. Sinai, Lorentz gas and random walks, in J Ehlers, K Hepp, R. Kippenhahn, H A Weidenmuller and J Zittartz (Editors), *Mathematical Problems in Theoretical Physics*, Springer-Verlag, Berlin, 1982, pp 12–14, *Theory of Probability and Applications*, 27 (1982) 256
6 H. Kesten, The limit distribution of Sinai's random walk in random environment, *Physica*, 138A (1986) 299–309
7 F Solomon, Random walks in a random environment, *Annals of Probability*, 3 (1975) 1–31
8 S Havlin, J E. Kiefer and G H Weiss, The trapping problem on a line with dichotomous disorder, *Physics Review*, B38 (1988) 4761–4764
9 J Machta, Renormalization group approach to random walks on disordered lattices, *Journal of Statistical Physics*, 30 (1983) 305–314
10 H Scher and M. Lax, Stochastic transport in a disordered-solid I Theory *Physics Review*, B7 (1973) 4491–4502, II Impurity conduction, 5402–4519
11 H Scher and E W Montroll, Anomalous transit-time dispersion in amorphous solids, *Physics*, B12 (1975) 2455–2477
12 G H Weiss, Chromatographic kinetics and the phenomenon of tailing, *Separation Science*, 17 (1982) 1609–1622, see ref 2 On a generalized transport equation for chromatographic systems pp 394–406
13 E.W Montroll and G H Weiss, Random walks on lattices II, *Journal of Mathematical Physics*, 6 (1965) 167–180
14 G H Weiss and R J Rubin, Random walks and selected applications, *Advances in Chemical Physics*, 52 (1983) 363–505
15 J Klafter and R Silbey, Derivation of the continuous-time random-walk equation, *Physics Review Letters*, 44 (1980) 55–58
16 G Pfister and H Scher, Dispersive (non-Gaussian) transient transport in disordered solids, *Advances in Physics*, 27 (1978) 747–798
17 M.F Shlesinger, Fractal time in condensed matter, *Annual Reviews of Physical Chemistry*, 39 (1988) 269–290
18 H Weissman, G H Weiss and S Havlin, Transport properties of the CTRW with a long-tailed waiting time, *Journal of Statistical Physics*, 57 (1989) 301–317

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 127–132
Elsevier Science Publishers B.V., Amsterdam

Low dimensional reaction kinetics and self-organization

R Kopelman *, L W. Anacker, E Clement, L. Li and L. Sander
Departments of Chemistry and Physics, The University of Michigan, Ann Arbor, MI 48109 (U S A)

(Received 8 November 1989, accepted 23 February 1990)

Abstract

Kopelman, R., Anacker, L.W., Clement, E., Li, L and Sander, L., 1991 Low dimensional reaction kinetics and self-organization *Chemometrics and Intelligent Laboratory Systems*, 10 127–132

Diffusion-limited reaction kinetics becomes anomalous not only for fractals, with their anomalous diffusion, but also for low-dimensional (one and two) and disperse media, where the random walk is compact We focus on annihilation, recombination and trapping reactions under non-equilibrium steady state (steady source) or batch (big bang) conditions The typical reactions are $A + A \rightarrow$ Products, $A + B \rightarrow$ Products and $A + C \rightarrow$ Products We are interested in the global rate laws, and their relation to particle–particle distributions (e g , pair-correlation and nearest-neighbor distribution functions) and in local rate laws (if definable) Anomalous reaction kinetics (more than classical kinetics) is particularly sensitive to initial conditions, source term structure, conservation laws (e g , equal densities for A and B), excluded volume effects, and medium size, dimensionality and anisotropy Analytical formalisms, scaling arguments, computer (and supercomputer) simulations and experiments (on chemical and physical reactions) all play an important role in the newly emerging picture

INTRODUCTION

This work can be viewed as a natural extension of the activity dealing with relaxation phenomena and transient kinetics problems in disordered media [1–4]. Its domain of application spans various areas of the physics and chemistry of condensed matter. For example, reactions of the type $A + A \rightarrow 0$ or $A + T \rightarrow T$ are models describing exciton kinetics in disordered molecular crystals or polymer blends. Reactions of the type $A + B \rightarrow 0$ are found in solid state physics in the case of electron–hole annihilation or defect fusion. A combination of experiments and Monte-Carlo simulations [5] has paved the way for a new theoretical understanding of steady-state rate laws and the kinetic self-organization of atoms, defects and elementary excitations in low dimensional media This theory is presented below.

Diffusion limited trapping is of particular interest in studies of energy migration and luminescence [1,5]. We present below some new simulations and their relation to theory. This includes both rate laws and self-ordering. Of particular interest is the resulting anomalously high partial order of reactions as a function of trap concentration.

'Big-bang' reaction models are simpler than steady-source models. The pioneering work has been done by Ovchinnikov and Zeldovich [6] and

0169-7439/91/$03 50

by Toussaint and Wilczek [7], with applications to fractals by Klafter et al. [8], Kang and Redner [9] and Klymko and Kopelman [10]. However, these ignored both finite size effects and finite correlation effects (at time zero). We demonstrate here that these finite extent effects give rise to new scaling effects, i e, anomalous time exponents and reaction orders. In particular, for the A + B reaction in one-dimension the time exponent rises from 1/4 (Zeldovich value) to 3/4 or 1 (depending on boundary conditions).

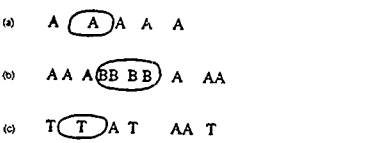


Fig 1 Schematic representation of the three cases of self organization on a one-dimensional system The circled domains represented here are of the order of Λ, the self organization scale. 1(a) is a depletion in the A + A → 0 case 1(b) is a segregation in the A + B → 0 case 1(c) is a trap-particle depletion in the A + T → T case

THEORY STEADY-STATE DIFFUSION CONTROLLED BIMOLECULAR REACTIONS

In the classical picture, all bimolecular reactions are the same and the distribution of reactants is at random Also, the reaction rate is proportional to the product of the reactant densities (overall order of reaction $X = 2$). Previous works show that the time dependence of such reactive systems, relaxing from an initial random situation, exhibits anomalous decay rates in low dimensions due to local fluctuations in reactant density [6–9]. Here we report the results of a theoretical investigation on the steady state properties of three different bimolecular diffusion limited reactions, taking place on regular Euclidean spaces and on fractal structures [11–13]. We show that the relevant parameter describing the steady state of the reaction kinetics is the spectral dimension d_s. The spectral dimension is an intrinsic parameter characterizing energy transfer properties, and in particular, diffusion in a medium. For Euclidean structures, d_s is the Euclidean dimension d, and the case of Euclidean spaces is viewed as an extension of the fractal case when we take $d = d_s$. The reason for the influence of the spectral dimension on reaction kinetics is due to the fact that d_s controls the time dependence of the number of distinct sites visited by a random walker. For spectral dimension $d_s < 2$ we show that a bimolecular reaction induces a self-organization of reactants up to a scale Λ such that:

$$\Lambda \approx \tau^{1-d_s/2} \qquad d_s < 2 \tag{1}$$

where τ is a characteristic time which is situation dependent. For $d_s > 2$, Λ is microscopic and independent of τ, therefore no large scale structure exists and the reaction kinetics is classical. The case $d_s = 2$ is found to be the critical dimension of the problem, where we find a marginal logarithmic dependence of Λ with τ. Below the critical dimension, large scale density fluctuations become relevant and each situation has its own phenomenology (see Fig. 1) In particular, we may find macroscopic reaction laws with anomalous reaction orders (larger than 2) or anomalous rate constants In all the cases investigated we found that the scaling behavior of the self organization length can be case in an interesting general way For every dimension we can write:

$$\Lambda/a \approx S_t/V_t$$

where a is the microscopic scale, S_t is the volume effectively explored by a particle during the time t (num. er of distinct sites visited) and V_t is the total (cumulative) volume swept out (proportional to τ).

In bimolecular diffusion limited processes the overall balance between reaction rates and steady state densities is accounted for by the Smoluchowski boundary condition:

$$Q \approx (\rho_1 \rho_2)/\Lambda$$

where ρ_1 and ρ_2 are, respectively, the steady state densities of reactants 1 and 2 (1 and 2 can be identical species). The scaling dependence of the self-organization scale Λ on τ is at the origin of the non-classical behavior.

In the case of homomolecular annihilation, $A + A \rightarrow 0$, Λ is a typical scale of depletion around each reactant and τ is the typical reactant life-time with:

$$\tau = \rho/Q$$

where ρ is the steady state density of A. We obtain an anomalous effective reaction order:

$$X = 1 + 2/d_s \qquad d_s < 2 \tag{2}$$

In the case of heteromolecular annihilation, $A + B \rightarrow 0$, λ is the scale of a self-organization phenomenon called segregation. At steady state, domains of identical species with sizes comparable to λ build up in the medium. The situation is more complex than in the homomolecular reaction case and τ is found to be dependent either on source conditions or on some intrinsic particle life time. We separated the source terms into two main categories. In the first category we consider sources for which at any time an identical number of As and Bs is conserved in the medium. If reactants are created at random, we find:

$$\tau \approx L^2$$

where L is the system size. We observe a size dependent segregation. With the same conservation constraint, if the particles are created as A–B pairs with A and B separated by a distance δ, we have.

$$\tau \approx \delta^2$$

The segregation scale becomes dependent on δ. It is important to notice that for geminate creation, we obtain a microscopic segregation scale and this situation becomes analogous to classical kinetics. In the second category, we consider sources where the conservation constraint is removed. If no other decay mechanism is present, fluctuations in particle difference grow until we have a complete saturation of the loop with one of the species. There is no reactive steady state. If an extra (first order) decay mechanism is considered, fluctuations grow up to a size defined by the intrinsic lifetime of the decay mechanism. In particular if we consider vertical annihilation with an external rate of particles R we have:

$$\tau \approx R^{-1}$$

In this case we obtain at low density an effective reaction order:

$$X = 4/d_s$$

On the other hand, if the decay is controlled by an intrinsic mechanism $A \rightarrow 0$ and $B \rightarrow 0$, with the same rate constant K, then we have

$$\tau \approx K^{-1}$$

We induce a K dependent segregation but no anomalous reaction order. These last three cases are important for practical applications because, besides geminate particle creation, it is difficult to find a source satisfying the exact conservation constraint. However, though the conservation is not exact, these cases lead to a mesoscopic segregation (or a total saturation).

For the trapping problem, $A + T \rightarrow T$, the fluctuation of the trap distribution is found to be unimportant for the leading scaling behavior of the self organization length Λ. The relevant fact is that we have, for $d_s < 2$, an organization of particles A around the traps. The typical lifetime at steady state is

$$\tau \approx \rho/Q$$

with ρ the density of A and Q the reaction rate. The scale of the trap–particle organization is

$$\Lambda \approx c^{1-2/d_s}$$

where c is the trap concentration. We have the anomalous rate law.

$$Q \approx \rho c^{2/d_s} \tag{3}$$

with an anomalous order relatively to the trap concentration.

$$X = 2/d_s$$

and we note that the overall reaction order is $1 + 2/d_2$, the same as for the $A + A \rightarrow 0$ case.

SIMULATIONS OF STEADY-STATE TRAPPING

We tested the trapping eq. (3). The Monte-Carlo simulations at the John von Neumann National Supercomputer Center give, for the Sierpinski gasket, a partial order $Y = 1.02 \pm 0.02$, with respect to the particle density ρ, and a partial order

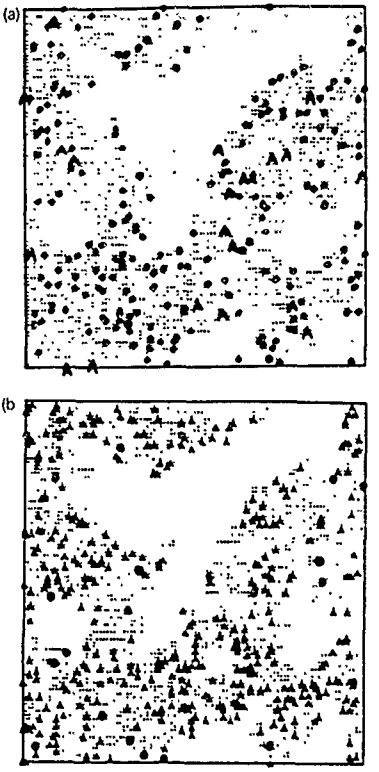


Fig 2 Distribution of reactants for two different trap concentrations on a percolation cluster at criticality The traps are the black circles On Fig 2a the trap concentration is 0 05 On Fig. 2b the trap concentration is 0 005

$X = 1.47 \pm 0\,02$, with respect to the trap density c. This is in excellent agreement with the predictions of eq. (2): $Y = 1$ and $X = 2/d_s = 1.465$ ($d_s = 1.365$). Similarly, the simulations for the critical percolation cluster are in excellent agreement with the eq. (2) predictions $Y = 1$ and $X = 2/d_s = 1.5$ ($d_s = 4/3$). In addition, the depletion zones around the traps can be seen qualitatively in Fig. 2.

SIMULATIONS OF A TRANSIENT A+A → 0 AND A+B → 0

We have employed three types of landing relationships: correlated, random and evenly spaced landings. When a particle is added to a site occupied by another particle, the landing particle may immediately try to land on another empty site, which is called 'forced landing'. Particles randomly move on a lattice

Correlated landing occurs when a pair of particles lands simultaneously, separated by a certain number of lattice spacings (η) One particle of the pair randomly finds an empty site on which to land; then the other particle chooses a site in a random direction at the correlation length distance from the first particle. If this selected site for the second particle is occupied, both particles of this pair will repeat the process described above until they find two empty sites at the correlated distance.

Random landing occurs when two particles of a pair are independent of each other, and all sites in a lattice have equal probability for a particle to land. Effectively there are no 'pairs'.

Evenly spaced landing is used only in simulations of transient reactions Particles are distributed throughout the lattice, and have an equal distance between each other. This interval is equal to L/N_0 and is chosen to be integer, where L is the lattice length and N_0 is the number of the particles at $t = 0$.

Since the kinetic equation can be written for long times,

$$\rho \sim t^{-\alpha} \tag{4}$$

the kinetic data is plotted as $\ln \rho$ vs $\ln t$. The least linear square fit is applied to find the slope of each part of each line, which is equal to $-\alpha$ in eq. (4).

Correlated landing for A + B → 0

A. For $\eta = 1$. Two kinds of landing are investigated. One is a pair of particles of AB with a definite orientation (e g , AB AB AB...). The other one is a pair of particles of AB with random orientations (e g , AB AB BA...) These two cases

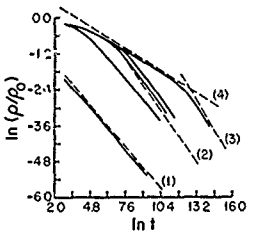


Fig 3. $\ln \rho/\rho_0$ vs. $\ln t$ for $A+B \rightarrow 0$ transient reaction on one-dimensional lattice (30000 sites) with $\rho_0 = 0.05$. From top to bottom, the correlated landing lengths are 1000, 100, 64, 16, and 1. The dashed lines are fitting lines. (1) with the slope 0.5, (2) with the slope 0.6, (3) with the slope 0.7, and (4) with the slope 0.25

have shown the same result — a straight line with a slope 0.5. It is important to notice that this result is the same as that in the $A + A \rightarrow 0$ case (see below).

B. For $\eta > 1$. The slopes of the lines increase (from a value 0.25) after $t > \eta^2$ (see Fig. 3), which is considered to be the effect of correlation in landing processes. As η increases, the slopes, at long times, increase toward the value 0.75.

For $\eta > 1$, there is no finite size effect, i.e., no second transition of the slope was found (see the bottom curve in Fig. 4).

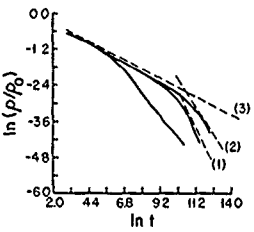


Fig. 4. $\ln \rho/\rho_0$ vs. $\ln t$ for $A+B \rightarrow 0$ transient reaction on one-dimensional lattice (1000 sites) with $\rho_0 = 0.20$. From top to bottom, they are random landing with reflecting boundary condition, random landing with periodic boundary condition, and correlated landing with the correlated landing lengths 50. The dashed lines are fitting lines. (1) with the slope 1.0, (2) with the slope 0.75, and (3) with the slope 0.25.

Random landing for $A + B \rightarrow 0$

Two types of boundary conditions are applied: periodic and reflecting boundary conditions. In both cases, the slopes increase (from the value 0.25) at long times (see Fig. 4), which is considered to be a finite size effect. However, important differences between these cases were observed. The α-value is higher with periodic boundary conditions (~ 1.0) than with reflecting boundary conditions (~ 0.75).

$A + A \rightarrow 0$

Both random landing and correlated landing processes are simulated. Under the periodic boundary conditions, neither the effect of correlated landing nor the finite size effect can be found in the $A + A \rightarrow 0$ case (see top two lines in Fig. 3), straight lines are found with the slope 0.50. However, under reflecting boundary conditions, at long time, a slight deviation from the slope 0.5 is observed.

Our results essentially agree with preliminary continuum models [14], replacing the Zeldovich–Kang–Redner time exponent $-d_s/4$ (for $d_s \leqslant 4$) with $-(d_s + 2)/4$ (for $d_s \leqslant 2$), for tightly correlated systems or finite-sized lattices. However, they emphasize the relative importance of the average interparticle distance and the finite scale of the lattice or of the correlation in the source. In particular, for geminate landing, we do not observe a change in slope at late times.

ACKNOWLEDGEMENTS

This work was supported by NSF Grants No. 8842001, DMR 8815008 and DMR 8801120

REFERENCES

1. R. Kopelman, Exciton percolation in molecular alloys and aggregates, *Topics in Applied Physics*, 15 (1976) 297–346.
2. D. L. Huber, D. S. Hamilton and B. Barnett, Time-dependent effects in fluorescent line narrowing, *Physical Review B*, 16 (1977) 4642–4650.

3 M D. Donsker and S R.S Varadhan, On the number of distinct sites visited by a random walk, *Communications in Pure and Applied Mathematics*, 32 (1979) 721–747.
4 J Klafter, A Blumen and G. Zumofen, Fractal behavior in trapping and reaction a random walk study, *Journal of Statistical Physics*, 36 (1984) 561–577.
5 R Kopelman, Fractal reaction kinetics, *Science*, 241 (1988) 1620–1626
6 A A Ovchinnikov and Ya B. Zeldovich, Role of density fluctuations in bimolecular reaction kinetics, *Chemical Physics*, 28 (1978) 215–218.
7 D Toussaint and F Wilczek, Particle-antiparticle annihilation in diffusive motion, *Journal of Chemical Physics*, 78 (1983) 2642–2647
8 J Klafter, A Blumen and G Zumofen, Fractal behavior in trapping and reaction a random walk study, *Journal of Statistical Physics*, 36 (1984) 561–577
9 K Kang and S. Redner, Scaling approach for the kinetics of recombination processes, *Physics Review Letters*, 52 (1984) 955–958
10 P W. Klymko and R Kopelman, Fractal reaction kinetics exciton fusion on clusters, *Journal of Physical Chemistry*, 87 (1983) 4565–4567
11 E Clément, L. Sander and R. Kopelman, Source-term and excluded-volume effects on the diffusion-controlled A+B →0 reaction in one dimension. rate laws and particle distributions, *Physical Review A*, 39 (1989) 6455–6465.
12 E. Clement, L Sander and R. Kopelman, Steady-state diffusion-controlled A+B→0 reactions in two and three dimensions rate laws and particle distributions, *Physical Review A*, 39 (1989) 6466–6471.
13 E. Clement, L. Sander and R Kopelman, Steady-state diffusion controlled A+A→0 reaction in Euclidean and fractal dimensions rate laws and particle self-ordering, *Physical Review A*, 39 (1989) 6472–6477
14 K Lindenberg, B J West and R Kopelman, Diffusion-limited A+B→0 reaction correlated initial condition, *Physical Review A*, 42 (1990) 890–894

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 133–140
Elsevier Science Publishers B V, Amsterdam

Universality laws in coagulation

D A Weitz *
Exxon Research and Engineering Co, Rt 22E, Annandale, NJ 08801 (U S A)

M.Y. Lin *
Department of Physics, Princeton University, Princeton, NJ 08544 (U S A)

H.M. Lindsay
Department of Physics, Emory University, Atlanta, GA 30322 (U S A)

(Received 8 November 1989, accepted 23 February 1990)

Abstract

Weitz, D A, Lin, M Y and Lindsay, H M, 1991 Universality laws in coagulation *Chemometrics and Intelligent Laboratory Systems*, 10 133–140

We show that the process of irreversible, kinetic colloid aggregation exhibits universal behavior, independent of the detailed chemical nature of the colloidal particles Modern methods of statistical physics, applied to a kinetic growth process provide a good basis to model the observed behavior Two limiting regimes of colloid aggregation are identified rapid aggregation, limited solely by the diffusion of the growing clusters, and slow aggregation, limited by the reaction rate that leads to the formation of bonds between the clusters In each regime the cluster structure is fractal, with fractal dimension $d_f \sim 1.8$ for diffusion-limited clusters and $d_f \sim 2.1$ for reaction-limited clusters A scaling method is used to compare dynamic light scattering data obtained from completely different colloids aggregated under the two limiting conditions. These data provide a critical comparison of the behavior of the different colloids, and confirm the universality of each limiting regime of colloid aggregation

INTRODUCTION

The aggregation of colloidal particles to form larger clusters is a process of wide technological importance and of great scientific interest. It has been the subject of serious scientific study for well over one hundred years. However, until recently the great complexity of the problem has limited the extent of our understanding of the process The structure of the clusters is highly random and disordered, making a quantitative analysis of their shape quite difficult. Furthermore, a wide variety of different types of behavior can seen for even a single colloid. This has precluded the development of a simple theoretical understanding of this complex, yet important process.

More recently, however, significant progress has been achieved in our understanding of irreversible

* Present address National Institute of Standards and Technology, React A106, Gaithersburg, MD 20899, U S A.

0169-7439/91/$03 50

colloid aggregation [1–3]. The impetus for much of this progress has been the recent developments in statistical physics Scaling concepts, which have found so much success in describing such reversible processes as phase transitions, have now also been applied with similar success to irreversible kinetic growth processes, such as colloid aggregation Indeed, recent work has shown that irreversible colloid aggregation exhibits universal behavior, which transcends the chemicals details of the particular colloid system, and which provides a unified, and relatively simple, description of this complex process [4,5] In this paper, we present a brief review of the recent applications of these concepts of modern statistical physics to colloid aggregation, and discuss the universal features that have emerged.

There are two general classes of colloid aggregation which have been widely studied [1]. Both begin with a monodisperse suspension of small, solid particles undergoing Brownian motion When the aggregation is initiated, the diffusive motion of the particles leads to collisions between them, causing them to stick together and form larger clusters In the first class of aggregation, the clusters, once formed, no longer diffuse, and all aggregation is due to the accretion of single particles This class is called single particle aggregation By contrast, in the second class, the clusters themselves continue to diffuse, collide and form yet larger clusters As the clusters grow, what began as a monodisperse distribution of single particles evolves into a very complex distribution of clusters of different sizes. This class is called cluster-cluster aggregation Both types of aggregation have been extensively studied theoretically. However, most experimental studies of colloid aggregation have focused on the cluster-cluster class, as it is by far the most commonly encountered.

Several key features characterize any aggregation process [3]. These include the structure of the clusters, the kinetics of the aggregation and the shape of the cluster mass distribution and its evolution in time. It is in the description of each of these features that the application of modern methods of statistical physics and the concepts of scaling has provided such progress. The first application of these techniques was to the description of the structure of the clusters. The cluster structure is highly random and disordered, and had long defied any quantitative description However, the cluster structure can, in fact, be quantitatively parameterized by means of a type of symmetry, that of invariance under a change in length scale, or dilation symmetry. Thus colloidal aggregates can be characterized as fractals [6], and their structure can be quantitatively parameterized by means of their fractal dimension [7] The aggregation kinetics, and the shape and time evolution of the cluster mass distribution can both be addressed through the application of scaling, in this case, in time The shape of the cluster mass distribution is found to be invariant in time, with all the time dependence described by the evolution of the average cluster mass [8,9].

The fundamental property which determines the nature of cluster–cluster aggregation is the form of the interaction potential between two colloidal particles as they approach one another [10] Colloidal particles which are stable against aggregation have some form of repulsive interaction which prevents two approaching particles from touching and sticking together This repulsion is often due to charged groups adsorbed on the surface of the colloidal particles, but can also arise from other sources, such as a thin coating of polymer on the particle surface The height of the resultant repulsive barrier, E_b, must be much greater than k_BT for the colloid to be stable against aggregation If E_b is reduced, colliding particles can surmount the barrier, and stick together, thus initiating the aggregation process. The rate of aggregation will be determined by the probability, P, that two particles will stick upon colliding. This is determined by the height of the remaining barrier, and is given by $P \sim \exp(-E_b/k_BT)$.

The exponential dependence of the sticking probability on E_b makes the aggregation rate very sensitive to the value of the repulsive energy barrier, and a very wide range of aggregation rates can be obtained with any colloidal suspension. However, there are two characteristic, limiting regimes of aggregation [11]. In the first, the repulsive barrier is removed completely, so that $E_b \ll kBT$ and $P \approx 1$. In this case, every collision results

in the particles or clusters sticking to one another, and the aggregation rate is limited solely by the time between diffusion-induced collisions This class of aggregation is called diffusion-limited colloidal aggregation (DLCA). In the second regime, the repulsive barrier is reduced only a small amount, so that $E_b > k_B T$, and P is very small. In this case, a large number of collisions are required before two particles or clusters stick to one another, which limits the aggregation rate. This regime is called reaction-limited colloid aggregation (RLCA) The two regimes lead to very rapid and very slow aggregation respectively, and have been recognized as such in the traditional colloid literature [10]. However, they also form two limiting types of behavior, with distinct, and universal features characteristic of each.

The 'rules' which determine the aggregation in each regime are quite simple. In DLCA, two clusters stick immediately upon contact, and the diffusive nature of the motion of the clusters plays an important role in determining both their structure and the aggregation kinetics. The diffusive motion ensures that the clusters always stick to one another at the edges, making the resultant aggregates significantly more tenuous. By contrast, in RLCA, the sticking probability is so low that, on an average, statistical basis, two clusters can adopt any bonding configuration that is physically possible, since the clusters have sufficient opportunity to explore all possible configurations Thus the diffusive nature of the cluster motion does not play a significant role in the aggregation process, and the clusters no longer stick solely at the edges, making their structure significantly less tenuous In both regimes, the bonds between particles, once formed, are assumed to be both permanent and rigid, so that no further change in their structure occurs as the aggregation proceeds.

The nature of the interparticle interactions determines the kinetics of the aggregation process, the kinetics in turn play a significant role in determining the structure of the clusters formed, and the shape of the mass distribution of clusters. Furthermore, since a very large number of clusters are involved in any aggregation process, and since the details of the structure of each clusters are not as important as the overall features, a statistical description is well suited to describing the physics The basic simplicity of the underlying physics facilitates modeling the aggregation process The models developed deal solely with the nature of the interaction and the resultant "rules" which determine how clusters move and stick to one another. Thus, these models are independent of the detailed chemical nature of each colloid, and should apply equally well to all colloids. It is in this sense that the description of colloid aggregation should be universal.

THEORY

The two limiting regimes of cluster–cluster aggregation have been studied extensively, and an elegant and detailed picture of their behavior has now been developed [1,3] The theoretical work has entailed two basic approaches the simplicity of the rules of the aggregation make computer simulation a very powerful method for studying both regimes, and considerable knowledge has been obtained about the structure of the clusters and the shape and time evolution of the cluster mass distribution [12] The aggregation kinetics and the cluster mass distribution have also been studied extensively through the use of the Smoluchowski equations [13]. These are a set of rate equations which assume that the aggregation rate between two clusters depends solely on their masses Scaling techniques have proven to be well suited to the study of these equations [8,9]. Experimentally, a wide range of colloid systems have been studied using many different techniques Excellent agreement is obtained between the experimental observations and the theoretical predictions [14,15].

Each regime is distinguished by several distinct characteristics. the clusters formed in each regime are fractal, so that their mass scales with their radius as $M = (R/a)^{d_f}$, where a is the radius of a single particle and d_f is the fractal dimension, which is non-integral and less than the dimension of space. For DLCA, $d_f \sim 1.8$ while for RLCA, $d_f \sim 2.1$. The cluster mass distribution in each regime exhibits dynamic scaling and can be written as $N(M) = \overline{M}^2 \psi(M/\overline{M})$, where the scaling

function, $\psi(M/\overline{M})$ describes the shape of the cluster mass distribution and is independent of time, while $\overline{M}$ is the mass of the average cluster and reflects all of the time dependence of the aggregation. For DLCA, $N(M)$ is slightly peaked around the average mass with an exponential cutoff at larger masses. For RLCA, the cluster mass distribution has a power-law form with an exponential cutoff at large mass, $N(M) \sim \overline{M}^{\tau-2}M^{-\tau}\exp(-M/\overline{M})$. The kinetics of the aggregation are determined by the time dependence of $\overline{M}$: for DLCA, $\overline{M}$ grows linearly with time, while for RLCA it grows exponentially with time.

EXPERIMENTAL

To experimentally demonstrate the universal features of colloid aggregation, we compare the behavior of three completely different colloids: gold, silica and polystyrene latex [4]. Each colloid is comprised of a different material, each colloid is initially stabilized by completely different functional groups on their surfaces; the aggregation for each colloid is initiated in a different manner, the interparticle bonds in the aggregates for each colloid are different, and each colloid has a different primary particle size. However, each colloid can be made to aggregate by either diffusion-limited or reaction-limited kinetics.

The colloidal gold has a particle radius of $a = 7.5$ nm and an initial volume fraction of $\phi_0 = 10^{-6}$. It is stabilized by citrate ions adsorbed on the surface. The aggregation is initiated by addition of pyridine, which displaces the charged ions, reducing the repulsive barrier between the particles. The amount of pyridine added determines the aggregation rate: for DLCA, the pyridine concentration is 10^{-2} M, while for RLCA, it is about 10^{-5} M. The interparticle bonds are metallic.

The colloidal silica used is Ludox SM obtained from DuPont. It has particles with $a = 3.5$ nm, and is diluted to $\phi_0 = 10^{-6}$. It is initially stabilized by OH^- or SiO^- on the surface. The pH is kept ≤ 11 by addition of NaOH and the aggregation is initiated by addition NaCl, which reduces the Debye–Hückel screening length, thereby reducing the repulsive barrier between the particles. For DLCA, the salt concentration is 0.9 M, while for RLCA, it is 0.6 M. The interparticle bonds are believed to be silica bonds.

The polystyrene latex has $a = 19$ nm and is diluted to $\phi_0 = 10^{-6}$. It is initially stabilized by charged carboxylic acid groups on the surface of the particles. Addition of HCl to a concentration of 1.2 M is used to neutralize the surface charges and decrease the screening length to initiate the aggregation for DLCA. For RLCA, NaCl is added to a concentration of 0.2 M, to reduce the screening length and initiate the aggregation. The particle surfaces deform on bonding leading to large Van der Waals interactions between the bound particles.

To study the aggregation of each colloid and to critically compare their behavior in the two regimes, we use light scattering [16]. Static light scattering is used to measure the fractal dimension of the clusters, while dynamic light scattering is used to follow the aggregation kinetics. In addition, the dynamic light scattering data obtained from each colloid in each regime can be scaled onto a single master curve. The shape of this master curve is very sensitive to the features of the aggregation process, depending on the detailed structure of the clusters and the shape of the clusters mass distribution. However, all features particular to the individual colloids are scaled out of the master curve, allowing the curves from the different colloids to be compared directly, with no free parameters, providing a critical test of the universality of colloid aggregation in each of the two limiting regimes [4].

RESULTS

Static light scattering measures the time averaged scattering intensity from the sample, $I(q)$, as function of the scattering wavevector, $q = (4\pi n/\lambda)\sin(\theta/2)$, where λ is the incident wavelength in vacuuo, n is the index of refraction of water, and θ is the scattering angle. Dynamic scattering measures the temporal autocorrelation function of fluctuations in the scattering intensity resulting from the diffusive motion of the clusters. We measure both the total scattered intensity and the

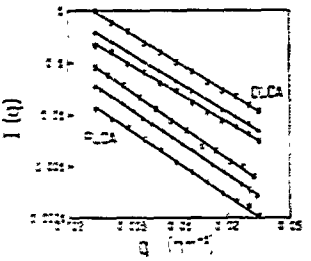


Fig. 1. Static light scattering from the three colloids aggregated in each of the limiting regimes. The linear behavior in the logarithmic plot demonstrates the fractal structure of the clusters, and the fractal dimensions determined from the slopes are: for DLCA, 1.86 for gold (○), 1.85 for silica (+) and 1.82 for polystyrene (•); for RLCA, 2.14 for gold (□), 2.07 for silica (+) and 2.09 for polystyrene (•).

autocorrelation function concurrently as functions of the scattering angle, and hence the scattering wavevector. The excitation source is the 488-nm line of an Ar^+ laser, and the accessible scattering vectors are $0.003 \le q \le 0.03\ nm^{-1}$.

Static light scattering probes the internal structure of the aggregates. Because the fractal clusters are self-similar in structure, the scattered intensity from each cluster depends only on the product qR_g, where R_g is the radius of gyration of the cluster. At low qR_g, the internal structure of the aggregate is not resolved, and the scattered intensity is isotropic, independent of q. At high qR_g, however, the internal fractal structure is resolved and the scattered intensity scales as $(qR_g)^{-d_f}$. The measured intensity is a weighted average over the cluster mass distribution. However, for aggregates that are sufficiently large, the total measured intensity also exhibits the fractal scaling in q, allowing d_f to be determined directly. The static light scattering obtained from all the colloids in each regime is shown in Fig. 1. In each case, the data were collected only after the clusters were sufficiently large that $q\bar{R} \gg 1$, where $\bar{R}$ is an average cluster size. The linear behavior in the double logarithmic plots confirms the fractal structure of the aggregates. The upper three data sets are obtained from clusters prepared under DLCA conditions; and have $d_f = 1.86$ for the gold, $d_f = 1.85$ for the silica and $d_f = 1.82$ for the polystyrene. To within the experimental error of roughly ± 0.05, these results are identical. By contrast, the lower three data sets, which are obtained from clusters prepared under RLCA conditions, have consistently higher values of the fractal dimensions, with $d = 2.14$ for the gold, $d_f = 2.07$ for the silica and $d_f = 2.09$ for the polystyrene. These values are again equal to within experimental error. Thus these results demonstrate the universal behavior of the structure of the fractal colloid aggregates in each of the two regimes.

Dynamic light scattering probes the diffusive motion of the clusters. When the clusters are large enough that their internal fractal structure can be resolved, both their translational and rotational diffusion contribute to the fluctuations [17]. Here, we consider only the first cumulant [18], or the initial logarithmic derivative of the autocorrelation function of the intensity fluctuations. This is given by $\Gamma_1 = q^2 D_{eff}(qR_g)$, where the effective diffusion coefficient reflects the contribution of both translational and rotational diffusion. When $qR_g \ll 1$, only translational diffusion contributes and $D_{eff}(qR_g) = D = \zeta/R_H$, where $\zeta = k_B T/6\pi\eta$ and η is the fluid viscosity. The hydrodynamic radius is related to the radius of gyration of the cluster, $R_H = \beta R_g$, with $\beta \sim 1$. For $qR_g \gg 1$, rotational diffusion also contributes and $D_{eff} \sim 2D$.

The effective diffusion coefficient determined from the measured first cumulant is again a weighted average over all the clusters in the distribution. It is given by

$$\bar{D}_{eff} = \frac{\sum N(M) I(qR_g) D_{eff}}{\sum N(M) I(qR_g)} \qquad (1)$$

In the limit of $q\bar{R} \to 0$, $\bar{D}_{eff} = \bar{D}$, providing a good measure of the average cluster size, $\bar{R} = \zeta/\bar{D}$.

The combination of the sensitivity to the cluster mass distribution and rotational diffusion leads to a pronounced q dependence in the measured $\bar{D}_{eff}$, and provides a very sensitive probe of the aggregation process [4,16]. However, to fully explore this q dependence at a single point in time during the aggregation process would require an experimentally inaccessible range of scattering angles. In-

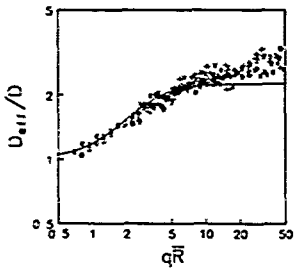


Fig. 2. Master curves obtained independently from dynamic light scattering data from each of the three colloids aggregated under diffusion-limited conditions. The curves are indistinguishable, demonstrating the universality of DLCA. The solid line is the calculated behavior. ○ = Gold, + = silica, × = polystyrene.

stead, we exploit the dynamic scaling of the cluster mass distribution to measure $\bar{D}_{eff}$ over a much wider range of $q\bar{R}$. Thus, we determine $\bar{D}_{eff}$ over the range of q experimentally accessible and repeat the measurements during the aggregation process, as $\bar{R}$ increases, while the shape of the cluster mass distribution remains unchanged. The values measured at each q are interpolated to obtain a series of data sets, each consisting of $\bar{D}_{eff}(q)$ evaluated at the same time. We normalize $\bar{D}_{eff}$ by $\bar{D}$, and plot the data as a function of $q\bar{R}$, where the required parameter, $\bar{D} = \xi/\bar{R}$, for each set is determined empirically by scaling the data onto a single master curve. With sufficient data, there is always a substantial overlap between data from different sets, making the scaling unambiguous. All material parameters are scaled out, so that these master curves provide a means to critically compare the behavior of completely different colloids

The master curve obtained for each colloid aggregated under DLCA conditions are shown in Fig. 2, while the master curves for each colloid aggregated under RLCA conditions are shown in Fig. 3. The shape of the master curve for DLCA is quite different from that of RLCA. This reflects the different shapes of $N(M)$ for each regime, with the power-law form for RLCA leading to a considerably stronger q-dependence of the master curve. In each regime, the master curves for the three colloids are indistinguishable. We emphasize that the master curves for each colloid are obtained independently, and there is no free parameter in comparing them. This is striking evidence of the universality of each of the regimes of colloid aggregation.

The solid lines drawn through the master curves are the calculated values using eq. (1), with the forms for $N(M)$ expected for each regime and a form for $I(qR_g)$ obtained from computer simulated clusters for the appropriate regime [19]. The agreement is very good, except for DLCA at large $q\bar{R}$. The calculation for the RLCA regime allows us to determine the cluster mass exponent, $\tau = 1.5$, which is in accord with theoretical predictions based on the Smoluchowski equations [20].

The scaling values of $\bar{R}$ also allow us to determine the aggregation kinetics of each colloid in each regime. We show the results for the DLCA regime in Fig. 4, where we plot $\bar{R}$ as a function of aggregation time t_a in a double logarithmic plot [14]. The linear behavior exhibited by each colloid confirms the power-law kinetics, the slopes, combined with the measured fractal dimensions, give the power law for the growth of the average mass In all cases, this exponent is 1 to within experimental error. The different offsets of the three curves reflect the differences in the initial con-

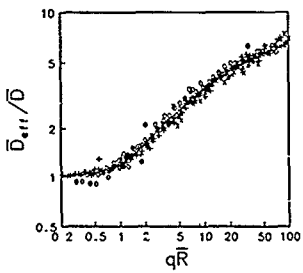


Fig. 3. Master curves obtained independently from dynamic light scattering data from each of the three colloids aggregated under reaction-limited conditions The curves are indistinguishable, demonstrating the universality of RLCA. The solid line is the calculated behavior ○ = Gold, + = silica, × = polystyrene

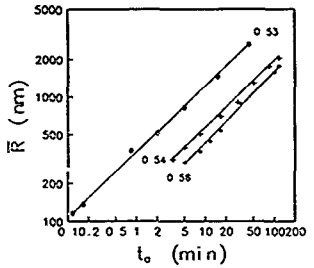


Fig. 4 The aggregation kinetics of the diffusion-limited aggregation of each of the three colloids obtained from the scaling of the data onto the master curves. The slopes of the power-law kinetics and the fractal dimensions show that the average cluster mass grows linearly with time in all cases ○ = Gold, + = silica, • = polystyrene.

centrations. The results for the RLCA regime for each of the colloids are shown in Fig. 5, where we now use a semilogarithmic plot to show the exponential growth observed for each colloid [15]. In this case, the different slopes reflect the different initial aggregation rates of each colloid, which do depend on the details of the chemistry Indeed, for the polystyrene, some time apparently elapses before the final aggregation rate is achieved. We believe that this is caused by the deformation of the particles which occurs on bonding and which

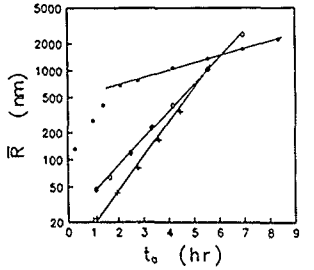


Fig 5 The aggregation kinetics of the reaction-limited aggregation of each of the three colloids, demonstrating the exponential kinetics in each case. ○ = Gold, + = silica, • = polystyrene

modifies the sticking probability at early time Nevertheless, all colloids display exponential growth of the radius of the average cluster, and hence of the mass, as expected.

CONCLUSIONS

In summary, we have shown experimental evidence to demonstrate the universal features of colloid aggregation. Two limiting regimes are observed· fast, diffusion-limited and slow, reaction-limited colloid aggregation Each regime follows universal laws that describe its behavior In many experimental situations, these limiting regimes are not achieved. Nevertheless, the overall aggregation behavior can usually be described in terms of these two regimes Typically the initial stages of the aggregation are controlled by some intermediate value of the sticking probability, and the aggregation is not strictly diffusion-limited. Instead, at the earlies times, it can be approximated as reaction-limited However, as the aggregation proceeds, and the concentrations of clusters decreases, their spacing increases, and diffusion becomes increasingly important as a rate limiting step Thus at longer times the aggregation crosses over to diffusion-limited Thus, these two limiting, and universal, regimes provide the basis for describing a large range of behavior for colloid aggregation

REFERENCES

1 P Meakin, The growth of fractal aggregates and their fractal measures, in C Domb and J L Liebowitz (Editors), *Phase Transitions and Critical Phenomena*, Vol 12, Academic Press, New York, 1988, pp 335–489
2 F Family and D P Landau (Editors), *Kinetic Aggregation and Gelation*, Elsevier, Amsterdam, 1984
3 D A Weitz, M Y Lin and J S Huang, Fractals and scaling in kinetic colloid aggregation, in S A Safran and N A Clark (Editors), *Physics of Complex and Supermolecular Fluids*, Wiley-Interscience, New York, 1987, pp 509–549
4 M Y Lin H M Lindsay, D A. Weitz, R C Ball, R Klein and P Meakin, Universality in colloid aggregation, *Nature (London)*, 339 (1989) 360–362
5 M Y Lin, H M. Lindsay, D A Weitz, R C Ball, R Klein and P Meakin, Universality of fractal aggregates as probed

by light scattering, *Proceedings of the Royal Society of London, Series A*, 423 (1989) 71–87
6 B B Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, CA, 1982
7 D A Weitz and M Oliveria, Fractal structures formed by kinetic aggregation of aqueous gold colloids, *Physical Review Letters*, 52 (1984) 1433–1436
8 T Vicsek and F Family, Dynamic scaling for aggregation of clusters, *Physical Review Letters*, 52 (1984) 1669–1672
9 P G J van Dongen and M H Ernst, Dynamic scaling in the kinetics of clustering, *Physical Review Letters*, 54 (1985) 1396–1399
10 E J W Verwey and J T G Overbeek, *Theory of the Stability of Lyophobic Colloids*, Elsevier, Amsterdam, 1948
11 D A Weitz, J S Huang, M Y Lin and J Sung, Limits of the fractal dimension for irreversible kinetic aggregation of gold colloids, *Physical Review Letters*, 54 (1985) 1416–1419
12 P Meakin, Fractal aggregates, *Advances in Colloid and Interface Science*, 28 (1988) 249–331
13 R J Cohen and G B Benedek, Equilibrium and kinetic theory of polymerization and sol–gel transition, *Journal of Physical Chemistry*, 86 (1982) 3696–3714
14 M Y Lin, H M Lindsay, D A Weitz, R Klein, R C Ball and P Meakin, Universal diffusion-limited aggregation, *Journal of Physics Condensed Matter*, 2 (1990) 3093–3113
15 M Y Lin, H M Lindsay, D A Weitz, R C Ball, R Klein and P Meakin, Universal reaction-limited aggregation, *Physical Reviews Section A*, 41 (1990) 2005–2020
16 H M Lindsay, M Y Lin, D A Weitz, R C Ball, R. Klein and P Meakin, Light scattering from fractal colloid aggregates, in *Proceedings on Photon Correlation Techniques and Applications*, Vol 1, Optical Society of America, Washington, DC, 1988, pp 122–131
17 H M Lindsay, R. Klein, D A Weitz, M Y Lin and P Meakin, Effect of rotational diffusion on quasielastic light scattering from fractal colloid aggregates, *Physical Reviews Section A*, 38 (1988) 2614–2626
18 B J Berne and R Pecora, *Dynamic Light Scattering*, Wiley, New York, 1976
19 M Y Lin, R Klein, H M Lindsay, D A Weitz, R C. Ball and P Meakin, The structure of fractal colloidal aggregates of finite extent, *Journal of Colloid and Interface Science*, in press
20 R.C Ball, D A Weitz, T A Witten and F Leyvraz, Universal kinetics in reaction-limited aggregation, *Physical Review Letters*, 58 (1985) 274–277

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 141–154
Elsevier Science Publishers B V, Amsterdam

Inference of mechanism from kinetic analysis of pulse voltammetric data

Janet Osteryoung
Department of Chemistry, SUNY University at Buffalo, Buffalo, NY 14214 (U S A)

(Received 14 November 1989, accepted 16 July 1990)

Abstract

Osteryoung, J., 1991 Inference of mechanism from kinetic analysis of pulse voltammetric data *Chemometrics and Intelligent Laboratory Systems*, 10 141–154

Voltammetry provides direct access to kinetic information in that the measured quantity, current, is itself the rate Kinetic analysis of voltammetric data generally focuses on the potential dependence of the current. For historical reasons, the most common method of analyzing data is to transform the data, often by very elaborate methods, to yield a potential-dependent rate constant, which is then plotted as a semilogarithmic function of potential. This procedure requires extrinsic normalization factors which easily can introduce systematic error In a few instances, statistically sound methods have been employed for analysis of data One approach employs a nonlinear least squares procedure equivalent to the method of maximum likelihood In addition to providing optimal values of kinetic parameters without recourse to other data, this method also provides confidence regions at a known level of confidence This method is implemented by the COOL algorithm, which has been described An important ancillary factor is that the COOL algorithm runs in 'real-time' for many problems This paper describes these alternative methods of analysis by using the particular example of slow charge transfer The sensitivity of the analysis to changes in values of parameters is examined by computation of confidence regions Then three specific kinetic problems are used to illustrate the types of questions which arise in the inference of mechanism The first involves the search for a second order dependence of current on potential, this having been predicted by theoretical treatments The second can arise in cases where two electrons are transferred Under what conditions is it possible to determine the rate parameters for both transfers? What criteria ensure that the variance in the data is explained by only one charge transfer step (i e, the other is too fast to see)? The third problem concerns heterogeneous charge transfers coupled by a homogeneous chemical step When the second charge transfer is more favored than the first, when does it take place through a homogeneous reaction route, and under what conditions can this be detected? The experimental examples include the reduction of Zn(II) and the reduction of *p*-nitrosophenol, both at mercury electrodes. The data are confounded to some degree by experimental artifacts, nonrandom distribution of residuals may arise from these artifacts or from choice of overly simple models

INTRODUCTION

Voltammetry comprises a suite of electrochemical techniques wherein the potential of an electrode is controlled and the resulting current is measured. Time is generally a parameter of the experiment Pulse voltammetry comprises a subset of voltammetric techniques in which potential is changed only in a stepwise fashion (changes in potential are instantaneous on the time scale of the experiment). The pulse mode has many advantages both experimentally and computationally when the experiment is carried out under the real-time control of a digital computer suitable for

0169-7439/91/$03 50

high speed calculations This paper deals only with pulse voltammetry. However, its main points apply to voltammetry in general

Voltammetry provides direct access to kinetic information in that the measured quantity, current, is itself the rate. Kinetic analysis of voltammetric data generally focuses on the potential dependence of the current. The purpose of kinetic analysis is generally two-fold, first to infer from the rate data the mechanism by which chemical transformation takes place, and second to obtain values of the rate constants or other parameters which characterize the system. Here this general problem is introduced by describing a straightforward example, the simple, first-order slow transfer of an electron.

The phenomenon of potential dependence of the rate is well-known and was first formulated empirically in the Tafel equation [1]

$$\eta = a + b \log i \tag{1}$$

where η is the overpotential (potential minus equilibrium potential), i is the steady-state current, and a and b are empirical constants. The experiments which gave rise to this observation employed large concentrations of oxidized and reduced forms in contact with an inert electrode, so that the equilibrium potential was well fixed, and so that small excursions of potential from the equilibrium value would not significantly change the concentrations near the electrode This mode of kinetic measurements dominated the study of electrochemical kinetics for the next 50 years.

It was not until the development of polarography by Heyrovsky in the '20s and '30s [2] that changes in concentration near the electrode and resulting diffusion were treated mathematically. After World War II, the confluence of mathematical expertise, the computational power offered by computers, and improved electronics permitted dynamic experiments in which potential could be changed rapidly and automatically. An appropriate formulation of the current arising from the reduction of reactant O to product R under these experimental conditions is

$$i/nFA = k_f C_O(0, t) - k_b C_R(0, t) \tag{2}$$

where

$$k_f = k_a^0 \exp[-\alpha n f(E - E^{\circ\prime})] \tag{3}$$

$$k_b = k_a^0 \exp[(1-\alpha) n f(E - E^{\circ\prime})] \tag{4}$$

and $C_O(0, t)$, $C_R(0, t)$ are the time-dependent concentrations of the oxidized and reduced forms, respectively, at the electrode surface, k_a^0 is the standard apparent heterogeneous charge transfer rate constant, referred to the formal potential, $E^{\circ\prime}$, for the reaction

$$O + ne^- \rightleftarrows R \tag{5}$$

α is the 'charge transfer coefficient', $f = F/RT = 38.9\ V^{-1}$ at 25°C, E is the electrode potential, i the current at an electrode of area A, and n the number of electrons transferred (eq (5)) This formulation ignores the effect of charge on the electrode and corresponding charge distribution in solution. For an elementary process, $n = 1$ In general it is found that even for more complicated processes, eqs (2)–(4) describe the experimental result, although the value of n in eqs. (3) and (4) may be less than that in eqs (2) and (5) (Electrons transferred after the rate-determining step do not contribute to n in eqs (3) and (4).) A complete description of a mechanism ideally consists only of elementary steps However here, for convenience and generality, we retain the symbol for n, and do not distinguish between the overall value and that which applies to the rate-determining step.

The technique of normal pulse voltammetry leads to a simple closed-form solution to the diffusion equation under linear, semiinfinite conditions with eqs (2)–(4) as a boundary condition. Thus quasireversible charge transfer under normal pulse voltammetric conditions provides a straightforward example of the types of questions which arise in kinetic analyses

The current which flows in response to the potential perturbation of normal pulse voltammetry for quasireversible charge transfer is given by [3]

$$i(t) = i_d(1+\epsilon)^{-1}\pi^{1/2}\lambda t^{1/2} \exp(\lambda^2 t)\operatorname{erfc}(\lambda t^{1/2}) \tag{6}$$

where

$$\lambda = \kappa(1+\epsilon)\epsilon^{-\alpha} \tag{7}$$

$$\kappa = k_a^0/D_O^{(1-\alpha)/2}D_R^{\alpha/2} \tag{8}$$

$$\epsilon = \exp\{nf(E - E'_{1/2})\} \tag{9}$$

$$E'_{1/2} = E^{\circ\prime} + (1/nf)\ln(D_O/D_R)^{1/2} \tag{10}$$

$$i_d = nFAD_O^{1/2}C_O^*/(\pi t)^{1/2} \tag{11}$$

D_O and D_R are the diffusion coefficients of the oxidized and reduced species (eq (5)), O and R, respectively, the initial uniform concentration of O is C_O^*, that of R is zero, and t is the time after the potential is applied at which current is measured The quantity i_d is the 'diffusion-controlled current', the maximum current which can be obtained under these conditions A typical result conforming to eq. (6) is presented in Fig 1

Eq (6) provides a means to calculate the current $i(t)$ at any potential and time, given the values of i_d, $E'_{1/2}$, α, k_a^0, D_O, and D_R Typically $i(t)$ can be measured, for various values of t, over the potential range of interest. The objective is to see whether the results conform to eq (6) and to obtain the values of the kinetic parameters, α and k_a^0.

The typical experimental procedure is as follows First note that as $\lambda^2 t \to \infty$, eq (6) approaches

$$i(t) = i_d(1+\epsilon)^{-1} \tag{12}$$

and when this is true (i e , k_f and k_b are large in comparison with the rate of diffusion), $E = E'_{1/2}$ when $i(t)/i_d = 1/2$. Thus $E'_{1/2}$ is measured in this way using data from an experiment at times sufficiently long that the kinetic effect is negligible.

When this regime is experimentally inaccessible, $E'_{1/2}$ may be obtainable through measurement of $E^{\circ\prime}$ (eq. (10)) This is done by preparing a solution containing high concentrations of both forms of the redox couple (O and R, eq. (5)) and measuring the potential of an inert electrode placed therein This route also has problems, in that the reduced form, R, may be unstable

When ϵ is small, that is, $E \ll E'_{1/2}$, i attains its limiting value of i_d (eq (12)) Thus, by carrying out the experiment at sufficiently negative potential, i_d is measured.

Depending on the value of $n\alpha$, the approach of i to the value i_d may be very slow, and a slight increase in i with increasingly negative potential may be confused with unwanted contributions to

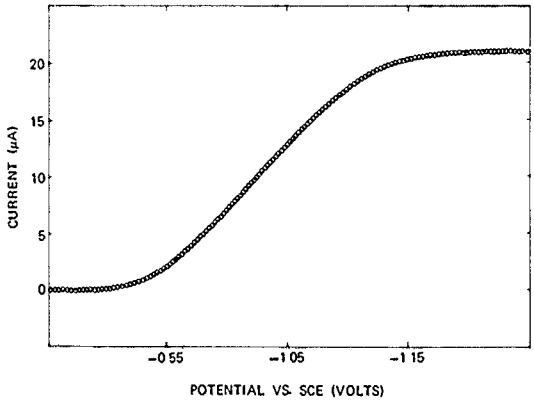


Fig. 1 Normal pulse voltammetric reduction of 0 99 mM Zn(II) in 1 0 M $NaNO_3$, SMDE (small drop), $t_d = 0\,5$ s, $t_p = 0\,01$ s (○) Experimental points, (——) optimal theoretical curve calculated for $E'_{1/2} = -0\,971$ V, $\alpha = 0\,23$, $\log(\kappa t_p^{1/2}) = -0\,81$

the current from other processes. In the example of Fig. 1, the current has not reached its limiting value at the most extreme potential

The third step is to obtain the current as a function of t and E over ranges of values for which the kinetic effect manifests itself. Using the data from these three steps, the quantity $i(E, t)[1 + \epsilon]/i_d$ is computed for each experimental current From eq (6),

$$i(E, t)[1 + \epsilon]/i_d = \pi^{1/2} x \exp(x^2) \operatorname{erfc}(x) = f(x) \quad (13)$$

where $x = \lambda t^{1/2}$. Having thus obtained values of $f(x)$, the function is inverted to obtain values of $\lambda t^{1/2}$, and thus λ. Comparing eqs. (3), (7), and (8),

$$k_f = D_O^{1/2} \lambda/(1 + \epsilon) \quad (14)$$

Measurement of i_d as a function of t or C_O^* allows one to determine D_O, provided n and A are known Thus k_f can be calculated The quantity $k_f(E)$ is then plotted as a function of E to obtain α from the slope according to eq. (7), and k_s^0 as the value of k_f at $E = E^{0\prime}$. Note from eq (10) that this also requires the value of D_R, which may be difficult to measure if R is unstable.

A plot of η vs log i is a 'Tafel plot' (cf eq (1)), and the similar plot of log i vs. E is usually given the same name. By extension, the plot of log k_f vs E is a 'Tafel' or 'Tafel-like' plot. This scheme for obtaining the potential dependence of the rate thus has arisen naturally from the earliest empirical observations.

DATA ANALYSIS

The measurement errors associated with this procedure have been described in detail [4]. Even without considering the experimental details, it should be apparent that this procedure and all other procedures which are similar in requiring normalizations and computation of k_f using data from different experiments are unsound. In particular, the result for α is very sensitive to the value of $E_{1/2}^f$.

Consider the following characteristics of the functional form. First, $\epsilon < 10^{-4}$ for $n(E - E_{1/2}^f) < -120$ mV Thus, as i approaches i_d, eq (13) becomes independent of ϵ Second, for large $x = \lambda t^{1/2}$, $f(x)$ (eq (13)) is insensitive to $\lambda t^{1/2}$. For example, for $x = 2$, $df(x)/dx = 0.073$ but for $x = 10$, $df(x)/dx = 0.00097$. In the range of large x, small errors in i_d, which cause only small errors in $f(x)$, result in large errors in x, and thus in k_f.

Third, $(1 + \epsilon)$ increases exponentially for $E > E_{1/2}^f$, and thus small errors in i at small i can cause $f(x)$ to be larger than its maximum value of unity When this is a problem, the analysis is always 'improved' by choosing a more positive value of $E_{1/2}^f$. These points are illustrated in Fig. 2, which displays the data of Fig. 1 according to the scheme presented here with four different values of $E_{1/2}^f$, the optimal value, obtained as described below, and both smaller and larger values

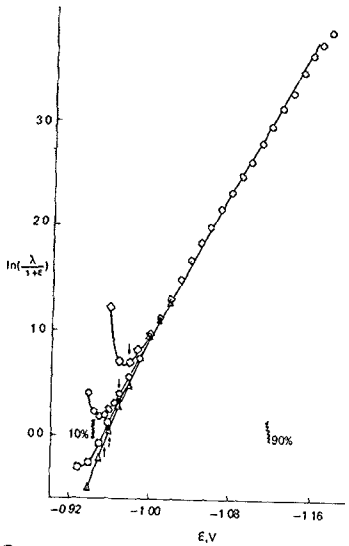


Fig 2 Semilogarithmic plot of data of Fig 1 according to eq (3) with various choices of $E_{1/2}^f$ (V) (○) −0 971, (◇) −0 981, (□) −0 966, (△) −0 961 These potentials are indicated by arrows on the figure The range $0.1 < i/i_d < 0.9$ is also indicated

For negative potentials of about −1.06 V, all of the points are the same, because $\epsilon \ll 1$ Even for the optimal value of $E'_{1/2}$, the value of $\ln[\lambda/(1+\epsilon)]$ deviates from the predicted linearity for potentials much more positive than $E'_{1/2}$, because small errors in $E'_{1/2}$ or in t are magnified by the large values of ϵ used in eq (13). More positive values of $E'_{1/2}$ increase the range of linearity, and thus appear to be 'better' values. Conventionally it is felt that experimental errors may dominate outside of the range $0.1 < t/t_d < 0.9$, which is indicated in Fig 2.

An alternative approach to the analysis of voltammetric data which is statistically sound has been developed and described in detail [5] To explain this approach, for simplicity we use as an example the kinetic problem just discussed. The model yields a dimensionless current function, ψ, here (cf eq. (6))

$$\psi = (1+\epsilon)^{-1}\pi^{1/2}\lambda t^{1/2}\exp(\lambda^2 t)\,\mathrm{erfc}(\lambda t^{1/2}) \tag{15}$$

Examining eqs. (15), (7), (8), and (9), the parameters sought are identified as α, κ, and $E'_{1/2}$. The experimental currents $i(E, t)$ are then analyzed according to the linear equation

$$i(E, t) = a\psi(\alpha, \kappa, E'_{1/2}) + c \tag{16}$$

by finding the optimal value $(\hat{\alpha}, \hat{\kappa}, \hat{E}'_{1/2})$ which maximizes the correlation of i with ψ (or minimizes the complement of the correlation coefficient, $(1-r)$). It is assumed that experimental errors are normally distributed with zero mean It has been shown that this procedure is equivalent to the method of maximum likelihood.

In addition, the confidence region for the quantity $(\hat{\alpha}, \hat{\kappa}, \hat{E}'_{1/2})$ is determined at a known level of confidence. The confidence ellipsoid may be described by the intervals I_α, I_κ, $I_{E'_{1/2}}$, where, for example, I_α is the size of the ellipsoid in the α dimension at $\kappa = \hat{\kappa}$, $E'_{1/2} = \hat{E}'_{1/2}$. The quantity I_α has endpoints α' and α''. The values α' and α'' are the values of α that lead to a correlation coefficient $r_2 = 1 - b(1-r_m)$ when the correlation is maximized as a function of the other two parameters, κ and $E'_{1/2}$. The interval I_α is not a confidence interval for α; it is the size of the confidence ellipsoid along a line passing through the optimum and parallel to the α-axis The value of b is given asymptotically by $\exp(\chi^2/m) = (1-r_2^2)/(1-r_m^2)$ When $r_m \simeq 1$, as is usually the case,

$$b = \exp(\chi^2/m) \tag{17}$$

where m is the number of experimental points and χ^2 is the chi-squared statistic for appropriate level of confidence and three degrees of freedom.

THE COOL ALGORITHM

This method has been implemented by means of an algorithm (called the COOL algorithm), which incorporates the modified simplex algorithm to search for the optimal values, and the secant algorithm to calculate the intervals of the confidence ellipsoid The important features of the procedure in applications to electrochemical kinetics are as follows.

(i) The treatment is independent of ψ any computational technique may be used to calculate any ψ for any model for use with the COOL algorithm.
(ii) The data are not transformed or manipulated prior to analysis
(iii) No normalizations are required; in particular, no data from other experiments are required
(iv) Offset in the current scale does not introduce bias.
(v) All of the data are used. There is no requirement that the experimenter truncate the data at some point.
(vi) Confidence regions may be calculated

Of course, there are other examples of statistically reasonable approaches to this problem. They are, however, remarkably sparse, considering the considerable mathematical sophistication required for any treatment of complex kinetic schemes studied by more complex voltammetric techniques. This general issue has been treated recently by Rusling [6]. From the point of view of the electrochemist,k focusing on the experiment, the features which distinguish this approach from those which are superficially comparable are the following. First, the COOL algorithm provides a uniform treatment for all mechanisms and all pulse

voltammetric experiments. Second, the separation of the linear and nonlinear parts of the problem according to eq. (16) not only avoids irritating experimental problems (the electrode area need not be known, for example), but is also efficient. Thus interesting problems can be solved in 'real-time', that is, times no longer than a few minutes Third, perhaps because the nonlinear problem is dealt with directly rather than through quadratic approximation near the optimum point, the application is surprisingly robust. The experimenter needs to provide only initial estimates of the parameters and the step sizes for the initial simplex. Even silly initial guesses do not significantly slow the approach to the optimum, and there seems to be no problem with false optima Thus it is a useful rather than a dangerous tool in the hands of a naive experimenter.

Determination of ψ in complex cases

Before presenting applications to kinetic problems we describe briefly the techniques employed to obtain the dimensionless current, ψ, for cases more complicated than the simple example of eq (15)

For any first order system and experiments with only stepwise changes in potential the dimensionless current function can be expressed in the form of an integral equation as

$$\psi(t) = \theta(t) - \phi(t)\int_0^t \left[\psi(y)/(t-y)^{1/2}\right] \mathrm{d}y \tag{18}$$

where θ and ϕ are functions of time. This is solved numerically using a simple linear quadrat re formula to yield an expression of the form

$$b_m = \left\{k_1(l, t) - \sum_{i=1}^{m-1} b_i s_j'\right\}/k_2(l, t) \tag{19}$$

where b_m is the estimate of $\psi(t)$, b_i is the estimate of $\psi(t)$ at $t = it_p/l$, t_p is the time over which potential is held constant, l is the number of subintervals employed by the quadrature in the interval t_p, $s_j' = j^{1/2} - (j-1)^{1/2}$, and $j = m - i + 1$.

EXAMPLES OF QUESTIONS ARISING IN DATA ANALYSIS

We now turn to the discussion of three examples of questions which arise in the analysis of kinetic data.

(i) For slow charge transfer, is the charge transfer coefficient (α) a function of potential?

(ii) For cases in which two electrons are transferred, is it necessary to consider both charge-transfer rate processes in the model?

(iii) For two charge transfers coupled by chemical reaction, under what conditions does the chemical cross reaction need to be considered?

We consider each of these questions in turn, keeping in mind the double objective of elucidating mechanism and measuring the values of kinetic parameters

Is the charge transfer coefficient (α) potential-dependent?

Modern theories of adiabatic charge transfer predict an explicit dependence of the rate on such parameters as the energy of reorganization of the molecule in going from the initial state to an excited state and from the excited state to the final state These theories predict that the rate of charge transfer should depend exponentially on a quadratic function of potential. By examination of eq. (3) it can be seen that this is equivalent to predicting that α depends linearly on potential

By comparing the theoretical treatment of Marcus with the phenomenological treatment of eqs (2)–(4) [7], one obtains

$$\alpha = 1/2 + (nF/4\lambda_a)(E - E^{\circ\prime} - \phi_2) \tag{20}$$

in which λ_a is the potential-independent standard free energy of activation and ϕ_2 is the potential drop across the diffuse charge layer in the electrolyte solution near the electrode. The experimental objective, then, is to test the proposition that for an appropriately constrained set of reactions the quantity α of eqs. (3) and (4) has the form given in eq. (20). It should be explicitly noted that eq. (2) does not display activity coefficients. Because

charge transfer necessarily involves change in net charge, the activity coefficients of reactant, product, and transition state will in general be different Provided that they are potential independent, activity considerations should not confound efforts to test relation (20) by the analysis of current–potential data

Consider first the graphical method of analysis based on eq (6), which assumes that the charge transfer coefficient is independent of potential. If instead α has the form of eq. (20), then a plot of $\ln(k_f)$ vs. E according to eq (3) will be curved. A common way of using this method to test eq. (20) is to define α by

$$\alpha = -(1/nf)\, \mathrm{d}\left[\ln\left(k_f(E)\right)\right]/\mathrm{d}E \qquad (21)$$

Then the slope of the curve $\ln(k_f(E))$ vs. E is determined numerically to give values of $\alpha(E)$, which are then plotted against E to test eq (20) and obtain the value of the coefficient of potential.

The values of α obtained point by point are obtained from the curves of Fig. 2 according to eq. (21) and plotted against potential as shown in Fig 3. Clearly the result for $E^{r}_{1/2} = -0.961$ V is 'best', that is, it is linear over the range $0.1 < i/i_d < 0.9$. By choice of range in each case, a slope, $\partial(\alpha nf)/\partial E$, can be determined. For the lines in Fig 3, the slopes are 26 1, 9.6, and 2 1 V^{-2} for $E^{r}_{1/2} = -0.961$, -0.971, and -0.981 V, respectively. A predicted value in this case is 70 V^{-2} [8] which, considering the uncertainties involved, agrees reasonably well with the value of 26 V^{-2}. Thus, a

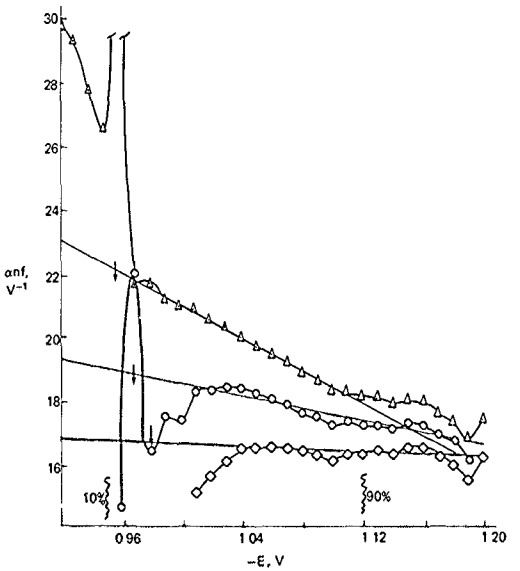


Fig. ^ Plot of α values obtained from Fig 2 by means of eq (7) against potential to test eq (20). Symbols as Fig. 2, range $0.1 < i/i_d < 0.9$ shown, and $E^{r}_{1/2}$ values indicated by arrows.

value of $E^r_{1/2}$ which is in error on the positive side (for a reduction) not only improves the linearity of the result but also yields a spurious potential dependence of the charge transfer coefficient.

The literature on this question is confused. There are papers describing charge transfer coefficients which are potential-dependent. Some of these, which report results in accord with theory, have been refuted. These potential dependencies have been inferred from data by methods similar to those described above, or by methods somewhat more sophisticated but containing the same fundamental flaws We have examined this question in detail using the COOL algorithm for analysis of data [9] Two models were employed, one equivalent conceptually to that described by eq (15) (but incorporating factors to take into account the interfacial charge distribution), thus having three parameters, and an alternative one with the formulation

$$\alpha = \alpha_0 + \alpha_1 nf\left(E - E^r_{1/2}\right) \qquad (21)$$

as suggested by eq. (20). Typical fits according to the four-parameter model yielded values of $(1 - r_m)$ and S/N no better than those of the three-parameter model, with a typical value of $\alpha_1 = 0.0002 \pm 0.0002$ (i.e. $I_\alpha = 0.004$). (Here S is the slope, a, of eq. (16) and N is the root mean square deviation of the experimental points from the optimal theoretical curve.) The predicted value of α_1 is 0.013, or $\alpha_1 n^2 f^2 = 79$ [8]. We conclude that α does not depend on potential, the experimental evidence provided by these authors notwithstanding.

The power of the COOL algorithm in this analysis rests in part on the identification of $E^r_{1/2}$ as a parameter. The resilience of the analysis to changes in the laboratory reference potential is illustrated in Fig 4, which presents results for four nominally identical experiments Rather than just presenting the confidence intervals, a more extensive calculation was employed, to compute the boundary of the confidence region in each of the three planes of the parameter space. The deviation of the optimal value of $E^r_{1/2}$ for the curve of panel

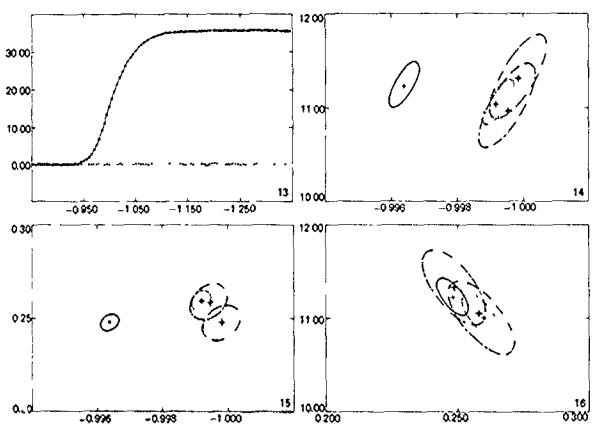


Fig 4 Normal pulse voltammogram for 1 mM Zn(II) in 0.3 M KNO_3, SMDE, medium drop size, potentials vs. saturated calomel electrode. Panel 13 Experimental points (○), best-fitting theoretical curve (———) and residuals (△). Panels 14, 15, 16· confidence regions at 95%; (———) data of panel 13; (· ·), (– · – ·), (· – · – ·) are for nominally identical experiments Optimal values (+) The axes are: (13) $i(\mu A)$ vs E(V), (14) $k°(10^{-3}$ cm/s) vs. $E^r_{1/2}$(V); (15) α vs. $E^r_{1/2}$(V); (16) $k°(10^{-3}$ cm/s) vs α

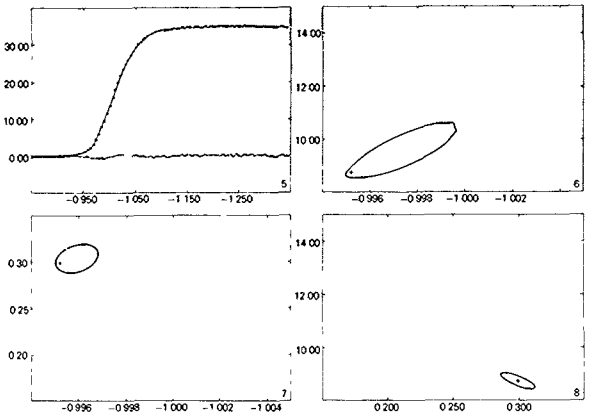


Fig 5 Normal pulse voltammogram analyzed with independent value of $E^{r}_{1/2}$ Panels 5–8 are equivalent to panels 13–16 of Fig 4, with the exception that $E^{r}_{1/2}$ is constrained to be 4 mV positive with regard to the optimal value found in Fig 4

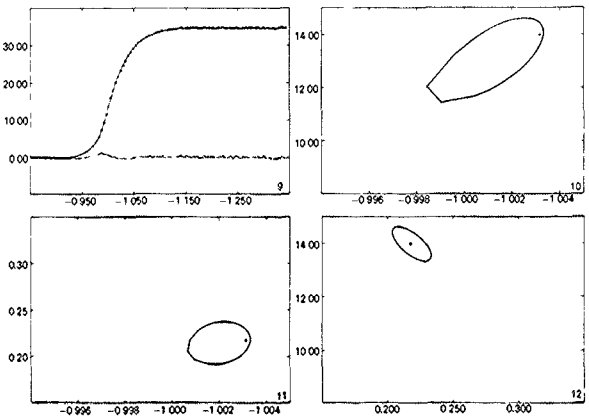


Fig. 6 Normal pulse voltammogram analyzed with independent value of $E^{r}_{1/2}$ Panels 9–12 are equivalent to panels 5–8 of Fig 5, respectively, with the exception that $E^{r}_{1/2}$ is constrained to be 4 mV negative with regard to the optimal value found in Fig 4

13 (of Fig. 4) is a systemic error caused by a change in the laboratory reference potential. This can be seen to have no effect on either the optimal values of the other parameters or the size and shape of the confidence regions.

The conventional procedure employs an independently measured value of $E'_{1/2}$, $(E'_{1/2})_m$. The effects of errors in this value on the analysis can be tested by analyzing the data of Fig. 4 by means of the COOL algorithm, but fixing the value of $E'_{1/2}$. In Fig. 4 the outlying value of $E'_{1/2}$ is about 4 mV from the mean value. Thus we analyze the data for one of the nominally identical experiments of Fig. 4, with the value of $E'_{1/2}$ fixed at $(E'_{1/2})_m = \hat{E}'_{1/2} + 0.004$ (V), where $\hat{E}'_{1/2}$ is the optimal value found in the optimization presented in

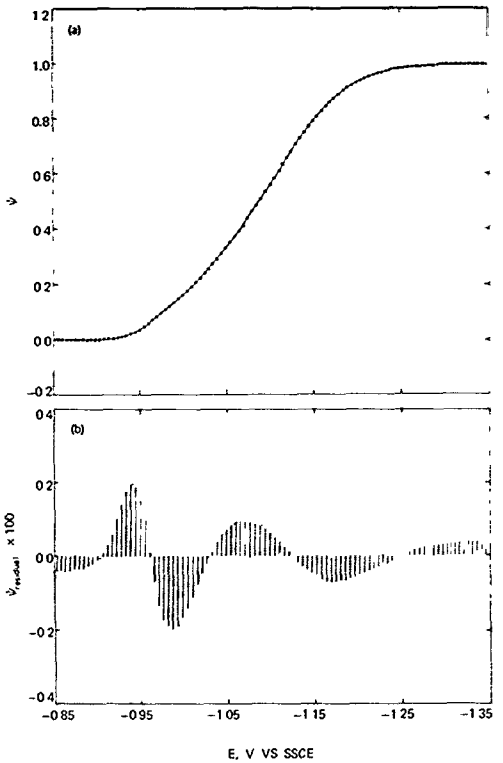


Fig. 7. (a) Normal pulse voltammogram calculated for two-step mechanism with $D_O = 6.6 \times 10^{-6}$ cm^2 s^{-1}, $D_R = 1.6 \times 10^{-5}$ cm^2 s^{-1}, $E^{0\prime} = -0.591$ V, $k^\circ_{s1} = 3.5 \times 10^{-3}$ cm s^{-1}, $k^\circ_{s2} = 7.1 \times 10^{-2}$ cm s^{-1}, $\alpha_1 = \alpha_2 = 0.40$ (○), optimal theoretical curve for one-step mechanism, $E'_{1/2} = -0.5966$ V, $an = 0.404$, $k^\circ_s = 3.51 \times 10^{-3}$ cm s^{-1} (———), (b) residuals, $\psi_r = (\psi_{np} - \psi_{th})/\psi_{np}$, where ψ_{np} is the current calculated for the two-step mechanism, and ψ_{th} the current for the optimal one-step mechanism $t_p = 5$ ms

Fig. 4. The value is fixed by setting the initial step size to zero. The result is shown in Fig. 5. The optimal theoretical curve (calculated from eq. (16) using $(\hat{\alpha}, \hat{k}, (E'_{1/2})_m)$) now displays noticible systematic variation from the experimental result (Fig. 5, panel 5). In addition, the confidence regions about $(\hat{\alpha}, \hat{k}, (E'_{1/2})_m)$ are substantially unsymmetrical. Similar results are obtained when $E'_{1/2}$ is fixed at the value $(E'_{1/2})_m = \hat{E}'_{1/2} - 0.004$ (V), as shown in Fig. 6. The change in optimal value of k^0_s resulting from change in $E'_{1/2}$ is expected, for k^0_s is just the rate constant at $E = E^{0\prime}$ (cf. eqs. (3), (4), and (10)). More striking is the large change in α. This demonstrates that $E'_{1/2}$ is properly a parameter of the experiment, and thus fixing $E'_{1/2}$ at some value determined in another experiment precludes the possibility of accurate kinetic analysis.

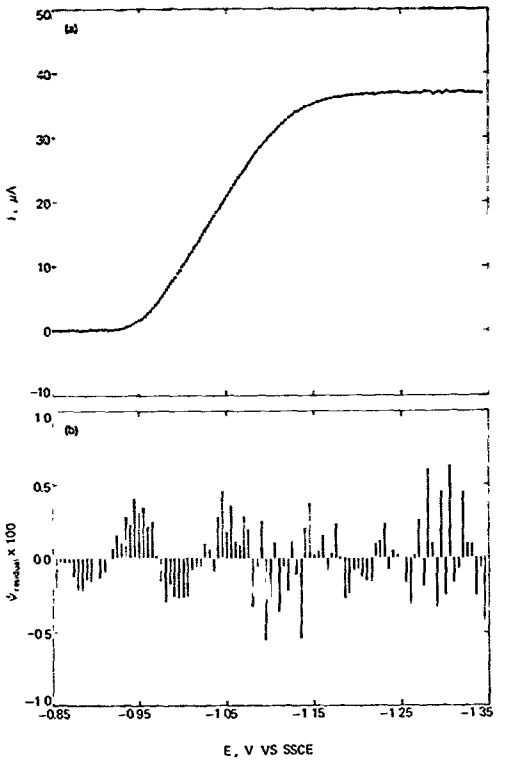


Fig. 8 As Fig. 7, but points are for an experimental voltammogram obtained under conditions nearly identical to those which yielded the values of rate parameters for the two-step mechanism.

Is it necessary to consider more than one charge transfer?

In the case discussed above [9] a second issue involves the detailed mechanism of charge transfer when $n = 2$. Is it necessary in that case to use the model incorporating two successive charge transfers and thus six rate parameters [4]? Or, inverting the question, can anything be learned about the faster of the two steps by this type of analysis? The issue here is more complex, for the more elaborate mechanism can be expected to exhibit non-monotonic changes in the shape of the response, and thus to produce a non-random pattern of residuals However, systematic errors in the experiments can have the same effect. Thus the non-random distribution of residuals cannot be attributed automatically to significant rather than adventitious or trivial failures of the model. A further problem arises when the model is available only numerically (cf. eq. (19)). The interpretation of humps or bumps in the response as arising from specific features of mechanism (the phrenologic school of kinetics), always risky, is foolhardy in this case, as minor changes in the values of parameters can produce quite striking changes in the appearance of the response

A typical illustration is given in Figs 7 and 8. Fig 7 displays the analysis of a calculated voltammogram The voltammogram was calculated from five parameters for two, one-electron transfers. Both rate constants are referred to the standard potential for the overall two-electron process The calculated voltammogram was then analyzed according to a model for a single slow electron transfer with $n = 2$ (three parameters). The obvious pattern in the residuals can be compared with those of the experimental example of Fig. 8 The experimental conditions of Fig 8 are nearly identical to those which produced the data on which the theoretical calculation of Fig 7 (5 parameters) is based In the experiment, noise and systematic experimental artifacts obscure the interpretation. The pragmatic conclusion is that the more simple model explains adequately the variance in the data This leaves open the question of whether the data contains information about the faster electron transfer step This might be obtained by analyzing the data of Fig. 8 according to the appropriate model for two successive slow electron transfers.

Is it necessary to consider more than one homogeneous reaction?

A much-studied mechanism is the so-called ECE sequence

$$O_1 + n_1e \rightleftarrows R_1 \qquad E_1^{\circ\prime} \tag{22}$$

$$R_1 \xrightarrow{k} O_2 \tag{23}$$

$$O_2 + n_2e \rightleftarrows R_2 \qquad E_2^{\circ\prime} \tag{24}$$

in which the heterogeneous charge transfers are linked by an intermediate homogeneous reaction, here taken to be irreversible. When $E_2^{\circ\prime} \gg E_1^{\circ\prime}$, reaction (24) is more favored than (22), and so the two reactions occur together at the potential for reduction of O_1. The reason for interest in this scheme is its potential catalytic significance. Many organic compounds, especially in aqueous solution, display the response expected for this sort of mechanism However, when $E_2^{\circ\prime} \ll E_1^{\circ\prime}$, the homogeneous reaction

$$O_2 + (n_2/n_1)R_1 \rightleftarrows O_1 + (n_2/n_1)R_2 \tag{25}$$

is highly favored and provides an alternative route to that of eq (24) for the transfer of electrons to R_2. The questions then arise, under what conditions is reaction (24) important in the overall process, and when it is important, can it be detected? Or to phrase the question somewhat differently, under what conditions does the model consisting of reactions (22)–(24) explain adequately the response?

A classical example is the reduction of *p*-nitrosophenol [10] Experimental results for *p*-nitrosophenol are presented in Fig 9 together with the optimal theoretical curves for the simple model comprising eqs. (22)–(24). To the eye it would appear that the correspondence is adequate For these data $r_m = 0.998$, and typical values of I_k are 0.2–1 s^{-1}, depending on the experimental conditions. There is considerable advantage in using this method to determine values of k, for the addition of the second order reaction, eq. (25), to the model complicates the mathematical formula-

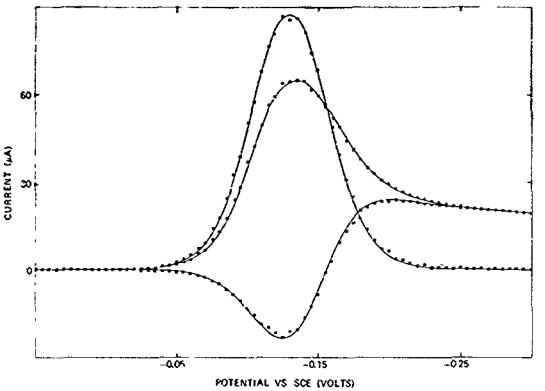


Fig. 9 Forward, reverse, and net experimental currents (○) and optimal theoretical curves (———) for square wave voltammetric reduction of *p*-nitrosophenol in 20% (v/v) aqueous ethanol, 0.1 *M* acetic acid, 0.1 *M* potassium acetate, 0.1 *M* potassium nitrate, 0.0005% Triton X-100 $\hat{E}^{\circ}_{1/2} = -0.1334$ V, $\hat{k} = 2.09\ s^{-1}$, $\epsilon_2 = 0$

tion enormously. Are the optimal values of k and the associated confidence regions reliable, even if the model is 'wrong' in that it does not incorporate eq (25)?

Unfortunately there does not appear to be a general answer to this question, even if it is restricted to problems involving only two parameters. For the case of Fig. 9 there is reason to believe on empirical grounds that this question has an affirmative answer [11]. This is an unsatisfactory conclusion, in that it relies on an intuitive argument based on example, rather than on objective criteria. The present statistical approach deals only with the description of phenomena, and thus cannot deal directly with questions of this type. It could be a useful tool, however, for computational investigations of this and related questions. Although the results could only serve as a guide, computation is so much less expensive than experimentation that this could well be the most efficient way to proceed with interpretation of kinetic measurements.

CONCLUDING REMARKS

These three examples raise issues commonly addressed ad hoc and qualitatively in electrochemical kinetic studies. The optimization technique presented here provides a rigorous evaluation of the correspondence between model and data in near-real-time. This may be used to discriminate between alternative models and to examine the power of the data to yield mechanistic information.

In favorable cases, the algorithm may be used to identify and quantify a minor feature of the mechanism. Equally important, and more difficult to demonstrate convincingly, this approach may

be used to show the absence of an effect Finally, after suitable computational investigation of various types of models, it may permit one to treat rather complex cases using the most simple model which incorporates the feature about which information is sought and which yields an acceptable signal-to-noise ratio (S/N).

ACKNOWLEDGEMENTS

This work was sponsored by the U.S National Science Foundation under Grant No. CHE 8521200. The author thanks John O'Dea and Winston Go for helpful discussions.

REFERENCES

1 J Tafel, Über die Polarisation kathodischer Wasserstoffentwicklung, *Zeitschrift fuer Physikalische Chemie*, 50A (1905) 641–712
2 R Brdicka, To the sixtieth birthday of Professor J Heyrovský An account of his scientific achievements to 1950, *Collection of Czechoslovak Chemical Communications*, 15 (1950) 691–698
3 J H Christie, E.P Parry and R A Osteryoung, The use of normal pulse polarography in the study of electrode kinetics, *Electrochimica Acta*, 11 (1966) 1525–1529
4 W Go and J Osteryoung, Alternative interpretations of kinetic data the reduction of zinc at mercury electrodes, *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, 233 (1987) 275–281
5 J. O'Dea, K Wikiel and J Osteryoung, Squarewave voltammetry for ECE mechanisms, *Journal of Physical Chemistry*, 94 (1990) 3628–3636.
6 J.F Rusling, Analysis of chemical data by computer modeling, *Critical Reviews in Analytical Chemistry*, 21 (1989) 49–81
7 D M Mohilner, Double layer effects in the kinetics of heterogeneous electron exchange reactions, *Journal of Physical Chemistry*, 73 (1969) 2652–2662
8 K Matsuda and R Tamamushi, Potential-dependent transfer coefficients of the Zn(II)/Zn(Hg) electrode reaction in aqueous solutions, *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, 100 (1979) 831–839
9 W Go, J O'Dea and J Osteryoung, Squarewave voltammetry for the determination of kinetic parameters, *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, 255 (1988) 21–44
10 R S Nicholson and I Shain, Theory for stationary electrode polarography for a chemical reaction coupled between two charge transfers, *Analytical Chemistry*, 37 (1965) 178–190
11 J O'Dea, J Osteryoung and T Lane, Determining kinetic parameters from pulse voltammetric data, *Journal of Physical Chemistry*, 90 (1986) 2761–2764

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 155–167
Elsevier Science Publishers B V., Amsterdam

Relating chromatographic data to measurements of wheat quality: Case studies in dimension reduction

D G. Simpson *

Department of Statistics and Institute for Environmental Studies, University of Illinois, Champaign, IL 61820 (U S.A)

S Guo and J. Sacks

Department of Statistics, University of Illinois, Champaign, IL 61820 (U S A)

J.A. Bietz, F Huebner and T. Nelsen

USDA Agricultural Research Service, Midwest Area, 1815 North University Street, Peoria, IL 61604 (U S A)

(Received 7 November 1989, accepted 12 July 1990)

Abstract

Simpson, D G , Guo, S , Sacks, J , Bietz, J A , Huebner and F , Nelsen, T , 1991 Relating chromatographic data to measurements of wheat quality case studies in dimension reduction *Chemometrics and Intelligent Laboratory Systems*, 10 155–167

Fractionating wheat proteins by reversed phase high-performance liquid chromatography yields extremely complex chromatograms The data they contain may relate to many characteristics of milled wheat such as the volume of a loaf of bread or the texture of the dough produced, but such relationships are not readily apparent from the raw data We report our experiences with two dimension reduction techniques that are widely cited in the chemometrics literature principal component analysis and partial least squares (PLS) Each of these methods replaces the original observation vectors by weighted averages of them components, where the weights are selected according to a data dependent criterion The analysis proceeds by operating on these weighted averages rather than the original, high-dimensional data In order to elucidate properties of significance tests and other inferences, we focus on the special case where only one factor is selected We show how to use simulation to compute the appropriate significance level of the regression on the PLS scores The common technique of using the F distribution to compute significance levels for PLS regression can be an extremely liberal procedure The interpretation of PLS weights requires considerable care

INTRODUCTION

With the advent of modern high-performance liquid chromatography (HPLC) for analyzing proteins from samples of wheat, there is considerable interest in developing the statistical technology for relating these chromatographic fingerprints to the attributes of milled wheat [1] Viewing a wheat

0169-7439/91/$03 50

sample as the basic experimental unit, it is typical that the number of independent observations (wheat samples) is small, but the number of characteristics available for study on each observation is large. For instance, there appears to be a multitude of active sites on the chromatogram that might potentially be included in a model for predicting various attributes of the milled wheat. Standard statistical methodology, e.g. multiple linear regression, cannot be applied directly to the raw data because the nominal dimension, that is, the number of measurements on each experimental unit, exceeds the number of independent observations, leading to ill-posed estimation problems.

Dimension-reduction techniques are based on the premise that much of the information collected on each observation is redundant, and that some lower-dimensional transformations of the data contain most of the information. If such transformations can be discovered, then one can in principle use standard statistical methodology on the constructed lower-dimensional data. Two dimension-reduction methods that are widely cited in the chemometrics literature are principal component analysis [2] and partial least squares (PLS) [3]. After describing these methods briefly, we illustrate their use on typical wheat protein chromatographic data, and offer some preliminary observations on the viability of these methods for investigating the relationships between HPLC patterns and attributes of milled wheat.

In principal component regression the predictor variables are reduced to a smaller number of projections that account for most of their variation [2]. Because the projections are selected independently of the response variable, this procedure has the advantage that classical regression theory may be applied to test for significance, to compute prediction intervals, and so on. On the other hand, there is no guarantee that the principal component projections contain adequate information about the relation between the predictor variables and the response. PLS has been proposed as a method for selecting projections that are more informative about the relationships between two sets of variables. It makes use of the covariances to select projections that account for the joint variation in the two sets. PLS regression, in particular, selects one-dimensional projections of the predictor variables that have large covariance with the response [4,5]. Because the projections depend on the response as well as the predictor variables, classical regression theory does not strictly apply. For instance, we demonstrate that comparing the PLS F test for the regression to the F distribution can be an extremely liberal procedure.

Both the principal component projections and the PLS projections are affected by the choice of scales for the different components of the raw data. Changing the scales differentially can drastically change the nature of the projections selected. For this reason many authors suggest standardizing the raw data componentwise prior to further analysis. In our examples we center but do not standardize, because the HPLC measurements at different sites on the chromatogram are in the same unit, and a change of units would affect them all simultaneously. Principal component and PLS factors are unaffected by common scale transformations of the raw components of the data, e.g. the results would be the same if we chose to express absorbance in different units. Applying a nonlinear transformation (e.g. a logarithm) does affect the results, and the selection of an appropriate transformation is an issue for further research. Such preprocessing of the data is often an important ingredient to the success of a dimension-reduction technique [6].

There are different versions of PLS and different recommendations about how to choose the number of projections for regression [7]. Our primary interest is in how to interpret the projections and in how to make inferences. For this reason we sidestep the other issues and focus on the special case where only one PLS projection of the predictor variables is to be selected. In our regression example this seems appropriate. Although each observation has many components, there are few observations, and one explanatory variable ought to be sufficient. The important issue of bias due to variable selection is clearly of broader scope, and our case study may be viewed as a telling example.

PRINCIPAL COMPONENTS

Principal component analysis is a method of investigating a multivariate dataset by looking at orthogonal one-dimensional projections [2]. By multivariate we mean that each experimental unit has a number of measurements associated with it. For instance, a given sample of wheat might be subjected to several different assessment of quality, in which case the different quality measurements constitute different components of the multivariate quality vector for that sample. Similarly, the HPLC pattern might consist of absorbance at 50 equally spaced points on the time scale, in which case the 50 measurements comprise a 50-dimensional vector associated with the given wheat sample. The usual goal in principal component analysis is to replace the large number of components on the original scale with a small number of new components consisting of the orthogonal projections that account for the largest portion of the variation in the dataset at hand.

A direction vector is a vector of unit length, where the length of an arbitrary vector $x = (x_1, \ldots, x_p)'$ is given by

$$\|x\| = \sqrt{x'x} = \sqrt{x_1^2 + \cdots + x_p^2}$$

If $\|x\| \neq 0$, then $u = x/\|x\|$ is the direction vector for x. If y is another vector with the same number of components, then its projection on x is

$$\frac{y'x}{\|x\|} = \frac{y_1x_1 + \cdots + y_px_p}{\|x\|} = y'u$$

The number $y'u$ is the component of y in the direction of u. For example, suppose u is the direction vector $(1, 0, 0, 0, \ldots, 0)'$. Then $y'u = y_1$, the first component of y.

A key idea in dimension-reduction is the projection of a dataset. Suppose a dataset consists of n vectors $x_1, \ldots, x_n$ each having p components.

$$x_i = (x_{i1}, \ldots, x_{ip})', \qquad i = 1, \ldots, n$$

Projecting each of these vectors on a direction vector u yields a new dataset of one-dimensional observations, the components of $x_1, \ldots, x_n$ in the direction of u:

$x_1'u, \ldots, x_n'u$

Given a set of numbers $\{x_1, x_2, \ldots, x_n\}$, a common measure of variation is the sample variance about the mean:

$$V(x_1, \ldots, x_n) = \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}$$

where

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}$$

The first principal component is obtained by finding the direction u_1 such that the projection of the dataset has maximal sample variance, that is,

$$V(x_1'u_1, \ldots, x_n'u_1) = \max_{\|u\|=1} V(x_1'u, \ldots, x_n'u)$$

The second principal component is obtained by maximizing the variance of the projections on directions orthogonal to u_1. In general, the kth principal component maximizes the variance of the projections on directions orthogonal to $\{u_1, \ldots, u_{k-1}\}$.

In using this construction for dimension-reduction the hope is that most of the relevant variation is accounted for by the first few principal components. For instance, it might be that most of the variation in a set of chromatograms is accounted for by a few peaks.

A number of software packages and programs have routines for principal components analysis including BMDP, Minitab, SAS, and Unscrambler. In addition, programs that perform the eigenvalue decompositions needed to get the principal components are widely available, e.g., LINPACK and S.

PLS PROJECTIONS

Principal component analysis attempts to produce a small number of directions that capture most of the variation in a single set of vector observations. An alternative dimension-reduction has been proposed in the chemometrics literature when the goal is to relate two sets of vectors. Given pairs of vectors $(x_1, y_1), \ldots, (x_n, y_n)$ the idea is to find directions u_1 and v_1 such that the projections of $x_1, \ldots, x_n$ on u_1 and the projections of $y_1, \ldots, y_n$ on v_1 have large coincident variation.

This is the basis of the PLS algorithm [3], which uses the projections on these directions as the input variables for least-squares regression.

Specifically, for pairs of numbers $(x_1, y_1), \ldots, (x_n, y_n)$ the sample covariance is given by

$$C(x_1, \ldots, x_n; y_1, \ldots, y_n) = \frac{\sum_{j=1}^{n} (x_j - \bar{x})(y_j - \bar{y})}{n}$$

and provides a measure of the extent to which the x and y values tend to vary together. PLS uses the covariance as a criterion for selecting the projection directions u_1 and v_1.

$$C(x_1'u_1, \ldots, x_n'u_1; y_1'v_1, \ldots, y_n'v_1)$$

$$= \max_{\|u\|=1} \max_{\|v\|=1} C(x_1'u, \ldots, x_n'u, y_1'v, \ldots, y_n'v)$$

As in principal component analysis, one can iterate the procedure and select additional direction vectors that maximize the covariance in directions orthogonal to previously selected projections. PLS has almost invariably been described in algorithmic form, but Frank [4] and Hoskuldsson [5] have pointed out that the algorithm selects covariance maximizing directions.

PLS provides a simultaneous dimension-reduction for x and y. For the special case with either x or y one-dimensional the solution can be written down explicitly Suppose $y_i = y_i$, a scalar, for $i = 1, \ldots, n$. Then the solution is given by

$$u_1 = \frac{\sum_{j=1}^{n} (y_j - \bar{y})(x_j - \bar{x})}{\left\| \sum_{k=1}^{n} (y_k - \bar{y})(x_k - \bar{x}) \right\|}, \qquad v_1 = 1$$

where $\bar{x}$ is the vector of componentwise sample means for $x_1, \ldots, x_n$. In this case u_1 may be recognized as the direction of the vector of slopes from the least-squares regression of x on y. In general the PLS algorithm is easily programmed. It has been implemented in the program Unscrambler, which is available for IBM-PC compatibles. We have programmed PLS regression in S and FORTRAN.

PLS bears a resemblance to canonical correlation analysis (CCA), in which projections of $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ are selected to maximize correlation [8]. The CCA directions are the ones with the strongest linear association for the data at hand, whereas the PLS directions have the highest coincident variation. Unfortunately, CCA is ill-posed in the present setting where the nominal dimension of the data exceeds the number of independent observations. One can achieve perfect sample correlation by weighting on any $n-1$ linearly independent columns of the data matrix.

REGRESSION ON CONSTRUCTED COMPONENTS

Consider the case where y has only one component, whereas x is of high dimension. This is the case in the examples below, where y is a particular attribute of milled wheat and x is the HPLC determination of protein composition Recall that the regression of y on x is ill-posed if the number of components of x exceeds the number of observations PLS attempts to circumvent this problem by regressing y on the linear combinations of x selected according to the maximum covariance criterion. Similarly, principal component regression involves regressing y on the linear combinations of x selected by principal component analysis. In each case one uses the constructed variables $z_1 = x'u_1$, $z_2 = x'u_2$, and so on as the regression variables for predicting y. In the case of principal components the ordinary theory of multiple linear regression can be used to compute standard errors and prediction intervals, because no information about y was used in the construction of z_1, z_2, etc. In the case of PLS the usual theory is inappropriate, because of the dependence of the constructed $z_1, z_2, \ldots$ on y. Further discussion of this point is given below. It is clear that ordinary principal component regression can fail if the linear combinations of x with the largest variability have little relation to y. PLS is an attempt to avoid this pitfall by selecting linear combinations that vary together with y.

A CLASSIFICATION EXAMPLE

The first example is a dataset consisting of HPLC runs of 43 samples of durum wheat. There

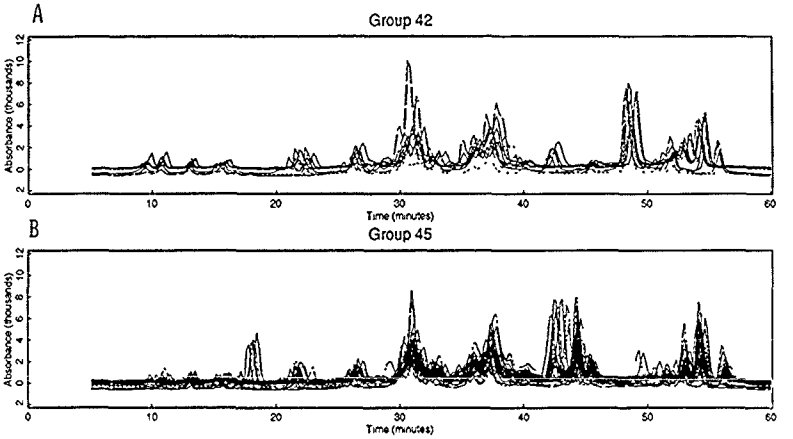


Fig. 1 Chromatograms for 43 samples of Durum wheat (A) Group 42, (B) group 45

are two groups labeled '42' and '45' depending on which of two proteins is present at a certain locus on the chromosome, as determined by electrophoresis It has been found that the presence of protein '42' indicates a variety with weak pasta quality, whereas protein '45' indicates strong variety. This example offers a test case for whether the dimension-reduction techniques can 'discover' this relationship. The experimental technique for the HPLC is described in ref. 1.

Figs. 1A and B show the chromatograms (absorbance versus time) for the group 42 and group 45 samples. Each chromatogram contains 330 equally spaced Measurements over the range 5–60 min. The most striking difference is that the group 42 samples have a sharp peak at 49 min that is absent from the group 45 samples. Conversely, group 45 has a large peak at 44 min that is absent in group 42. Presumably this difference in HPLC results for the two groups is a reflection of the two proteins identified by electrophoresis, Burnouf and Bietz [1] cited it as evidence that HPLC could be used to identify strong and weak varieties. There is a minor peak evident at 18 min for group 45 but not for group 42. This peak was present only in five analyses of one variety (Langdon), so its appearance in group 45 seems coincidental.

As the difference between the two groups is obvious in Fig 1, any reasonable procedure ought to be able to recover it. We employed composite classification rules in which we first selected one or two orthogonal weight vectors by principal components or PLS, and then applied Fisher's linear discriminant rule [9] to the scores obtained by projecting the data on the weight vectors. The effect of this composite rule is to select a single direction vector, say w, that is a linear combination of the original direction vectors selected by principal component analysis or PLS The composite discriminant rule is equivalent to assigning a candidate chromatogram to the group whose mean projection on w is closest to its own.

Fig. 2A shows the first two eigenvectors from principal component analysis. Fig. 2B shows the first PLS weighting vector and the weighted average of the first two principal components selected by the two-dimensional PC linear discriminant. The components of a weight vector $u =$

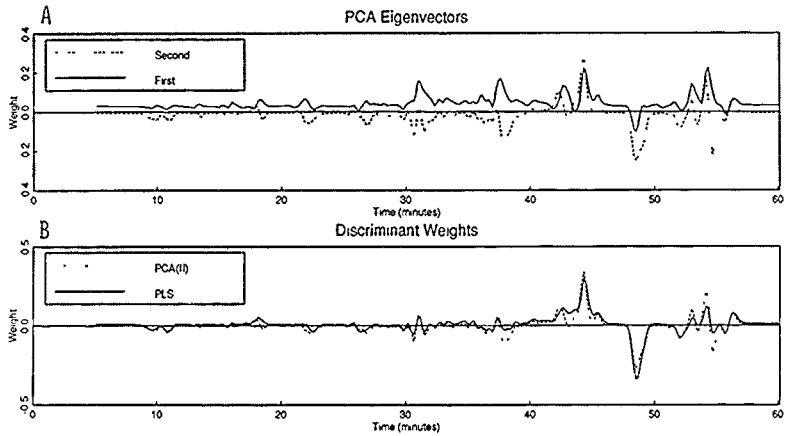


Fig 2 Weight vectors for centered chromatograms of Durum wheat samples (A) First two eigenvectors from PCA, (B) PLS projection and linear discriminant projection based on first two principal components

$(u_1, \ldots, u_{330})'$ give the weights for the time-ordered sites on the chromatogram in the constructed variables

$$z_i = x_i'u = u_1x_1 + u_2x_2 + \cdots + u_{330}x_{330}$$

As described above, the principal component weights do not use the classification information, but simply give the direction of the most variable projection of the chromatograms. On the other hand the PLS weights give the projection direction having the largest covariance with the group labels, coded, for instance, as 0s and 1s (If there were more than two groups we would have to introduce a vector of binary variables for group labeling.) It can be shown that if y is binary the PLS weight vector is simply the direction of the difference between the componentwise averages for the two groups, in the present case, the difference between the mean chromatograms for the two groups.

The first principal component weights in Fig. 2A appear to confound the two peaks noted above with several other sites on the chromatogram. The second component appears to cancel out most of the other sites, allowing us to recover the difference between the two main peaks of interest with a bivariate linear discriminant It is clear from Fig 2B that the PLS factor is weighting primarily on the difference between the two major peaks noted previously. The weighting vector that results from applying the bivariate linear discriminant to the first two principal components is similar to the PLS weighting vector except that the former gives more weight to gliadins eluting beyond 50 min.

Fig 3 is an indication of the effectiveness of the constructed classification variables The vertical axis is the group label. The horizontal axis is the value of the score, $z_i = x_i'u$, for each of the 43 samples. In each plot the vertical line is the cutoff value for the linear discrimant rule, which is given by $(\bar{z}_1 + \bar{z}_2)/2$, where $\bar{z}_1$ and $\bar{z}_2$ are the mean scores for the two groups. The first principal component scores, shown in Fig. 3B, are not very effective for classifying the two groups. Adding the second component reduces the error date dramatically. The PLS scores, shown in Fig 3C, provide a complete separation of the two groups The apparent error rates and leave-one-out cross-

validation (CV) estimates of the error rates [10] are as follows.

Method	Apparent error rate	CV error rate
Principal component analysis (I)	12/43	13/43
Principal component analysis (II)	1/43	2/43
PLS	0/43	1/43

The apparent error rate is known to be optimistic, the CV estimate is generally considered to be more reliable.

When only one PLS projection is selected, applying the linear discriminant rule to the PLS scores is equivalent to using a rule that assigns a new observation to the group whose mean is closest in Euclidean norm [11], that is, it assigns a variety with chromatogram $x = (x_1, \ldots, x_p)'$ to the group with mean vector $\bar{x}_g$ for which $\| x - \bar{x}_g \|$ is smallest This procedure, known as Euclidean distance classification [12,13], has an obvious generalization to several groups

The PLS classification is highly effective in this example, and it identifies the gliadin peaks associated with pasta quality. Classification by PCA can achieve nearly the same results but it requires two components and a bivariate linear discriminant, so it takes a bit more effort. An alternative principal component method is to use SIMCA, which takes the grouping into account by finding separate principal component projections for the different groups in the training data [14] It is not clear that there is much to gain by using more complex methods in the present example In other examples, e g. when there is doubt that all of the observations fall in known groups, other methods might well yield superior results.

A REGRESSION EXAMPLE

The second example concerns a dataset containing measurements on twelve varieties of hard red spring wheat. For each variety we have HPLC

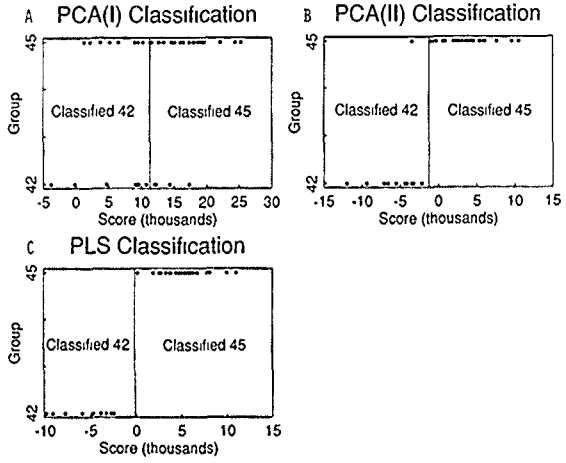


Fig 3 Linear discriminant classification of Durum wheat samples using (A) first principal component, (B) first two principal components, and (C) first partial least-squares projection

results of proteins extracted with 80% ethanol. In addition, a number of different kinds of measurements were made of the physical properties of the grain, its milling properties, and its mixing and baking properties as described by Nolte et al. [15]. We selected three for detailed study. (i) loaf volume, the volume of a baked loaf of bread from a given amount of flour, (ii) mix time, the amount of mixing required for the dough to achieve a certain consistency, and (iii) percentage wheat ash, a measure of the mineral content Loaf volume and mix time were selected because they are known to be related to the proteins of wheat. Ash was selected as a negative control in our data analysis experiment, since it is a variable thought to be unrelated to the protein composition.

Fig 4A shows the chromatograms for the twelve varieties. Each chromatogram contains 514 measurements at the rate of 12 per minute starting at 5 min For the purpose of relating protein content to the various attributes an important issue is the variability at the different sites, which can be seen more clearly from the mean-centered chromatograms in Fig. 4B. For instance, the raw chromatograms in Fig. 4A have a strong peak around 26 min that shows very little variation across samples

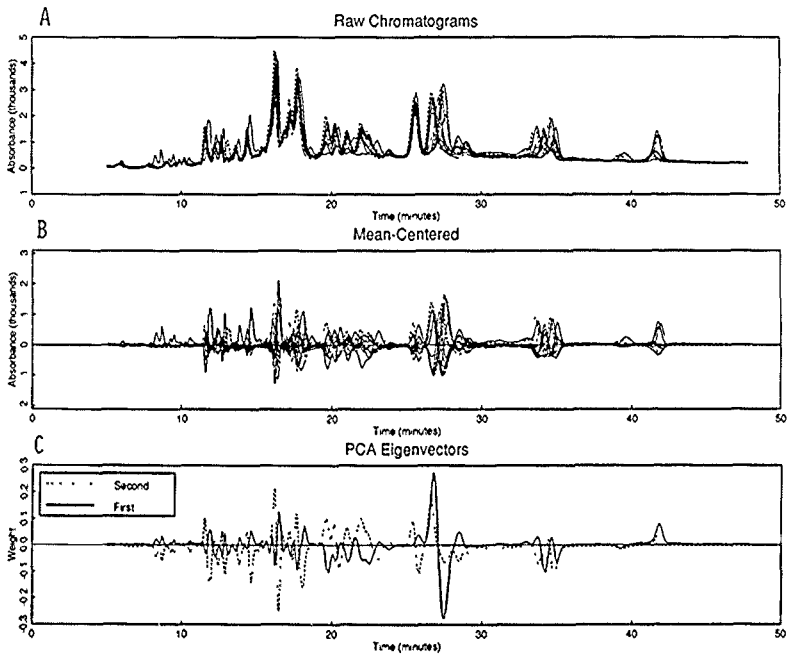


Fig. 4 (A) Chromatograms for twelve samples of wheat grown in Mesa, AZ, (B) mean-centered chromatograms, (C) first two eigenvectors from PCA

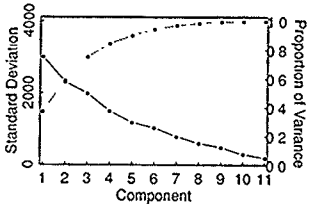


Fig 5. Standard deviations (solid line) and cumulative proportions of variance (dashed line) for principal components with nonzero eigenvalues

and appears only as a small bump in Fig 4B. Centered chromatograms were computed as follows:

1. Compute the vector of componentwise means $\bar{x}' = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{514})$ where $\bar{x}_j$ is the mean of the twelve absorbance measurements for the jth time point.
2. Subtract the components of $\bar{x}$ from the corresponding components of each of the twelve individual chromatographs.

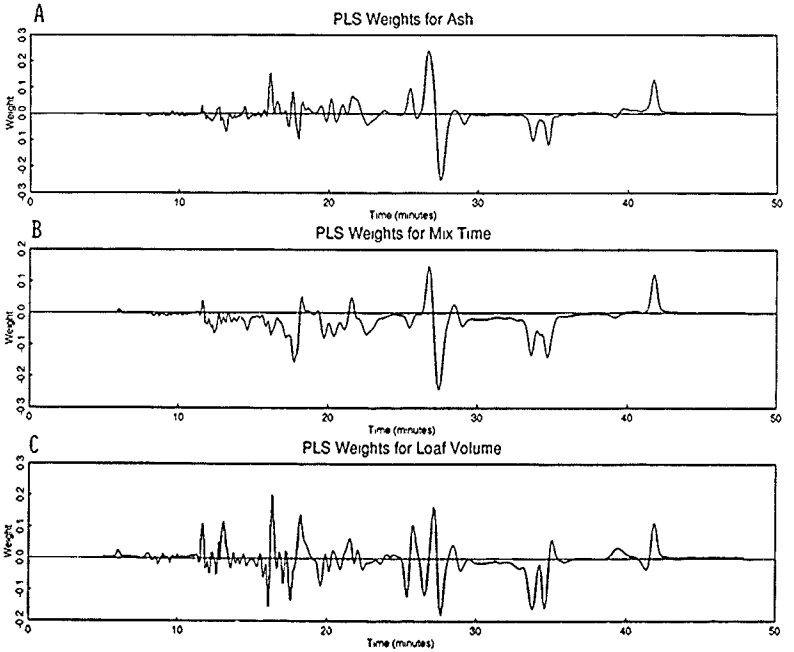


Fig 6 Weight vectors for centered chromatograms of twelve wheat samples. (A) PLS weights for ash (B) PLS weights for mix time, (C) PLS weights for loaf volume

In some instances there may be unusual chromatograms that have a large effect on the mean centering. In such cases it is useful to plot median-centered chromatograms for comparison.

The first two components from principal component analysis are shown in Fig. 4C. The largest source of variation is a peak or pair of peaks eluting at 27–28 min. The first component is essentially a difference across this region of the chromatogram. The second component has contributions from many sites, with no apparent dominant contributor. Three or four components are required to account for the bulk of the variation in the chromatograms. Fig. 5 shows the standard deviations (solid line) and cumulative proportions of total variance (dashed line) for the principal components

We next carried out the PLS computations to relate ash, mix time and loaf volume to the HPLC profiles. Although one can treat linear combinations of the three quality variables using PLS, we treated them one at a time because we wished to compare the predictability of these three attributes using the different methods. Fig. 6a–c show the first PLS weight vectors for ash, mix time and loaf volume. The magnitudes of the weights indicate the relative importance of the different sites on the chromatogram according to the criterion used to select the projection.

We were initially surprised at Fig. 6A for ash, which seemed to indicate that proteins eluting at 27–28 min were important for predicting ash. However, an explanation can be found by comparison with the first principal component in Fig.

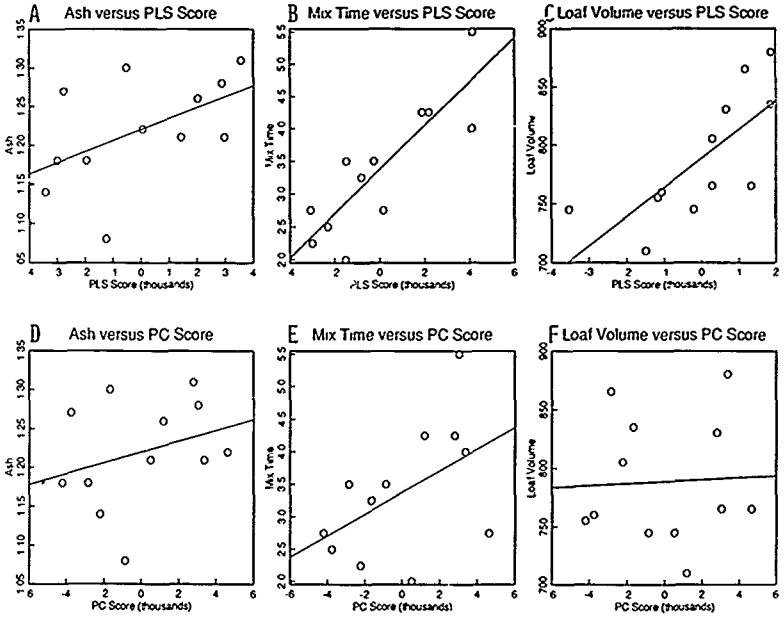


Fig 7 Scatter plots for response variables versus PLS scores (A–C) and PC scores (D–F)

4C, which is quite similar. Recall that the PLS factor is the direction with the largest covariance with ash. It is plausible that ash varies as the total protein content varies (more protein means less ash). Variation in total protein content would in turn be connected with the variation in the principal component. It so happens that peaks in the indicated region show substantially greater variation that the other sites, so these show up in both the PLS factor for ash and the principal component. Contrary to the impression conveyed by Fig. 6A, it is doubtful that the proteins eluting at 27–28 min have any causitive relationship with ash content. Instead, it is quite likely that they receive the highest weights simply because they account for the largest portion of the variability in the chromatograms and, consequently, the variation in total protein content.

Fig. 7A–C are scatter plots of the three response variables ash, mix time and loaf volume versus their respective PLS scores. Fig. 7D–F show the same response variables plotted against the principal component scores. The least-squares lines for regression on PLS and Principal components analysis scores are included as well. All of the PLS scatter plots suggest some positive relationship; however, there is a hidden bias in these plots because each PLS direction was selected to have a relationship with the corresponding response. One manifestation of this bias is inflation of the false positives rate for the so-called *F*-test for the regression on the PLS scores. The *F*-test provides a means for assessing the statistical significance of the apparent regression relationship [16]. For simple linear regression, including as a special case regression on the first PCA component, the test statistic has an F distribution with 1 and $n-2$ degrees of freedom under the zero-slope hypothesis. This is making the standard assumption that the noise terms in the regression model are independent and normally distributed with mean zero and a common variance. For regression on the PLS direction this reference distribution is no longer correct, because of the dependence of the direction on the response variable

To get an approximation to the correct reference distribution we generated 5000 random samples of size 12 from the normal distribution, using

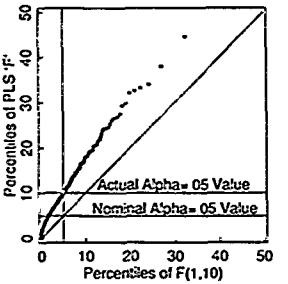


Fig. 8. Percentile–percentile plot of 5000 Monte Carlo-generated F statistics for regression on PLS factor versus the F distribution.

each sample of responses to get the PLS direction for the twelve observed chromatograms in our example. Uniform deviates were generated using a multiplicative congruential generator with modulus $2^{31}-1$ and multiplier 7^5 [17]. Normally distributed deviates were obtained via the Box–Muller transformation. We assumed unit variance for the responses, but this has no bearing on the results, because the PLS direction vector and the F statistic are invariant to scale multiples of the response [11]. For each of the 5000 samples we computed the F statistic for the regression on the PLS direction. The ordered values are plotted against percentiles of the F distribution with 1 and 10 degrees of freedom in Fig. 8. If this were the correct reference distribution the points should fall very close to the diagonal, however, there is a clear upward bias that results from the way the PLS direction is selected. The figure allows us to correct for this bias. For instance, with our design a PLS F of 10 is equivalent to an ordinary F of 5, which has significance level 0.05. If instead we were to look up the PLS F value in the ordinary F table we would erroneously conclude that the significance level is 0.01.

The simulated distribution of the test statistic provides estimates of the significance levels for the regressions of ash, mix time and loaf volume on their respective PLS directions count the number of times the simulated values exceed the observed

values for the data at hand, and divide by the number of random samples generated. The following table shows the correct significance levels and the values that result from using the F distribution for regressing ash, mix time and loaf volume on their PLS directions. The numbers in parentheses are estimated standard deviations that arise from the Monte Carlo sampling technique.

Response	Observed F	F-level	True level
Ash	3.8	0.080	0.38 (±0.007)
Mix time	26.8	$4.1 \cdot 10^{-4}$	0.002 (±0.0006)
Loaf volume	12.0	0.0061	0.03 (±0.0024)

Höskuldsson [5] and others have suggested to use the F-test for the regression on the PLS component as an approximation. Because of the upward bias, comparing the PLS F statistic to the F distribution is a liberal procedure; 'non-significance' according to the F distribution implies non-significance according to the correct distribution of the PLS F statistic, but significance according to the F does not imply significance according to the correct distribution. The above computations show that the difference between the F-level and the true level can be quite dramatic.

Unlike the ordinary regression F statistic, the PLS F statistic has a null distribution that depends on the distribution of the predictor variables. Hence, this statistic has to be recalibrated for each new regression design. Monte Carlo simulation offers a means for performing this calibration. The exact distribution for some very special designs has been worked out in ref. 11.

The fact that certain peaks are given large weight by PLS or principal components does not prove that they are strongly related to the response of interest. Some direction will always be selected, and in high dimensions it is quite possible to obtain a striking plot of the PLS weights that is simply an artifact. From the preceding calculations we conclude that, despite the impressive loadings plot, there is little evidence of a relationship between ash and protein composition. On the other hand mix time appears to have a rather strong relationship with protein composition; however, further experimentation would be needed to determine whether the peaks indicated by PLS and principal component analysis, which are virtually identical in this case, have a causitive relationship or were selected simply because they show the greatest variation. Loaf volume is an intermediate case, showing a moderately significantly relationship with protein composition. The corresponding PLS direction differs somewhat from the principal component, and we have an indication that proteins eluting at 17–19 min might be important. Further experimentation would be needed before we could say anything conclusive. Such information is, however, of great potential value, as it gives a tentative indication of specific proteins that, through subsequent isolation and characterization, might explain various attributes or serve as the basis for sensitive and rapid tests.

DISCUSSION

Data analysis in high dimensions is a tricky business. There is considerable latitude for the selection of 'factors' that appear to demonstrate striking relationships. In order to separate the artificial relationships from the real ones, great care should be taken to employ proper statistical inference methods that account for the multiplicity of directions available. One method that we have demonstrated is the use of simulations to get the correct null distribution of the F statistic for regression on the PLS direction. This provides a useful screening procedure for spurious directions.

Our goal in the present investigation is an ambitious one. In addition to classifying or predicting from the chromatogram we attempt to interpret the weighting vectors produced by the dimension reduction. This is the most difficult aspect of the analysis and the one that is most likely to give spurious results. There is less of a problem if one merely wants to predict or classify without attempting to interpret the weighting vectors. In such instances the PLS dimension reduction is likely to be a useful one, because it chooses projections with maximal covariance with the response. Nevertheless, as we have demonstrated, the standard regression tests and prediction intervals require adjustment for the variable selection.

ACKNOWLEDGEMENTS

The authors thank J. Alho, H.D Petersen and the referees for helpful comments and additional references. This research was supported by United States Department of Agriculture Contract USDA-58-5114-8-1026, National Security Agency Grant NSA-MDA904-89-H-2011, and Air Force Office of Scientific Research Grant AFOSR-87-0041.

REFERENCES

1 T Burnouf and J A Bietz, Reversed phase high-performance liquid chromatography of durum wheat gliadins relationships to durum wheat quality, *Journal of Cereal Science* 2 (1984) 3–14
2 I T Jolliffe *Principal Components Analysis*, Springer, New York 1986
3 S Wold, H Martens and H Wold, The multivariate calibration problem in chemistry solved by the PLS method in A Ruhe and B Kågstrom (Editors), *Matrix Pencils, Proceedings of a Conference held at Pite Havsbad, March 22–24 1982, Lecture Notes in Mathematics, 973* Springer-Verlag, Heidelberg, 1983, pp 286–293
4 I E Frank Intermediate least squares regression method *Chemometrics and Intelligent Laboratory Systems*, 1 (1987) 233–242
5 A Hoskuldsson, PLS regression methods, *Journal of Chemometrics*, 2 (1988) 211–228
6 R Sundberg and P J Brown, Multivariate calibration with more variables than observations, *Technometrics*, 31 (1989) 365–371
7 P Geladi and B R Kowalski, Partial least-squares regression a tutorial, *Analytica Chimica Acta*, 185 (1986) 1–17
8 H Hotelling, Relations between two sets of variables, *Biometrika*, 28 (1936) 321–377
9 B Flury and H Riedwyl, *Multivariate Statistics A Practical Approach*, Chapman and Hall, London, 1988
10 P Lachenbruch and M Mickey, Estimation of error rates in discriminant analysis, *Technometrics*, 10 (1968) 1–11
11 S Guo, *Inferences on high-dimensional data, Ph D Thesis*, Department of Statistics, University of Illinois, Urbana-Champaign, IL, 1990
12 S Raudys and V Pikelis, On dimensionality, sample size, classification error, and complexity of classification algorithm in pattern recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2 (1980) 242–252
13 V R Marco, D M Young and D W Turner, The Euclidean distance classifier an alternative to the linear discrimination function, *Communications in Statistics — Simulation*, 16 (1987) 485–505
14 S Wold, Pattern recognition by means of disjoint principal components models *Pattern Recognition*, 8 (1976) 127–139
15 L L Nolte, V L Youngs, R D Crawford and W H Kunerth Computer program evaluation of durum and hard red spring wheat, *Cereal Foods World*, 30 (1985) 227–229
16 N Draper and H Smith *Applied Regression Analysis, Second Edition*, Wiley, New York, 1981
17 B D Ripley, *Stochastic Simulation*, Wiley, New York, 1987

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 169–184
Elsevier Science Publishers B V, Amsterdam

Source apportionment with one source unknown

Karen Bandeen-Roche *
School of Hygiene and Public Health, Department of Biostatistics, The Johns Hopkins University, 615 N Wolfe St, Baltimore, MD 21205 (U S A)

David Ruppert
School of Operations Research and Industrial Engineering, Cornell University, Ithaca, NY 14853 (U S A)

(Received 7 November 1989; accepted 5 July 1990)

Abstract

Bandeen-Roche, K and Ruppert, D, 1991 Source apportionment with one source unknown *Chemometrics and Intelligent Laboratory Systems*, 10 169–184

Attribution of local pollution to area sources is essential to effective management of the environment Source apportionment addresses the problem by statistical inference of source contributions to total pollution from observations of ambient air chemical composition Mass balance methods of source apportionment use linear models with chemical composition vectors of sources as covariates Historically, mass balance methods have assumed that at least a proxy of each covariate is available and has been accounted for

We attempt to adapt the mass balance method to the case in which unidentified sources may exist by estimating an unknown, possibly 'background', source Further, we allow source contributions to pollution to vary over time, creating a model with a 'structural' parameter and infinitely many 'incidental' parameters We treat the 'incidental' source contribution parameters as random quantities Investigating the properties of the distribution governing relative source contributions is then of interest Reasonable identifiability constraints are required in this context Nonparametric estimation of the unknown source is possible under such constraints but is impractical for small samples which are measured with error Therefore, we develop a parametric model for the distribution of the observations and examine estimates based on this model

INTRODUCTION

One of the important problems of environmental engineering is to identify major sources of pollution and determine their relative effects upon the surrounding ('ambient') air, water, or some other medium Attempts have been made to predict cumulative effects based on chemical measurements taken at individual source locations However, factors such as meteorology, topography, and multiplicity of sources make predicting the effects of sources at a removed location difficult. An alternative approach is to measure samples of the ambient medium. Source contributions to pollution levels are then inferred using statistical methods. The body of methods which has been developed to achieve such inference is known as source apportionment.

0169-7439/91/$03 50

The chemical mass balance (CMB) method of source apportionment, developed for study of atmospheric pollution, assumes a linear model for the chemical composition of ambient air. The chemical composition vectors of area pollution sources, called source profiles, are used as covariates, and mass contributions of sources to pollution are considered to be the parameters of interest Typically, source profiles are given in terms of mass fractions — for instance, milligrams of particulate matter with a given chemical property per gram of particulate source output Source contributions, then, are often parameterized as concentrations — particulate mass of a given filtering specification contributed by each source per unit mass of that specification, or per unit volume of ambient air. Linearity arises from the assumption that mass is conserved from sources to the ambient air sampler, so that the composition of the observed sample is just a sum of the parameters multiplied (in a vector sense) by the corresponding covariates Parameter estimation has usually been achieved using variations on standard least-squares methodology It is important to note that the traditional CMB model treats each ambient profile observation, perhaps time-averaged, as a distinct sample In this context, vector elements provide repeat observations, and time variation is not accounted for in any explicit way.

As useful as CMB models have proven to be in practice, the methodology has significant shortcomings Perhaps chief among them is the fact that they require both awareness of all possible sources and knowledge of their chemical compositions, as is illustrated by an example described by Aldershof and Ruppert [1] Researchers at EPA were interested in the relative contributions of woodstoves and vehicular emissions to local environments. A source profile for woodstove smoke was carefully constructed, but unfortunately the source profile for vehicular emissions was not available at the time. A chemical engineer involved in the study suggested that the profile of the unknown source might be considered as a stable parameter, and that thereby a well posed model for the composition of area pollution might be formulated. As in usual CMB models, the parameters of interest are the contributions of all sources to pollution However, their estimation requires estimation of the unknown source profile.

The existence of problems such as that which we have just described has led us to develop methodology which generalizes the traditional CMB model in two ways Firstly, it allows for the possibility that all sources have not been determined by estimating an unknown source. We will allow an arbitrary number of known sources but only one unknown source The case of one unknown source is interesting in its own right, as the woodstove example shows Moreover, in some situations where there are several unknown sources, investigators will be willing to aggregate all unknown sources into a general 'background' unknown For example, this would be sensible if the relative contributions of the unknown sources were stable over time After this aggregation of unknown sources our methodology can be applied, though of course only the distribution of the aggregate contribution from the unknown sources will be estimated

A second, more subtle modification is that our models are formulated for source profiles given in a form which is proportional with respect to a fixed set of chemical species, rather than in mass fraction form In particular, we define a profile vector by taking the particulate mass per unit of source output due to each member of the fixed set and dividing by the particulate mass per unit attributable to the entire set of species. This is a generalization in the sense that transforming mass profiles to proportional profiles is always possible, whereas the information necessary to perform the converse operation may not be available in some applications Although this course of action was taken chiefly to accommodate cases when profiles are only given in proportional form — the woodstove data set is such a case — we remark that it is often possible to obtain proportional profiles which are much more accurate than mass fraction profiles (see Kowalczyk et al [2]) An important spinoff of using proportional profiles, however, is that the total mass contributions of sources to pollution are no longer estimated directly Instead, source contributions of only those chemical species actually measured and used to define the profile — a quantity of interest in its

own right — are estimated Happily, one may deduce total contributions from the proportional profile parameters if source profiles are available in terms of amounts

Studying the estimation of an unknown source in the context of the CMB method has led us to consider two other limitations of the CMB model The first arises when one attempts to estimate the unknown source namely the inability of the CMB method to deal with the variations of source contributions over time. To understand the problem, consider the fact that ambient sample composition is determined by the compositions of known sources, a constant unknown source parameter, and source contribution parameters that differ with each observation This creates, in the terminology of Kiefer and Wolfowitz [3], two classes of parameters a finite-dimensional 'structural' parameter (the profile of the unknown source) and an infinite sequence of 'incidental' parameters (daily proportional source contributions) Any reasonable estimator of the structural parameter must include observations corresponding to distinct incidental parameters. However, it is well known that estimation is often impossible if incidental parameters are deterministic In order to address this difficulty, we have chosen to treat daily source contributions as random quantities In this context, the distribution of the incidental parameters (source contributions) rather than the individual parameters is estimated.

We will address a second limitation by exploring error structures which are more natural to nonnegative vector observations than the additive, Gaussian error structure implicitly assumed by CMB models

Henceforth, random-proportion, unknown-source CMB models will be referred to as source apportionment, one source unknown (SASU) models, and we will consider source contributions to be those resulting from a proportional profile formulation unless otherwise specified. We will develop our model, which is no longer linear, and examine its relationship to the traditional CMB model in the next section. To make the exposition simpler, in this paper only the case of a single known source will be treated explicitly. In addition to the nonlinearity of the model, the one-source-unknown case differs fundamentally from the case in which all source profiles are known in that its parameters are not identifiable without the addition of constraints It will be helpful to examine the case in which observations are made without measurement error — in other words, day-to-day differences in source contributions provide the only random variation In this case, a simple constraint allows consistent estimation of the parameters of interest, and asymptotic distributions for the estimates are available Measurement error complicates estimation considerably — so much so that nonparametric estimation becomes extremely and perhaps prohibitively difficult in a small sample context. Consequently, we will propose an appropriately constrained parametric model and study its behavior

Source apportionment and CMB models have been discussed by many authors, including Cooper and Watson [4], Gordon [5], and Henry et al [6] Introduction of unknown source estimation into CMB methodology was done following ref 1 Estimation of structural parameters in the presence of incidental parameters was first discussed by Neyman and Scott [7] and has since been a topic of continuing interest Relevant papers include refs 3 and 8–10 Campbell and Mosimann [11] provide insight into parametric models for proportional data.

SETUP OF THE PROBLEM

Observations in CMB models are generally the total amounts of various chemical species collected during ambient air sampling, perhaps given as concentrations When source profiles are proportional, an equivalent, geometrically intuitive formulation of the problem results by standardizing observations to proportions as well Both formulations prove to be useful in what follows

The SASU model

Although focus soon shifts to the case in which only one source is known, we will state the model for the general case of m known sources. Let $x_1, \ldots, x_m$ be p-dimensional, deterministic co-

variates, let θ be an unknown, p-dimensional parameter. In the SASU model, x_k, $k = 1,\ldots,m$, are profiles of known sources, θ is the profile of the unknown source, and each has been standardized to proportions. We impose the resulting constraints

$$\sum_{j=1}^{p} x_{kj} = \sum_{j=1}^{p} \theta_j = 1 \quad (k = 1,\ldots,m) \tag{1}$$

$$x_k \geq 0 \forall k, \quad \theta \geq 0$$

In addition, let α_i $(i = 1,\ldots,n)$ be independent and identically distributed (iid), m-dimensional random vectors whose components are nonnegative and sum to at most 1. The vectors α_i represent the daily contributions of the known sources, so that the scalars $(1 - \sum_k \alpha_{ik})$ correspond to proportional contributions of the unknown source. Let G be the joint distribution function for the components of α_i, that is,

$$G(c) = \mathrm{P}\{\alpha_i \leq c\}$$

where the inequality holds for each component. Analogously, let γ_i^x be iid, m-dimensional random vectors with nonnegative components and γ_i^θ a nonnegative, real-valued random variable $(i = 1, \ldots, n)$. γ_i^x is the corresponding vector to α_i, given in terms of amounts, so that γ_i^θ represents the mass contribution of the unknown source to the set of chemical species defining the profiles. We will denote the joint distribution function of the components of γ_i^x and γ_i^θ to be F, that is,

$$\mathrm{F}(c) = \mathrm{P}\{\gamma_i^{x'}, \gamma_i^{\theta'} \leq c\}$$

where V′ stands for the transpose of the vector V. Again, the inequality holds for each component.

Random variables of interest are

$$y_i = \sum_{k=1}^{m} \alpha_{ik} x_k + \left(1 - \sum_{k=1}^{m} \alpha_{ik}\right)\theta \tag{2}$$

and

$$s_i = \sum_{k=1}^{m} \gamma_{ik}^x x_k + \gamma_i^\theta \theta$$

so that

$$y_{ij} = \frac{s_{ij}}{\sum_j s_{ij}} \quad \text{and} \quad \alpha_{ik} = \frac{\gamma_{ik}^x}{\sum_k \gamma_{ik}^x}$$

Components of the vectors y_i and s_i represent true ambient air chemical proportions and amounts, respectively, on day i. We observe $\mathbf{Y}_i$ and $\mathbf{S}_i$ which are measured values of y_i and s_i. In the next section we examine the simple case where $\mathbf{Y}_i = y_i$ and $\mathbf{S}_i = s_i$. We develop below a parametric model for the measurement errors. Notice that in the case of present interest, $m = 1$, $y_i = \alpha_i x + (1 - \alpha_i)\theta$ and $s_i = \gamma_i^x x + \gamma_i^\theta \theta$.

Transformation to CMB model

The traditional CMB model is as follows

$$s_i = \sum_{k=1}^{m+1} c_{ik} a_k \tag{3}$$

where c_{ik} = total particulate mass contributed by source k per unit volume of ambient air on day i, a_k = mass profile of source k.

The subtle difference between the CMB parameters, c_{ik}, and the SASU parameters, γ_{ik}^x — equivalently, α_{ik} — occurs because information regarding the relative amounts of source outputs not accounted for by the set of measured chemical species is lost in the transformation from mass profiles to proportional profiles. In this section we show how the parameters in our formulation as given in the SASU model are related to the parameters in eq. (3).

Suppose one profile — say, a_{m+1} — is unknown. Let $\xi_i = \sum_j s_{ij}$. Clearly,

$$x_k = \frac{a_k}{\sum_j a_{kj}}, \quad \theta = \frac{a_{m+1}}{\sum_j a_{m+1\,j}}, \quad \text{and} \quad y_{ij} = \frac{s_{ij}}{\xi_i}$$

Therefore, eq. (2) is equivalent to

$$\frac{s_i}{\xi_i} = \sum_{k=1}^{m} \left(\frac{\alpha_{ik}}{\sum_j a_{jk}}\right) a_k + \frac{\left(1 - \sum_{k=1}^{m} \alpha_{ik}\right)}{\sum_j a_{m+1\,j}} a_{m+1}$$

which implies that $c_{ik} = \xi_i \alpha_{ik} / \sum_j a_{kj}$ and $c_{ik} a_k = \xi_i \alpha_{ik} x_k$, $k = 1,\ldots,m$ (similarly for $k = m+1$, θ).

Knowledge of ξ_i, θ, and α_i, then, are sufficient to determine $c_{ik} a_k$, $k = 1,\ldots,m+1$. (Physically, $c_{ik} a_{kj}$ represents the amount of the jth chemical species contributed by source k to the ambient air

sample on day i.) However, we will be estimating only the distribution of the α_i values because it is impossible to consistently estimate the α_i values themselves in the presence of measurement error (see below). It is possible to get around this inconvenience in the following way:

Let $q_{kj} = c_{ik} a_{kj}$. Then

$$\frac{q_{kl}}{q_{kj}} = \frac{q_{kl}/\sum_n q_{kn}}{q_{kj}/\sum_n q_{kn}} = \frac{a_{kl}/\sum_n a_{kn}}{a_{kj}/\sum_n a_{kn}} = \frac{x_{kl}}{x_{kj}} \tag{4}$$

Letting $x_{m+1} = \theta$, we have a system of independent linear equations in $p(m+1)$ unknowns. Eq. (3) implies

$$\sum_{k=1}^{m+1} q_{kj} - s_{ij} = 0, \quad j = 1, \ldots, p \tag{5}$$

Eq. (4) implies

$$q_{kl} x_{kj} - q_{kj} x_{kl} = 0, \quad k = 1, \ldots, m+1, \quad j = 2, \ldots, p \tag{6}$$

There are a total of $p + (m+1)(p-1) = p(m+1) + (p-m-1)$ equations. Hence, we may solve for q_k, $k = 1, \ldots, m+1$ if $p \geq m+1$ — that is, if there are more species than known sources.

If, in addition, the mass source profiles, a_k, $k = 1, \ldots, m$, are known, solution for the corresponding source contributions, c_{ik}, follow immediately. Notice that these solutions are corrected for the contribution of the unknown source.

NONPARAMETRIC MODELS

No measurement error

For now, we will use the formulation of the SASU model for which observations are proportional. In the case where observations occur without measurement error,

$$\mathbf{Y}_i = y_i = \alpha_i (x - \theta) + \theta \tag{7}$$

In this section we take a nonparametric approach in that the distribution G of α_i is not assumed to be in a parametric family. Without the natural constraints mentioned above and an additional restriction on G, the model (eq. (7)) is not identifiable. To understand why, consider a simple transformation:

Let $\mathbf{Y}_i - x = \mathbf{Z}_i$, $\theta - x = \phi$, $(1 - \alpha_i) = \lambda_i$. It is clear that eq. (7) is equivalent to

$$\mathbf{Z}_i = \lambda_i \phi \tag{8}$$

Let $\tilde{\lambda}_i = \lambda_i/2$ and $\tilde{\phi} = 2\phi$. $\mathbf{V}_i = \tilde{\lambda}_i \tilde{\phi}$ has exactly the same distribution as $\mathbf{Z}_i$. This means that the parameters of $\mathbf{Z}_i$ are not identifiable from its distribution. In fact, the model (7) implies that $\mathrm{Corr}[y_{ij}, y_{il}] = 1 \ \forall\ j, l$ and, hence, that the p-dimensional system effectively reduces to one dimension.

Realizing that nonidentifiability occurs because our model allows too much scaling suggests an appropriate constraint confining allowable distributions for α_i to those whose left boundary of support is exactly 0 — that is, $G(a) > 0$ for each $a > 0$. It follows that

$$\lim_{n \to \infty} \min_{1 \leq i \leq n} \alpha_i = 0 \quad \text{with probability 1}$$

or

$$\lim_{n \to \infty} \min_{1 \leq i \leq n} \lambda_i = 1 \quad \text{with probability 1} \tag{9}$$

which means that if enough samples are taken, eventually one will be composed almost entirely of chemicals contributed by the unknown source. Our motivating example, described in the introduction, provides some insight on summer days, one would not expect people to be using woodstoves, and condition (9) appears reasonable.

Several results follow under condition (9). When there are no measurement errors, the observations $\mathbf{Y}_i$ lie on the line segment in p-dimensional space connecting the known x and the unknown θ. This suggests a simple estimator of θ — the observation farthest from x. In fact, it is not difficult to prove the following.

Proposition 1 Define $\mathbf{Y}_n^*$ to be the observation $\mathbf{Y}_m$ such that $\max_{1 \leq i \leq n} \|\mathbf{Y}_i - x\|_2 = \|\mathbf{Y}_m - x\|$. Then $\mathbf{Y}_n^*$ is a consistent estimator of θ if and only if condition (9) holds.

Once we estimate θ, we can then estimate the contributions, α_i, of the known source. The basic idea is that α_i is the distance between $\mathbf{Y}_i$ and θ expressed as a fraction of the distance between x

and θ. Note that Y_n^* is a monotone sequence and, thus, $\tilde{\theta}_{jn}$ equal to the jth component of Y_n^* satisfies the conditions of the following result.

Proposition 2 Let θ_j be such that $|x_j - \theta_j| \neq 0$. Let $\tilde{\theta}_{jn}$ be any monotone sequence whose limit is θ_j. Define

$$F_n(x) = \frac{1}{n}\sum_i \mathbf{I}(Y_{ij} \leqslant x) = \text{e d.f.}(Y_{ij})$$

where $\mathbf{I}$ is the indicator function and e d f. stands for empirical distribution function. Define

$$G_{jn}(z) = 0 \quad \text{if } z \leqslant 0$$

$$F_n\left[z\left(x_j - \tilde{\theta}_{jn}\right) + \tilde{\theta}_{jn}\right] \quad \text{if } 0 < z \leqslant 1$$

$$1 \quad \text{if } 1 < z$$

Under (7), $G_{jn}(z) \xrightarrow{\text{wp1}} G(z)$

Define $\hat{\alpha}_{in} = (Y_{ij} - \tilde{\theta}_{jn})/(x_j - \tilde{\theta}_{jn})$. Then G_{jn} is the empirical distribution function of $\{\hat{\alpha}_{1n}, \ldots, \hat{\alpha}_{nn}\}$. Because there are no measurement errors, $(\hat{\alpha}_{in} - \alpha_i) \to 0$ with probability 1. As it is a matter of algebra to show that $\max 1 \leqslant i \leqslant n |\hat{\alpha}_{in} - \alpha_i| \leqslant |\theta_j - \tilde{\theta}_{jn}|/|x_j - \tilde{\theta}_{jn}|$, the stronger statement $\max 1 \leqslant i \leqslant n |\hat{\alpha}_{in} - \alpha_i| \xrightarrow{\text{wp1}} 0$ also holds.

We have seen that θ can be estimated because when α_i is close to 0, then Y_i consists mostly of the contribution from the unknown source. The rate at which Y_n^* converges to θ depends on how fast $\min 1 \leqslant i \leqslant n \alpha_i$ approaches 0, which in turn depends on the behavior of G near 0. In fact, extreme value theory provides an asymptotic distribution for Y_n^*.

Proposition 3 Suppose that support $\{\alpha\} = [0, c]$, $c \leqslant 1$.
Suppose also that $G(\alpha) = K\alpha^\beta$ on $0 \leqslant \alpha \leqslant K^{-(1/\beta)}$. Then

$$\lim_{n\to\infty} P\left\{|Y_{jn}^* - \theta_j| \leqslant \frac{y}{(nK)^{1/\beta}}\right\} = 0 \quad \text{if } y < 0$$

$$1 - \exp\left\{-\left(\frac{y}{x_j - \theta_j}\right)^\beta\right\} \quad \text{if } y \geqslant 0$$

The key fact in the proof of Proposition 3 is as follows: given independent observations from G, $\alpha_1, \ldots, \alpha_n$, extreme value theory dictates that

$$P\left\{(nK)^{1/\beta} \min_{1\leqslant i\leqslant n} \alpha_i \leqslant a\right\} \xrightarrow{w} H(a)$$

where $\xrightarrow{w}$ denotes weak convergence and H is the distribution such that $H(a) = 1 - \exp\{-a^\beta\}$ for nonnegative values of a and 0 otherwise. Leadbetter et al. [12] provide an excellent reference to extreme value theory.

With measurement error

Given the results attainable in the case of no measurement error, one might hope that under condition (9) similar results might hold in the case with measurement error. Unfortunately, this does not appear to be the case. of course condition (9) is still necessary, but introducing measurement error makes the problem 'much' harder. For one thing, it makes consistent estimation of the individual α_i values impossible. In the case of no measurement error, it is possible to write α_i as a function of the observations and the structural parameter. Since every observation contributes to the estimation of θ, every observation contributes to the estimation of α_i through the function (recall the estimator $\hat{\alpha}_{in}$ discussed following Proposition 2). In the presence of measurement error, it is no longer possible to write α_i as a function of the observations and the structural parameter (we no longer see the true value of the observation). Hence in effect only finitely many observations (1 vector observation or p scalar observations) contribute to the estimation of α_i, so consistent estimation is impossible.

Estimating the structural parameter is much harder, as well. It is still helpful to think of the problem in terms of estimating the endpoint of the line segment between the known x and the unknown θ. When observations are made with measurement error, however, they appear as a 'cloud' of points about the line segment rather than being confined to the segment itself. The estimator Y_n^* defined in the previous section, then, will eventually overshoot θ if the cloud extends far enough: formally, Y_n^* converges to a support boundary of the distribution of Y_i rather than to θ, the support boundary of the distribution of y_i. If one were to assume additive errors for the observations given in terms of amounts, S_i (or more appropriately for some transformation of S_i), the method of decon-

volution may be used, in effect, to account for the measurement error Such an approach is capable of estimating θ consistently. Unfortunately, convergence rates of nonparametric deconvolution estimators are inherently very slow. Carroll and Hall [13] have shown for a large class of distributions that in the case of normal error, no deconvolution estimator can achieve a rate higher than a factor of $(\log n)^{-1}$. It is possible, however, that a higher rate may obtain for distributions confined to a bounded support. Also, it is known that certain functionals of deconvolution estimators converge significantly faster than the estimators themselves, and it is not unreasonable to expect that an estimator of θ could be one of them Further research is necessary to investigate these possibilities

A PARAMETRIC MODEL

In order to produce estimators which achieve reasonable rates of convergence for moderate sample sizes, it appears that parametric models are required for both G and the measurement error. The discussion in the previous section indicates that any reasonable model for time variation must satisfy condition (9). In keeping with the spirit of maximum generality, we will model for proportional observations, which suggests that we utilize distributions inherently appropriate for proportional data.

With these considerations in mind, we have chosen to model both time variation and measurement error with the Dirichlet distribution. A generalization of the Beta distribution, the Dirichlet distribution is especially well suited to modeling proportional vectors created by dividing amounts observations by their sum. Such vectors are exactly Dirichlet-distributed whenever amounts are independent of each other and the proportions which result from dividing the amounts by their sum are independent of the sum, whenever amounts are independent gamma random variables with common scale, and in certain cases when amounts are positively correlated, the vector of amounts divided by sum is Dirichlet. In addition, Dirichlet random variables satisfy some very convenient properties Let Y be a p-dimensional Dirichlet random vector with p-dimensional parameter vector, δ (see below). Any permutation of Y is Dirichlet with parameter equal to the corresponding permutation of δ Also, suppose Z is an amalgamation over some partition, $A = \{a_1, \ldots, a_q\}$, of the coordinates of Y — in other words,

$$\mathbf{Z} = \left\{ \sum_{j \in a_1} Y_j, \ldots, \sum_{j \in a_q} Y_j \right\}$$

with $q < p$. Then Z is Dirichlet with parameter equal to the corresponding amalgamation of δ. These properties will allow us to combine and permute coordinates of observations in order to improve estimates without changing the underlying model for estimation, see below. Campbell and Mosimann [11] provide a basic summary of these and other properties of the Dirichlet distribution

In general, the Dirichlet density has the form

$$f_0(y \mid \delta) = \frac{\Gamma(\Delta)}{\prod_{j=1}^{p} \Gamma(\delta_j)} \prod_{j=1}^{p} y_j^{\delta_j - 1}$$

where $\Delta = \sum_{j=1}^{p} \delta_j$

It follows that the first two moments of a Dirichlet random variable, Y, with distribution $f_0(y \mid \delta)$ are:

$$E[Y_j] = \mu_j = \frac{\delta_j}{\Delta} \tag{10}$$

$$\mathrm{Var}[Y_j] = \frac{\delta_j(\Delta - \delta_j)}{(\Delta + 1)\Delta^2} = \frac{\mu_j(1 - \mu_j)}{(\Delta + 1)} \tag{11}$$

In general, the kth moment of Y_j is

$$E[Y_j^k] = \frac{\prod_{m=0}^{k-1} (\delta_j + m)}{\prod_{m=0}^{k-1} (\Delta + m)} \tag{12}$$

Notice also that the coefficient of variation of Y is

$$\mathrm{CV}[Y_j] = \sqrt{\frac{(1 - \mu_j)}{\mu_j(\Delta + 1)}} \tag{13}$$

We will model the measurement error process by assuming that the conditional distribution of Y given α has a Dirichlet distribution with mean $\alpha x + (1-\alpha)\theta$ and scale independent of α. Therefore, assume that Y_i is Dirichlet with parameter $\delta_{ij} = \Delta[\alpha_i(x_j - \theta_j) + \theta_j]$ for some constant $\Delta > 0$ (note that $E[Y_i|\delta_i] = \alpha_i x + (1-\alpha_i)\theta = y_i$ and that $\sum_j \delta_{ij} = \Delta$ for each i). Now, the marginal distribution of Y_i is obtained by integrating f_0 over the distribution of α, which we hypothesize to be the Beta(λ_1,λ_2) = Dirichlet(λ_1,λ_2) distribution. In other words, the density of the marginal distribution of y has the form:

$$f(y_i) = \int_0^1 f_0(y_i|\alpha)\frac{\Gamma(\lambda_1+\lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \times \alpha^{\lambda_1-1}(1-\alpha)^{\lambda_2-1}\,d\alpha \qquad (14)$$

where $f_0(y_i|\alpha) = f_0(y_i|\Delta[\alpha(x-\theta)+\theta])$ Eq (14) corresponds to taking an average of the densities f_0 at y_i given each possible value of α, weighted by the probability of α

In the development which follows, we will be using the quantity $1-\alpha$ rather than α From the permutation property mentioned above, it is clear that $(1-\alpha)$ has a Beta(λ_2,λ_1) distribution. Letting $\lambda = \lambda_2$ and $\Lambda = \lambda_1 + \lambda_2$, we may re-parameterize the beta parameters from $\{\lambda_1,\lambda_2\}$ to $\{\lambda,\Lambda\}$. Certain functions of the source contribution parameters are of at least as much interest as the parameters themselves: for example, $\lambda/\Lambda = E[1-\alpha]$ and $[\lambda(\Lambda-\lambda)]/[\Lambda^2(\Lambda+1)] = \mathrm{Var}[1-\alpha] = \mathrm{Var}[\alpha]$. From now on we will refer to Δ as the error parameter and $\{\lambda,\Lambda\}$ as the source contribution parameters.

Given $p \geq 3$, all of the parameters of this model are identifiable from its moments (of order 3 and less). In other words, these moments completely determine Δ, λ, Λ, and θ. Therefore, method of moments estimators for the parameters are consistent. The moments equations may be developed as follows. Let Y_i be an observation from eq. (14), where δ_i is as defined above. Recall that $Z_i = Y_i - x$ and $\phi = x - \theta$. Using eqs. (10), (12), and the fact that, for any random variables U and V, $E[U] = E[E[U|V]]$, we may write for each coordinate j, $j = 1,\ldots,p$:

$$\begin{aligned} m_{1j} = E[Z_{ij}] &= E\left[E\left[Z_{ij}|(1-\alpha_i)\right]\right] \\ &= E\left[\frac{\Delta\left((1-\alpha_j)\phi_j + x_j\right)}{\Delta} - x_j\right] \\ &= \phi_j E[1-\alpha_i] \\ &= \frac{\phi_j\lambda}{\Lambda} \end{aligned} \qquad (15)$$

Similarly,

$$\begin{aligned} m_{2j} = E\left[Z_{ij}^2\right] &= \left\{\Delta\phi_j^2 E\left[(1-\alpha_i)^2\right] + \phi_j(1-2x_j)E[1-\alpha_i] \right. \\ &\quad \left. + x_j(1-x_j)\right\}\{\Delta+1\}^{-1} \\ &= \frac{1}{\Delta+1}\left\{\frac{\Delta\phi_j^2\lambda(\lambda+1)}{\Lambda(\Lambda+1)} + \frac{\phi_j(1-2x_j)\lambda}{\Lambda} \right. \\ &\quad \left. + x_j(1-x_j)\right\} \end{aligned} \qquad (16)$$

$$\begin{aligned} m_{3j} = E\left[Z_{ij}^3\right] &= \frac{1}{(\Delta+1)(\Delta+2)}\left\{\frac{\Delta^2\phi_j^3\lambda(\lambda+1)(\lambda+2)}{\Lambda(\Lambda+1)(\Lambda+2)} \right. \\ &\quad + \frac{3\Delta(1-2x_j)\phi_j^2\lambda(\lambda+1)}{\Lambda(\Lambda+1)} \\ &\quad + \frac{\left[3x_j(1-x_j)(\Delta-2)+2\right]\phi_j\lambda}{\Lambda} \\ &\quad \left. + 2x_j\left(2x_j^2 - 3x_j + 1\right)\right\} \end{aligned} \qquad (17)$$

Method of moments estimators are formed by substituting the sample moments,

$$M_{q,r} = \frac{1}{n}\sum_{i=1}^{r}(Y_{ir} - x_r)^q \qquad (18)$$

for $m_{q,r}$ in eqs. (15)–(17) and solving for the parameters of interest. As the moments equations overdetermine the parameters, however, moments estimators are not unique. In the following sections we will develop several different estimators, examine their performance under the model in a simulation study, and test them on a famous

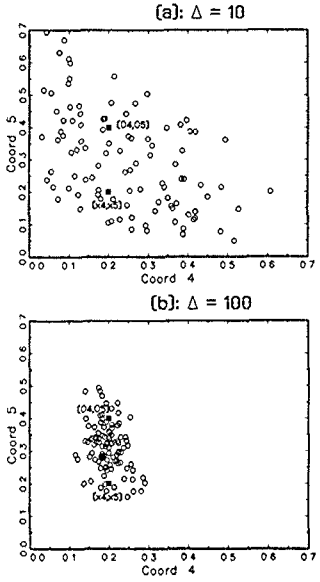


Fig 1 Dirichlet mixture data Observations are generated from model (14) with parameters $\lambda = (2, 2)'$ and $\theta = \{0, 0.05, 0.1, 0.2, 0.4, 0.25\}$, $x = \{0.2, 0.2, 0.2, 0.2, 0.2, 0\}$ Contrast of case (a), $\Delta = 10$, with case (b), $\Delta = 100$, illustrates role of Δ parameter 5th data component is plotted against 4th

simulated source apportionment data set. Computation of estimates and error measures, as well as generation of 'random' observations, were performed on an AST PC with a 286 processor using the GAUSS system, version 2 0.

Description of the simulation

Simulations each consisted of 100 runs at 100 observations per run generated from the model (14). The sample size of 100 was chosen to be comparable to that of a typical source apportionment data set. Observations were six-dimensional with $x = (0.2, 0.2, 0.2, 0.2, 0.2, 0)'$ and $\theta = (0, 0.05, 0.1, 0.2, 0.4, 0.25)'$. A simulation was performed for each of three pairs of source contribution parameters, $\{\lambda, \Lambda\}$. (i) {4,5}, (ii) {2,4}, and (iii) {1,5}. Recalling that $E[1 - \alpha_i] = 1 - E[\alpha_i] = \lambda/\Lambda$ (see eq 10) and noting that

$$E[Y_i] = (E[\alpha_i])x + (E[1 - \alpha_i])\theta$$

it is clear that (i) represents a very favorable estimation scenario — one in which observations tend to be close to the unknown source profile, θ Similarly, (ii) and (iii) represent increasingly less favorable scenarios.

In addition, each simulation described above was performed at two values of Δ· $\Delta = 10$ and $\Delta = 100$. Fig 1a and b display plots of data simulated under (ii) for the two values of Δ. Examination of the plots and review of eq. (13) makes it clear that $\Delta = 100$ represents middling measurement error while $\Delta = 10$ produces very severe error.

As a measure of performance, median and worst 90th-percentile distance of each estimate from its true value are given

Estimation of error scale, Δ

Although Δ, the error scale parameter, is in effect a nuisance parameter, its estimation is important because estimates of source contribution parameters, λ and Λ, and the location parameter, θ, depend directly upon Δ. Also, since severity of measurement error varies inversely with Δ, that parameter is itself of measure of how well we may expect to estimate the parameters of interest.

It happens that each pairwise combination of observation coordinates — say, (j,k) — produces an estimate $\hat{\Delta}_{jk}$ of Δ. Define for each coordinate j, $j = 1, \ldots, p$:

$$A_j := (1 - 2x_j)m_{1_j} + x_j(1 - x_j) \tag{19}$$

$$B_j := (\Delta + 1)m_{2_j} - A_j \tag{20}$$

$$\begin{aligned} C_j := {} & (\Delta + 1)(\Delta + 2)m_{3_j} - \{3(1 - 2x_j)B_j \\ & + m_{1_j}[3(x_j(1 - x_j)(\Delta - 2) + 2] \\ & + 2x_j(2x_j^2 - 3x_j + 1)\} \end{aligned} \tag{21}$$

where m_{1_j}, m_{2_j}, and m_{3_j} are as defined in eqs. (15)–(17). It is straightforward, if algebraically

painful, to verify that for any pair of coordinates $\{j,k\}$,

$$\Delta = \frac{m_{1j}^2 A_k - m_{1k}^2 A_j}{m_{2k} m_{1j}^2 - m_{2j} m_{1k}^2} - 1 \tag{22}$$

The estimator $\hat{\Delta}_{jk}$ results by substituting the sample moments, M_{qr} (see eq (18)), into eq (22). However, the first-order bias and variance of $\hat{\Delta}_{jk}$ increase with $|\theta_j - x_j|$ and $|\theta_k - x_k|$. From a heuristic viewpoint, one would expect the best estimates to result from coordinates for which $\theta_j = x_j$ — in other words, for which the only variation is due to measurement error. Since $E[Y_{ij} - x_j] = c(\theta_j - x_j)$ (c constant over i and j), the M_{1j} should contain information about the relative sizes of the quantities $|\theta_j - x_j|$. With these things in mind, we examined four estimates of Δ, each a weighted average of the pairwise estimates with weights w_{jk} on pair (j,k) as follows:

DEST w_{jk} all equal — i.e., unweighted average of pairwise estimates

DAWEST. $w_{jk} = \dfrac{1}{|M_{1j}| + |M_{1k}|}$

DMWEST: $w_{jk} = \dfrac{1}{|M_{1j} M_{1k}|}$

DBEST: $w_{jk} = 1$ for pair j,k such that $|M_{1j}|$, $|M_{1k}|$ are minimum (in other words, such that one coordinate of the pair has the smallest value of $|M_{1i}|$, $i = 1,\ldots,p$, and the other has the second smallest)

$w_{jk} = 0$ otherwise

A summary of the simulation results is given in Table 1. DMWEST and DBEST clearly outperform DEST and DAWEST. It is harder to distinguish between DMWEST and DBEST; although DBEST generally outperforms DMWEST slightly in terms of standard deviation and 90% distance, DMWEST tends to have a smaller 50% deviation from the true parameter value. In both cases, reasonable estimates seem to be produced regardless of model parameterization.

Estimation of source contribution parameters

In this section we develop estimators first of $\{\lambda, \Lambda\}$ and then of θ. Given Δ, each coordinate of the observations provides information sufficient to estimate $\{\lambda, \Lambda\}$. In particular, it happens that for any coordinate j,

$$B_j = \frac{\Delta m_{1j}^2 \Lambda(\lambda+1)}{\lambda(\Lambda+1)} \tag{23}$$

$$C_j = \frac{\Delta m_{1j} B_j (\lambda+2)\Lambda}{\lambda(\Lambda+2)} \tag{24}$$

Clearly we could substitute the sample estimates (18) in (23), (24), and the definition of B_j and C_j, and then solve the above system of equations for λ and Λ, for any j. Reasoning that some coordinates may produce more reliable estimates than others, however, we may write

$$\sum_{j=1}^{p} w_j B_j = \frac{\Delta\lambda(\lambda+1)}{\Lambda(\Lambda+1)} \sum_{j=1}^{p} w_j m_{1j}^2 \tag{25}$$

$$\sum_{j=1}^{p} w_j C_j = \frac{\Delta\Lambda(\lambda+2)}{\lambda(\Lambda+1)} \sum_{j=1}^{p} w_j B_j m_{1j} \tag{26}$$

for any system of weights, w. We will create estimates based on the solution of eqs (25) and (26) for λ and Λ, using several choices of weights and sample-based substitutions.

In order to identify coordinates which should produce more reliable information than others, note that since $\mathrm{Var}[Y_{ij} - x_j] = (\mathrm{Var}[\alpha_i])(\theta_j - x_j)^2 + [(m_{1j} + x_j)(1 - m_{1j} - x_j)]/(\Delta + 1)$,

$$\left(\mathrm{CV}\left[Y_{ij} - x_j\right]\right)^2 = \frac{\mathrm{Var}[\alpha_i]}{\left(E[1-\alpha_i]\right)^2} + \frac{(m_{1j} + x_j)(1 - m_{1j} - x_j)}{m_{1j}^2(\Delta+1)} \tag{27}$$

As the first term is constant and exactly what would result if there were no measurement error, the coordinate-wise CVs measure how much variation is due to measurement error relative to each other. Theoretically, the most reliable information should be obtained from the coordinates having the highest proportion of its variation due to source contribution randomness — in other words, the coordinates with the lowest CVs. One approach might be to calculate source contribution esti-

TABLE 1

Estimation of measurement error parameter

Median and 90% absolute deviation of estimated error parameter from Δ

Parameter value		Median distance from Δ = 10			
λ	Λ	DEST	DAWEST	DMWEST	DBEST
4	5	0.902	0.852	0.802	0.859
2	4	1.71	1.74	1.21	1.48
1	5	2.68	2.14	1.93	1.99
		90% distance from Δ = 10			
		DEST	DAWEST	DMWEST	DBEST
4	5	3.78	3.19	1.98	2.30
2	4	10.4	7.80	6.00	2.89
1	5	12.4	12.7	10.1	7.00
		Median distance from Δ = 100			
		DEST	DAWEST	DMWEST	DBEST
4	5	12.1	10.6	8.53	8.65
2	4	20.0	13.6	8.28	9.62
1	5	14.3	12.9	9.40	10.7
		90% distance from Δ = 100			
		DEST	DAWEST	DMWEST	DBEST
4	5	43.7	37.4	21.1	20.7
2	4	91.7	62.8	26.2	23.5
1	5	69.5	57.1	26.3	24.7

mates based only on the coordinate with the lowest sample CV. However, examination of eq. (27) suggests an approach which includes all of the data The second term of the sum tends to decrease as $|m_{1,j}|$ increases — in other words, as the distance between x_j and θ_j increases. As the dimension of the observation increases, the $|m_{1,j}|$ values will tend to decrease. However, amalgamating the observations to a few favorable dimensions can provide several coordinates with large values of $|m_{1,j}|$ while retaining a correct parametric form We chose to amalgamate to three coordinates, the smallest number which allows identification of the entire parameter space, and chose the particular amalgamation for which the coordinate with the lowest sample CV is retained and the others are added in such a way as to maximize the resulting sample $|m_{1,j}|$ values Better amalgamations might well exist

Given p variate observations Y_i, $i = 1, \ldots, n$, the estimators we examined are as follows:

(1) AMAL-MW:

(a) For each Y_i, create R_i as follows:
$R_{i1} = Y_{im}$, where the sample CV of Z_{im} is minimum among all coordinates of Y_i;
$R_{i2} = \sum_{\substack{j \in P \\ j \neq m}} Y_{ij}$, where $P := \{j$ such that $M_{1,j} \geq 0\}$;
$R_{i3} = \sum_{\substack{j \in N \\ j \neq m}} Y_{ij}$, where $N := \{j$ such that $M_{1,j} < 0\}$.
(if N is empty, let N = the coordinate of the second least sample CV and delete that coordinate from P, perform analogous operation if P is empty).

(b) Create the analogous amalgamation of x, u.

(c) Substitute the DMWEST estimator of Δ for Δ in eqs. (25) and (26), and the definitions of B_j and C_j.

(d) Substitute the sample expectations of $(R_{ij} - u_j)^q$, $\frac{1}{n}\sum_{i=1}^{n}(R_{ij} - u_j)^q$, for m_{qj} in (25), (26) and the definitions of B_j and C_j.

(e) Solve (25) and (26) for λ and Λ using $w = (1/3, 1/3, 1/3)$, resulting in estimators $\hat{\lambda}$ and $\hat{\Lambda}$, respectively.

(2) AMAL-BP: Same as AMAL-MW, except substitute DBEST for DMWEST in step (c)

(3) BCV-MW:

(a) Substitute M_{qj} for m_{qj} in eqs. (25) and (26) and the definitions of B_j and C_j.

(b) Substitute the DMWEST estimator of Δ for Δ in eqs. (25) and (26) and the definitions of B_j and C_j.

(c) Solve eqs (25) and (26) for λ and Λ using $w_j = \{1$ if coordinate j has least sample CV, 0 otherwise$\}$, resulting in estimators $\hat{\lambda}$ and $\hat{\Lambda}$, respectively

(4) BCV-BP. Same as BCV-MW, except substitute DBEST for DMWEST in step (b)

TABLE 2

Estimation of source contribution parameter

Median and 90% absolute deviations of estimates from $E[\alpha]$.

$\Delta = 10$

Parameter value		Median distance from $E[\alpha]$			
λ	Λ	Amal-MW	Amal-BP	BCV-MW	BCV-BP
4	5	0.198	0.199	0.220	0.241
2	4	0.546	0.504	0.478	0.479
1	5	0.676	0.663	0.468	0.311
		90% distance from $E[\alpha]$			
		Amal-MW	Amal-BP	BCV-MW	BCV-BP
4	5	0.346	0.294	0.745	0.953
2	4	2.06	1.46	0.905	0.742
1	5	1.27	1.23	1.25	1.17

$\Delta = 100$

Parameter value		Median distance from $E[\alpha]$			
λ	Λ	Amal-MW	Amal-BP	BCV-MW	BCV-BP
4	5	0.041	0.041	0.031	0.030
2	4	0.070	0.069	0.091	0.089
1	5	0.105	0.111	0.068	0.061
		90% distance from $E[\alpha]$			
		Amal-MW	Amal-BP	BCV-MW	BCV-BP
4	5	0.169	0.194	0.117	0.121
2	4	0.264	0.213	0.400	0.280
1	5	0.350	0.321	0.162	0.162

Parameter value				2% trimmed mean of $E[\alpha]$ estimates			
λ	Λ	Δ	$E[\alpha]$	Amal-MW	Amal-BP	BCV-MW	BCV-BP
4	5	10	0.2	0.070	0.025	−0.047	−0.116
4	5	100	0.2	0.193	0.193	0.221	0.221
2	4	10	0.5	−0.066	0.101	0.105	0.188
2	4	100	0.5	0.518	0.523	0.564	0.551
1	5	10	0.8	0.347	0.236	0.536	0.459
1	5	100	0.8	0.772	0.833	0.789	0.790

In each case, having produced estimators $\tilde{\lambda}$ and $\tilde{\Lambda}$ of λ and Λ, we estimate $E[\alpha]$ by

$$\tilde{\mu}_\alpha = 1 - (\tilde{\lambda}/\tilde{\Lambda})$$

and θ by

$$\tilde{\theta} = \mathbf{M}_1 \tilde{\Lambda}/\tilde{\lambda} + x$$

where $\mathbf{M}_1 = (M_{11}, \ldots, M_{1p})'$.

Primary results of the simulation study — performance of the estimators as measured by median and 90% deviation from true values — are summarized in Table 2. For many scenarios the performance of the estimators was virtually indistinguishable, although relative performance of the BCV estimators to the AMAL estimators seemed to improve as the estimation scenario worsened. All of the functions estimated $E[\alpha]$ reasonably well in the case $\Delta = 100$, with only slight decreases in performance (especially from the BCV estimators) as the parameterization favorability decreased. Both estimators performed badly in the

TABLE 3

Estimation of location parameter

Euclidean distance to estimated location from θ

$\Delta = 10$

Parameter value		Median distance from {0, 0.05, 0.1, 0.2, 0.4, 0.25}			
λ	Λ	Amal-MW	Amal-BP	BCV-MW	BCV-BP
4	5	0.078	0.076	0.090	0.094
2	4	0.224	0.217	0.198	0.197
1	5	0.312	0.317	0.314	0.294
		90% distance from {0, 0.05, 0.1, 0.2, 0.4, 0.25}			
		Amal-MW	Amal-BP	BCV-MW	BCV-BP
4	5	0.150	0.120	0.255	0.432
2	4	0.526	0.589	0.354	0.512
1	5	0.440	0.556	0.880	1.15

$\Delta = 100$

Parameter value		Median distance from {0, 0.05, 0.1, 0.2, 0.4, 0.25}			
λ	Λ	Amal-MW	Amal-BP	BCV-MW	MCV-BP
4	5	0.020	0.020	0.020	0.020
2	4	0.057	0.057	0.076	0.082
1	5	0.195	0.168	0.137	0.136
		90% distance from {0, 0.05, 0.1, 0.2, 0.4, 0.25}			
		Amal-MW	Amal-BP	BCV-MW	BCV-BP
4	5	0.111	0.132	0.080	0.080
2	4	0.346	0.333	0.410	0.430
1	5	1.04	1.72	0.598	0.614

case $\Delta = 10$; indeed, only in the most favorable (i) scenario did estimates approach being passable, which is not surprising given the severity of the error.

For each estimator it is useful to know not only how much $\tilde{\mu}_\alpha$ varies about $E[\alpha]$ but also whether $\tilde{\mu}_\alpha$ tends to overestimate or underestimate $E[\alpha]$. Perhaps the most common measure of the tendency to overestimate or underestimate is bias ($E[\tilde{\mu}_\alpha] - E[\alpha]$), which we could estimate by taking the average of the $\tilde{\mu}_\alpha$ values generated in each 100-replication simulation of a parameter scenario and subtracting the corresponding $E[\alpha]$ values. The sample averages of the $\tilde{\mu}_\alpha$ estimates turned out to be highly unstable, however, invariably because of one or two outlandish observations. Instead, we give in Table 2 the 2% trimmed mean of the $\tilde{\mu}_\alpha$ values — the average of the 96 central values — for each estimator and parameter scenario. (In other words, we discarded the two greatest and the two least estimates and took the average of the remaining values.) For $\Delta = 10$, the estimators all have a severe tendency to underestimate $E[\alpha]$. For $\Delta = 100$, on the other hand, the estimators exhibited only mild and somewhat sporadic bias. In fact, at least one estimator had 2% trimmed mean bias less than 0.02 in each parameter scenario. Because of their ratio form, the $\tilde{\mu}_\alpha$ estimators will be biased for most parameter scenarios, but this bias does not appear to be serious when measurement error is not too severe.

Median and 90% distance of estimated source profiles from the true source profile, θ, are given in Table 3. Although these results should conform generally to results for estimating $E[\alpha]$, it is interesting to note that the AMAL estimators performed relatively better than one would expect from that criterion alone. All estimators reflect the increasing difficulty of estimation with worsening of parameter scenario.

Application to simulated source apportionment data

Currie et al. [14] describe the generation of three simulated data sets which were made available to participants of the Mathematical and Empirical Receptor Models Workshop (Quail Roost II). Each was constructed from reported source profiles and real meteorological data from St. Louis over a 40-day period in 1976. We sum-

TABLE 4

Estimation of $E[\alpha]$

Quail Roost II Data Set 1

Estimator	Known source			
	Road	Steel	Coal	Wood
True value, $E[\alpha]$	0.172	0.002	0.063	0.102
Estimate ± standard deviation				
AMAL-MW	0.143 ± 0.041	0.016 ± 0.017	0.042 ± 0.026	0.314 ± 71.1
AMAL-BP	0.143 ± 0.046	0.017 ± 0.026	0.036 ± 0.087	0.556 ± 33.9
BCV-MW	0.163 ± 0.372	0.095 ± 0.073	0.128 ± 0.324	0.238 ± 0.069
BCV-BP	0.163 ± 0.314	0.095 ± 0.073	0.122 ± 0.430	0.237 ± 0.066
95% Confidence interval				
AMAL-MW	(0.084, 0.248)	(0.008, 0.031)	(0.023, 0.442)	(0, 1)
AMAL-BP	(0.085, 0.252)	(0.009, 0.351)	(0.011, 0.122)	(0.179, 1)
BCV-MW	(0, 0.245)	(0.082, 0.374)	(0.034, 0.216)	(0.144, 0.426)
BCV-BP	(0.082, 0.316)	(0, 0.115)	(0, 0.214)	(0.141, 0.402)
Distance between estimated location, true θ				
AMAL-MW	0.034	0.009	0.012	0.262
AMAL-BP	0.034	0.008	0.013	0.886
BCV-MW	0.025	0.034	0.032	0.146
BCV-BP	0.025	0.034	0.030	0.144

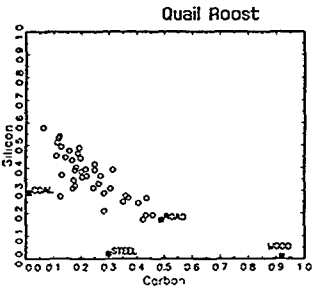


Fig. 2 Quail roost data. 'Ambient' proportional data are represented by circles, with silicon component plotted against a carbon component. Squares represent corresponding values for proportional profiles of known sources.

marize here the performance of our moments estimators on the first of the data sets, which was based on eight source profiles and observations contaminated by normal error. For each of the source profiles coal, road, steel, and wood, we fixed one profile as 'known' and attempted to estimate its influence with respect to the rest, which were aggregated as described below and treated as a single unknown. Since the 'unknown' is really known, we can test how well our methodology estimates it. Eighteen chemical species were used to define proportions — all of the species from which profiles were constructed in ref. 14 with the exception of As and CC (contemporary carbon)

Results are summarized in Table 4. Standard errors and confidence intervals were determined by bootstrap methods (1000 resampling replications) described by Efron [15,16] Actual values of source contributions and, therefore, of the α_i values are given in ref 14; $E[\alpha]$ is taken to be the sample average of the α_i values. θ is obviously not, in fact, a constant parameter However, using actual source contributions, one may calculate a composite source profile for each day The 'actual value' of θ is taken to be the average of the daily composite profiles.

Estimation of $E[\alpha]$ ranged from excellent in the best case ('road') to poor in the worst case ('wood'), with reasonable estimates resulting in the other two cases. The algorithm does, in fact, appear to estimate the average composite profile as the location parameter, θ. Fig. 2 may cast some light on the behavior of the estimates. The road parameter is on the 'edge' of and in line with the bulk of the data, almost as if it were one of two contributing sources, Steel and coal, in the center of the data, appear to be 'in between' other sources, and neither is in line with the data. One would not expect to estimate either one as well as road. Wood, finally, is extremely far from the observed data, which would certainly be expected to cause problems. Estimates of 'Δ' also shed some light on the situation; for road and steel, all estimates were large and stable (> 1000 in the case of road). In the case of wood, especially, estimates were unstable, perhaps an indication that model assumptions are in severe violation (Recall that the simulated errors are additive and Gaussian rather than from our Dirichlet model.) It is reassuring to notice that bootstrap standard errors and intervals identify the poor estimators as being unreliable.

In addition, we attempted to estimate traditional CMB parameters using the principles outlined in the section on the CMB model, above. The most simple transformation to the CMB model may be carried out by substituting observed ambient mass profiles, S_i, for s_i and an estimated source profile, $\hat{\theta}$, for θ above and solving the appropriate equations When the dimension of the observations is greater than the number of sources, as is the case here, one may select a subset of the eqs (5) and (6) to determine the parameters In an attempt to base as many equations on 'known' data as possible, we chose to use all of the eqs (5) and all of the eqs (6) based on the 'known' source profile. Only one equation remained to identify the parameters, we chose the equation from (6) based on the components of the unknown profile for which observed CV was smallest and second-smallest Using this method, we were able to estimate the total source contributions for 'road' quite adequately, indeed, with one exception, we were able to estimate contributions to within a factor of 2 whenever road accounted for more than 0.6% of the total mass. (The exception was within a factor of 3, and estimated values were

generally much closer to true values than a factor of 2 when road accounted for more than 5% of total mass.) Estimation of total source contributions for the other profiles was much less successful. The fact that the composite of the remaining sources behaved very much like a single, second source in the case of road whereas it did not in the other cases accounts for much of this effect. However, estimation of total source contributions in this manner will be difficult whenever measurement error is severe enough to push a sizeable number of the observations 'beyond' the profiles x and θ in the sense described in the section on nonparametric models with measurement error, above.

CONCLUSION

Limitations of standard CMB models led us to introduce SASU models — source apportionment with one source unknown In this paper, we have considered the case of one source known and one source unknown. Inherent to this situation are at least two interesting statistical problems estimation of a structural parameter in the presence of infinitely many incidental parameters and estimation of a parameter which is not, in general, identifiable. The latter problem is easily addressed in the case of no measurement error by requiring that the unknown source is a support boundary (which is eventually attained) of the observation distribution In the case of measurement error, it would appear that deconvolution methods are required in order to identify the unknown source in a completely nonparametric model. We have begun research in this area, but more work is needed before making recommendations.

Parametric models may present a reasonable, practical alternative to the nonparametric approach The Dirichlet model examined appears promising, as an added benefit, it is easily generalizable to the case when there is more than one known source A number of issues need to be considered, however Given model (14), the source contribution parameters λ and Λ not only identify $E[\alpha]$ but all higher moments and, indeed, the exact shape of the distribution of α. The role of the individual parameters λ and Λ if observations do not satisfy eq. (14) is unclear. For the Quail Roost data, the magnitudes of best-CV estimates of λ and Λ appeared reasonable, but amalgamation estimators seemed to underestimate the magnitude quite severely. Research into this phenomenon is necessary. Study of sensitivity to model assumptions is needed, in general. Modifications of the model may be warranted — for example, allowing Δ to vary either with time or with chemical species. Alternat. to moments estimates, such as maximum likelihood estimates, should be available given enough computing power. However, computation of maximum likelihood estimators requires accurate estimators as starting values, so the method of moments estimators should be useful even if maximum likelihood estimators prove to be superior. Finally, other parametric models should be investigated.

The question of how best to transform from the SASU model to the sta..dard CMB model when enough data are available to do so remains open. One may always substitute observed ambient air mass profiles, S_i, for s_i and an estimated source profile, $\hat{\theta}$, for θ. However, the presence of measurement error guarantees that the resulting estimates will not be consistent. We will continue to investigate this question.

A complete approach to the SASU problem will eventually require investigation of numerous complications to the model, including the case when the x_k values are measured with error and the case when observations are correlated. One might like to include observable covariates such as weather or seasonal variables in a reasonable model. Estimation in the case of more than one known source presents an interesting problem as well. While some analog of eq. (9) is probably necessary in order to identify the problem [a Dirichlet model imposes eq (9) naturally], the geometric nature of the problem is somewhat different than in the one-source-known case Research into these issues is underway.

When the unknown 'source' is actually an aggregate of several unknown sources, then it is questionable whether one should model its profile θ as fixed Instead, one might model θ_t, the unknown profile at time t, as a stochastic process,

either stationary or with a time trend depending upon the nature of the unknown sources. In some situations it would be sensible to model θ_i as depending on a covariate.

ACKNOWLEDGEMENTS

David Ruppert was partially supported by NSF Grant DMS-8800294. Both authors were supported by the Army Research Office through the Mathematical Sciences Institute at Cornell. We thank Ray Merrill for introducing one of us (D R.) to this area of research and suggesting the possibility of indirectly estimating the unknown source. This work includes material from the Ph D. Dissertation of K B.-R.

REFERENCES

1 B Aldershof and D Ruppert, A statistical analysis of woodstove PAH emissions and source apportionment of ambient air samples, Unpublished report prepared for EPA, Research Triangle Park, NC, 1987
2 G S Kowalczyk, C.E. Choquette and G E. Gordon, Chemical element balances and identification of air pollution sources in Washington, D C., *Atmospheric Environment*, 12 (1978) 1143–1153
3 J Kiefer and J Wolfowitz, Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *The Annals of Mathematical Statistics*, 27 (1956) 887–906
4 J A Cooper and J G Watson, Receptor oriented methods of air particulate source apportionment, *Journal of the Air Pollution Control Association*, 30 (1980) 1116–1125
5 G E. Gordon, Receptor models, *Environmental Science and Technology*, 22 (1988) 1132–1142
6 R.C. Henry, C.W. Lewis, P K Hopke and H.J. Williamson, Review of receptor model fundamentals, *Atmospheric Environment*, 18 (1984) 1507–1515
7 J Neyman and E.L. Scott, Consistent estimates based on partially consistent observations, *Econometrica*, 16 (1948) 1–32.
8 B G Lindsay, The geometry of mixture likelihoods A general theory, *The Annals of Statistics*, 11 (1983a) 86–94
9 B G. Lindsay, The geometry of mixture likelihood II The exponential family, *The Annals of Statistics*, 11 (1983b) 783–792.
10 J M. Begun, W.J Hall, W. Huang and J.A Wellner, Information and asymptotic efficiency in parametric-nonparametric models, *The Annals of Statistics*, 11 (1983) 432–452
11 G Campbell and J E. Mosimann, Dirichlet covariate models for random proportions, in R M Heiberger (Editor), *Computer Science and Statistics Proceedings of the 19th Symposium on the Interface*, ASA, Alexandria, VA, 1987, pp. 93–101
12 M R Leadbetter, G Lindgren and H Rootzen, *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York 1983
13 R.J. Carroll and P Hall, Optimal rates of convergence for deconvolving a density, *Journal of the American Statistical Association*, 83 (1988) 1184–1186
14 L.A Curne, R W Gerlach, C W Lewis, W D Balfour, J A Cooper, S L Dattner, R T De Cesar, G E Gordon, S L Heisler, P K Hopke, J.J Shah, G D Thurston and H J Williamson, Interlaboratory comparison of source apportionment procedures results for simulated data sets, *Atmospheric Environment*, 18 (1984) 1517–1537
15 B Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1982
16 B Efron, Better bootstrap confidence intervals, *Journal of the American Statistical Association*, 82 (1987) 171–185

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 185–187
Elsevier Science Publishers B.V., Amsterdam

Comments on "Source apportionment with one source unknown" by K. Bandeen-Roche and D. Ruppert

P.K. Hopke * and M.D. Cheng
Department of Chemistry, Clarkson University, Potsdam, NY 13699-5810 (U S A)

An initial remark we would like to make is to note the interest in the receptor modeling problem by statisticians. This paper along with the one elsewhere in these proceedings by L. Gleser provide some of the first efforts to explore the receptor modeling problem as a statistical problem. We think that there are a number of interesting aspects to this particular form of the mixture resolution problem because of the lack of constancy in the source profiles and the errors in the sampling and analyses that make receptor modeling different from mixture resolution using spectrometric data Thus we welcome more statistical inputs and insights into the exploration of sources of airborne pollutants.

The next aspect of this paper that needs to be discussed is that of facilitated communication. It is clear from the paper that receptor modelers have not defined their terminology sufficiently clearly such that people entering the field can immediately adopt our jargon. The paper suggests that the problem they are solving is that of the chemical mass balance (CMB) However, as this term is commonly used within the receptor modeling community, it refers to the resolution of a single sample into its components based on a set of source profiles that are known a priori In the approach outlined here, a number of samples are used to deduce the profile of the 'unknown' source when one or more profiles are known and then obtain the mass contributions of the known sources This method requiring multiple samples then falls into the multivariate methods category as outlined by Cooper and Watson [1] As such, it seems that this new method should be compared with other methods that attempt to deduce source profiles including absolute principal components analysis [2], target transformation factor analysis (TTFA) [3] and SAFER [4]

The model presented in this paper suffers from the need for a basic assumption that the 'unknown' source is constant in composition However, if the 'unknown' source is really a combination of sources, then it is unlikely that this assumption will be valid In a complex, urban airshed, wind direction shifts can drastically alter the number and types of sources [5] and even at more remote sites, there can be highly significant seasonal variations in composition of emissions from various sources so that the applicability and utility of this approach relative to the traditional multivariate approaches is not at all clear.

Before getting into other more detailed comments on the source apportionment with one source unknown (SASU) methods, we would like to raise some other issues regarding communications. This paper is written by statisticians for statisticians and has therefore been written in 'statistics'. However, for us armchair statisticians, it becomes very difficult to read and digest because we first have to translate it from symbolic notations into terms we can follow. We realize that this paper takes advantage of commonly (for the statistics literature) used symbols such as $\subset$, $\forall$, and

[0, 1] We would suspect that most readers of this journal are not going to be able to easily follow the arguments because they get lost in the symbols, and we suggest that although it is cumbersome to do so, these symbols should generally be avoided in papers that are written for non-statisticians to read.

We also would urge that theorems, propositions and the likes be relegated to Appendices rather than breaking the flow of the reasoning in the text. We recognize the heresy of this proposal, but offer it notwithstanding in order to improve communications to the non-statistician

There are a number of other aspects of this paper that we would like to discuss The authors suggest the CMB model cannot deal with time varying source contributions. CMB analysis does not deal with time variation at all because it is performed on only one sample. Time variations in source contributions could only be found by performing a series of CMB analyses on a sequence of samples. Time variation in the source profiles is normally not incorporated because multiple source samples are not often taken at the same time as the ambient samples However, only the financial and access constraint that often plague field studies preclude the incorporation of time variation of the source profiles in the CMB calculations. An alternative approach to incorporate systematic time variations would be to use Kalman filtering It would appear feasible to utilize this method to take such time variation into account. Although it has not yet been studied in the context of the receptor modeling problem, the Kalman filter appears to be a method worthy of further exploration.

There is a statement that the use of additive, Gaussian error structures may be a limitation to a CMB analysis because the observations should be non-negative and may be constrained as when the measurements are proportions summing to one One of the continuing problems in air quality data handling is that of the compulsion to left truncate data Most of our chemical analytical methods have demonstrably symmetric error bands on the results even if the errors are not truly Gaussian For many airborne particle analyses based on photon spectroscopy such as neutron activation or X-ray fluorescence, we know that the count data on which the concentrations are determined have a Poisson distribution and the additivity of the uncertainties can be explicitly calculated. Thus, it is certainly possible that if the sample does not contain the analyte of interest, a measured value less than zero is a valid result. Too many people will then set the value to zero because of their misunderstanding of the effects of the measurement error. Thus, some of the starting premises of this work seem to be in error.

In the non-parametric model, they suggest that in the limit of sufficiently large numbers of samples being taken and analyzed, there will be one that will be composed almost entirely of the species contributed by the 'unknown' source This assumption again raises the problem of the constancy of the mixture of unknown sources that constitute the 'unknown' source Although the wood stove would not be burned in the summer, there may be other sources that are on in the summer but not in the winter. The real situation is not likely to be as simple as portrayed here.

It also appears that it is necessary to know the probability distribution of the 'unknown' source contributions $G(\alpha)$. It has not yet been done for any source to the extent that the distribution of values is known Thus, at this time, this approach does not appear to provide practical help to the receptor modeler particularly in light of the other problems that arise when measurement error is introduced into the model.

One of the problems with the use of proportional data is that ultimately the results will need to be back transformed into absolute concentrations ($\mu g/m^3$) to be used by air quality managers. It will be necessary to provide a method to give such values with associated error bounds if the method is to be applied to real air quality management problems

In the parametric model studies, the stimulated data were assumed to have identical and constant errors for all chemical species from all of the sources The authors note this is unrealistic We would encourage further study with more realistic error structures so that any possible points at which the analysis shows problems can be identified.

Finally in the analysis of the Quail Roost II data set, it is interesting that SASU was able to estimate the STEEL source even though it was well below the 'detection limits' as defined by Currie et al. [6] It seems surprising that "WOOD" was so poorly estimated as it could be found relatively well using the other multivariate methods [6] It would be interesting to know if the choices of Si and C are unique in showing the results presented in Fig. 2 or whether there are other pairs of variables that show the same pattern. The results on the Quail Roost data also suggest that the Dirichlet distribution was not a very good representation of the needed distribution for these data sets. Since these sets are based on a reasonably realistic data generation model, it suggests that there is a need to explore other distributions beside the Dirichlet to find one that better represents air quality data distributions

In conclusion, we welcome the increased input of statisticians into the receptor modeling field We hope that we can open lines of communications so that the problems examined can better relate to actual receptor modeling problems and ask the indulgence of the statistics community to be patient with those of us who are not fluent in symbolic logic symbols and thus find great difficulty in reading and understanding the work that is being presented

REFERENCES

1 J A Cooper and J G Watson, Receptor-oriented methods of air particulate source apportionment, *Journal of Air Pollution Control Association*, 30 (1980) 1116–1175
2 G D Thurston and J D Spengler, A quantitative assessment of source contributions to inhalable particulate matter pollution in metropolitan Boston, *Atmospheric Environment*, 19 (1985) 9–24
3 P K Hopke, Target transformation factor analysis as an aerosol mass apportionment method a review and sensitivity analysis, *Atmospheric Environment*, 22 (1988) 1777–1792
4 R C Henry and B M Kim, Extension of self-modeling curve resolution to mixtures of more than three components Part 1 Finding the basic feasible region, *Chemometrics and Intelligent Laboratory Systems*, 8 (1990) 205–216
5 S W Rheingrover and G E. Gordon, Wind-trajectory method for determining compositions of particles from major air pollution sources, *Aerosol Science & Technology*, 8 (1988) 29–61
6 L.A Currie, R W Gerlach, C W Lewis, W D Balfour, J A. Cooper, S L Dattner, R T DeCesar, G E Gordon, S L Heisler, P K Hopke, J J Shah, G D Thurston and H J Williamson, Interlaboratory comparison of source apportionment procedures results for simulated data sets, *Atmospheric Environment*, 18 (1984) 1517–1537

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 189–198
Elsevier Science Publishers B V., Amsterdam

Mathematical topics in combustion

J. Buckmaster
University of Illinois, Urbana, IL 61801 (U S A)

(Received 8 November 1989; accepted 4 January 199ᵥ)

Abstract

Buckmaster, J , 1991 Mathematical topics in combustion *Chemometrics and Intelligent Laboratory Systems*, 10 189–198

A number of mathematical approaches that are currently of interest in theoretical combustion are briefly described These are (1) activation energy asymptotics — flame-sheets and hot-spots, (2) bifurcations and routes to chaos, (3) turbulent premixed flames — fractals and renormalization, (4) reduced chemistry and rate-ratio asymptotics, (5) nonlinear high-frequency acoustics and combustion

PROLOGUE

With rare exceptions, combustion is fluid mechanics with the addition of highly exothermic, temperature-sensitive chemical reaction Progress in combustion theory has therefore been closely linked to tools that have been developed to deal with the reaction terms, and this is apparent in the topics discussed here. Section 1 briefly describes a successful asymptotic treatment based on the idea of extreme sensitivity of the reaction rate to temperature variations This can lead to flamesheet models in which reaction is confined to thin layers, and this provides a powerful tool for examining flame stability, the subject of Section 2. At high Reynolds numbers the role of chemistry is reduced to generating a hydrodynamic flame, a temperature and density discontinuity separating two inviscid flow fields (Section 3). More subtle aspects of the chemical kinetics play a role in Section 4, which describes a rational procedure for reducing complex kinetic systems to reduced sets involving three or four reaction steps. Our discussion concludes in Section 5 with the interaction of high frequency acoustic waves and a combustion field. Of particular interest is the fact that a small-amplitude nonlinear periodic wavetrain can accelerate a temperature-sensitive reaction

1 ACTIVATION ENERGY ASYMPTOTICS — FLAME-SHEETS AND HOT-SPOTS

It is commonplace in combustion theory to adopt a simple one-step kinetic model characterized by Arrhenius kinetics For premixed flames this might have the form

mixture → products

at a rate

$$\Omega = DYe^{-\theta/T} \tag{1.1}$$

0169-7439/91/$03 50

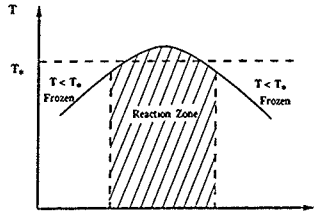


Fig 1 Flame-sheet separating two regions of frozen flow. This is typical of the structure seen in diffusion flames [1]

where Y is the mixture fraction, T the temperature; for diffusion flames,

fuel + oxygen → products

at a rate

$$\Omega = DXYe^{-\theta/T} \qquad (1.2)$$

where X (Y) is the oxygen (fuel) mass fraction θ is a nondimensional activation energy or activation temperature.

Asymptotic treatments are possible in the limit $\theta \to \infty$ and have proven to be of great value in elucidating a wide range of combustion phenomena [1–3] For some problems the asymptotics lead to flame-sheets, thin regions in which there is a balance between diffusion and reaction; beyond the flame-sheet reaction is negligible. This comes about by considering the distinguished limit

$$D \to \infty, \quad \theta \to \infty, \quad D = e^{\theta/T^*}, \quad T^* \text{ fixed} \qquad (1.3)$$

This immediately leads to a partition of the flow-field into regions where $T < T^*$ so that $\Omega \to 0$ (frozen chemistry), and regions where $T > T^*$ so that $Y \to 0$ or $XY \to 0$ (equilibrium chemistry), and again $\Omega \to 0$ for the irreversible kinetics of eqs. (1.1) and (1.2) *.

The thin reaction zone or flame-sheet is characterized by

$$T = T^* + O(1/\theta) \qquad (1.4)$$

* In a special but important case, the plane deflagration, an unbounded region of equilibrium gas exists where $T = T^*$.

e g. Fig 1. These flame-sheet structures are well understood and the approach is a well established and proven tool

A quite different class of problems involves hot-spot formation and ignition Consider the following simple model for homogeneous thermal ignition.

$$dT/dt = e^{-\theta/T}, \qquad T(0) = T_0 \qquad (1.5)$$

Adopting the ansatz

$$T = T_0\left(1 + \frac{1}{\theta}\phi + \ldots\right) \qquad (1.6)$$

the perturbation function ϕ satisfies the initial-value problem

$$d\phi/d\tau = e^{\phi}, \qquad \phi(0) = 0 \qquad (1.7)$$

where τ is a scaled time This has solution

$$\phi = -\ln(1 - \tau) \qquad (1.8)$$

valid for $0 \leqslant \tau < 1$. Thermal runaway occurs at $\tau = 1$. In nonhomogeneous problems, runaway is confined to a small region called a hot-spot. A well-known example occurs in a certain type of deflagration-to-detonation transition [4] A weak shock is generated by the accelerating flame, and in the shocked gas a hot-spot forms and gives rise to an expanding shock which interacts with, and reinforces, the lead shock (Fig. 2).

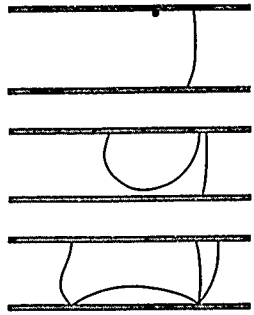


Fig 2 Hot-spot formation and initiation of a shock in deflagration-to-detonation transition (cartoon based on plate 5 of ref 4)

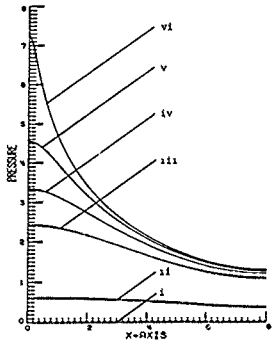


Fig. 3 Pressure distribution at different times in an interior hot-spot. From ref 5 with permission

Only recently have these hot-spots been analyzed for a compressible gas, and Fig. 3 shows the early pressure rise for an interior hot-spot one not next to a wall) [5]. Density changes are shown in Fig. 4; the process is so rapid that no significant mass flux can occur, and these changes are small (inertial confinement). Recent efforts have been concerned with the consequences of hot-spot formation [6,7].

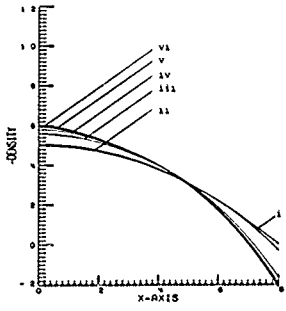


Fig. 4 Density distribution at different times in an interior hot-spot From ref 5 with permission

2 BIFURCATIONS AND ROUTES TO CHAOS

The constant density model for premixed flames can take the form (see ref. 2, p. 25):

$$\frac{\partial T}{\partial t} = \Delta T + \Omega, \qquad \frac{\partial Y}{\partial t} = \frac{1}{Le}\,\Delta Y - \Omega \tag{2.1}$$

with Ω given by eq. (1.1). Here Le is the Lewis number and values of Le different from 1 can give rise to Turing instabilities [8].

As noted in Section 1, in the limit $\theta \to \infty$ reaction is confined to a thin flame sheet. Indeed, for deflagrations that are nominally plane and adiabatic, Ω behaves like a Dirac δ-function of strength $\sim e^{-\theta/2T*}$ where $T*$ is the flame temperature. It is then not difficult to construct a stationary solution (unchanging flame propagation), whose linear stability can be explored using a modal analysis If the flame-sheet displacement is

$$x_f = -W_{ad}t + \epsilon e^{\alpha t + iky}, \qquad \epsilon \to 0 \tag{2.2}$$

where the unperturbed flame propagates to the left at the adiabatic flame speed, the stability diagram Fig. 5 can be constructed [9].

In the neighborhood of P long wavelength disturbances grow very slowly and weak nonlinearities can be incorporated into the analysis by means of a bifurcation analysis. In this way the Kuramoto-Sivashinsky equation can be derived [10] for $\phi \sim x_f + W_{ad}t$, and when corrugations in the z direction are also admitted this is

$$\phi_t + \tfrac{1}{2}(\nabla\phi)^2 = -\nabla^2\phi - 4\nabla^4\phi \tag{2.3}$$

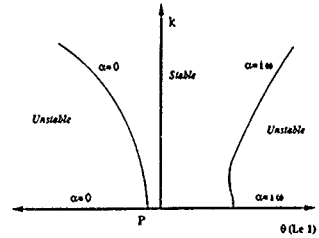


Fig 5 Stability boundaries in the wavenumber-scaled Lewis number plane

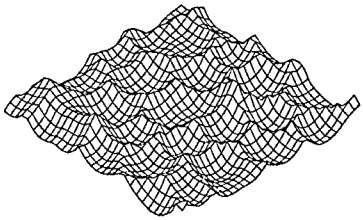


Fig. 6. Numerical solution of the Kuramoto–Sivashinsky equation. From ref 2

The first term on the right is viscous-like with a negative viscosity coefficient and is strongly destabilizing; the ∇^4 term stabilizes short waves. Numerical simulations show that the flame-sheet adopts an irregular, unsteady, cellular configuration (Fig. 6). Physical flames in mixtures with $Le < 1$ can display similar behavior (Fig. 7) [11] (see also ref. 1, p. 194).

Fig. 5 shows the stability boundaries for an unbounded flame. If we consider flames that are attached to burners, accounting for the heat flux to the burner, the left stability boundary is modified (Fig. 8). If at the same time the burner geometry restricts the wave number k to discrete values, discrete points on this boundary are defined, each of which is a potential bifurcation point from which can spring a nonplanar solution. These various solutions can interact (e.g. bimodal bifurcations) and display interesting dynamical behavior. Analysis [12–14] can explain the behavior of polyhedral flames, multiple-sided flames sometimes seen on Bunsen burners (Fig. 9). These are sometimes stationary, sometimes they spin, and the number of sides can be changed by varying the combustion parameters (mass flow-rate, mixture strength).

The right stability boundary of Fig. 5 is relatively inaccessible to physical mixtures but has a counterpart in the analysis of thermites, which are solids that burn to form solids and so have $Le = \infty$. In the k–θ plane (θ is no longer asymptotically large [15]), and again with k restricted to discrete values, possible bifurcation points are identified in Fig. 10).

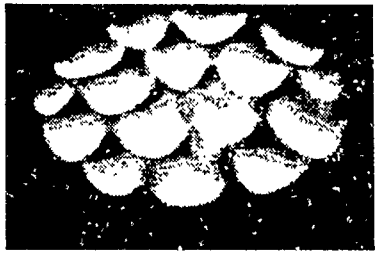


Fig. 7 Cellular flames, courtesy of M. Gorman [11]. An optical illusion can make these look like liquid drops on the underside of a plate, with the white regions corresponding to convex surfaces. In fact these are top views of the flame with concave or cup-like white regions, each cup being surrounded by a multiple-sided sharp ridge. As with many optical illusions, persistence will cause the image to 'flip'

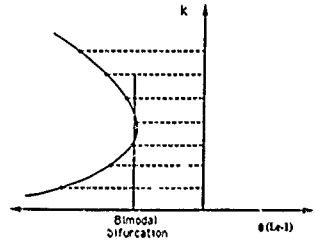


Fig. 8. Modification of the left stability boundary of Fig. 5 by heat losses, showing possible bifurcation points when k is restricted to discrete values

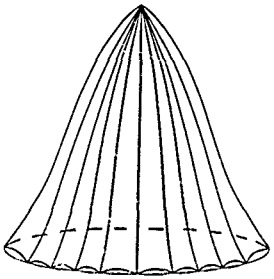


Fig 9 Polyhedral flame A cartoon based on a photograph in ref 1

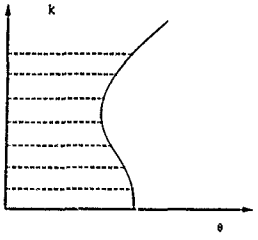


Fig 10. Possible bifurcation points corresponding to planar corrugations of thermite flames. A similar figure can be constructed for cylindrical geometry

A rich dynamic structure is associated with bifurcations from the right stability boundaries [16–18]. Fig. 11 shows variations of the flame speed with time for a problem discussed in ref. 18 and exhibits $2-T$ periodic behavior. Fig. 12 corresponds to slightly different parameter values from those of Fig 11 and apparently displays chaotic behavior.

3 TURBULENT PREMIXED FLAMES – FRACTALS AND RENORMALIZATION

Fig. 13 shows premixed flame images obtained in a laboratory engine at Princeton University

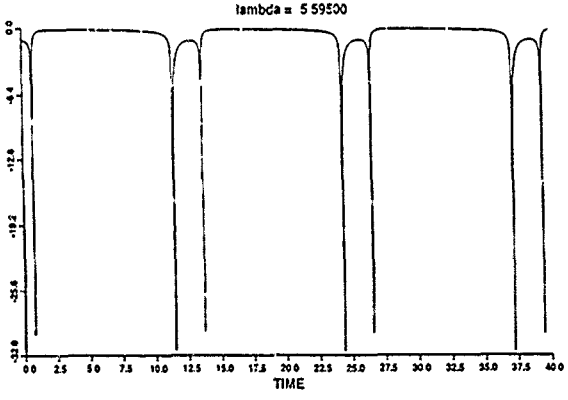


Fig 11 Flame front velocity vs. time in thermite burning. From ref 18 with permission This displays $2T$ periodic behavior

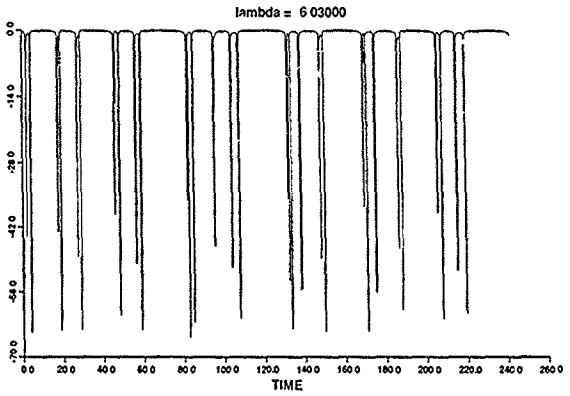


Fig 12 Flame front velocity vs. time in thermite burning. From ref 18 with permission This appears to display chaotic behavior

[19,20]; the flame is the boundary between the products (white) and the reactants (black) These images are typical of turbulent flames, and one may ask whether or not the flame is a fractal surface. To answer this question it is necessary to measure the surface area using 'rulers' of different size, plotting the area vs. the 'ruler' length on a log–log plot (Fig. 14). Between large- and small-

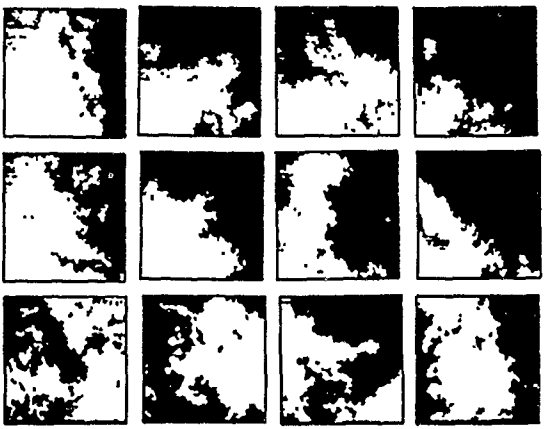


Fig. 13. Flame images in an internal combustion engine at 2400 rpm. From ref. 19 with permission The equivalence ratios are 0 9, 0 8, 0 7 (top to bottom).

scale cutoffs, a fractal surface is characterized by a straight line of slope $(2-D)$, $2 \leqslant D < 3$, where D is the fractal dimension. Note that fractal behavior is only observed over a 1-decade range of length scales, and this may be too small for the concept to be of value.

Turbulent flames travel faster (W_{turb}) than laminar flames (W_{lam}) because of the enhanced average burning area generated by the wrinkling. Discarding other effects (e.g. flame-stretch, ref. 1, p. 146),

$$\frac{W_{turb}}{W_{lam}} = \frac{A_t}{A_0} = \left(\frac{\lambda_t}{\lambda_0}\right)^{2-D} \tag{3.1}$$

(see Fig. 14), and Gouldin [22] has used this idea to predict turbulent flame speed as a function of turbulent intensity. Fig. 15 shows some of his results. For other mixtures the agreement is not as good; moreover Gouldin's choice of λ_t (the Kolmogoroff length scale) has been questioned [23]. Nevertheless, the agreement is encouraging.

Some related mathematical treatments have dealt with the kinematic flame equation

$$\frac{\partial G}{\partial t} + (\tilde{v}\cdot\tilde{\nabla})G = W_{lam}|\tilde{\nabla}G| \tag{3.2}$$

which governs a scalar function $G(\tilde{x}, t)$ where the surface $G = 0$ represents the flame. This surface is convected by the flow field $\tilde{v}$ and propagates relative to the fluid at the laminar flame speed. Given a turbulent flow $\tilde{v}$ we can ask what turbulent flame speed will be predicted by this equation [24–26].

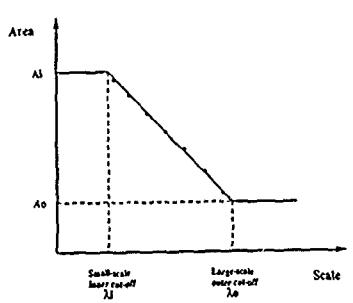


Fig. 14. Area vs. scale for a fractal surface. The data points are obtained from ref. 21, corresponding to a tube-burner flame.

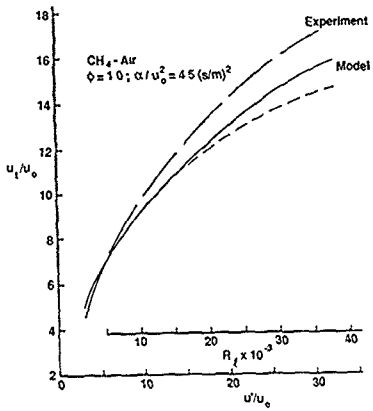


Fig. 15 Turbulent flame-speed vs turbulent intensity. From ref. 22 with permission

The turbulent field is characterized by a wide range of scales $\{l\}$ where $l_0 > l > l_i$ (outer and inner cut-offs), and we define

$$G(l_1) = \langle G(\tilde{x}, t)\rangle_{l_1} \tag{3.3}$$

the average of G over all length scales $l_1 > l > l_i$. Similarity on the different length scales implies that

$$\frac{\partial G(l_1)}{\partial t} + (\tilde{v}(l_1)\cdot\tilde{\nabla})G(l_1) = W_{turb}(l_1)|\tilde{\nabla}G(l_1)| \tag{3.4}$$

where $W_{turb}(l_1)$ is a 'partial' turbulent flame-speed associated with wrinkling on the scales smaller than l_1. By definition

$$W_{turb}(l_1) \to W_{turb} \text{ as } l_1 \to l_0 \tag{3.5}$$

so that W_{turb} can be calculated if the averaging procedure leading to eq. (3.4) can be carried out. Existing analyses yield (ibid.)

$$W_{turb} \sim u_{rms} \tag{3.6}$$

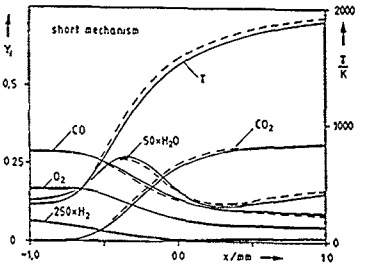


Fig 16 Calculated structure of a wet CO flame using the complete mechanism and the short mechanism From ref 27 with permission

or

$$W_{\text{turb}} \sim u_{\text{rms}} \left[\ln \frac{u_{\text{rms}}}{W_{\text{lam}}} \right]^{1/2} \tag{3 7}$$

4 REDUCED CHEMISTRY AND RATE-RATIO ASYMPTOTICS

The chemistry of physical flames is extremely complicated, presenting an insurmountable obstacle to analysis and a severe challenge to numerical simulations unless substantial simplifications are introduced. Consider, for example, wet CO flames [27]. A complete description of the kinetics involves 67 steps with rates characterized by 162 nonzero parameters and a commensurate number of reactants. Even after unimportant reactions are discarded, 21 steps remain governing 10 species. (The accuracy of such short mechanisms can be checked by comparing the flame structures they yield with exact calculations, Fig. 16). Clearly additional simplification is necessary and two simple ideas play an important role in this connection: the steady-state approximation for an intermediate and the quasi-equilibrium approximation for a reaction.

Consider the ith species. Its variation due to reaction can be written in the form

$$\frac{dc_i}{dt} = w_i^+ - w_i^- \tag{4.1}$$

where w_i^+ refers to the positive contribution from the various reactions (production) and w_i^- refers to the negative contribution (consumption). The steady state approximation, valid if c_i is small compared to each term on the right, is

$$w_i^+ \simeq w_i^- \tag{4.2}$$

If we just examine the change in c_i due to the jth reaction, then

$$\left.\frac{dc_i}{dt}\right|_j = k_{f_j} - k_{r_j} \tag{4.3}$$

where k_f (k_r) is the forward (reverse) reaction rate, and the quasi-equilibrium approximation is

$$k_{f_j} \simeq k_{r_j} \tag{4.4}$$

When these approximations are correctly applied, substantial simplification is possible and yet reasonable accuracy is maintained. As an example, for stoichiometric methane/air flames a four-step scheme can be deduced [28],

$$CH_4 + O_2 \xrightarrow{k_1} CO_2 + H_2 + H_2O$$

$$CO + H_2O \leftrightarrow CO_2 + H_2$$

$$O_2 + 2H_2 \xrightarrow{k_3} 2H_2O \tag{4 5}$$

Additional approximations are sometimes possible permitting analytical treatment of flame-structure. Thus, in eq (4.5), $k_3 \ll k_1$, so that we can define the parameter

$$\delta = \frac{k_3}{k_1} \tag{4 6}$$

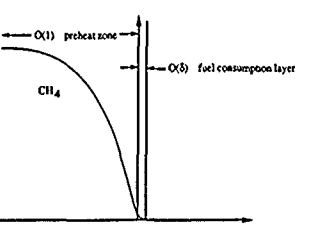


Fig. 17. Example of a structural simplification arising from rate-ratio asymptotics After a figure in ref. 28.

and examine the limit $\delta \to 0$ (rate–ratio asymptotics). In this limit the fuel-consumption layer is of vanishing thickness and its structure can be analyzed using the ansatz

$$x = 0(\delta), \quad [CH_4]_o = 0, \quad T = T_o + 0(\delta),$$
$$C_t = C_{t_o} + 0(\delta) \tag{4.7}$$

(see Fig. 17) Further details may be found in ref. 28.

5 NONLINEAR HIGH-FREQUENCY ACOUSTICS AND COMBUSTION

Auto-ignition is important in many combustion problems. In Section 1 we indicated the role that it can play in one type of deflagration-to-detonation transition, and it is central to engine knock in which point ignition occurs ahead of the primary flame front. High-frequency waves (generated by turbulence or inhomogeneities) might have a significant impact on this process, and recently there have been some interesting extensions of nonlinear high-frequency acoustic theory to the problem of propagation through reacting gases [29,30].

A periodic sound wave propagating to the right through a uniform time-independent medium (constant background) is described by

$$u = u^{\mathrm{II}} + \epsilon e^{i w(x - t\sqrt{T^{\mathrm{II}}})T}(1, \sqrt{T^{\mathrm{II}}}, 0, 0) + \ldots \tag{5.1}$$
$$u = {}^{T}(\rho, v, S, Y)$$

(ρ = density, v = velocity, S = entropy, Y = mass fraction, $\sqrt{T^{\mathrm{II}}}$ = speed of sound, $(\ )^{\mathrm{II}}$ = background).

If, instead, the background is homogeneous but nonconstant, corresponding to a homogeneous explosion, so that

$$\frac{1}{\gamma(\gamma-1)}\frac{\mathrm{d}T^{\mathrm{II}}}{\mathrm{d}t} = -\beta\frac{\mathrm{d}Y^{\mathrm{II}}}{\mathrm{d}t} = Y^{\mathrm{II}}e^{-A/T^{\mathrm{II}}} \tag{5.2}$$

[cf. eq. (1.5), an early-time approximation valid when reactant depletion can be neglected], then we adopt the ansatz [30]

$$u = u^{\mathrm{II}} + \epsilon a\left(x, t, \frac{\phi}{\epsilon}\right)^{T}(1, \sqrt{T^{\mathrm{II}}}, 0, 0) + \epsilon^2 u_2 + \ldots \tag{5.3}$$

for small-amplitude high-frequency waves. When substituted into the governing equations, with attention restricted to a single right-moving wave, solutions $\sigma(x, t, \theta)$ valid as $\epsilon \to 0$ satisfy

$$\frac{\partial\phi}{\partial t} + \sqrt{T^{\mathrm{II}}}\frac{\partial\phi}{\partial x} = 0 \tag{5.4}$$

$$\frac{\partial\sigma}{\partial t} + \sqrt{T^{\mathrm{II}}}\frac{\partial\sigma}{\partial x} + \frac{(\gamma+1)}{2}\sqrt{T^{\mathrm{II}}}\sigma\frac{\partial\sigma}{\partial\theta} = \frac{T^{\mathrm{II}}}{T^{\mathrm{II}}}\left(\frac{1}{2\gamma} - \frac{3}{4} + \frac{(\gamma-1)A}{2\gamma T^{\mathrm{II}}}\right)\sigma \tag{5.5}$$

An appropriate solution of eq (5.4) is

$$\phi = x - \int_0^t \sqrt{T^{\mathrm{II}}}\,\mathrm{d}\xi \tag{5.6}$$

and it may be noted that in the limit $\sigma \to 0$, $T^{\mathrm{II}} \to 0$ (vanishing amplitude, constant background) the solution

$$\sigma = e^{i\theta} \tag{5.7}$$

recovers eq (5.1) with $w = \epsilon^{-1}$. The nonlinear term in eq (5.5) will cause dissipation if (and only if) shocks form, but the term on the right can lead to a growth in amplitude

Nonlinear feedback can occur, with the acoustic signal affecting the mean field (background) if the activation energy is large and

$$\epsilon \to 0, \quad A \to \infty, \quad \epsilon A \text{ fixed} \tag{5.8}$$

As an example, during the induction phase of an explosion [when the ansatz (1.6) is valid], and for a left-moving wave [29],

$$\sigma_1 = \bar{\sigma}_1(x,t) + \sigma(\theta, x, t), \qquad \theta = \frac{x+t}{\epsilon}$$
$$\bar{T}_1 = (\gamma-1)(\bar{\sigma}_1 + \bar{\sigma}_2 + \bar{\sigma}_3)$$
$$2\gamma(\bar{\sigma}_{1_t} - \bar{\sigma}_{1_x}) = \gamma(\gamma-1)\bar{\sigma}_{2_t} = 2\gamma(\bar{\sigma}_{3_t} + \bar{\sigma}_{3_x})$$
$$= e^{\bar{T}_1}\left\{e^{(\gamma-1)\sigma}\right\}$$
$$\sigma_t - \sigma_x - 2\{(\gamma+1)\bar{\sigma}_1 + (\gamma-1)\bar{\sigma}_2 + (\gamma-3)\bar{\sigma}_3\}\sigma_\theta$$
$$- \tfrac{1}{4}(\gamma+1)(\sigma^2)_\theta = \frac{1}{2\gamma}e^{\bar{T}_1}\left\{e^{(\gamma-1)\sigma} - \left[e^{(\gamma-1)\sigma}\right]\right\} \tag{5.9}$$

Here, all the perturbation quantities can be written in terms of σ_1, $\bar{\sigma}_2$ and $\bar{\sigma}_3$ (e.g. $S = \epsilon\bar{\sigma}_2$); $\bar{T}_1$ is

the perturbed mean field temperature, nonzero because of the nonlinear feedback; and the average is taken over the θ variable. Using these equations it can be shown that ignition [i.e. thermal runaway as identified with the result (1.8)] will occur earlier because of the presence of the acoustic wave.

ACKNOWLEDGEMENTS

This work was supported by AFOSR. I am grateful to A. Kapila, B. Matkowsky, M Gorman, F. Gouldin, P Felton, and F. Williams for providing some of the figures; and to C.J. Lee for drawing the others, especially Figs. 2 and 9.

REFERENCES

1 J D Buckmaster and G S S Ludford, *Theory of Laminar Flames*, Cambridge University Press, New York, 1982.
2 J D Buckmaster and G S S Ludford, *Lectures on Mathematical Combustion*, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM Press, Philadelphia, PA, 1983
3 F Williams, *Combustion Theory*, Benjamin/Cummings, Menlo Park, CA, 2nd ed , 1985
4 R Urtiew and A.K. Oppenheim, Experimental observations of the transition to detonation in an explosive gas, *Proceedings of the Royal Society of London Series A*, 295 (1966) 13–28
5 T L. Jackson, A K Kapila and D S Stewart, Evolution of a reaction center in an explosive material, *SIAM Journal on Applied Mathematics*, 49 (1989) 432–458
6 A K Kapila and J W Dold, A theoretical picture of shock-to-detonation transition in a homogeneous explosive, paper presented at the *Ninth Symposium (International) on Detonation, Portland, OR, August 1989.*
7 J W. Dold and A K Kapila, Asymptotic analysis of detonation initiation for one-step chemistry· I — emergence of a weak detonation, submitted for publication
8 A M. Turing, The chemical basis of morphogenesis, *Philosophical Transactions of the Royal Society*, B237 (1952) 37–72
9 G I. Sivashinsky, Diffusional-thermal theory of cellular flames, *Combustion Science and Technology*, 15 (1977) 137–145.
10 G I. Sivashinsky, Instabilities, pattern formation, and turbulence in flames, *Annual Review of Fluid Mechanics*, 15 (1983) 179–199.
11 N. el-Hamdi, M. Gorman and K. Robbins, *A Picture Book of Dynamical Modes of Flat, Laminar Premixed Flames*, Report, Department of Physics, University of Houston, Houston, TX, 1990.
12 J D Buckmaster, Polyhedral flames — An exercise in bimodal bifurcation analysis, *SIAM Journal on Applied Mathematics*, 44 (1984) 40–55.
13 S B. Margolis and G I Sivashinsky, On spinning propagation of cellular flames, submitted for publication
14 D O. Olagunju and B.J. Matkowsky, Polyhedral flames, *SIAM Journal on Applied Mathematics*, in press
15 B J Matkowsky and G I. Sivashinsky, Propagation of a pulsating reaction front in solid fuel combustion, *SIAM Journal on Applied Mathematics*, 33 (1978) 465–478.
16 B.J. Matkowsky, Interaction of pulsating and spinning waves in condensed phase combustion, *SIAM Journal on Applied Mathematics*, 46 (1986) 801–843.
17 S B Margolis and B.J. Matkowsky, New modes of quasi-periodic combustion near a degenerate hopf bifurcation point, *SIAM Journal on Applied Mathematics*, 48 (1988) 828–853
18 A Bayliss and B.J Matkowsky, Two routes to chaos in solid fuel combustion, *SIAM Journal on Applied Mathematics*, 50 (1990) 437–439
19 J Mantzaras, Three-dimensional visualization of premixed charge engine flames. *Ph D Thesis*, Princeton University, Princeton, NJ, December 1989, No 1878-T
20 J Mantzaras, P G Felton and F.V Bracco, Fractals and turbulent premixed engine flames, *Combustion and Flame*, 77 (1989) 295–310
21 M Murayama and T Takeno, Fractal-like character of flamelets in turbulent premixed combustion, *Twenty-Second Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, PA, 1988 pp 551–559.
22 F C Gouldin, An application of fractals to modeling premixed turbulent flames, *Combustion and Flame*, 68 (1987) 249–266.
23 N. Peters, personal communication.
24 V Yakhot, Propagation velocity of premixed turbulent flames, *Combustion Science and Technology*, 60 (1988) 191–214
25 V Yakhot, Scale invariant solutions of the theory of thin turbulent flame propagation, *Combustion Science and Technology*, 62 (1988) 127–129
26 G.I Sivashinsky, Cascade-renormalization theory of turbulent flame speed, *Combustion Science and Technology*, 62 (1988) 77–96.
27 B. Rogg and F A Williams, Structures of wet CO flames with full and reduced kinetic mechanisms, *Twenty-Second Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, PA, 1988, pp. 1441–1451.
28 N. Peters and F A. Williams, The asymptotic structure of stoichiometric methane–air flames, *Combustion and Flame*, 68 (1987) 185–207.
29 A Majda and R.R Rosales, Nonlinear mean field-high frequency wave interactions in the induction zone, *SIAM Journal on Applied Mathematics*, 47 (1987) 1017–1039.
30 Y.S Choi and A Majda, Amplification of small amplitude high-frequency waves in a reactive mixture, *SIAM Review*, 31 (1989) 401–427

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 199–210
Elsevier Science Publishers B.V., Amsterdam

Stochastic aspects of turbulent combustion processes

G.M. Faeth *, M.E. Kounalakis and Y.R. Sivathanu
Department of Aerospace Engineering, The University of Michigan, Ann Arbor, MI (U.S.A.)

(Received 8 November 1989, accepted 4 January 1990)

Abstract

Faeth, G.M., Kounalakis, M.E. and Sivathanu, Y.R., 1991. Stochastic aspects of turbulent combustion processes. *Chemometrics and Intelligent Laboratory Systems*, 10: 199–210.

Methods of using stochastic simulations to treat nonlinear interactions in turbulent combustion processes are described — emphasizing the use of statistical time-series techniques to analyze the turbulence–radiation interactions of nonpremixed flames. Three aspects of the problem are considered, as follows: the statistics of scalar properties in turbulent flames, the formulation of algorithms to simulate flame radiation based on flame statistics, and evaluation of the methodology using recent measurements for nonluminous flames. It is shown that the process becomes tractable through the laminar flamelet approximation whereby all scalar properties are taken to be solely functions of a conserved scalar like the mixture fraction. Thus, the simulations are designed to generate realizations of mixture fractions along radiation paths with the radiation properties of each realization found using a narrow-band radiation model. An autoregressive process that reproduces probability density functions and spatial and temporal correlations of mixture fractions was found to yield reasonably good predictions of the statistical properties of spectral radiation intensities measured for turbulent carbon monoxide and hydrogen jet flames burning in still air. Although the approach appears to be promising, additional development is needed in order to treat some of the unique statistical features of turbulence that are not encountered during conventional use of statistical time-series techniques.

INTRODUCTION

Stochastic simulations are promising for treating a variety of nonlinear interactions in turbulent flows. Recent studies along these lines include the turbulent dispersion of particles and bubbles [1–5], the motion and transport of drops in evaporating and combusting sprays [6,7], and the turbulence–radiation interactions of nonpremixed flames [8–13]. The objective of the present paper is to describe the application of this methodology to processes encountered in turbulent combusting flows. In order to control the scope, the discussion will focus on turbulence–radiation interactions of nonpremixed (diffusion) flames, since this problem involves the most significant features of stochastic simulations of turbulent combustion processes.

Initially, methods of simulating turbulent processes were relatively ad hoc [1,2], however, more systematic techniques currently are being emphasized. This includes full stochastic simulation of the turbulent field, along the lines of Kraichnan [14], to study the turbulent dispersion

0169-7439/91/$03.50

of particles in an isotropic turbulent field [15], and adapting statistical time-series techniques, analogous to methods described by Box and Jenkins [15], for problems of turbulent dispersion of particles [3,5] and turbulence–radiation interactions [13] The present discussion will be limited to statistical time-series techniques since they have modest computational requirements and provide reasonable flexibility for treating a variety of practical turbulent flows.

The main reason for interest in turbulence–radiation interactions is that radiation levels of turbulent flames are generally higher (often 2–3 times higher) than estimates based on mean scalar properties within the flames [8–12] The bias of mean radiation levels is caused by nonlinear relationships between scalar and radiation properties in flames This precludes averaging scalar properties first and then computing radiation properties, instead, the radiation properties of realizations of the scalar field must be found first and then averaged Properties other than mean radiation levels are also of interest, for example, fire and flame detectors often use the temporal properties of flame radiation fluctuations to distinguish flames from background radiation. Furthermore, maximum (rather than average) flame radiation levels provide the most conservative estimate of flame radiation properties for fire safety considerations. Finally, studying the temporal properties of radiation fluctuations (moments, probability density functions, and temporal power spectral densities) provides information to better understand turbulence–radiation interactions, analogous to the information provided by he temporal properties of velocity and concentration fluctuations to better understand turbulent mixing. Thus, the general problem of turbulence–radiation interactions involves both the mean and fluctuating radiation properties of turbulent flames [11,12].

Statistical time-series simulations of the radiation properties of turbulent flames are based on simulation of scalar properties within the flames. Therefore, the paper begins with a description of the statistics of scalar properties in turbulent flames. The formulation of typical stochastic simulations is then considered. The paper concludes with evaluation of the methodology using measurements from turbulent hydrogen and carbon monoxide jet flames burning in still air

SCALAR PROPERTIES OF DIFFUSION FLAMES

Scalar property correlations

Assuming equal exchange coefficients of all species and heat, negligible effects of potential and kinetic energies and radiation, and reaction occurring at an infinitely-thin flame sheet, Burke and Schumann [16] showed that scalar properties in laminar nonpremixed flames were functions (called state relationships) of any one of a number of conserved scalars. Although the formal requirements are rather restrictive, state relationships have been found for many laminar flame systems and are widely used for analysis of flame structure and radiation properties The use of state relationships has also been extended to turbulent nonpremixed flames, since they generally can be approximated as wrinkled laminar flames The use of state relationships for turbulent nonpremixed flames has come to be called the conserved-scalar formalism under the laminar flamelet approximation [17,18].

Typical state relationships are illustrated in Fig 1. This involves measurements of the concentrations of major gas species and temperature, T, for radial traverses at various heights, x, above a burner having diameter, d, as well as axial traverses, within laminar nonpremixed carbon monoxide/air flames having various burner Reynolds numbers, Re. In this case, the conserved scalar is the local fuel-equivalence ratio (the mass fraction of fuel elements irrespective of species divided by the stoichiometric mass fraction of fuel elements). Predictions based on the assumption of local thermodynamic equilibrium for an adiabatic flame, using the Gordon and McBride [19] algorithm, are also shown on the figure. Aside from temperature (where radiative heat losses and errors of uncorrected temperature measurements are a factor) the measured state relationships are seen to be in excellent agreement with equilibrium predictions. Thus, the tendency of reactive systems to approach equilibrium provides a physical justifica-

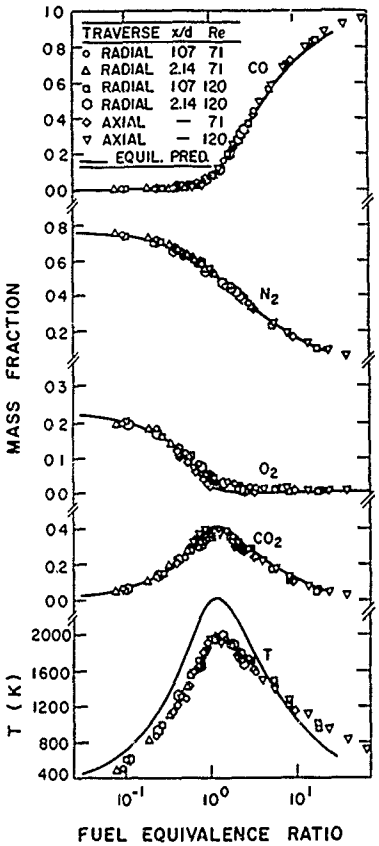


Fig 1 State relationships for carbon monoxide/air diffusion flames From Gore et al. [8]

tion for the laminar flamelet approximation in this instance.

State relationships for the concentrations of major gas species and temperature, adequate for estimates of structure and radiation properties, have been found from measurements in laminar flames for a variety of fuels burning in air: hydrogen [9,17], methane [18,20,21], propane [22], *n*-heptane [17,23], acetylene [11] and ethylene [10] Hydrocarbons exhibit significant departures from local thermodynamic equilibrium at fuel-rich conditions due to effects of finite-rate chemistry associated with soot processes; however, these departures are still relatively universal so that adequate state relationships are still found except near points of flame attachment. Finally, generalized state relationships have been found for hydrocarbon/air flames so that tedious measurements to find state relationships for specific fuels can be avoided [22].

Application of the conserved-scalar formalism and the laminar flamelet approximation to find the structure of turbulent flames has been reasonably successful for virtually all the materials for which state relationships are available [8–13,17,24,25] Recent studies also suggest that state relationships for soot volume fractions, an important property for estimates of continuum radiation from soot, exist in turbulent flames having sufficiently long residence times [26,27]. This implies that scalar properties needed to estimate radiation are strongly correlated through their state relationships and can be simulated by simulating a conserved-scalar alone

Mixture fraction statistics

Mixture fraction, f, defined as the fraction of elemental mass that originated from the fuel, is the conserved scalar most commonly used to find the scalar structure of turbulent nonpremixed flames. Turbulence models under the conserved-scalar formalism are designed to provide estimates of the mean value and variance of mixture fractions [17,28]. Methods used to estimate the other statistical properties needed to simulate mixture fraction distributions along radiation paths — probability density functions and correlations — will be considered in the following.

A fuel burning in air involves instantaneous properties at any point that can be pure air, pure fuel or some mixture of the two with scalar properties given by the state relationships. Several

probability density functions (PDFs) of mixture fraction, $P(f)$, have been proposed to accommodate these possibilities but the clipped-Gaussian PDF has received the most attention [28]. This involves a Gaussian function defined in range $0 < f < 1$ with the tails of the distribution replaced by Dirac delta functions at $f = 0$ and 1 that have weights equal to the probability of $f < 0$ and $f > 1$ for the original Gaussian distribution, respectively. Thus, the air intermittency of the flame at any point, defined as the fraction of time spent in ambient air, is given by the weighted Dirac delta function at $f = 0$.

Recent measurements in noncombusting and combusting turbulent flows suggest that the clipped-Gaussian PDF of mixture fraction is reasonable [29,30]. Some typical results are illustrated in Fig. 2 for turbulent carbon monoxide jet flames burning in still air. The measurements in the figure are fitted with clipped-Gaussian PDFs having the same mean values and variances Results are shown for various radial positions, r, before and after the flame tip ($x/d = 30$ and 50). The air intermittency spike is prominent for these conditions but the fuel intermittency spike can only be seen in the fitted PDFs near the axis at $x/d = 30$. The main deficiency of the clipped-Gaussian fits is that they fail to represent the broadened air intermittency spike caused by direct mixing between turbulent fluid and air near the edge of the flow (the air superlayer). A PDF having additional moments is needed to correct this problem; however, the complication of finding addition moments has not been pursued pending evaluation of the performance of the two-moment PLF. Notably, the functions used for mixture fraction PDFs normally do not have a strong effect upon predictions of scalar properties in turbulent flames [28].

Correlations of mixture fraction fluctuations, f', have been measured for turbulent jet-like flows for both noncombusting [31,32] and combusting [30] conditions. Some typical spatial correlations are illustrated in Fig. 3 for a carbon monoxide jet diffusion flame burning in still air. These results involve two-point spatial correlations of mixture fraction fluctuations for horizontal radial paths through the flame axis at positions before, near, and after the flame tip ($x/d = 30$, 40, and 50).

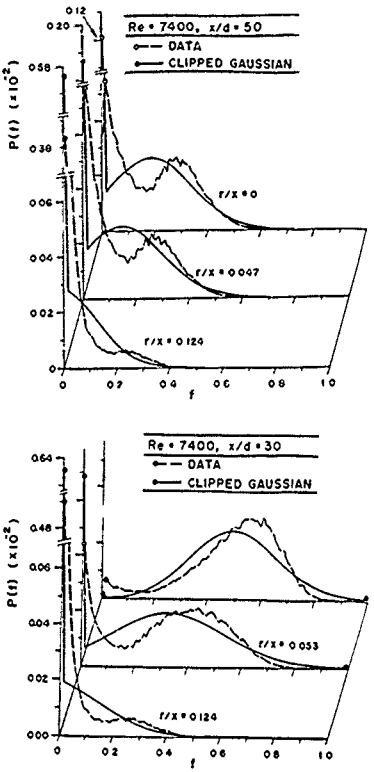


Fig 2 Typical probability density functions of mixture fraction for a turbulent carbon monoxide/air diffusion flame From Kounalakis and Faeth [30]

The correlations are plotted as a function of $\Delta r / \Gamma_r$, where Δr is the distance between the points and Γ_r is the spatial integral scale in the radial direction. The spatial correlations exhibit remarkably little variation with either radial or axial position when plotted in this manner. A simple exponential

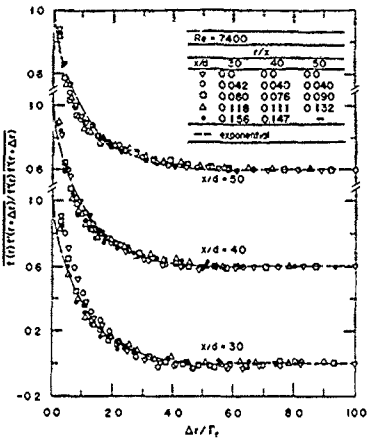


Fig 3 Spatial correlations of mixture fraction fluctuations for a turbulent carbon monoxide/air diffusion flame From Kounalakis and Faeth [30]

fit of the spatial correlation:

$$\overline{f'(r)f'(r+\Delta r)}\,/\left(\overline{f'^2(r)f'^2(r+\Delta r)}\right)^{1/2} = \exp(-\Delta r/\Gamma_f) \qquad (1)$$

is also shown in the figure. The exponential function is seen to provide a reasonably good fit of the measurements, as illustrated in Fig. 3. This is partly due to experimental limitations, since the spatial resolution was not sufficient to resolve the smallest scales of the flow which are expected to modify the correlation near $\Delta r = 0$ [30]. Nevertheless, the exponential expression provides a good representation of the larger scales that contain most of the signal energy and are expected to have the greatest influence on turbulence–radiation interactions. It should be noted, however, that these results differ from earlier findings in nearly constant density jets where radial correlations of mixture fraction fluctuations had the shape of a Frenkiel function [31,32] — these differences between combusting and noncombusting conditions must still be resolved.

Temporal correlations of mixture fraction fluctuations have been measured for the turbulent carbon monoxide jet diffusion flames as well [30] These results were also relatively independent of position and could be correlated by an exponential function analogous to eq. (1) with time differences, Δt, normalized by the integral time scale, τ_I (subject to the same limitations as eq. (1) near $\Delta t = 0$) The exponential form of the low-resolution temporal correlation measurements agrees with earlier findings for noncombusting flows [32]

With exponential functions established as reasonable approximations of spatial and temporal correlations of mixture fraction fluctuations, the next problem is specification of integral scales Measurements of these scales for turbulent carbon monoxide jet diffusion flames are illustrated in Fig 4. The scales are normalized as Γ_f/x and as $\tau_I u_m/(x - x_0)$, where u_m is the average velocity at the burner exit and x_0 is a virtual origin at $x_0/d = 13$ When correlated in this manner, the measurements tend to collapse to single curves for a range of flame positions The spatial integral scales are relatively independent of radial position and can be correlated as $\Gamma_f/x = 0.017$. In contrast, τ_I is smallest at the axis. This behavior can be explained through Taylor's hypothesis, e.g., $\tau_I \sim \Gamma_f/\bar{u}$, where $\bar{u}$ is the local time-averaged stream-

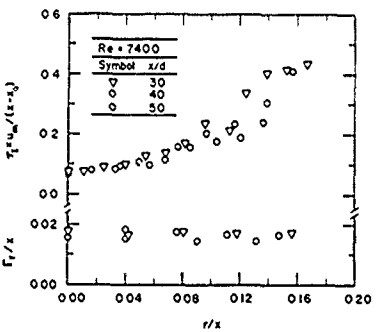


Fig. 4. Temporal and spatial integral scales in a turbulent carbon monoxide/air diffusion flame. From Kounalakis and Faeth [30].

wise velocity, while Γ_t is nearly independent of radial position and $\tilde{u}$ is a maximum at the axis.

Results concerning mixture fraction statistics in Figs. 2–4 were generally preserved as the Reynolds number of the carbon monoxide flames was increased [30]. Nevertheless, this only represents fragmentary findings for a single reactant combination, and generalization is needed to treat other flame systems. One proposal has been to assume that the radial spatial integral scale is proportional to the local dissipation length scale [12,13], as follows:

$$\Gamma_t = C_\epsilon C_\mu^{3/4} k^{3/2}/\epsilon \tag{2}$$

where C_ϵ is an empirical constant having a value in the range 5–7, C_μ is a turbulence modeling constant having a value of 0.09, and k and ϵ are mass-weighted (Favre) averaged turbulence kinetic energy and dissipation found from structure predictions using a turbulence model. The temporal integral scale was then estimated using Taylor's hypothesis while assuming that streamwise and radial scales were the same, as follows:

$$\tau_I = \Gamma_t/\tilde{u} \tag{3}$$

where $\tilde{u}$ is the mass-weighted (Favre) averaged mean velocity in the streamwise direction. Eqs. (2) and (3) are consistent with the results illustrated in Fig. 4 but additional study of the approximations is certainly needed. For lack of an alternative, eqs (2) and (3) will be used to find integral scales in the following.

STOCHASTIC SIMULATION

Formulation

The stochastic simulation provides realizations of mixture fraction distributions along radiation paths through the flow. Given the mixture fractions, the state relationships provide all other scalar properties so that spectral radiation intensities can be calculated from a narrow-band radiation model for each distribution. The resulting ensemble, or time series, of spectral radiation intensities is then used to compute moments, PDFs, correlations and power spectra of spectral radiation intensities in the usual manner. The formulation of the simulation and the narrow-band radiation model will be discussed in the following.

It will be assumed that the statistical properties of mixture fractions are known along the radiation path. This includes $P(f)$ (taken to be a clipped-Gaussian function), spatial correlations, and temporal correlations if temporal properties are needed (both taken to be exponential functions). Aside from isolated cases where measurements are available [30], these properties must be estimated from a model of the turbulent combustion process. For flows having relatively high Reynolds numbers, this is generally done using a turbulence model. Fortunately, for relatively simple flame geometries, like buoyant jet flames, turbulence models provide reasonably good estimates of scalar properties, including mean and fluctuating mixture fractions [8–13,24,25]. The necessary statistical properties of mixture fractions are then found as described earlier.

Due to the exponential form of the mixture fraction correlations, it is most convenient to carry out the simulation as an autoregressive process [15]. This involves finding the mixture fraction fluctuation at any point as a weighted sum of fluctuations at other points and a random shock. A procedure of this type encounters difficulties with any finite range PDF, since the fluctuation algorithm can easily generate a value of the variable which is beyond the range of the PDF. This is handled by transforming the simulation from f, which has a clipped-Gaussian PDF, to a corresponding Gaussian random variable z, with appropriate moments to match $P(f)$, so that

$$f = z,\ 0 \le z \le 1; \quad f = 0,\ z < 0; \quad f = 1,\ z > 1 \tag{4}$$

Since the PDFs of f and z are not the same, correlations of f and z differ as well. Methods to find the appropriate correlations for z will be taken up later.

Values of z are simulated at a number of points along the radiation path. Following Box and Jenkins [15], the value of the fluctuation of z at point i, z_i', is found as a weighted sum of fluctua-

tions found earlier, z'_j, where $j = i-1, \ldots, p$, and a random shock, a_i, as follows:

$$z'_i = \sum_{j=p}^{i-1} \phi_{i,j} z'_j + a'_i; \qquad 1 \le p \le i-1 \tag{5}$$

The index p is selected to eliminate points having small correlation coefficients with respect to point i. The $\phi_{i,j}$ are weighting factors so that the simulation satisfies correlations between fluctuations at various points appearing in eq (5). The parameter a_i is an uncorrelated Gaussian random variable having a mean value of zero and a variance selected so that the simulation satisfies $P(z_i)$

Box and Jenkins [15] derive expressions for the $\phi_{i,j}$ and the variance of a_i, $\bar{a}_i^2$, as follows:

$$\overline{z'_i z'_k} = \sum_{j=p}^{i-1} \phi_{i,j} \overline{z'_j z'_k}; \qquad k = p, \ldots, i-1 \tag{6}$$

$$\bar{a}_i^2 = \bar{z}_i'^2 - \sum_{j=p}^{i-1} \phi_{i,j} \overline{z'_j z'_j} \tag{7}$$

With the correlations between the various points known, eqs (6) provide i-p linear equations, called the Yule–Walker equations, needed to find the $\phi_{i,j}$. This system of equations has a symmetric positive definite matrix and can be solved readily using Cholesky factorization. Given the $\phi_{i,j}$, $\bar{a}_i^2$ can be found since all quantities on the right-hand side of eq (7) are known.

A time-independent simulation is initiated by making a random selection for point 1, noting that $z'_1 = a_1$ from eq. (5) and $\bar{a}_1^2 = \bar{z}_1'^2$ from eq. (7). The regression relationships are then successively applied to find the remaining z'_i along the radiation path. Finally, the f_i are found from eq (4), noting that $z_i = \bar{z}_i + z'_i$, followed by computation of spectral radiation intensities for this realization, as described earlier. This process is repeated a sufficient number of times to obtain statistically significant radiation properties.

The previously computed points in the regression process of eq. (5) only enter the calculations through their correlations, therefore, time-dependent simulations are essentially the same as time-independent simulations after appropriately numbering points to keep track of them in space and time. This involves realizations of f along the radiation path at times Δt apart. The simulation is initiated by finding a realization using the time-independent solution. Realizations are then found at subsequent times considering correlations with all previous realizations, until temporal correlations are properly represented. Subsequently, the points at the earliest time are dropped when calculations for the next time are begun, for computational efficiency.

The main new difficulty with the time-dependent simulation is that two-point–two-time correlations are needed. Information of this type is not available; therefore, the following ad hoc approximation has been adopted for lack of an alternative [13]

$$\overline{z'_i(t) z'_j(t - k\,\Delta t)} = R_i(k\,\Delta t) \overline{z'_i z'_j} \tag{8}$$

where $R_i(k\,\Delta t)$ is the temporal correlation coefficient of z_i fluctuations at a time delay of $k\,\Delta t$. Naturally, it would be just as plausible to use $R_j(k\,\Delta t)\overline{z'_i z'_j}$ on the right-hand side of eq. (8) for a stationary turbulent flow. The differences between these possibilities provides a measure of potential errors resulting from the use of eq. (8). Since Γ_t is nearly constant over a cross-section of the flow, eq. (3) indicates that errors are greatest in regions where $\tilde{u}$ varies rapidly. Fortunately, spatial correlations become small for separation distances of Γ_t and $\tilde{u}$ does not vary significantly over such distances, providing some justification for the approximation.

When temporal correlations are exponential, use of eq. (8) for two-point–two-time correlations leads to substantial simplification of time-dependent simulations. Carrying out a derivation similar to that of Box and Jenkins [15] for a pure time series with stationary statistics and an exponential temporal correlation yields similar results for the combined spatial/temporal simulation with temporal correlations varying according to eq. (8), namely the $\phi_{i,j} = 0$ for all points at times less than $t - \Delta t$. Thus, only the realization at $t - \Delta t$ must be retained while developing the realization at t, vastly reducing the storage and computational requirements of the simulation.

Another useful simplification is that radiation predictions are relatively insensitive to the func-

tional form of the spatial correlation, since they are found by integrating properties along a radiation path [13,33] Thus, temporal simulations using statistically independent points spaced a distance Γ_t apart along the radiation path yielded results that were essentially the same as simulations that satisfied twenty-point fits of spatial correlations along the radiation paths [13]. This simplification reduces the simulation to a first-order (Markov) process in time at each point, for an exponential temporal correlation, yielding [15]:

$$z_t'(t) = R_t(\Delta t) z_t'(t - \Delta t) + a_t \tag{9}$$

where

$$\overline{a_t^2} = \left(1 - R_t(\Delta t)^2\right)\overline{z_t'^2} \tag{10}$$

Correlation corrections

Initial time-series simulations of mixture fraction distributions involved the approximation that correlations of f and z were the same [12,13]. This was adequate in most regions of the flames but discrepancies between actual and simulated correlations of mixture fraction fluctuations were significant in regions where either air or fuel intermittencies were high [13]. The cause of the difficulty is the transformation from f to z, since z has an infinite range while $0 \leq f \leq 1$. This implies that the correlations of the fluctuations of z must be corrected in order to properly simulate the correlations of the fluctuations of f.

A generalized correction of the z correlations has been developed for any two points, $i > j$, having identical mean and fluctuating mixture fractions, $\tilde{f}_i = \tilde{f}_j = \tilde{f}$ and $\overline{f_i'^2} = \overline{f_j'^2} = \overline{f'^2}$, i.e., for temporal correlations at stationary conditions. The simulation is carried out with the z variable where $\tilde{z}_i = \tilde{z}_j = \tilde{z}$ and $\overline{z_i'^2} = \overline{z_j'^2} = \overline{z'^2}$ can be found from the transformation of eq. (4). In order for the simulation to yield the correct correlation, $\overline{f_i' f_j'}$, the value of $\overline{z_i' z_j'}$ must be corrected so that the following equation is satisfied

$$\overline{f_i' f_j'} + \overline{f'^2} = \int_{-\infty}^{\infty} f(z_i) P(z_i) \int_{-\infty}^{\infty} f(z_j) P(z_j : z_i)\, \mathrm{d}z_j\, \mathrm{d}z_i \tag{11}$$

where $f(z_i)$ and $f(z_j)$ are obtained from eq. (4) and $P(z_j, z_i)$ is the probability density function of z_j given z_i. Now, the correct correlation for the z variables can be found by considering an autoregressive process between the two points under the present approximations, as follows

$$z_j' = z_i'\left(\overline{z_i' z_j'}/\overline{z'^2}\right) + a_j \tag{12}$$

where a_j has a Gaussian PDF with

$$\overline{a_j^2} = \overline{z'^2} - \left(\overline{z_i' z_j'}\right)^2 / \overline{z'^2} \tag{13}$$

Then, for any realization of z_i', $P(z_j : z_i)$ is a Gaussian distribution having a mean value of $\tilde{z} + z_i'(\overline{z_i' z_j'}/\overline{z'^2})$ and a variance of $\overline{a_j^2}$, while $P(z_i)$ is a Gaussian distribution having a mean value of $\tilde{z}$ and a variance of $\overline{z'^2}$. Substituting these expressions, along with $f(z_i)$ and $f(z_j)$ from eq. (4) into eq. (11) yields an expression relating $\overline{f_i' f_j'}$ and $\overline{z_i' z_j'}$. This expression must be evaluated numerically for a clipped-Gaussian $P(f)$ The procedure was to select values of $\tilde{z}$, $\overline{z'^2}$ and $\overline{z_i' z_j'}$ and then find the corresponding values of $\tilde{f}$, $\overline{f'^2}$, and $\overline{f_i' f_j'}$. Present results were found by integrating over the region within 5 standard deviations from the mean of the PDFs.

Since the temporal correlations of f are exponential, it was convenient to fit the correlations of z in the same manner and to express the corrections of the correlations as ratios between the temporal integral scales of f and z, τ_f/τ_z. This ratio is plotted in Fig. 5 as a function of $(\overline{f'^2})^{1/2}$ with $\tilde{f}$ as a parameter. The results are symmetric with respect to $\tilde{f} = 0.5$. The plots of τ_f/τ_z at a particular value of $\tilde{f}$ are terminated at the maximum possible value of $(\overline{f'^2})^{1/2}$, i.e., where $P(f)$ degenerates to Dirac delta functions at $f = 0$ and 1. The ratio of τ_f/τ_z decreases from unity as $(\overline{f'^2})^{1/2}$ increases and $\tilde{f}$ approaches either 0 or 1. Thus, there is no correction when z remains in the range 0–1 where $z = f$. Whenever $z < 0$ or $z > 1$, however, $z^2 > f^2$ and the correlation for f generally is less than the correlation for z so that τ_f/τ_z is less than unity.

Simulations using corrected correlations for z were evaluated for $\overline{z_i' z_j'}/\overline{z'^2} > 0.1$. Using 10^4 reali-

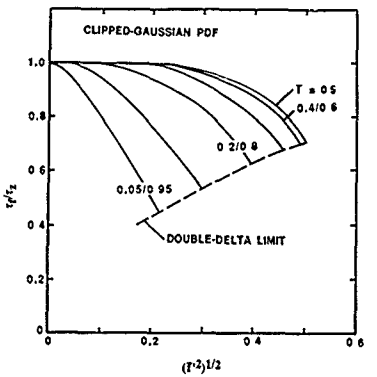


Fig. 5 Ratio of simulated and original integral scales for exponential correlations of functions having clipped-Gaussian probability density functions.

zations, values of $\tilde{f}$ and $\overline{f''^2}$ were satisfied within 1% while values of $f_1'f_2'/f''^2$ were satisfied within 3%. Analogous calculations to find the corresponding corrections of the correlations when $\tilde{f}$ and $\overline{f''^2}$ are not the same at the two points are straightforward on a case-by-case basis.

Narrow-band radiation model

Given the distribution of scalar properties along a radiation path, through the stochastic simulation of mixture fractions and the state relationships, spectral radiation intensities are found by solving the equation of radiative transfer along the path. Present results involved using a narrow-band model, ignoring scattering, due to Ludwig et al. [34]. The procedure uses the Goody statistical narrow-band model, with the Curtiss–Godson approximation to account for absorption along inhomogeneous gas paths. This model accounts for the infrared gas bands of water vapor, carbon dioxide, carbon monoxide, and methane, as well as continuum radiation from soot. Radiation contributions of other species in hydrogen, carbon monoxide, and hydrocarbon flames burning in air are generally negligible since these species have small concentrations in regions where temperatures are high.

RESULTS AND DISCUSSION

Some comparisons between simulated and measured radiation properties will be considered in order to illustrate the nature and effectiveness of the simulations. The discussion will be limited to results reported by Kounalakis et al. [12,13] for vertical turbulent hydrogen and carbon monoxide jet flames burning in still air. Spectral radiation intensities, i_λ were measured for horizontal radiation paths through the axis of the flames. Predictions were based on the present formulation of the stochastic simulation of mixture fraction distributions. As noted earlier, twenty-point fits of spatial correlations in the simulation yielded essentially the same results as the simplified formulations of eqs. (9) and (10); therefore, the following results are based on the simplified formulation. Mixture fraction statistics were estimated based on structure predictions using a turbulence model. This introduces uncertainties although the turbulence model yielded reasonably good predictions of scalar structure for the same flames during earlier studies [8,9].

Predicted and measured probability density functions of i_λ are illustrated in Fig. 6 for positions before, near, and after the tip of a hydrogen jet flame ($x/d = 50$, 90, and 130). These results are for a wavelength $\lambda = 2520$ nm which is within a prominent infrared gas radiation band for water vapor. Near the burner, the PDFs are relatively symmetric but they become increasingly skewed as distance from the burner exit increases. This is an effect of air intermittency as mean radiation levels become small, since the spectral intensity can never be negative while the mean value is generated by occasional periods of high radiation levels. The stochastic predictions represent the measurements reasonably well, particularly for the small path diameter which more closely approximates the negligible path diameter of the simulation.

Predicted and measured temporal power spectra of spectral radiation intensities, $E_\lambda(n)$, are illustrated in Fig. 7 for positions before, near, and

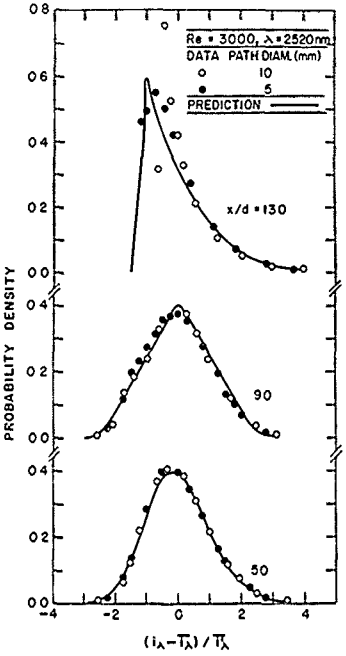


Fig 6 Measured and predicted probability density functions of spectral radiation intensities for a turbulent hydrogen/air diffusion flame From Kounalakis et al [12]

after the tip of a carbon monoxide jet flame ($x/d = 35$, 50, and 65). The power spectra are plotted as a function of frequency, n, both normalized by the characteristic frequency, $\bar{u}_c/x$, where $\bar{u}_c$ is the mean velocity at the flame axis. The spectra exhibit a break frequency with an energy-containing region having a nearly constant $E_\lambda(n)$ at low frequencies, followed by decay of $E_\lambda(n)$ with increasing frequency beyond the break frequency. Normalized break frequencies increase somewhat with increasing distance from the burner. This follows since the high temperature region that contributes most to radiant emission is located off axis near the burner and moves toward the axis with increasing distance above the burner. Since temporal integral scales are smallest near the axis (see Fig. 4), this implies a corresponding increase in the break frequency when normalized by properties at the axis.

The predictions provide reasonable estimates of break frequencies and signal properties in the energy-containing region for the results illustrated in Fig. 7. The main deficiency of the predictions is that they underestimate the rate of decay of $E_\lambda(n)$ at high frequencies. Two main reasons can be advanced for this behavior. First of all, spectral intensities were measured for a finite diameter radiation path. This tends to average out high-frequency effects over the cross-section of the radiation path in comparison to predictions which represent an infinitely thin path. An indication of this effect can be seen by comparing measurements for 5- and 10-mm-diameter paths appearing in Fig. 7, which show that the spectra decay more rapidly for the larger-diameter path. Secondly, the

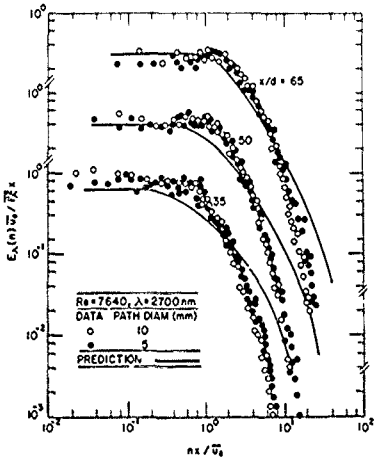


Fig. 7. Measured and predicted temporal power spectral densities of spectral radiation intensities for a turbulent carbon monoxide/air diffusion flame From Kounalakis et al. [13]

exponential correlation function used in the stochastic simulation does not properly truncate high-frequency fluctuations as turbulent microscales are approached, as noted earlier. This causes the predictions to overestimate high frequency signal levels Resolving these problems would require extension of the stochastic simulation, to allow simulation of groups of parallel radiation paths so that they can be summed over a finite-diameter path and to accommodate high-frequency cut-offs associated with turbulence microscales when simulating correlations However, since spectral intensity signal energies are relatively small when the discrepancy becomes significant, such extensions are not needed for most applications.

CONCLUSIONS

The use of statistical time-series techniques to treat nonlinear interactions during turbulent combustion processes was described. Turbulence–radiation interactions were used to illustrate the method, however, other turbulence interaction problems for combusting flows require a similar treatment of scalar properties. Existing evidence suggests that scalar properties are strongly correlated through state relationships in turbulent diffusion flames and can be simulated by only simulating a conserved-scalar like mixture fraction. The statistics of mixture fractions in turbulent diffusion flames can be approximated by a clipped-Gaussian PDF and exponential spatial and temporal correlations, at least for the large-scale features that dominate radiation properties. Stochastic simulations using statistical time-series techniques must be modified to account for the finite-range PDF of mixture fraction. This involved transformation to a new variable having a Gaussian PDF and finding appropriate corrections for the correlations in terms of the new variable. An autoregressive process that reproduced the PDFs and spatial and temporal correlations of mixture fractions yielded an effective simulation to find the statistical properties of spectral radiation intensities from turbulent jet diffusion flames. Thus, additional development and application of the method appears to be warranted.

ACKNOWLEDGEMENT

This research was supported by the Center for Fire Research of the National Institute of Standards and Technology (formerly the National Bureau of Standards), Grant No. 60NANB8D0833 with H.R. Baum serving as Scientific Officer.

REFERENCES

1 J.-S Shuen, A S P. Solomon, Q.-F Zhang and G M. Faeth, Structure of particle-laden jets. measurements and predictions, *American Institute of Aeronautics and Astronautics Journal*, 23 (1985) 396–404
2 T.-Y. Sun and G M Faeth, Structure of turbulent bubbly jets, *International Journal of Multiphase Flow*, 12 (1986) 99–126
3 A. Picart, A. Berlemont and G Gouesbet, Modeling and predicting turbulence fields and dispersion of discrete particles transported by turbulent flows, *International Journal of Multiphase Flow*, 12 (1986) 237–261
4 M R Maxey, The gravitational settling of aerosol particles in homogeneous turbulence and random flow fields, *Journal of Fluid Mechanics*, 174 (1987) 441–465
5 R N. Parthasarathy and G M Faeth, Turbulent dispersion of particles in self-generated homogeneous turbulence, *Journal of Fluid Mechanics*, in press
6 A S P Solomon, J.-S Shuen, Q.-F Zhang and G M Faeth, Measurements and predictions of the structure of evaporating sprays, *Journal of Heat Transfer*, 107 (1985) 679–686
7 J S Shuen, A S P Solomon and G M Faeth, Drop–turbulence interactions in a diffusion flame, *American Institute of Aeronautics and Astronautics Journal*, 24 (1986) 101–108
8 J P Gore, S.-M Jeng and G M Faeth, Spectral and total radiation properties of turbulent carbon monoxide/air diffusion flames, *American Institute of Aeronautics and Astronautics Journal*, 25 (1987) 339–345
9 J P Gore, S.-M Jeng and G M Faeth, Spectral and total radiation properties of turbulent hydrogen/air diffusion flames, *Journal of Heat Transfer*, 109 (1987) 165–171
10 J P Gore and G M Faeth, Structure and spectral radiation properties of turbulent ethylene/air diffusion flames, *Twenty-First Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, PA, 1986, pp 1521–1531
11 J P Gore and G M Faeth, Structure and radiation properties of luminous turbulent acetylene/air diffusion flames, *Journal of Heat Transfer*, 110 (1988) 173–181
12 M E. Kounalakis, J P Gore and G M Faeth, Turbulence/radiation interactions in nonpremixed hydrogen/air flames, *Twenty-Second Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, PA, 1988, pp 1281–1290.
13 M E. Kounalakis, J P Gore and G M Faeth, Mean and fluctuating radiation properties of turbulent nonpremixed

carbon monoxide/air flames, *Journal of Heat Transfer*, 111 (1989) 1021–1030
14 R H Kraichnan, Diffusion by a random velocity field, *Physics of Fluids*, 13 (1970) 22–31.
15 G E P. Box and G M. Jenkins, *Time Series Analysis*, Holden Day, San Francisco, CA, revised edition, 1976, pp. 47–84
16 S P Burke and T.E W. Schumann, Diffusion flames, *Industrial and Engineering Chemistry*, 20 (1928) 998–1004.
17 R W Bilger, Turbulent jet diffusion flames, *Progress in Energy and Combustion Science*, 1 (1976) 87–109.
18 R W. Bilger, Reaction rates in diffusion flames, *Combustion and Flame*, 30 (1977) 277–284
19 S Gordon and B.J. McBride, *Computer Program for Calculation of Complex Chemical Equilibrium Compositions, Rocket Performance, Incident and Reflected Shocks, and Chapman–Jouguet Detonations*, NASA SP-273, Washington, DC, 1971
20 K C Smyth, J H Miller, R C Dorfman, W G Mallard and R.J Santoro, Soot inception in a methane/air diffusion flame as characterized by detailed species profiles, *Combustion and Flame*, 62 (1985) 157–181.
21 K. Saito, F A Williams and A S Gordon, Structure of laminar coflow methane-air diffusion flames *Journal of Heat Transfer*, 108 (1986) 640–648
22 Y R Sivathanu and G M Faeth, Generalized state relationships for scalar properties in nonpremixed hydrocarbon/air flames, *Combustion and Flame*, in press.
23 J H Kent and F A Williams, Extinction of laminar diffusion flames for liquid fuels, *Fifteenth Symposium (International) on Combustion*, The Combustion Institute, Pittsburgh, PA, 1974, pp 315–325
24 S-M Jeng and G M Faeth, Species concentrations and turbulence properties in buoyant methane diffusion flames, *Journal of Heat Transfer*, 106 (1984) 721–727
25 Y.R. Sivathanu, J P. Gore and G M. Faeth, Scalar properties in the overfire region of sooting turbulent diffusion flames, *Combustion and Flame*, 73 (1988) 315–329.
26 Y.R. Sivathanu and G M. Faeth, Soot volume fractions in the overfire region of turbulent diffusion flames, *Combustion and Flame*, 81 (1990) 133–149.
27 Y R. Sivathanu and G.M. Faeth, Temperature/soot volume fraction correlations in the fuel rich region of buoyant turbulent diffusion flames, *Combustion and Flame*, 81 (1990) 150–165.
28 F.C. Lockwood and A S Naguib, The prediction of the fluctuations in the properties of free, round-jet, turbulent diffusion flames, *Combustion and Flame*, 24 (1975) 109–124.
29 M-C. Lai and G M Faeth, Turbulence structure of vertical adiabatic wall plumes, *Journal of Heat Transfer*, 109 (1987) 663–670.
30 M E. Kounalakis and G.M Faeth, Measurement of mixture-fraction correlations in turbulent jet diffusion flames, *Proceedings of Fall Technical Meeting*, Eastern Section of the Combustion Institute, Pittsburgh, PA, 1989
31 S Corrsin and M S. Uberoi, *Spectra and Diffusion in a Round Turbulent Jet*, NACA Report No 1040, Washington, DC, 1951
32 H A. Becker, H C. Hottel and G C Williams, The nozzle-fluid concentration field of the round, turbulent, free jet, *Journal of Fluid Mechanics*, 30 (1967) 285–303
33 W.L. Grosshandler and P. Joulain, The effect of large-scale fluctuations on flame radiation, *Progress in Astronautics and Aeronautics*, 105(II) (1986) 123–152
34 C B Ludwig, W Malkmus, J.E. Reardon and J A Thomson, *Handbook of Infrared Radiation from Combustion Gases*, NASA SP-3080, Washington, DC, 1973

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 211–212
Elsevier Science Publishers B V., Amsterdam

Nonequilibrium chemistry and flamelet modeling of nonpremixed turbulent reacting flows

Mitchell D. Smooke

Department of Mechanical Engineering, Yale University, New Haven, CT 06520 (U S A)

Practical combustion systems often involve the burning of nonpremixed fuel/air systems in a turbulent flow environment. While the ultimate modeling of such nonpremixed systems will inevitably involve the direct solution of the three-dimensional time-dependent conservation equations of mass, momentum, species balance and energy, such a task is computationally infeasible on even the largest supercomputers at the current time. The primary difficulty with such an approach is that there are large variations (orders of magnitude) in the length scales present in the reacting flow. The ability to resolve the relevant solution structure requires computational resources that currently do not exist. As a result, the modeling of nonpremixed turbulent reacting flows requires the introduction of a number of simplifying assumptions to make the problem more tractable. One of these methods, the laminar flamelet model, considers a turbulent flame to be composed of an ensemble of thin laminar diffusion flames. It can be shown that these flamelets have a one-dimensional structure normal to the surface of the stoichiometric mixture [1]. The model is applicable if the length scales of the turbulent eddies are much larger than the reaction zone thickness of the flamelets. The structure of these flames are often described in terms of a conserved scalar Z called the mixture fraction. The mixture fraction can be considered to be the fuel element mass fraction in the system. Variations of the laminar flamelet approach center primarily in terms of the chemical approximations used in describing the flamelets. In some situations local thermodynamic equilibrium chemistry is appropriate. In other cases finite rate chemistry is needed. In the discussion that follows we consider the incorporation of finite rate chemistry into flamelet models of nonpremixed turbulent combustion.

Due to the spatial variation in the stretching of the turbulent flame, the flamelets are subjected instantaneously to a certain rate of strain. This can be represented in terms of the scalar dissipation χ_{st} at the point of stoichiometry [2]

$$\chi_{st} = 2a\left(\frac{C}{\mathrm{Pr}}\right)_{st}\left(\frac{\mathrm{d}Z}{\mathrm{d}\eta}\right)^2_{st} \tag{1}$$

where a is the strain rate, C is the Chapman–Rubesin parameter, Pr is the Prandtl number and η is a density weighted coordinate. The implications of this model are that at any point of space the instantaneous local composition of the turbulent flame is that of the diffusion flamelet. Local conditions may be viewed as corresponding to a flamelet in a flamelet family that is parameterized by the degree of stretching χ_{st}. The structure of the flamelet provides a unique relationship between any scalar S and Z. We write this in the form

$$S = S(Z, \chi_{st}) \tag{2}$$

We treat Z and χ_{st} as random variables whose joint probability density function (PDF) $\tilde{P}(Z, \chi_{st})$

0169-7439/91/$03 50

must be determined. In practice the PDF is factored such that

$$\tilde{P}(Z,\chi_{st}) = \tilde{P}(Z)\tilde{P}(\chi_{st}) \quad (3)$$

Ordinarily, $\tilde{P}(Z)$ is taken to be the beta function and $\tilde{P}(\chi_{st})$ is taken to be the log normal distribution [3]. The mean and variance of the log normal distribution are computed from the first moment of χ_{st} and the Favre averaged turbulent dissipation and turbulent kinetic energy, respectively. With this formalism established, scalar properties are determined by postulating a set of burned and unburned states [2]. In particular, we can write the Favre averaged value of the burned contribution to the scalar S as

$$\tilde{S} = \int_0^{\chi_{ext}} \int_0^1 S(Z,\chi_{st})\tilde{P}(Z)\tilde{P}(\chi_{st})\,dZ\,d\chi_{st} \quad (4)$$

where χ_{ext} represents the maximum value of the scalar dissipation at which a flame exists A similar integral can be written for the unburnt states The joint dependence of S on Z and χ_{st} is parametric in χ_{st} and is characterized by a limited number of data files that constitute a flame library [4] Evaluation of the properties in (4) are carried out by replacing the integrals by numerical quadratures. The Favre averaged properties are then utilized in a boundary layer k–ϵ turbulent flow model.

The individual laminar diffusion flamelets in the flamelet library are modeled by considering counterflowing streams of fuel and oxidizer in either a Tsuji or a Seshadri-type burner [5,6]. A similarity solution is sought for the two-dimensional governing conservation equations The flamelet problem is then reduced to solving a nonlinear two-point boundary value problem along the stagnation point streamline. Individual flamelet calculations can be made for a given chemical mechanism, transport approximation and jet velocities. Once the computation is completed, the solution can be stored as a function of the mixture fraction with each flamelet characterized by the scalar dissipation at the point of stoichiometric mixture.

REFERENCES

1 N Peters, Laminar diffusion flamelet models in nonpremixed turbulent combustion, *Progress in Energy and Combustion Science*, 10 (1984) 319–339.
2 S K. Liew, K.N C. Bray and J B Moss, A stretched laminar flamelet model of turbulent nonpremixed combustion, *Combustion and Flame*, 56 (1984) 199–213
3 N. Peters, Local quenching due to flame stretch and nonpremixed turbulent combustion, *Combustion Science and Technology*, 30 (1983) 1–17
4 B. Rogg, F. Behrendt and J. Warnatz, *Twenty-First Symposium (International) on Combustion*, Reinhold, New York, 1986, p 1533.
5 G Dixon-Lewis, T David, P.H. Haskell, S Fukutani, H Jinno, J A Miller, R.J Kee, M D. Smooke, N Peters, E. Effelsberg, J Warnatz and F Behrendt, *Twentieth Symposium (International) on Combustion*, Reinhold, New York, 1984, p 1893
6 M D. Smooke, I K Puri and K Seshadri, *Twenty-First Symposium (International) on Combustion*, Reinhold, New York, 1986, p. 1783

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 213–227
Elsevier Science Publishers B.V., Amsterdam

Novel graph theoretical approach to heteroatoms in quantitative structure–activity relationships *

Milan Randić

Department of Mathematics and Computer Science, Drake University, Des Moines, IA 50311 (U.S.A.)

(Received 8 November 1989; accepted 3 July 1990)

Abstract

Randić, M., 1991. Novel graph theoretical approach to heteroatoms in quantitative structure–activity relationships. *Chemometrics and Intelligent Laboratory Systems*, 10: 213–227.

A novel approach to characterization of heteroatoms in graph theoretical approaches to quantitative structure–activity relationships (QSAR) is outlined. The basis of the approach is the use of diagonal entries of the adjacency matrix as variable parameter, in full analogy to the well known generalization of the Hückel Molecular Orbitals (HMO) method when extended to heteroconjugated systems. The approach is illustrated on clonidine-like compounds where carbon and chlorine atoms are discriminated by using $\alpha = -0.20$ as the diagonal entry for chlorine atoms. Derived weighted path numbers are used as descriptors and a multiple regression based on three descriptors resulted in the correlation coefficient $R = 0.977$ and the standard error $S = 0.233$. This represents a substantial improvement over the best traditional QSAR analysis which involves five descriptors (in a nonlinear correlation equation with $R = 0.964$ and $S = 0.301$). A detailed comparison is made with available QSAR results, and the advantages (as well as limitations) of graph theoretical descriptors are discussed.

INTRODUCTION

In contrasting graph theoretical schemes [1] to traditional quantitative structure–activity relationship (QSAR) methods [2] one cannot fail to observe the complementarity of the two approaches. Traditional QSAR is mostly based on a large number of empirical parameters. The graph theoretical approaches use a rather small set of structural invariants, graph invariants in particular. In traditional QSAR one uses statistical methods in order to select critical descriptors and derive a structure–activity correlation. In graph theory one manipulates structures algebraically, using partial order and ranking based on selected standards. Of course, graph theoretical descriptors also lead to structure–property and structure–activity correlations based on statistical analysis [3–6]. The applications of graph theory [7] to QSAR cover a variety of topics, from the study of various physicochemical data to biological activities and toxicities (refs. 1, 2 and 5–7, and references cited therein, and refs. 8–24), including even the use of graph theoretical descriptors in pattern recognition [25]. But the prime distinction between graph theoretical schemes and traditional QSAR is that the former is 'structure-explicit' while the latter is 'structure-cryptic' [1]. The former uses well defined mathematical invariants which have a direct structural interpretation, while the latter are mostly

* This paper is dedicated to Professor Dušan Hadži from Boris Kidrič Institute in Ljubljana, Slovenia, Yugoslavia

0169-7439/91/$03.50

expressed as physicochemical properties that remain to be interpreted structurally. For example, the molar refraction (MR) has frequently been used as a descriptor in traditional QSAR, but how MR depends on molecular structure, so that it can be predicted once the chemical structure is known, still remains to be understood. The distinction can be illustrated by reference to a particular study of selected physicochemical properties of over a hundred compounds by Cramer [26]. Using principal component analysis Cramer has shown that aqueous solvation or the activity coefficient in water, partition coefficient (octanol/water), boiling points, molar refraction, liquid state molar volumes and heats of vaporization, which mutually show variable pairwise correlations, from nonexistent to very high correlations, can all be well explained (at 95% variance) by two variables This illustrates well the presence of structural factor, as yet to be identified, on which all the studied properties critically depend. According to Cramer [22] ". it seems possible that molecular connectivity indices may represent alternative axes for compound subsets within 'BC(DEF) space'." Uncertainty here reflects upon the intrinsic difficulty associated with attempts to express mathematical properties (graph invariants) as a combination of physicochemical variables, instead of the other way round. If Cramer is correct in identifying the connectivity indices as alternative axes of physicochemical space, the two major variables being identified as 'bulkiness' and cohesiveness', that would only idicate that 'bulkiness' and 'cohesiveness' as molecular properties, will corelate with the connectivity indices.

LIMITATIONS OF GRAPH THEORETICAL APPROACHES

Graphs depict molecular connectivity and as such are devoid of information on heteroatoms and the spatial arrangements of atoms. It is not then surprising that to uninitiated people graph theoretical schemes appear at best unpromising, if not doomed to failure. Equally, graph theory does not produce numerical data, analogous, say, to quantum mechanical computations. It can nevertheless lead to quantitative results when information on selected standards is available. As long as the molecules considered are structurally closely related (e.g. they have the same heteroatoms in similar locations and have the same stereochemistry) graphs can be employed and useful correlations derived [28–36]. A neglect of heteroatoms and spatial molecular architecture may appear to be severe limitations of graph theoretical models. However, for QSAR studies concerned with a search for optimal compounds, once lead compounds are known, graph theoretical schemes were found to be quite successful, not only in suggesting a more potent compound, but in providing assurance that the compound thus found is the best possible one within the given family [37].

An extension of molecular graphs to molecular structures by embedding graphs on a regular three-dimensional grid has only recently been considered [38–40] By using topographic (geometrical) matrices, rather than topological (graph theoretical) adjacency matrices, one can differentiate between different conformers, such as *cis* and *trans, boat* and *chair*, between individual rotational isomers, etc. Importantly, the derived molecular descriptors are quite analogous to molecular connectivity indices, weighted path numbers and other graph-related invariants, except that now they are sensitive to precise molecular geometry. Moreover, the indicated generalization from adjacency (connectivity) to topography (geometry) suggests further generalizations of graphs in which structural invariants are derived from other matrices associated with molecules, such as the bond order matrix, the bond polarizability matrix and even the Hamiltonian matrix [41]. It appears that we are only at the beginning of new directions in our search for useful molecular descriptors. However, here we will restrict our attention to another generalization of graphs: to the problem of treating heteroatoms.

Applications of graph theoretical methods in QSAR to molecules with heteroatoms in more general positions lead to a number of generalization of simple graphs. Kier and Hall [3] introduced the concept of valence connectivities in which they associated different 'correction' factors with different heteroatoms. Kupchik [42] consid-

ered the use of Van der Waals radii of heteroatoms as a source of their discrimination by suitably modifying the connectivity indices. Hansen [43] similarly considered purely empirical correction factors for heteroatoms. In this paper we will outline yet another alternative approach to heteroatom which has some analogy with generalizations of Hückel Molecular Orbital (HMO) methods from hydrocarbons to heteroconjugated compounds [44] and represents an extension of an earlier work on sensitivity of path numbers to variation in bonds involving oxygen and nitrogen [45].

CLONIDINE-LIKE IMIDAZOLIDINES — AN ILLUSTRATION OF A GRAPH THEORETICAL APPROACH TO HETEROATOMS

We have selected clonidine and clonidine-like imidazolidines — compounds having a hypotensive action — because in these molecules chlorine (as heteroatom) appears in different locations and therefore the compounds offer a suitable test if the suggested novel descriptor for a heteroatom is adequate. The clonidine-like compounds examined here have been extensively studies in the past [46–48], including a particularly detailed study by Timmermans and Van Zwieten [49] based on the traditional QSAR. Thus it will be possible to make a detailed comparison between the correlations based on molecular properties as descriptors and our results derived from the use of graph theoretical indices as descriptors. Moreover, the data set used includes two extreme potency values which would be expected to give trouble in curve fittings and cross-validation, and hence the data enables a critical test of a modelling of biological activity to be made.

A need for a novel appraoch to heteroatoms in graph theoretical approaches becomes apparent from a comparison of the biological activity of the following:

Compound	*Activity*
2,4-Dimethyl	810
2-Methyl,4-chloro	275
2,4-Dichloro	61
2-Chloro, 4-methyl	53

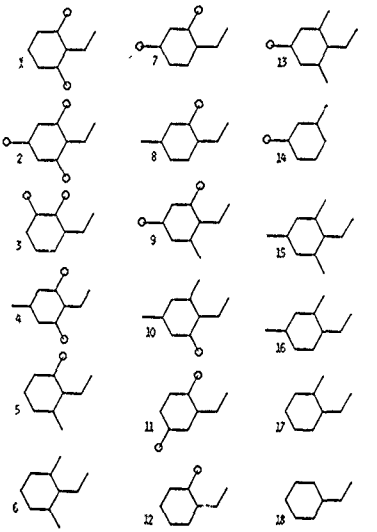


Fig. 1 Numbering of the compounds and diagrams of the variable fragment of 2-(arylimino)imidazolidines considered Chlorine atoms are indicated as small circles

The four compounds selected illustrate a lack of a bond additivity for the biological activity (experimental ED_{50} values in μg/kg obtained from dose–response curves following intravenous administration to anesthetized, normotensive rats, i.e., in-vivo effective dose which produces in 50% of population anesthesia). Any bond additive scheme should interpolate data on derivatives with a single methyl and single chlorine between the dimethyl and the dichloro derivative, but this apparently is not possible here.

TRADITIONAL QSAR CORRELATIONS BETWEEN PROPERTIES AND ACTIVITY

In Fig. 1 we depicted molecular skeletons of the 18 imidazolidines from a collection of 27 reported in the study of Timmermans and Van Zwieten

[49]. We have restricted our attention only to clonidine derivatives with chlorine as heteroatom The nine compounds not considered here involve bromine, fluorine, nitrogen and oxygen and offer too small a sample to allow one to determine empirically the graph theoretical parameters that discriminate between these heteroatoms. The QSAR parameters considered by Timmermans and Van Zwieten include:

(a) d pK_a, which refers the substituent effect on the dissociation of the imidazolidines in water expected to prevail under psychological conditions;
(b) π-electron charge densities, from quantum chemical calculations derived for free bases and protonated species;
(c) the energies of the highest occupied molecular orbital (HOMO) and lowest empty (unoccupied) molecular orbitals (LEMO or LUMO), in particular those of protonated species were considered as molecular descriptors;
(d) the lowest electronic excitation energies of the molecules (given by the difference of HOMO and LUMO energies);
(e) log P' (apparent partition coefficient) from the octanol–0.1 M phosphate buffer, pH 7.4, system;
(f) the hydrophobic constant π (in fact the summation over the substitutent π values) adopted as a measure of hydrophobic interactions,
(g) parachor, defined by Sudgen [50] as the product of the molecular volume and the fourth root of the surface tension, a measure of molecular size (along the series where surface tension is constant) perhaps related (via surface tension) to an overall lipophilic behavior of the molecules;
(h) Taft's steric constant [51], as expanded by Hansch [52], to account for the steric properties;
(i) the molar refractions at the wavelength of the sodium D doublet line, MR as a representation of the molecular volume.

Observe the rather lengthy list of molecular properties, experimental or computed, used in the search for the structure–activity correlations. This kind of QSAR should be termed property–activity, because the analysis is confined mostly to property–activity relationships. Let us point out the difficulties in counting the parameters used in such analyses. A lack of information on the degrees of freedom (i.e the number of independent parameters) involved leads to ambiguities about the reported statistics. Part of the problem originates with difficulties in tracing underlying assumptions and the number of parameters used there. For instance, what variant of MO calculations is used, and what assumptions and approximations are involved there? How does one estimate molecular volume? What is involved in determining the numerical magnitude of the volume? How does one scale various contributions? To what extent are selected QSAR parameters internally consistent and to what extent are individual parameters independent? How does a change in a choice of one descriptor influence changes of other parameters in order to preserve internal consistency of the model? It may be difficult to answer these questions. It is this accumulation of many small steps, each perhaps well defined, which eventually makes it difficult to identify the degrees of freedom used in subsequent correlations. The situation may be contrasted to the use of graph theoretical descriptors, the number of which is always known and which are defined a priori.

The correlations reported by Timmermans and Van Zwieten [49] are summarized in Table 1. We give the statistics and the correlation equation corresponing to a set of 18 methyl and chloro derivatives which we selected from the initial set

TABLE 1

Summary of the correlations based on eighteen 2-(arylimino)imidazolidine compounds having only chlorine as heteroatoms

Regression	R	S
0.546 log P − 0.222 (log P)^2	0.786	0.629
− 0.004 (Par)^2 + 0.119 Par − 0.534 pK		
+ 2.707 HOMO + 4.984 EE − 15.583	0.964	0.301
− 0.717 pK − 0.057	0.675	0.726
0.111 (Par)^2 − 0.0003 Par − 8.842	0.731	0.691
− 0.885 pK + 6.687 HOMO + 7.238 EE + 22.651	0.789	0.646
− 0.0003 (Par)^2 + 0.096 Par − 0.572 pK − 7.849	0.902	0.454

of 27 compounds. The revised correlations gave slightly better statistics, as expected, in view of the fact that now the sample of compounds studied is more homogeneous.

We may briefly summarize the main results of Timmermans and Van Zwieten as follows:

(a) correlation coefficient R of over 0.950 (and accompanying standard deviation, S, of less than 0.350) require five molecular descriptors;
(b) single descriptor, log P (as an indicator of drug transport processes), gives the correlation with $R = 0.529$ ($S = 0\,864$);
(c) the major single variable of the best correlation is d pK_a with $R = 0.482$ and $S = 0.892$;
(d) the best two-parameter correlation involves parachor (linear and quadratic terms) and increases the correlation coefficient to $R = 0.656$ ($S = 0\,784$);
(e) the best three-term expression (based on parachors and d pK_a) achieves the somewhat respectively correlation coefficient of $R = 0\,853$ ($S = 0.544$).

Timmermans and Van Zwieten [49] concluded their study by examining the role of the hydrophobic constant π and the role of the steric substituent parameter. Each case, in a comparison with the best five-parameter correlation, shows a slightly reduced correlation coefficient ($R = 0.912$ and $R = 0.943$, respectively) and an increased standard deviation ($S = 0.455$ and $S = 0.369$, respectively). The traditional QSAR study of Timmermans and Van Zwieten well illustrates the various choices in multiple regression analysis, resulting in a correlation equations using five descriptors with a high coefficient of multiple regression.

How would graph theoretical schemes fare in comparison?

THE CONNECTIVITY INDEX FOR HETEROATOMS

In order to consider the above question we have first to consider an adequate graph theoretical approach to heteroatoms. The index, initially called the branching index [53] and subsequently, quite appropriately, renamed by Kier et al. [54] as the connectivity index, was designed fron an analysis of selected physicochemical properties of alkanes. Firstly one orders isomers with respect to a property of interest. Thus, for example, in the case of hexanes and their boiling points we obtain the following sequence.

2,2-dimethylbutane < 2,3-dimethylbutane
< 2-methylpentane < 3-methylpentane
< n-hexane

By differentiating bond types involved the above ordering leads to inequalities, shown below, where (m, n) represents CC bond type with m and n being neighboring carbon atoms:

$$[(1, 2) + 3(1, 4) + (2, 4)] < [4(1, 3) + (3, 3)]$$
$$< [(1, 2) + 2(1, 3) + (2, 2) + (2, 3)]$$
$$< [2(1, 2) + (1, 3) + 2(2, 3)]$$
$$< [2(1, 2) + 3(2, 2)]$$

Similar inequalities follow from ordering of other alkanes. The bond type contributions

$(1, 2)$, $(1, 3)$, $(1, 4)$, $(2, 2)$, $(2, 3)$, $(2, 4)$, $(3, 3)$, $(3, 4)$ and $(4, 4)$

are viewed as unknown variables, which will need to be determined. Instead of searching for individual (m, n) values one finds that a simple algorithm: $1/\sqrt{(m, n)}$ generates an acceptable solution. Hence, this single assumption defines bond contributions to the connectivity index [53].

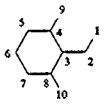
It may appear amazing that a simple ad hoc mathematical construction, the connectivity index, performs so well. But there ought to be no surprise, because the index was constructed to be a solution to an ordering of structures, an ordering which parallels the relative magnitudes for selected properties. The success of the connectivity index is in its design. One can interpret the variable bond weights as relative contributions of bonds in a typical molecular additivity, when bonds are differentiated according to the number of the nearest neighbors. The bond types (1, 2), (1, 3) and (1, 4), for example, correspond to bonds between primary and secondary, primary and tertiary, and primary and quaternary carbon atoms, respectively. The

connectivity index attributes different 'volumes' and 'surface' contributions to these different bond types, simulating the relative volume and surface fragment contributions.

WEIGHTED PATH NUMBERS AS MOLECULAR DESCRIPTORS

A single molecular descriptor will not suffice in many applications. When extending the basis of descriptors one can (i) either consider a collection of additional, structurally unrelated descriptors or (ii) design a number of different but structurally related descriptors. The higher connectivity indices [55] represent an illustration of the latter. They were defined by extending the bond as a fragment to pairs of bonds and several consecutive bonds as larger molecular fragments. In this way not only one increases the number of descriptors available for regression, but also facilitates the use of sequences as mathematical objects to represent structures. Other choices of structurally related descriptors include extended connectivities [56], path numbers [57], weighted path numbers [58] and distance sums [59]. We will use here weighted path numbers (to be subsequently briefly outlined).

In Table 2 we illustrate weighted path numbers on a ten-atom common fragment for compounds of Fig. 1:



which represents a variable fragment of the graph of clonidine-like molecules The weighting factors for the individual bond types are the same ones introduced in the definition of the connectivity index.

Let us emphasize the wealth of data in Table 2. Firstly, for each atom separately we obtain path sequences. these are the numbers listed in separate rows. As a sum of atomic path sequences we show in the last row the corresponding sequence for the molecule. The first number gives the number of atoms, but alternatively this can be replaced by the 'molecular' zero-order connectivity index of Kier and Hall [3]. The second number in the molecular sequence is the connectivity index, which can be viewed as the molecular path number associated with paths of length one, i.e. bonds The successive path counts correspond to higher connectivity indices, although they differ some-

TABLE 2

Weighted path numbers for a ten-atom fragment of the 2,6-dimethyl derivative of clonidine

Rows give weighted paths for individual atoms, the last row (obtained by summing atomic contributions) represents a characterization of the molecule (molecular fragment) as a whole.

Atom	P_1	P_2	P_3	P_4	P_5	P_6	P_7	Atomic ID
1	0 817	0 272	0 181	0.179	0 037	0 019	0 008	2 516
2	1.150	0 222	0 219	0 045	0 023	0 001	0 001	2 674
3	1	0.929	0.136	0 068	0 028	0 016		3 177
4, 8	1.319	0 426	0.302	0 064	0 049	0 001		3.170
5, 7	0.908	0 622	0.193	0 175	0 032	0 016	2 945	
6	1	0 408	0 372	0 091	0 082			2 953
9, 10	0 577	0 428	0 246	0. 75	0 037	0 029		2 497
Molecule								Molecular ID
	4.788	2.392	1.195	0 605	0 203	0 071	0 013	19 271

what in the definition in that here the weight factors of the 'connecting' atoms are used twice. But that is a minor difference that changes the results quantitatively, not qualitatively, and we may continue to refer to these as 'higher' connectivity indices. In addition to the quantities already mentioned we may also consider adding atomic contributions, not along columns as was the case with deriving the molecular path numbers, but along the rows. We then obtain a characteristic number for each atom, the so-called atomic identification (ID) number. As one can see these atomic ID numbers are sensitive to the atomic environment and tend to be different for atoms even in highly similar atomic environments However, significantly, smaller changes in the environment are accompanied by smaller variations in atomic ID numbers. By adding all atomic ID numbers (or alternatively by adding the molecular path numbers, proper account of the role of the zero-connectivity index), one obtains the molecular ID number [60] These molecular ID numbers, which in a way encode the molecular volume, have already been used in some structure–activity clusterings and correlations [58,61]. One ought to view Table 2 as a pool of various molecular descriptors.

Is it possible to incorporate heteroatoms in some analogous way in the path count schemes?

The quantities in Table 2 were calculated (by a program ALL PATH [62,63] from the graph adjacency matrix, which have zero everywhere (including diagonal entries) except on places corresponding to any pair of connected atoms when the entry is 1. Heteroatom X can be discriminated by settting the corresponding diagonal elements of a matrix to be different from zero. This is fully analogous to the treatment of heteroatoms in HMO theory. Spialter [64,65], attempted in this way to record heterosystems in chemical documentation, and even earlier Balandin [66] used the same technique to identify heteroatoms Dugundji and Ugi [67], in a similar manner, recorded the number of valence electrons of non-carbon atoms in their BE (bond–electron) matrices used to follow chemical reactions. Thus it appears natural to use variable diagonal entry to discriminate among heteroatoms, a practice which apparently is not novel.

In Table 3 we list a weighted path numbers for the same ten-atom fragment of clonidine, but now the atoms 9 and 10, corresponding to chlorines in compound *1*, have been assigned a non-zero diagonal entry in the adjacency matrix The ALLPATH program recognizes the non-zero diagonal entries and modifies the weighted path count accordingly Hence, if we compare Table 2 and Table 3 we can observe the differences induced by the two chlorine atoms. Thus, Table 2 represents the 2,6-dimethyl derivative, compound *6*, while Table 3 corresponds to the 2,6-dicholoro derivative, compound *1*. In the next section we will consider correlations between the eighteen im-

TABLE 3

Weighted path numbers for the ten-atom fragment with chlorine atoms as heteroatom substituents (labels 9, 10), corresponding to the 2,5-dichloro derivative of clonidine

Observe a similarity between the corresponding path numbers of Tables 2 and 3

Atom	p_1	p_2	p_3	p_4	p_5	p_6	p_7	Atomic ID
1	0 816	0 272	0 181	0 191	0 037	0 019	0 008	2 529
2	1 150	0 222	0 234	0 045	0 023	0 009	0 006	2 690
3	1	0.975	0 136	0 068	0 028	0 018		3 225
4, 8	1 387	0 426	0 310	0 064	0 052	0 005	0 004	3 248
5, 7	0 908	0 650	0 193	0.185	0 032	0 017		2 984
6	1	0 408	0 400	0 091	0 085			2 984
9, 10	0 646	0 479	0 275	0 200	0 042	0 034	0 003	2 680
Molecule								Molecular ID
	4 924	2 493	1 254	0 647	0 212	0 078	0 014	19 625

idazolidines of Fig. 1 using the graph theoretical descriptors from Table 2 and Table 3, and similar data for other compounds of interest.

GRAPH THEORETICAL CORRELATION OF THE ACTIVITIES OF IMIDAZOLIDINES

The emphasis in this article is on advantages of graph theoretical descriptors in comparison with the traditional QSAR descriptors. Table 3 illustrates how a graph theoretical scheme naturally incorporates heteroatoms, but the task of finding optimal 'diagonal' contributions for various heteroatoms or even the same heteroatom in different environments remains to be studied in greater detail. The preliminary examination reveals that positive diagonal elements decrease the path counts while negative elements increase the magnitude of the weighted path counts This suffices for our purpose of generating preliminary connectivity indices that discriminate positional isomers with variable heteroatom location. In Table 4 we listed the leading connectivity indices for the eighteen compounds of interest as derived by the ALLPATH program with assumed $X = -0.20$ entry for each chlorine present. In addition there is also an option to change C–Cl bond weights but at this stage we decided to keep the number of variables to a minimum. In order to emphasize the role of substitution pattern, because we are dealing with compounds of different number of atoms, we focused attention on the eight-atom skeleton

TABLE 4

Leading connectivity indices for the eighteen compounds considered

No	Compound	$1-X$	$2-X$	$3-X$
1	2,6-Cl_2	4 278	2 015	0 978
2	2,4,6-Cl_3	4 418	2 120	0 969
3	2,3-Cl_2	4 278	2 015	0 963
4	2,6-Cl_2-4-Me	4 384	2 092	0 957
5	2-Cl-6-Me	4 244	1 989	0 964
6	2,6-Me_2	4 210	1 964	0 949
7	2,4-Cl_2	4 262	2 036	0 982
8	2-Cl-4-Me	4 228	2 008	0 970
9	2,4 Cl_2-6-Me	4 384	2 095	0 955
10	2,4-Me_2-6-Cl	4 350	2 067	0 944
11	2,5-Cl_2	4 262	2 036	0 982
12	2-Cl	4 122	1 939	0 982
13	2,6-Me_2-4-Cl	4 350	2 069	0 942
14	2-Me-4-Cl	4 228	2 011	0 968
15	2,4,6-Me_3	4 316	2 042	0 931
16	2,4-Me_2	4 194	1 983	0 956
17	2-Me	4 088	1 914	0 966
18	Unsubstituted	3.966	1 869	0 979

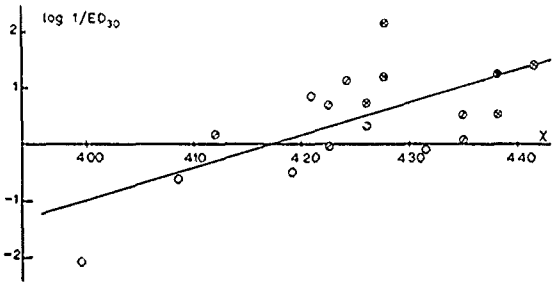


Fig. 2. Plot of the connectivity index against log $1/ED$ Open circles indicate compounds without chlorine substituents, singly crossed circles indicate compounds with single substituted chlorine, doubly crossed circles indicate compounds with two chlorines, and triply crossed circle indicates the compound with three chlorine heteroatoms

shown below, which is common to all the structures and is sensitive to substitutions

The reported connectivity indices in Table 4 therefore represent fragment connectivities, i e they include only contributions from the above common eight atoms. The computed path numbers, however, involve from eight to eleven atoms, depending on the substitution pattern. In this way we have separated the combined influences of the two structural features — the size and the shape — which will enable us to focus attention on the 'shape', i.e. the substitution pattern, and its role on the relative bioactivities of the compounds

With a single graph theoretical descriptor, the connectivity index $1 - X$ of Table 4, we obtain the correlation shown in Fig. 2. The correlation coefficient is $R = 0.690$ and the standard error estimate is $S = 0.712$ This is visibly better than the best single property-based QSAR correlation, the log P', with $R = 0.529$ and $S = 0.846$ The correlation equation

$$\log(1/ED) = 5\ 781X - 24.1643$$

explains almost 50% of the variance in hypotensive activity and equally shows that bond additivity (implied in the connectivity index) is not the only aspect of this particular structure–activity relationship. The above may be contrasted with the log P' correlation, which account for only 30% of the variance in hypotensive activity and shows that hipophilic behaviour is not the dominant contributor to the biological activity of clonidine-like imidazolidines.

What is the next best descriptor that will improve the correlation? A way to proceed is to examine the correlation predictions more closely and see if a well characterized subset of the compounds show greatr departure from the correlation. By inspection of Fig. 2, which gives a plot of $\log(1/ED)$ against the connectivity index, we observe the average values of $\log(1/ED)$ significantly increase with the number of chlorine atoms as substitutents. Hence we may expect that inclusion of the molecular weight, which increases with the number of substituted chlorines, will improve the correlation. Alternatively, we may consider the count of chlorines (which parallels molecular weight as a descriptor) to improve the correlation) This observation leads to the following two-descriptor correlation equation:

$$\log(1/ED) = 3.508X + 0.440N - 15.0101$$

with $R = 0.764$ and $S = 0\ 656$, where N is 0, 1 or 2. The improvement is not dramatic, but the correlation is significantly better than the quadratic correlation based on log P' (with $R = 0.647$ and $S = 0.793$) or the quadratic correlation based on parachor (with $R = 0.656$ and $S = 0\ 784$), which similarly involve three terms in the correlation equations. Hence, again, we see that simple graph theoretical considerations produce visibly better results.

Another look at the compounds which show a greater departure from the correlation line in Fig 2 suggests that bond dipoles may play some role. Among the isomers having a same number of C–Cl bonds those with bonds in the *ortho* position have greater activity than those with C–Cl bonds in *meta* or *para* positions Consequently, one can visualise the resultant dipole vectors as pointing to the direction of the 'shift' of the points in the correlation By using the magnitudes D of the dipoles (which are sensitive to the substitution mode) as a parameter we also expect to improve the structure–activity correlation Implementation of this observation leads to the expression.

$$\log(1/ED) = 5.854X + 0\ 679D - 24.508$$

with $R = 0.799$ and $S = 0\ 611$ This particular two-descriptor correlation compares well with the two-descriptor correlation of Timmermans and Zwieten.

The graph theoretical approaches not only have their quantitative value, they also provide novel qualitative structural insights. In the above case we identified molecular weight and bond dipoles as potentially useful descriptors. Nevertheless, we should add that the quantitative results, impressive as they are, are not necessarily the best which the particular graph theoretical approach may yield. We have not attempted to optimize our

heteroatom parameter (the diagonal entry in the associated adjacency matrix) That there is a room for improvement can be seen by considering another choice for diagonal entry parameter of chlorine. The value of $x = -0.40$ gives a better (single descriptor) correlation: $\log(1/ED) = 4.860X - 20.531$ with $R = 0.750$ and $S = 0.651$, as compared to the correlation derived with $x = -0\,20$ ($R = 0\,690$ and $S = 0.712$). This result, which is also not optimal, shows that a single graph theoretical descriptor can capture dominant structural features well. If a single graph theoretical descriptor can produce correlations which are better than alternatives using two and more traditional descriptors, it seems worthwhile to explore further the possibilities based on graph theoretical descriptors. Recently an approach to the construction of better single descriptors has been considered [68] It appears that further improvements in structure–property and structure–activity studies are possible by following similar promising directions in modifying the functional dependence of the topological indices used Traditional QSAR approaches lack this flexibility by virtue of being limited to molecular properties as a source of structural characterizations

MULTIPLE REGRESSION USING HIGHER CONNECTIVITIES

A single best descritor allows one to model a structure–activity study by considering the role of various 'correction' factors, as outlined above. An alternative approach is to use 'higher-order' descriptors, such as higher-order connectivities, paths of longer length, extended connectivities, etc. If, for a collection of compounds considered, such descriptors are not strongly intercorrelated they may span the structure space adequately, and hence produce impressive correlations. We want to end this exposition by showing correlations of antihypertensive activities of clonidine-like imidazolines using longer (weighted) paths involving the particular encoding of chlorine heteroatoms. In Table 5 we collected the information on the correlations using paths of length one (the connectivity index $1 - X$), paths of length two (denoted as $2 - X$ and corresponding to the connectivity index of second order) and paths of length three (denoted as $3 - X$ and corresponding to connectivity indices of order three). The three connectivity indices $1 - X$, $2 - X$ and $3 - X$ have not been selected as the best three from a pool of possible indices, their reciprocal and other combinations, as sometimes has been the case in multiple regression analyses. They have rather been selected as the leading members of a sequence of weighted paths (higher connectivities).

TABLE 5

Predicted antihypertensive activities based on the connectivity indices $1 - X$, $2 - X$ and $3 - X$ derived from multiple regression and cross-validation

No	Compound	Regression	Cross-validation	Experiment
1	2,6-Cl_2	2 034	1 977	2 14
2	2,4,6-Cl_3	1 460	1 478	1 41
3	2,3-Cl_2	1 298	1 286	1 37
4	2,6-Cl_2-4-Me	1 061	1 035	1 22
5	2-Cl-6-Me	1 372	1 451	1 12
6	2,6-Me_2	0 697	0 627	0 85
7	2,4-Cl_2	0 566	0 536	0 68
8	2-Cl-4-Me	0 111	0 061	0 68
9	2,4-Cl_2-6-Me	0 850	0 901	0 57
10	2,4-Me_2-6-Cl	0 459	0 448	0 52
11	2,5-Cl_2	0 548	0 607	0 32
12	2-Cl	0 259	0 285	0 15
13	2,6-Me_2-4-Cl	0 249	0 300	0 04
14	2-Me-4-Cl	−0 080	−0 084	−0 05
15	2,4,6-Me_3	−0 150	−0 193	−0 07
16	2,4-Me_2	−0 532	−0 527	−0 56
17	2-Me	−0 448	−0 406	−0 61
18	Unsubstituted	−2 076	−2 047	−2 10
R		0 9773	0 9676	
S		0 2223	0 2475	

Connectivity indices $1 - X$, $2 - X$ and $3 - X$ lead to quite impressive regressions A stated earlier, $1 - X$ already accounts for close to 50% of the variance. In combination with $2 - X$ the two-descriptor characterization of the compounds (three-parameter correlation equation) account for 60% of the variance (correlation coefficient $R = 0\,781$). This is better than any two-parameter correlation, based on traditional QSAR parameters, even including correlations using bond dipoles or molecular weights as descriptors in conjunction

with the connectivity index. The improvement by including $2 - X$ to the already existing correlation based on $1 - X$ is substantial, even if not dramatic. In part a reason for the achievement of only a partial improvement is that $1 - X$ has already absorbed much of the correlation variance. However, if we now include $3 - X$, in addition to $1 - X$ and $2 - X$, we obtain a correlation equation which accounts for 95.5% of the variance (a correlation coefficient of 0 977). The regression is also accompanied by an impressive reduction of the standard deviation to $S = 0.223$. This particular result is better than a correlation based on four and five descriptors using any combination of apparently plausible physicochemical descriptors, such as log P, d pK_a, parachor, Taft's steric constants, molar refractions etc., supplemented by quantum chemical parameters, such as HOMO and LUMO parameters and their derivatives.

The central finding — that the X indices provide a superior correlation of the antihypertensive clonidine data for the eighteen compounds chosen — appear to be correct, providing that a chance correlation does not play a role. In order to confirm this finding we undertook to examine whether the result would be upheld by cross-validation. In Table 5 we also report the outcome of the cross-validation, which leads to the overall coefficient of correlation of 0.968 with the standard error of estimate of 0.247. The result is particularly striking for this data set, because there are two extreme potency values which would be expected to give much trouble in cross-validation. The suspicion with which many people in the QSAR community regard graph theoretical approaches is based on misunderstandings of graphs, on a feeling that there is no physicochemical basis for connectivity correlations. Since "receptors surely do not perform edge counting", skeptics feel that correlations with graph indices which do exist are actually a consequence of correlations with some more 'meaningful' physicochemical property which the graph indices happen to correlate with. However, the result reported here cannot be understood in this way. With new statistical methods, such as the partial least-squares method, inclusion of many sets of highly intercorrelated parameters is no longer a problem, and combining graph indices with physicochemical indices in a single study is practical.

COMBINED USE OF PHYSICOCHEMICAL AND GRAPH THEORETICAL DESCRIPTORS

While the traditional QSAR parameters may have apparent advantages in some applications in this particular study, where several factors contribute to the overall molecular behavior, it is difficult even to speculate on the importance of individual physicochemical descriptors On the other hand graph theoretical descriptors can not only do the same job, they can accomplish it impressively better. Successful graph theoretical correlations, of course, do not signal the termination, or even a diminution of the importance of traditional approaches; rather, they indicate the beginning of a novel alternative, sending a signal for attention. Certainly, one needs to accumulate more experience and additional insights into the potential of the outlined approach. We do not even claim any general suitability of the approach outlined for the study of structure–activity phenomena involving heteroatoms. Even less do we want to leave the impression that the traditional approaches have no considerable, as yet untapped, potential along with graph theoretical approaches in QSAR. In fact, we believe that combined approaches using molecular properties, quantum chemical parameters and well selected graph theoretical descriptors are likely not only to produce superior correlations but are likely to do so in a most efficient way. While this paper has demonstrated some advantages of mathematical descriptors as opposed to physicochemical descriptors in this particular application, the advocation of one set of descriptors does not preclude the use of other sets of descriptors. Moreover, any claim to a general superiority of one kind of descriptors over another kind, even if based on a larger body of results, overlooks the possibility that yet unexplored descriptors (properties or graph invariants) may surpass in quality those considered hitherto It seems that the most pragmatic approach at this time is to combine physicochemical descriptors with graph theoretical descriptors, a course which

already has received some support [69–71]. This then represents a generalization of a more common current practice in which physicochemical descriptors are combined with quantum chemical descriptors. Such generalized approaches are likely to result not only in better but also in simpler correlations than the approaches using one type of descriptor only, if used separately.

To illustrate a relationship between properties and connectivities as descriptors for the eighteen compounds considered we report in Table 6 correlations using traditional QSAR descriptors against the connectivity index X. Such correlations may assist one in selecting graph theoretical and physicochemical descriptors in 'admixture'. We find that X and parachor produce quite a good correlation ($R = 0.965$), not quite unexpectedly, in view of the interpretation of the parachor in terms of molecular surface. The magnitudes of molecular surface area are well simulated by the relative magnitudes of the connectivity index [52]. Also a quite good correlation (with $R = 0.950$) was obtained between X and hydrophobic constants (summation over the substituent π values). The correlation between X and the Taft substituent steric constants produced a fair correlation, not as good as hydrophobic constants or parachor, but still suggesting that over 75% variance is accounted for by X ($R = 0.881$). On the other hand, the correlation between X and quantum chemical HOMO parameters (as well as the derived EE parameters) are nonexistent ($R = 0.114$ and $R = 0.070$, respectively). These molecular orbital descriptors (for the set of structures considered) have apparently 'nothing in common' with the bond

TABLE 6

Correlations between the various physicochemical descriptors and the connectivity indices for the eighteen compounds considered

Descriptor	R	S	Coefficient	Constant
Parachor	0.965	8.92	2809	−1032.8
π	0.950	0.169	4.397	−17.365
E_s	0.881	0.463	−7.350	28.953
log P	0.715	0.734	6.397	−27.603
pK_a	0.430	0.837	−3.392	13.757
HOMO	0.114	0.161	0.158	−12.330
EE	0.071	0.146	0.089	7.2519

TABLE 7

Two-parameter correlations combining the connectivity index and selected physicochemical descriptors

Regression	R	S
4.154 X − 0.489 pK_a − 17.574	0.808	0.599
2.607 X + 0.500 log P − 10.465	0.786	0.628
6.181 X − 2.212 HOMO − 51.654	0.782	0.633
5.604 X + 1.988 EE − 38.636	0.751	0.671
11.052 X − 1.385 π − 44.661	0.701	0.725

additivities implied by the connectivity index. Hence, they illustrate descriptors which, figuratively speaking, are 'orthogonal' to the connectivity index. They supply additional 'directions' in correlations if, on their own, they show some correlation with a property considered. We should emphasize that use of R, the coefficient of regression, or R^2, the coefficient of determination, as a sole criterion for a quality of a regression, as is known, is deficient and can be downright misleading. Hence conclusions based on R or R^2 have to be taken with due reservation. It is desirable to substantiate such correlations with other independent statistical criteria, such as are given by magnitudes of the standard errors, F-tests, cross-validation, etc.

In Table 7 we show several 'mixed' correlations based on the connectivity index X and a selected property as descriptors. We see that when X is combined with quantum chemical descriptors HOMO and EE fair correlations result ($R = 0.782$ and $R = 0.751$, respectively). Comparisons of the correlations in Table 6 and Table 7 give insight into the role that some physicochemical descriptors play in multiple regressions. We see that there is a fair, but not satisfactory, correlation between log P' and X, the correlation coefficient being $R = 0.715$. Combined log P' and X then give a better correlation, though the improvement appears not to be dramatic ($R = 0.786$). Because log P' alone does not perform well ($R = 0.529$) it seems, then, that in this particular application to clonidine-like compounds, log P' owes its correlation 'power' to partial parallelism with X. However, the parts in which X differ from log P' appear relevant for the particular correlation. The situation can be contrasted to the use of d pK as

an additional physicochemical descriptor We see that d pK combined with X produces a good correlation ($R = 0\,808$), the improvement in the correlation, however, in this case is more substantial This should not be surprising in view of the limited correlation between d pK and X ($R = 0.430$) It implies a lesser 'duplication' between X and d pK on one side, while the improved correlation coefficient in the combined regression points to a role of d pK, which alone shows poor correlation ($R = 0.482$), as complementary descriptor, rather than competitive to X; i.e. they differ in structurally relevant features.

CONCLUDING REMARKS

The complexity of structure–activity studies is enormous, and different methodologies, even if addressing limited aspects of the QSAR problem, ought to be exhaustively explored and combined if possible. We have demonstrated, albeit on a single case of hypotensive clonidine-type compounds, that graph theoretical descriptors not only have the potential to describe structural variations in molecules with 'floating' heteroatoms, but that the accompanying descriptors are superior to any well-tested combination of traditional QSAR descriptors. The result ought to draw attention to mathematical descriptors, while at the same time the use of physicochemical descriptors is not discouraged. It should be superfluous to add that mathematical descriptors, of which graph theoretical invariants are illustrations, have an important advantage — an explicit structural interpretation. By contrast, many quantum chemical descriptors and molecular properties as descriptors are highly convoluted, without pointing to simple structural features directly as the dominant components of a correlation.

Pragmatism suggests that, at least at the present time, before we fully understand the intricate interrelationship of structure and properties, the best results may follow when both sets of descriptors are combined, by 'mixing' the two points of view. Be that as it may, it is opportune to end this article with a quote from Max Planck [72], intended for those who continue to be skeptical regarding graph theoretical methods:

" ..the experimenter cannot afford to close his eyes to a new discovery, obtained from another point of view, which will not fit his own ideas, nor must he treat it as unimportant, if not incorrect ".

One should not need to add that graph theoretical indices — being mathematical constructions — cannot be incorrect! They can be useful or useles, but not incorrect, and we leave it to readers to decide which is the case

ACKNOWLEDGEMENT

I would like to thank Professor A.T. Balaban for valuable comments which have led to an improved presentation of the results.

RFFERENCES

1 N Trinajstić, M Randić and D J Klein, On the quantitative structure–activity relationship in drug research, *Acta Pharmaceutica Jugoslavia*, 36 (1986) 267–279
2 C Hansch, A quantitative approach to biochemical structure–activity relationships, *Accounts of Chemical Research*, 2 (1969) 232–239
3 L.B Kier and H L Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976
4 L.B Kier and H L. Hall, *Molecular Connectivity in Structure–Activity Analysis*, Research Studies Press, Letchworth, 1986
5 A Sabljić and N Trinajstić, Quantitative structure–activity relationships the role of topological indices, *Acta Pharmaceutica Jugoslavia*, 31 (1981) 189–214
6 D H Rouvray, The prediction of biological activity using molecular connectivity indices, *Acta Pharmaceutica Jugoslavia*, 36 (1986) 239–252
7 N Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, FL, 1985
8 R Kaliszan and H Foks, The relationship between the R_m values and the connectivity indices for pyrazine carbothioamide derivatives, *Chromatographia*, 10 (1977) 346–349
9 G R. Parker, Correlation of log P with molecular connectivity in hydroxyureas influence of conformational system on log P, *Journal of Pharmaceutical Science*, 67 (1978) 513–516
10 A. Cammarata, Molecular topology and aqueous solubility of aliphatic alcohols, *Journal of Pharmaceutical Science*, 68 (1979) 839–842

11 C. Mercier and J. Dubois, Comparison of molecular connectivity and Darc/Pelco methods. performance in antimicrobial, halogenated phenol QSARs, *European Journal of Medicinal Chemistry*, 14 (1979) 415–423.

12 T.R. McGregor, Connectivity parameters as predictors of retention in gas chromatography, *Journal of Chromatographic Science*, 17 (1979) 314–316

13 D R. Henry and J H Block, Pattern recognition of steroids using fragment molecular connectivity, *Journal of Pharmaceutical Science*, 69 (1980) 1030–1034

14 K. Altenburg, Eine Bemerkung zu dem Randićschen ' Molekularen Bindungs-Index (Molecular Connectivity Index)" (Note on the Randić molecular connectivity index), *Zeitschrift fur Physikalische Chemie (Leipzig)*, 261 (1980) 389–393

15 A Sabljić, N Trinajstić and D. Maysinger, Molecular connectivity and biological activity in a series of isatin derivatives, *Acta Pharmaceutica Jugoslavia*, 31 (1981) 71–76

16 A Sabljić and M Protic, Molecular connectivity a novel method for prediction of bioconcentration factor in hazardous chemicals, *Chemical and Biological Interactions*, 42 (1982) 301–310.

17 D Maysinger, M Movrin and M Ljubić, Structure–activity relationship in opioid peptides, *Acta Pharmaceutica Jugoslavia*, 32 (1982) 177–184.

18 S C Basak, D P Gieschen, D K Harriss and V.R Magnuson, Physicochemical and topological correlates of the enzymatic acyltransfer reactions, *Journal of Pharmaceutical Science*, 72 (1983) 934–937

19 L Buydens, D Coomans, M Vanbelle, D L. Massart and R Vanden Dnessche, Comparative study of topological and linear free energy-related parameters for the prediction of GC retention indices, *Journal of Pharmaceutical Science*, 72 (1983) 1327–1329

20 W.J Spillane, G McGlinchey, I O. Muircheartaigh and G A Benson, Structure–activity studies on sulfamate sweeteners III Structure–taste relationships for heterosulfamates, *Journal of Pharmaceutical Science*, 72 (1983) 852–856

21 Y. Takahashi, Y Miyashita, Y. Tanaka, H Hayasaka, H Abe and S.-I Sasaki, Discriminative structural analysis using pattern recognition techniques in structure–taste problem of penillartines, *Journal of Pharmaceutical Science*, 73 (1984) 737–741.

22 A Robbat, N.P Corso, P.J Doherty and D. Marshall, Multivariate relationship between gas chromatographic retention index and molecular connectivity of mononitrated polycyclic aromatic hydrocarbons, *Analytical Chemistry*, 58 (1986) 2072–2077

23 D. Vasilescu and R. Viani, Molecular similarity in aminothiol radioprotectors. a Randić graph approach, *International Journal of Quantum Chemistry Quantum Biology Symposium*, 14 (1987) 149–165.

24 S C. Basak, V R Magnuson, G.J Niemi and R R. Regal, Determining structural similarity of chemicals using graph-theoretic indices, *Discrete Applications in Mathematics*, 19 (1988) 17–44.

25 T. Okuyama, Y. Miyashita, S. Kanaya, H Katsumi, S-i Sasaki and M Randić, Computer assisted structure–taste studies on sulfamates by pattern recognition method using graph theoretical invariants, *Journal of Computational Chemistry*, 9 (1988) 636–646.

26 R.D. Cramer, III, BC(DEF) Parameters 1 The intrinsic dimensionality of intermolecular interactions in liquid state, *Journal of the American Chemical Society*, 102 (1980) 1837–1849

27 R.D. Cramer, III, Errata, *Journal of the American Chemical Society*, 103 (1981) 2143

28 C.L. Wilkins and M Randic, A graph theoretical approach to structure–property and structure–activity, *Theoretica Chimica Acta*, 58 (1980) 45–68

29 M. Randić and C.L. Wilkins, Graph theoretical study of structural similarity in benzomorphans, *International Journal of Quantum Chemistry Quantum Biology Symposium*, 6 (1979) 55–71.

30 B Jerman-Blažič and M Randić, Modelling molecular structures for computer-assisted studies of drug structure–activity relations, *Proceedings of the International AMSE Conference on Modelling and Simulation, Nice, Sept 12–14, 1983*, AMSE Press, Tassin, 1983, Vol 5, pp 161–174

31 M Randić, G A Kraus and B Dzonova-Jerman-Blažič, Ordering of graphs as an approach to structure–activity studies, *Studies in Physical and Theoretical Chemistry*, 28 (1983) 192–205.

32 C L. Wilkins, M. Randić, S M Schuster, R S Markin, S Steiner and L Dorgan, A graph-theoretic approach to quantitative structure–activity/reactivity studies, *Analytica Chimica Acta*, 133 (1981) 637–645

33 M. Randić, B Jerman-Blažič, D.H Rouvray, P G Seybold and S C Grossman, The search for active substructures in structure–activity studies, *International Journal of Quantum Chemistry Quantum Biology Symposium*, 14 (1987) 245–260

34 M Randić, S C Grossman, B Jerman-Blažič, D H Rouvray and S El-Basil, An approach to modeling of mutagenicity of nitroarenes, *Mathematical Computation and Modelling*, 11 (1988) 837–842.

35 M Randić, B. Jerman-Blažič, S C Grossman and D.H Rouvray, A rational approach to the optimal design of drugs, *Mathematical Computation and Modelling*, 8 (1986) 571–582.

36 M. Randić, Graph theoretical approach to structure–activity studies' search for optimal antitumor compounds, in R Rein (Editor), *Molecular Basis of Cancer, Part A*, Alan R Liss, New York, 1985, pp 309–316

37 M Randić, in M A Johnson and G M Maggiora (Editors), *Concepts and Applications of Molecular Similarity*, Wiley, New York, 1990, pp. 77–145

38 M Randić, Molecular topographic descriptors, *Studies in Physical and Theoretical Chemistry*, 54 (1988) 101–108

39 M Randić, On characterization of three-dimensional structures, *International Journal of Quantum Chemistry Quantum Biology Symposium*, 15 (1988) 201–208

40 M. Randić, B Jerman-Blažič and N Trinajstić, Computation of 3-dimensional molecular descriptors, *Computers and Chemistry*, 14 (1990) 237–246.

41 M Randić, The nature of chemical structure, *Journal of Mathematical Chemistry*, 4 (1990) in press.
42 E.J Kupchik, General treatment of heteroatoms with the Randić molecular connectivity index, *Quantitative Structure–Activity Relationships*, 8 (1989) 98–103.
43 P.J. Hansen, Northwestern College, Iowa, personal communication
44 A Streitwieser, Jr., *Molecular Orbital Theory for Organic Chemists*, Wiley, New York, 1961.
45 S C. Grossman, B Jerman-Blažič Dzonova and M Randić, A graph theoretical approach to quantitative structure–activity relationship, *International Journal of Quantum Chemistry Quantum Biology Symposium*, 12 (1986) 123–139
46 F Avbelj and D Hadži, Potential energy functions and the role of the conformational entropy of clonidine-like imidazolidines in determining their affinity for alpha-adrenergic receptors, *Molecular Pharmacology*, 27 (1985) 466–470
47 A Carpi, J M Leger, G Leclerc, N Decker, B Rouot and C G Wermuth, Comparison of chrystallographic and quantum mechanical analysis with biological data on clonidine and some related analogues, *Molecular Pharmacology*, 21 (1982) 400–408
48 C M Meerman-van Benthem, K van der Meer, J.J C Mulder, P B M W Timmermans and P A van Zwieten, Clonidine base evidence for conjugation between both ring systems, *Molecular Pharmacology*, 11 (1975) 667–670
49 B M W M Timmermans and P A van Zwieten, Quantitative structure-activity relationship in centrally acting imidazolidines structurally related to clonidine, *Journal of Medicinal Chemistry*, 20 (1977) 1636–1644
50 S. Sudgen, A relation between surface tension, density and chemical composition, *Journal of the Chemical Society*, 125 (1924) 1177–1189
51 R W Taft, Separation of polar, steric and resonance effects in reactivity, in M S Newman (Editor), *Steric Effects in Organic Chemistry*, Wiley, New York, 1956, pp 556–675
52 C Hansch, Quantitative approaches to pharmacology quantitative structure relationship, in C J Cavallito (Editor), *Structure–Activity Relationship*, Vol 1, Pergamon, Oxford, 1973, pp 75–165
53 M Randić, On characterization of molecular branching, *Journal of the American Chemical Society*, 97 (1975) 6609–6615
54 L B Kier, L H Hall, W J Murray and M Randić, Molecular connectivity I: Relationship to nonspecific local anesthesia, *Journal of Pharmaceutical Science*, 64 (1975) 1971–1974
55 L B Kier, W J Murray, M Randić and L.H Hall, Molecular connectivity V Connectivity series concept applied to density, *Journal of Pharmaceutical Science*, 65 (1976) 1226–1230
56 M Randić, in preparation.
57 M Randić and C.L. Wilkins, Graph theoretical approach to recognition of structural similarity in molecules, *Journal of Chemical Information and Computer Science*, 19 (1979) 31–37
58 M. Randić, Nonempirical approach to structure–activity studies, *International Journal of Quantum Chemistry Quantum Biology Symposium*, 11 (1984) 137–153
59 A T. Balaban, Chemical graphs Part 48 Topological index J for heteroatom containing molecules taking into account periodicities of element properties, *MATCH*, 21 (1986) 115–122
60 M Randić, On molecular identification numbers, *Journal of Chemical Information and Computer Science*, 24 (1984) 164–175.
61 B Bogdanov, S Nikolić, A Sabljić, N Trinajstić and S Carter, On the use of weighted identification numbers in the QSAR study of the toxicity of aliphatic ethers, *International Journal of Quantum Chemistry Quantum Biology Symposium*, 14 (1987) 325–330
62 M Randić, G M Brissey, R B Spencer and C L Wilkins, Search for self-avoiding paths for molecular graphs, *Computers and Chemistry*, 3 (1979) 5–13
63 M Randić, G M Brissey, R.B. Spencer and C L Wilkins, Use of self avoiding paths for characterization of molecular graphs with multiple bonds, *Computers and Chemistr* 4 (1980) 27–32
64 L. Spialter, The atom connectivity matrix (ACM) and its characteristic polynomial (ACMCP) a new computer-oriented chemical nomenclature, *Journal of the American Chemical Society*, 85 (1963) 2012–2013
65 L Spialter, The atom connectivity matrix (ACM) and its characteristic polynomial (ACMCP), *Journal of Chemical Documentation*, 4 (1964) 261–269
66 A.A Balandin, Strukturnaya algebra v khimii (Structural algebra in chemistry), *Uspekhi Khimii*, 9 (1940) 390–400
67 J Dugundji and I Ugi, An algebraic model of constitutional chemistry as a basis for chemical computer programs, *Topics in Current Chemistry*, 39 (1973) 19–64
68 M Randić, P J Hansen and P.C Jurs, Search for useful graph theoretical invariants of molecular structure, *Journal of Chemical Information and Computer Science*, 28 (1988) 60–68
69 S. Basak, D P Gieschen and V R Magnuson, A quantitative correlation of the LC 50 values of esters in pimephales promelas using physicochemical and topological parameters, *Environmental Toxicology and Chemistry*, 3 (1984) 191–199
70 T.R Stouch and P C Jurs, Computer-assisted studies of molecular structure and genotoxic activity by pattern recognition techniques, *Environmental Health Perspectives*, 61 (1985) 329–343
71 P C Jurs, T R Stouch, M Czerwinski and J N Narvaez, Computer-assisted studies of molecular structure–biological activity relationships, *Journal of Chemical Information and Computer Science*, 25 (1985) 296–308
72 Max Planck, *Survey of Physical Theory*, Dover, New York, 1960 (republication of *A Survey of Physics*, Methuen & Co)

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 229–237
Elsevier Science Publishers B.V., Amsterdam

The ligand-field regime

M. Gerloch
University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW (U.K.)

(Received 1 March 1990; accepted 3 July 1990)

Abstract

Gerloch, M., 1991. The ligand-field regime. *Chemometrics and Intelligent Laboratory Systems*, 10: 229–237.

The ligand-field regime defines the domain of applicability and underlying reasons for the empirical success of ligand-field analysis. This article reviews the structural connections between quantum chemistry at large and the phenomenology of the ligand-field method. These connections provide a sound basis for the chemical interpretation of ligand field parameters. Differences between ligand-field and molecular-orbital approaches are identified.

THE LIGAND-FIELD FORMALISM

Ligand-field theory (LFT) addresses the spectroscopic and paramagnetic properties associated with open d or f electron shells in transition-metal complexes. It is parametric. We require of the models of such a theory that all appropriate electronic properties be reproduced essentially quantitatively for object systems regardless of molecular geometry, coordination number, or d^n (f^n) configuration, on the same footing, and that the optimized parameters affording that reproduction be relatable, both empirically and structurally, to chemical concepts established by other means. Hundreds of ligand-field analyses of paramagnetic susceptibility, electron-spin-resonance g values, 'd-d' and 'f-f' transition energies, intensity distributions, and their natural or magnetic circular dichroism have satisfied these criteria. It is crucial to observe that, within its proper or 'regime', LFT works, because, at first sight, it ought not to.

LFT developed from crystal-field theory (CFT). Within that approach, d (f) electron energies (say) are calculated by diagonalization of the appropriate d (f) basis under the crystal-field Hamiltonian,

$$\mathcal{H}_{CF} = \sum_{i<j}^{N} \frac{e^2}{r_{ij}} + V_{CF} \qquad (1)$$

in which two-electron energies are accounted for by the Coulomb operator and one-electron energies by the crystal-field potential, V_{CF}. Various models of the electrostatic, classical potential have been entertained, ranging from ligands as point-charges or point-dipoles to spatially extended charge distributions. In each case, all operators of eq. (1) are explicit, involving real bond lengths and charges. The d (f) basis is equally explicit for example one might employ the 3d functions of Clementi et al. [1] for cobalt as a dipositive cation. While the qualitative, symmetry aspects of CFT remain as useful as ever, the quantitative predictions of splitting parameters and accounts of the spectrochemical series, for example, were recognized to be hopeless almost from the beginning. 1935 marks the year in which Van Vleck [2,3]

0169-7439/91/$03.50

resolved intriguing conflicts in the contemporary literature and introduced amendments to CFT that defined the birth of LFT. In essence, acknowledging the covalency that undoubtedly exists in all transition-metal complexes, he proposed LFT as an isomorphous approach to CFT in which the operators of the ligand-field Hamiltonian,

$$\mathscr{H}_{\mathrm{LF}} = \sum_{i<j}^{N} U(i, j) + V_{\mathrm{LF}} \tag{2}$$

are to be taken as effective operators and ligand-field splittings to be regarded as parameters. Today, we refer to the two-electron energies as computed with an effective, or screened, Coulomb operator, $U(i, j)$, and the one-electron energies as ligand-field parameters of the effective ligand-field potential, V_{LF}. It is also to be recognized that the only part of the basis functions that is employed explicitly in ligand-field calculations is the angular property. Matrix elements of functions built from pure d ($l = 2$) or f ($l = 3$) orbitals under $\mathscr{H}_{\mathrm{LF}}$ are manipulated within LFT: any differences between the first and second row of the d block, for example, are left to emerge in the parameters of the system. Altogether, therefore, in LFT we employ effective operators within a basis whose radial character is left implicit. One immediate consequence of these differences between CFT and LFT is the change from (calculable) free-ion, two-electron energies — like B_0 and C_0, using Racah's notation — to parametric quantities like B, C and the nephelauxetic effect.

LFT and CFT are isomorphous in the way they formally separate one- and two-electron effects and by their operation within a pure d (or f) basis. No explicit recognition is made of metal s or p functions, or of ligand orbitals. They are thus quite unlike molecular-orbital (MO) theory. Despite Van Vleck's illustration [2] of the effects of covalency upon splitting factors by reference to MO theory in his famous 1935 paper, it is quite incorrect to view LFT as MO theory applied to transition-metal complexes. LFT and MO theory do not map onto one another. Over the past ten years, Woolley and Gerloch [4–7] resolved to uncover the underlying reasons for the successes of LFT and so to provide a defensible physical basis for the interpretation of its parameters. These interrelated aims are best reviewed separately, first in terms of a many-electron basis and then with respect to the one-electron matrix elements that define ligand-field parameters.

PROJECTION ONTO A d ORBITAL BASIS

The focus on a d or f basis is sharpened by a review of Löwdin's partitioning theory [8]. The Schrodinger equation for some full many-electron problem is written.

$$\mathscr{H}\Psi = E\Psi \tag{3}$$

Expanding the eigenvectors within a freely chosen basis $\{\Phi\}$ of infinite size,

$$\Psi_i = \sum_{k}^{\infty} c_{ik}\Phi_k; \qquad c_{ik} = \langle \Phi_k \mid \Psi_i \rangle \tag{4}$$

and defining

$$H_{kl} = \langle \Phi_k \mid \mathscr{H} \mid \Phi_l \rangle \tag{5}$$

we obtain the Heisenberg matrix representation of eq. (3):

$$\mathbf{Hc} = E\mathbf{c} \tag{6}$$

Suppose we partition the basis $\{\Phi\}$ into two groups, a and b of dimension N_a and N_b, respectively. N_b will be infinitely large, in general. The infinitely numerous eqs (6) may be partitioned similarly:

$$\mathbf{H}_{aa}\mathbf{c}_a + \mathbf{H}_{ab}\mathbf{c}_b = E\mathbf{c}_a \tag{7}$$

$$\mathbf{H}_{ba}\mathbf{c}_a + \mathbf{H}_{bb}\mathbf{c}_b = E\mathbf{c}_b \tag{8}$$

where $\mathbf{c}_a$ is a vector of dimension N_a and $\mathbf{H}_{aa}$ a square matrix of that dimension. The vector $\mathbf{c}_b$ and matrix $\mathbf{H}_{bb}$ are both of (infinite) dimension. $\mathbf{H}_{ab}$ is rectangular. Provided the inverse may be defined, we rewrite eq. (8) as

$$\mathbf{c}_b = (E \cdot \mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1}\mathbf{H}_{ba}\mathbf{c}_a \tag{9}$$

and substitute it into eq. (7) to give

$$\mathbf{H}_{aa}\mathbf{c}_a + \mathbf{H}_{ab}(E \cdot \mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1}\mathbf{H}_{ba}\mathbf{c}_a = E\mathbf{c}_a \tag{10}$$

This comprises a set of N_a equations of the form

$$\overline{H}_{aa}\mathbf{c}_a = E\mathbf{c}_a \tag{11}$$

with

$$\bar{\mathbf{H}}_{aa} = \mathbf{H}_{aa} + \mathbf{H}_{ab}(E \cdot \mathbf{1}_{bb} - \mathbf{H}_{bb})^{-1}\mathbf{H}_{ba} \quad (12)$$

where $\mathbf{1}_{bb}$ is a unit matrix of dimension $N_b \times N_b$. Solution of the secular determinantal equation,

$$|\bar{\mathbf{H}}_{aa} - E\mathbf{1}_{aa}| = 0 \quad (13)$$

yields N_a eigensolutions whose energies are identically equal to N_a eigenvalues of eq. (3) with eigenvectors expressed as (finite) combinations of the sub-basis $\{\Phi^a\}$.

The same formal manipulations may be expressed within the Schrodinger representation [4,7] by defining a projection operator P_a onto the subspace $\{\Phi^a\}$:

$$P_a = \sum_{i}^{N_a} |\Phi_i\rangle\langle\Phi_i| \quad (14)$$

together with Q_b onto the orthogonal, complementary subspace $\{\Phi^b\}$:

$$Q_b = 1 - P_a \quad (15)$$

and thence by construction of a finite-dimensional Schrödinger equation,

$$(\bar{\mathscr{H}} - E)\Phi^a = 0 \quad (16)$$

with

$$\bar{\mathscr{H}} = \mathscr{H} + \Delta\bar{\mathscr{H}}(E) \quad (17)$$

where

$$\Delta\bar{\mathscr{H}}(E) = \mathscr{H}Q_b(E \cdot Q_b - Q_b\mathscr{H}Q_b)^{-1}Q_b\mathscr{H} \quad (18)$$

We recognize, of course, that the formal manipulations that produced eqs. (16)–(18) involve no approximation of the full many-electron problem (eq. (3)) whatever. They merely project the infinitely large problem onto a finite basis $\{\Phi^a\}$ while 'folding in' all contributions from the complementary subspace $\{\Phi^b\}$ into the operator $\Delta\mathscr{H}(E)$. Furthermore, this reformulation does nothing to assist the solution of the many-electron Schrodinger equation, for the computation of $\Delta\mathscr{H}(E)$ is every bit as formidable a task as the original problem. It can, however, suggest a useful avenue for approximation.

We can, for example, make the identity

$$E_i = \frac{\langle\Psi_i|\mathscr{H}|\Psi_i\rangle}{\langle\Psi_i|\Psi_i\rangle} = \frac{\langle\Phi_i^a|\bar{\mathscr{H}}|\Phi_i^a\rangle}{\langle\Phi_i^a|\Phi_i^a\rangle} \quad (19)$$

for the ith eigenvalue. Here we recognize that such is the tactic of LFT if we take $\{\Phi^a\}$ as functions built from pure $d(f)$ orbitals and $\bar{\mathscr{H}}$ as $\mathscr{H}_{\mathrm{LF}}$:

$$E_i^d \approx \frac{\langle\Phi_i^d|\mathscr{H}_{\mathrm{LF}}|\Phi_i^d\rangle}{\langle\Phi_i^d|\Phi_i^d\rangle} \quad (20)$$

However, $\bar{\mathscr{H}}$ of eq. (18) is an energy-dependent operator so that the identities represented by eq (19) are different for each eigensolution (each i), that is $\bar{\mathscr{H}}$ in eq. (19) is different for each solution. By contrast, the procedures of LFT are such that one implicitly considers one and the same effective operator $\mathscr{H}_{\mathrm{LF}}$ throughout the manifold of d-based states that co-define the 'ligand-field regime' Were it otherwise, one would not exploit a single set of parameters (matrix elements of $\mathscr{H}_{\mathrm{LF}}$) throughout the regime And the whole point of the ligand-field parametric approach is to account for the splittings (and associated properties) of the manifold of $d(f)$ states simultaneously with one set of variables. So here is the root of one's surprise that LFT works. That it does indeed work — that one may employ some mean ligand-field Hamiltonian and thence a mean parameter set with remarkably consistent efficacy — must be attributed to Nature providing suitable and particular circumstances Their provision is not within the power of the user.

Rather similar circumstances ensure the success of π electron theory in delocalized organic systems. There, one projects the many-electron problem onto a subspace of π functions No explicit reference is made to the σ bonding framework or atomic core functions. In the manner of eq (18), these are folded into an effective, mean Hamiltonian. Matrix elements of that mean Hamiltonian are parameterized in the Huckel model by the so-called Coulomb and resonance integrals, α and β. So LFT is to transition-metal chemistry what π electron theory is to delocalized organic systems. That both models work so well in their own domains is to be ascribed to the functions of their

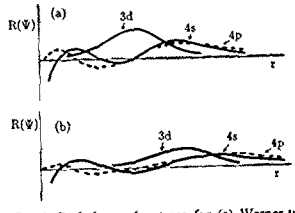


Fig. 1 Radial wavefunctions for (a) Werner-type complexes and (b) low-oxidation-state complexes, of the first transition series.

appropriate subspaces being largely uncoupled from all else.

THE CHEMICAL SIGNIFICANCE OF THE LF EFFICACY

In chemical terms, one sees that natural 'separation' of the d basis in transition-metal complexes from the complementary subspace in terms of an effective removal of the d functions from the valence shell. This is proposed strictly as a 'zeroth order' viewpoint, for some mixing with the d orbital takes places, as evidenced for example by the (small) breakdown of Laporte's rule for 'd–d' intensities. Furthermore, this separation is proposed for Werner-type complexes — those involving metals in higher oxidation states and which form suitable objects for ligand-field study — but not for carbonyl chemistry or low-oxidation state complexes. Radial forms of $3d$, $4s$ and $4p$ functions are sketched in Fig. 1 for both types of complex. The view we take here of the Werner-type systems is that, rather like the way the $4f$ orbitals in lanthanide(III) complexes are well buried and uninvolved in bonding, the d-orbitals are relatively 'inner' functions that overlap very poorly with functions offered by the ligands Chemically, this view accords well with the stability of open d shells in these systems consider, for example, the absence of free-radical behaviour of unpaired electrons in such complexes By contrast, the much greater mixing between d, s and p orbitals in the more expanded electron clouds of very low-oxidation state complexes define a valence shell with all

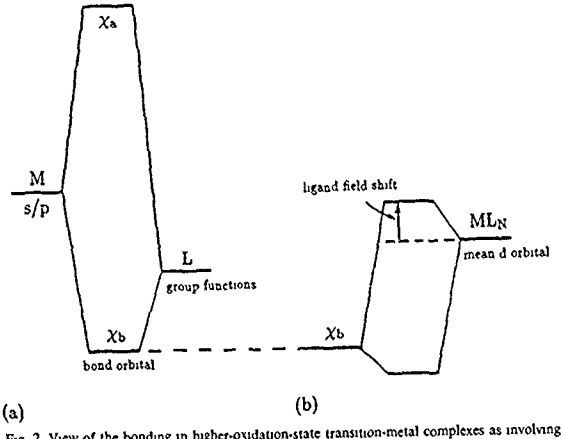


Fig. 2 View of the bonding in higher-oxidation-state transition-metal complexes as involving two notional steps (a) primary bond formation between metal and complete group of ligands (b) secondary perturbation of the mean d orbitals by bond orbitals

nine metal orbitals and chemistry dominated by the 18-electron rule.

We therefore commend a view [9] of electron interactions in Werner-type complexes as involving two notional steps In the primary step, firm bonds between metal and ligands are formed by overlap of metal s and/or p orbitals together with appropriate ligand functions. The second, or smaller, perturbation is the interaction of the d shell with the bonding functions so formed, as sketched in Fig. 2 LFT and the experimental properties it addresses are to be seen as part of this second step. Of course, the very interactive nature of this process means that while d orbitals energies and d electron distributions are affected by the bonding electrons, the bond orbitals are affected by the d electrons Fig. 2 is to be seen as the end product of such a cyclic process In this way, the exigencies of the electroneutrality principle, for example, will have been satisfied and thence probed or reflected by the effects upon the d orbitals that we analyse by LFT.

ONE-ELECTRON LIGAND-FIELD PARAMETERS

Parameters of the effective ligand-field potential are one-electron integrals. In order to gauge their chemical significance we review an attempt to forge a link between one-electron theory and the many-electron formalisms above.

One-electron theory begins with the selection of a basis The total freedom available in making this choice is not limited to the technical question of preferring hydrogenic functions to Slater-type orbitals (STOs) or to Gaussians but includes the extent to which exchange and correlation effects are included at the outset. The basis functions are defined as eigenfunctions of the one-electron Hamiltonian,

$$\mathscr{H} = T + U \tag{21}$$

where T is the usual kinetic energy Laplacian and U is some form of potential energy operator. In MO calculations, various forms of U have been adopted: in early Hartree computations U excluded all reference to exchange and correlation; in Hartree–Fock, a particular scheme for inclusion of exchange is included, in X_α calculations, a quite different approach defines a basis which includes some account of both exchange and correlation effects. Subsequent computation of many-electron molecular properties in terms of the various orbital bases require varying — and usually extremely extensive — 'corrections' to provide an acceptable account of all exchange and correlation.

In one sense, however, there exists a 'best' choice of orbital basis which, apart from trivial unitary transformations, is unique That such a choice exists is established by density functional theory [10,11], the central theorem of which shows that there exists a set of orbitals $\{\zeta\}$ for the system ground state from which one may compute the exact total electron density simple by forming the sum $\Sigma_i \zeta_i^* \zeta_i$ over populated orbitals. no further 'corrections' are required. Unfortunately, the theorem provides no practical help in calculating what these 'best orbitals' are, so the many-electron problem remains as difficult as ever. However, their existence provides the basis of a structural analysis of a model like LFT.

Let us suppose we have the form of the potential energy operator in eq (21) that leads to the 'best orbitals' for the system. it takes the form of a functional of the total electron density ρ:

$$U = U(\rho) \tag{22}$$

and, for the ground state at least, the one-electron Hamiltonian (eq. (21)) defines the solution to the given problem entirely. Now we must recall that the ligand-field procedures and eq. (2) explicitly separate d–d interactions from all others In mimicing this artificial but established structure of LFT, we define a new potential energy operator V as a functional of the total electron density minus that of the d electrons·

$$V = U(\rho - \rho_d) \tag{23}$$

That d electron density remains to be defined, cyclically, in a moment. We thus construct an orbital basis of ligand-field orbitals (LFO) as notional solutions to the one-electron Hamiltonian,

$$\mathscr{H} = T + V \tag{24}$$

The LFO is then expressed as a linear combination of fragment orbitals, rather as molecular

orbitals may be expanded in the linear combination of atomic orbitals (LCAO) system However, the fragment orbitals are chosen here in a different way We divide up V into spherical and aspherical parts, $\langle V \rangle$ and V', respectively.

$$V = \langle V \rangle + V' \tag{25}$$

Then, solutions of the mean one-electron Hamiltonian, $\mathscr{H}^{(0)}$,

$$\mathscr{H}^{(0)}\phi = (T + \langle V \rangle)\phi = \epsilon\phi \tag{26}$$

take the usual central-field form,

$$\phi = R(\tau)Y_m^l(\theta, \varphi) \tag{27}$$

so spanning a series of functions we may label as s, p, d, f.... We select the d function of the mean Hamiltonian $\mathscr{H}^{(0)}$ — which we henceforth call the mean d orbitals of the system — as one part of the fragment orbitals of the LFO, which latter are exact solutions of the hamiltonian $\mathscr{H}$ of eq. (24). So

$$\psi_{\mathrm{LFO}} = \phi_d + \phi_r \tag{28}$$

where ϕ_r represents all other functions required to span the rest of $\mathscr{H}^{(0)}$ as well as $\mathscr{H}^{(1)} = V'$, the aspherical part of $\mathscr{H}$ It is the electron density in these $\{\phi_d\}$ that is subtracted in the definition of V' in eq (23) Though notional, the procedures so far are exact. However, to make contact with the reality of LFT, we must now approximate and presume that the 'best orbitals' for all excited ligand-field states (but not for others) are somewhat similar to each other and to those of the ground state in short that the 'mean d orbitals' are also a mean throughout the ligand-field regime. Insofar as this assumption is satisfactory, LFT should 'work'; insofar as LFT works, the assumption may be deemed to be satisfactory. At this stage, notice that the precise radial form of the mean d orbitals (or, of course, the mean f orbitals if one is dealing with a lanthanide problem), though unknown to us in practice, is determined by and for the system in question In this connection recall that the radial part of the ligand-field d basis is left implicit in ligand-field procedures.

In principle we now have the basis for interpreting one-electron ligand-field parameters through the relationship

$$\langle \phi_d | V_{\mathrm{LF}} | \phi_d' \rangle = \langle \psi_{\mathrm{LFO}} | \mathscr{H} | \psi_{\mathrm{LFO}}' \rangle \tag{29}$$

However, little chemical transparency would derive from a study of this relationship, for the LFOs refer to the molecule as a whole At this point, one recognizes that one of the most powerful ideas throughout chemistry is the notion of the functional group. The power of modern ligand-field analysis is only realized when this notion is blended with the theoretical structure we have outlined above: this blend defines so-called cellular ligand-field (CLF) theory [5,6].

In the CLF model, we consider the space around the metal as divided up into N contiguous volumes or 'cells'. In general — though there is an important exception we have no space to discuss here — we arrange these cells so as to enclose one M–L ligation each We then consider the total molecular effective ligand-field potential as a simple sum of all cellular potentials Part of that supposition is the idea that the sources of the effective potential in any one cell are physically located in that cell Such is not the case in CFT, for the potential of any point charge is sensed in all regions of space. Here we presume that dielectric screening by all electrons in the bonds and cores is such as to render effective ligand-field potentials spatially local. Consider then the effects of this local effective potential upon the metal mean d orbitals in a given cell

After some simple algebra [6,7], which we do not review here, analysis of the relationship (29) within a single cell yields an expression for the energy shift of orbital d_λ as.

$$e_\lambda \sim \langle d_\lambda | \mathscr{H} | d_\lambda \rangle + \sum_\chi \frac{\langle d_\lambda | \mathscr{H}^{(1)} | \chi_\lambda \rangle \langle \chi_\lambda | \mathscr{H}^{(1)} | d_\lambda \rangle}{\epsilon_d - \epsilon_{\chi_\lambda}} \tag{30}$$

These orbital energies, $\{e_\lambda\}$, are the parameters of the CLF model. Here we write d for the ϕ_d of eq (29) and χ for functions built from the 'rest' functions ϕ_r of eq. (28). All functions are referred to the local, cellular frame and transform with symmetry λ with respect to the local pseudosymmetry. ϵ_d is the energy of the mean d orbitals and

$\bar{\epsilon}_{\chi\lambda}$ the mean, or expectation value energy of χ_λ. The first term in eq. (30) is called the 'static' contribution and the second, the 'dynamic' contribution It is sufficient for the present illustration to focus on an M–L ligation with local C_{2v} pseudosymmetry. lower local ligation symmetries have been studied in detail and reviewed [12]. In C_{2v} symmetry, $\lambda = \sigma$, π_x or π_y: δ interactions are neglected. It has been shown that for $\lambda = \sigma$, the static contribution is likely to be several times smaller than the dynamic and, for $\lambda = \pi$, that the static contribution should be negligible. Our discussion focuses, then, upon just the dynamic part of eq (30). Both total, $\mathscr{H}$ and aspherical, $\mathscr{H}^{(1)}$, parts of the Hamiltonian within the given cell transform totally symmetrically and so ensure the identical symmetry speciation of d_λ and χ_λ in eq (30) In C_{2v} symmetry, therefore, a d_σ orbital interacts with χ_σ orbitals exclusively, $d_{\pi x}$ with $\chi_{\pi x}$, and $d_{\pi y}$ with $\chi_{\pi y}$ as represented in Fig 3. In short, the local cellular potential matrix is diagonal:

$$\begin{array}{c} \\ d_\sigma \\ d_{\pi x} \\ d_{\pi y} \end{array} \begin{array}{c} \begin{array}{ccc} d_\sigma & d_{\pi x} & d_{\pi y} \end{array} \\ \begin{pmatrix} e_\sigma & 0 & 0 \\ 0 & e_{\pi x} & 0 \\ 0 & 0 & e_{\pi y} \end{pmatrix} \end{array} \tag{31}$$

with the local cellular parameters,

$$e_\lambda = \langle d_\lambda | v^c_{LF} | d_\lambda \rangle; \qquad \lambda = \sigma, \pi_x, \pi_y \tag{32}$$

where v^c_{LF} is the effective ligand-field potential in cell c. Taking eq (31) together with eq (30) and remarks above, we have

$$e_\lambda \sim \sum_\chi \frac{|\langle d_\lambda | \mathscr{H}^{(1)} | \chi_\lambda \rangle|^2}{\epsilon_d - \bar{\epsilon}_{\chi\lambda}} \tag{33}$$

and

$$v^c_{LF} \sim \sum_\chi \frac{\mathscr{H}^{(1)} | \chi_\lambda \rangle \langle \chi_\lambda | \mathscr{H}^{(1)}}{\epsilon_d - \bar{\epsilon}_{\chi_\lambda}} \tag{34}$$

Observe, in passing, how the effective ligand-field operator is energy dependent but that this is explicitly built into the ultimate parameterization. Further energy dependence, which is ignored, is implicit within the ~ sign and in the concept of mean d orbitals.

Now one can invoke the simple chemical reasoning to simplify these sums for the purpose of interpretation. Thus, we observe that the dominant contributions to e_λ in eq. (33) will be those with larger numerators and smaller denominators. $\mathscr{H}^{(1)}$ is the aspherical part of the Hamiltonian (potential) in that cell and so maximizes away from the metal core. Furthermore, it relates to the

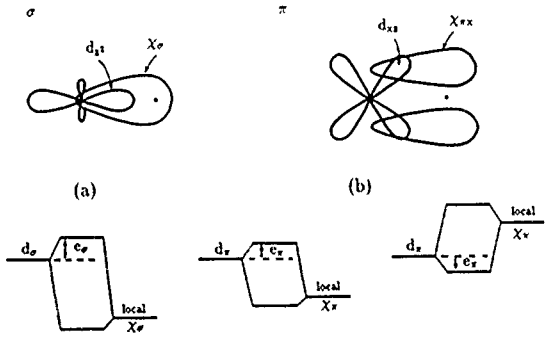


Fig. 3. Second step of Fig. 2 within the local CLF scheme. (a) for σ bonding, (b) for πx bonding (πy, in the plane normal to the paper, is similar).

electron density of the complementary set (the 'rest') and so to all occupied 'rest' orbitals. Numerators in eq. (33) will therefore be largest when χ_λ maximizes near these regions. Denominators will be smallest for χ_λ closest in energy to the mean of orbitals. All in all, we expect e_λ to be dominated by those χ which are most proximate to the d orbitals in both space and energy, that is, by the bond orbitals. We conclude that the sources of effective ligand-field potential are the bonding electrons and, in this sense, assert that LFT and observable ligand properties probe the underlying chemical bonds

It is worth emphasizing the main points and cyclic nature of the arguments summarized in this article Both the many- and one-electron constructions refer to the projection of the full many-electron problem onto a d basis. In principle, a complete description of all exchange and correlation effects are built ('folded') into the structure though in practice, of course, averaging is implicit within the process, manifested first within the mean d orbitals basis and secondly within the interpretation of the e parameters as being dominated by one or two bond functions. Subsequent rationalizations relating empirical e parameters to bond polarization or shape, atomic polarizabilities or whatever, are qualitative and must be judged by the insight they bring to the enterprise. The schemes discussed above have never been offered as routes for quantitative ab initio computation of ligand-field properties, though they could be They have the virtue, however, of making formal connections between the phenomenological ligand-field procedures of eq. (2) and accepted quantum mechanical principles and of so providing, via eq (33), a defensible basis for parameter interpretation. The whole structure, is of course, predicated on the assertion that the ligand-field method 'works'. One further aspect of the cyclic nature of our exposition is that part of the justification for that assertion is provided by the chemical consistency of the interpretations that have emerged from scores of CLF analyses

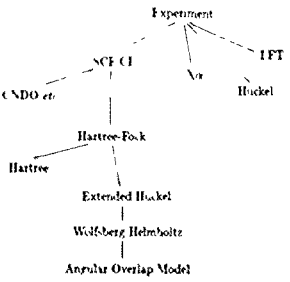


Fig. 4. Relationships between computational methods.

THE PLACE OF LFT IN COMPUTATIONAL CHEMISTRY

LFT does not have the purpose of providing a model for the computation of molecular properties in general Its domain is restricted to the spectroscopic and magnetic electronic properties of open d or f shells in transition-metal complexes of the Werner type Furthermore it is parametric Nevertheless, its underlying structure is such as to separate d (f) electron properties from all else and so to probe the chemical bonding that surely should be its central object By being excused the tasks of bonding theory it leaves to Nature the formidable tasks of accounting for the exchange and correlation effects that are so vexatious for computational chemistry at large. Bonds are formed, the electroneutrality principle is satisfied, the cut and thrust of balancing electron distribution is enacted; and LFT probes the end result. That is why LFT is so effective in reproducing experiment — far more so than even the best ab initio computational techniques — but only within its proper domain.

In Fig. 4 a tree-like scheme is represented [13] showing the relationship of one computational method with another: it is not intended to be comprehensive. It shows for example how conventional MO schemes do not map onto LFT and how the angular overlap model (a precursor to the CLF), being an MO scheme at root, is of a quite different ilk to that of the CLF.

ACKNOWLEDGEMENTS

It is a pleasure to thank Drs. Cliff Spiegelman and Larry Falvello for their invitation to participate in this conference

REFERENCES

1 E Clementi, D L Raimondi and W P Reinhardt, Atomic screening constants from SCF functions II Atoms with 37 to 68 electrons, *Journal of Chemical Physics*, 47 (1967) 1300–1307
2 J.H Van Vleck, The group relation between the Mulliken and Slater–Pauling theories of valence, *Journal of Chemical Physics*, 3 (1935) 803–806
3 J H Van Vleck, Valence strength and the magnetism of complex salts, *Journal of Chemical Physics*, 3 (1935) 807–813
4 M Gerloch, J H Harding and R G Woolley, The context and application of ligand-field theory, *Structure and Bonding*, 46 (1981) 1–46
5 R G Woolley, The angular overlap model in ligand field theory, *Molecular Physics*, 42 (1981) 703–720
6 M Gerloch and R G Woolley, The functional group in ligand-field studies the empirical and theoretical status of the angular overlap model, *Progress in Inorganic Chemistry*, 31 (1984) 371–446
7 M Gerloch, *Magnetism and Ligand Field Analysis*, Cambridge University Press, Cambridge, 1983
8 P O Löwdin, The calculation of upper and lower bounds of energy eigenvalues in perturbation theory by means of partitioning techniques, in C H Wilcox (Editor), *Perturbation Theory and its Applications in Quantum Mechanics*, Wiley, New York, 1966, pp 255–294
9 M Gerloch, The rôles of *d* orbitals in transition metal chemistry, a new emphasis, *Coordination Chemistry Review*, 99 (1990) 117–136
10 P Hohenberg and W Kohn, Inhomogeneous electron gas, *Physics Review*, B136 (1964) 864–871
11 V Heine, Electronic structure from the point of view of the local atomic environment, *Solid State Physics*, 35 (1980) 1–127
12 M.J Duer, N D Fenton and M Gerloch, Bent bonds probed by ligand-field analysis, *International Reviews of Physical Chemistry*, 9 (1990) 227–280
13 M Gerloch, The cellular ligand-field model, in J S Avery, J -P Dahl and A Hansen (Editors), *Understanding Molecular Properties*, Reidel, Dordrecht, 1987, pp 111–142

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 239–240
Elsevier Science Publishers B V., Amsterdam

Discussion of "The ligand-field regime" by M. Gerloch

Larry R Falvello

Texas A & M University, Laboratory for Molecular Structure and Bonding, College Station, TX 77843-3255 (U S A)

In these remarks I will attempt to place a perspective on the validity and the domain of applicability of the ligand field theory that Dr Gerloch has discussed.

It is easy for a working chemist to be drawn ashore by computational sirens, since many theoretical computational methods are so attractive from a distance and so easily misinterpretable as offering methodology with a hint of permanence. When stoichiometry was new, chemistry had its first 'reduce-the-entirety-of-chemistry-to-computation' tool With the discovery of quantum mechanics, the goal of computing molecular properties from first principles was *conceptually* achieved. Putting this result into practice has turned out to be a formidable task, and is today a major area of on-going chemical research. And once it has come to fruition, it will face the equally challenging requirement of reducing the complex molecular orbital descriptions to results in a paradigm useful to the working chemist.

Ligand field theory, as we know it today, is a conceptual development purely within the realm of transition-metal chemistry. It does not belong to, nor is it derived from, the molecular orbital theory Like the molecular mechanics used in organic chemistry, today's ligand field theory is based on concepts derived from a large body of knowledge within its own chemical domain.

Although one may feel a loss of satisfaction at first, in using a bonding theory not derivable from physical cosmology, the benefits of using the ligand field theory are immediately obvious and allow a fuller appreciation of the purpose of chemical theory Ligand field theory is rigorously valid within its domain The results are directly pertinent to bonding And perhaps most importantly, the theory can be used by a chemist 'in the lab'

Now, just what are the results that one obtains? The cellular ligand field theory is used to describe bonding in mononuclear transition-metal complexes. The parameters describing bonding between each ligand and the central metal, are variable, they are adjusted to produce the best agreement between the observed properties of the complex, and those calculated from the theory When a computation is finished, the user has a set of parameters describing the strengths of the various bonding interactions between ligand and metal Each parameter represents a particular component of a particular bond — for example, there will be separate parameters for the sigma and pi bonds between each ligand and the central metal (And the pi interaction can be further divided by direction, if this is appropriate)

The immediate utility of such a scheme is clear It is indeed convenient to compare one complex to another in terms of local bonding interactions Most importantly, it is possible within this regime to speak of computational results directly in terms of bonding properties. And there is a lagniappe. This sort of calculation is efficient.

I want to touch on the limitations of ligand field theory. I think that the cellular ligand field theory, although rather mature in its treatment of the first transition series, can still benefit from

0169-7439/91/$03.50

further exploration of the second and third rows of the *d*-block, and from a treatment of the *f*-series There seems still to be un-tapped potential, both for development of the theory and for better understanding of complexes of the heavier elements. It is difficult to say — even to speculate — whether the fundamental concepts underlying ligand field theory might usefully be applied to non-Wernerian inorganic chemistry It is appropriate to add at this point that the numerical algorithms used in these calculations are both mature and robust, and should not need major development unless the theory itself or its realm of applicability changes significantly.

When one considers the panorama of computational chemistry today, it is clear that the variety of the types of calculation provides one of the field's richest properties The theory that Dr. Gerloch has described is among those modern theories that provide useful bonding information to chemists Inorganic chemistry would be poorer without it — and, I believe, richer with further development of it

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 241–243
Elsevier Science Publishers B.V., Amsterdam

Discussion of "Maximum entropy as a phasing tool in macromolecular crystallography"

Larry R. Falvello

Texas A&M University, Laboratory for Molecular Structure and Bonding, College Station, TX 77843-3255 (U.S.A.)

The algorithm that Dr Prince has described once again opens the possibility that large-molecule structure determinations will one day be done with something approaching the facility now enjoyed only by small-molecule diffractionists.

It has been true until quite recently that the major practical and theoretical advances in the science behind crystallography have been applied easily and naturally to the purpose of facilitating small structure determinations, while macromolecular crystallography has received less benefit. The means of solving structures via Patterson synthesis [1,2], the discovery and development of the direct methods [3,4], and the invention of the four-circle diffractometer [5] have all had far greater facilitating influences on small molecule science than on large. The maximum-entropy methods may come to be an important facilitating influence in macromolecular work.

In putting a context around the maximum entropy method as a phasing tool, it is worthwhile to examine the phasing tools used in small-molecule work, as they would be viewed in importance by a practitioner in the field. (Professor Hauptman has described the solution of the theoretical problem of determining phases from a set of amplitudes [6]. It is interesting to see that practical and theoretical advances can follow different, though related, courses.) Before the advent of the direct methods, one could attempt to determine phases by model building, or by application of the Patterson function, a self-convolution of the structure which can be calculated in a phaseless transformation. These methods, as viewed by today's practitioner, rely on one of more of the following (1) a non-uniform distribution of electron density; (2) the presence of useful symmetry elements, and (3) a priori chemical knowledge of the contents of the asymmetric unit. In practice, these methods often depend on the skill and experience of the practitioner.

The phase problem was solved in principle (for large and small systems) with the discovery of the Hauptman–Karle determinants, the non-negativity of which is a necessary consequence of the non-negativity and atomicity of electron density within a crystal. Of course, solving the problem in practice was another matter. The determinants, in their most general form, simply were computationally too difficult at the time of their discovery to yield a closed form numerical solution for a given crystal structure. They did, however, yield the means for achieving a practical solution to the phase problem.

The third order determinant, D_3 (eq. 1), yields an expression on the basis of which certain values of the combination $(\phi_{-h} + \phi_k + \phi_{h-k})$ can be ruled out if the amplitudes are large enough. (In the case of a centrosymmetric crystal the phases ϕ are restricted to values of zero and π, and the theory develops slightly differently.) However, even when the three-phase combination cannot be de-

0169-7439/91/$03.50

termined with certainty, one can still apply probability theory to establish an expected distribution for its value [7–10].

$$D_3 = \begin{bmatrix} 1 & U_h & U_k \\ U_{-h} & 1 & U_{k-h} \\ U_{-k} & U_{h-k} & 1 \end{bmatrix} \geq 0 \qquad (1)$$

The application of probability theory thus becomes an important area of work in the phase problem. The result of prime importance for practical application was the tangent formula (eq. 2), which gives an indication for a phase of a reflection h in terms of the phases and amplitudes of other reflections which can participate with h in third-order Hauptman–Karle determinants. The tangent formula is used in conjunction with its variance [11], from which inferences are drawn about the reliability of the indicated phase

$$\phi_h = \tan^{-1}\left[\frac{\sum_k |E_k||E_{h-k}|\sin(\phi_k + \phi_{h-k})}{\sum_k |E_k||E_{h-k}|\cos(\phi_k + \phi_{h-k})}\right] \qquad (2)$$

The tangent formula served as the launch pad for the next important practical developments — the multiple tangent method [12] and the popular computer program (MULTAN) employing it [13] This was the development which finally allowed a rapid growth in the number of laboratories conducting crystal structure analyses, and the concomitant growth in the importance of crystallography to chemists Further refinements in methodology and more efficient algorithms and programs [14,15] led to further rustication of X-ray structure determination, as the esoteric aspects of the phase problem became buried in packaged protocols.

Meanwhile, the probability theory that allowed the direct methods to stimulate the flowering of small-molecule diffraction work, proved initially to be its undoing in large-molecule work, since the reliability of a phase indication changes inversely with the square root of the number of atoms in the cell. So macromolecular diffractionists have not been able to share fully in the practical benefits of the solution of the mathematical phase problem. Rather, the labor-intensive multiple isomorphous replacement method has remained a workhorse for protein structure determination.

Now, where does the principle of maximum entropy fit in with all of this? The important conceptual property of maximum entropy is that, like the Hauptman–Karle determinants, it is consistent with the analysis of data arising from a non-negative electron density distribution [16,17] The principle of maximum entropy has also had a practical problem in common with the determinantal equations — useful, widely applicable numerical algorithms for diffraction analysis have not appeared as obvious consequences of theory. Thus, the use of the dual method that Dr Prince has described here and elsewhere [18] is a practical development which has merited a thorough test. The examples we have seen today show the usefulness of the algorithm While the clarification of noisy electron density maps is valuable and itself would justify full exploration of the method, it is in the a priori determination of phases that I believe the maximum entropy method can be most profoundly exploited.

The maximum entropy method has its roots in probability theory, as Jaynes has explained in detail [19] The modern developments by Shannon [20] (for information theory) and Jaynes represent the climax of a long conceptual development. While closely tied to probability theory, the maximum entropy method, in its most basic notion, formalizes prior ignorance of a system and allows experimental data as constraints It does not employ conditional probability distributions, and apparently does not suffer from a loss of efficacy with increasing size of the problem. Considering all of this it is natural to regard the maximum entropy method as a logical and potentially powerful extension of the direct methods with promise for macromolecular diffraction studies

REFERENCES

1 A L. Patterson, A direct method for the determination of the components of interatomic distances in crystals, *Zeitschrift fur Kristallographie*, 90 (1935) 517–542
2 D. Harker, The application of the three-dimensional Patterson method and the crystal structures of proustite, Ag_3AsS_3, and pyragynte, Ag_3SbS_3, *Journal of Chemical Physics*, 4 (1936) 381–390.
3 J. Karle and H Hauptman, The phases and magnitudes of

the structure factors, *Acta Crystallographica*, 3 (1950) 181–187.
4 M M Woolfson, Direct methods — from birth to maturity, *Acta Crystallographica*, A43 (1987) 593–612.
5 T C Furnas and D Harker, Apparatus for measuring complete single-crystal X-ray diffraction data by means of a Geiger-counter diffractometer, *Review of Scientific Instruments*, 26 (1955) 449–453
6 H A Hauptman, History of X-ray crystallography, *Chemometrics and Intelligent Laboratory Systems*, 10 (1991) 13–18
7 W Cochran and M M Woolfson, The theory of sign relations between structure factors, *Acta Crystallographica*, 8 (1955) 1–12
8 W Cochran, Relations between the phases of structure factors, *Acta Crystallographica*, 8 (1955) 473–478
9 H A Hauptman and J. Karle, *Solution of the Phase Problem. I The Centrosymmetric Crystal*, A C A Monograph Number 3, American Crystallographic Association, Ann Arbor, MI, 1953
10 E.F Bertaut, La méthode statistique en cristallographie I, *Acta Crystallographica*, 8 (1955) 537–543
11 J Karle and I.L Karle, The symbolic addition procedure for phase determination for centrosymmetric and noncentrosymmetric crystals, *Acta Crystallographica*, 21 (1966) 849–859
12 G Germain and M M Woolfson, On the application of phase relationships to complex structures, *Acta Crystallographica*, B24 (1968) 91–96
13 G Germain, P Main and M M Woolfson, The application of phase relationships to complex structures III The optimum use of phase relationships, *Acta Crystallographica*, A27 (1971) 368–376
14 G M Sheldrick, *SHELX-76 – Program for crystal structure determination*, Cambridge University, 1970
15 G M. Sheldrick, *SHELXS-86 – FORTRAN-77 program for the solution of crystal structures from diffraction data*, Göttingen University, 1986
16 D M Collins and M Mahar, Electron density an exponential model, *Acta Crystallographica*, A39 (1983) 252–256
17 G Bricogne, Maximum entropy and the foundations of direct methods, *Acta Crystallographica*, A40 (1984) 410–445
18 E. Prince, The maximum entropy distribution consistent with observed structure amplitudes, *Acta Crystallographica*, A45 (1989) 200–203
19 E.T Jaynes, Where do we stand on maximum entropy?, in R.D Rosenkrantz (Editor), *E T Jaynes Papers on Probability, Statistics and Statistical Physics*, Reidel, Dordrecht, 1983, pp 210–314.
20 C E Shannon, The mathematical theory of communication, *Bell System Technical Journal*, (1948) 379–423, 623–656

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 245–257
Elsevier Science Publishers B.V., Amsterdam

Determining chemical characteristics of plutonium solutions using visible spectrometry and multivariate chemometric methods

W. Patrick Carey and Lawrence E. Wangen *
Chemical and Laser Sciences Division, G740, Los Alamos National Laboratory, Los Alamos, NM 87545 (U S A)

(Received 10 November 1989; accepted 1 September 1990)

Abstract

Carey, W P and Wangen, L E., 1991 Determining chemical characteristics of plutonium solutions using visible spectrometry and multivariate chemometric methods *Chemometrics and Intelligent Laboratory Systems*, 10 245–257

Two chemometric analysis approaches for rapidly screening samples are presented The first method is for determining Pu(III) and nitric acid concentrations by using the multivariate calibration technique of partial least squares (PLS) regression Quantitation of plutonium using its visible spectrum is straightforward, however, the effects of nitric acid on the Pu(III) absorption spectra are subtle, and nitric acid quantitation from the absorbance spectrum is more difficult In this study PLS regression is successfully applied to quantitate both plutonium and nitric acid by using the information contained in the absorption spectra of appropriate solutions Evaluation of the calibration models, using test samples that span the range of the calibration concentrations, gave predictions consistent with the standard error of the calibration models.

Secondly, pattern recognition methods are used to investigate the effects of various amounts of nitric acid, fluoride, or oxalate on visible spectra of Pu(IV) solutions. The methods enable qualitative estimates of the solution composition, which can potentially be used to adjust solution properties to desired specifications. The main pattern recognition methods employed are nearest neighbor classification and principal components analysis

DETERMINATION OF Pu(III) AND NITRIC ACID

Plutonium can be precipitated from nitric acid solutions by forming an insoluble oxalate salt of Pu(III). However, the concentrations of both total nitric acid ($CHNO_3$) and oxalic acid affect the solubility of the Pu(III) oxalate product [1,2]. Pu(III) oxalate solubility is at a minimum between 0.5 to 1.0 *M* nitric acid and with a 0.05 to 0.1 *M* stoichiometric excess of oxalic acid. At these concentrations the solubility of Pu(III) ranges between 2 and 20 mg/l. At higher nitric acid concentrations, the solubility of Pu((III) increases; for example in 2.0 *M* nitric acid, the Pu(III) concentration increases tenfold. There are also indications that increasing the oxalic acid concentration above 0.2 *M* will lead to increased solubility of the plutonium. To assist in optimizing solution conditions for the precipitation reaction of Pu(III) oxalate, it would be beneficial to have a rapid analytical method for estimating the concentrations of plutonium and nitric acid.

In this study we evaluated a method based on partial least squares (PLS) regression for predicting both Pu(III) and nitric acid concentrations using the visible absorption spectra of solutions containing the species of interest Several techniques based on visible absorption spectroscopy have been developed for estimating Pu(III), and quantitation is fairly straightforward [3–6] However, determination of the nitric acid concentration from the visible absorption spectra is more difficult because of the subtle effects of nitric acid on the spectrum. In this paper we demonstrate the use of PLS for extracting the small signal of the nitric acid effect in the presence of a much larger signal caused by the Pu(III) absorption. This information provides a measure of nitric acid concentration that can be used in studying the precipitation reaction.

The fundamental theory and applications of PLS have been investigated by several researchers [7–11] PLS uses a large part or all of the spectral data points to develop linear combinations of the spectral absorbances that correlate with the analyte concentration vector. The PLS regression procedure is based on an algorithm in which the scores are orthogonal. This method is similar to principal component regression in that the spectral response matrix is factor analyzed into orthogonal vectors based on the variance. However, it includes information from the analyte concentration vector in the matrix decomposition procedures The model built by the PLS algorithm between the spectral and concentration variables during calibration is different for each analyte in so far as their effects on the spectra are different. Two separate PLS models were developed, one each for Pu(III) and nitric acid Using the models developed during calibration, we predicted analyte concentrations in several solutions not used in calibration.

Experimental

All chemicals were reagent grade, except for the plutonium nitrate stock solutions. Plutonium nitrate stock solutions were obtained by dissolving PuO_2 in $CHNO_3$/HF, followed by fluoride removal using ion exchange. The concentrations of these stock solutions were determined by standard radiochemical methods based on gamma-ray spectroscopy with a relative standard deviation of 0.5% [12]. We prepared a 25-sample calibration set and a 6-sample test set by performing volumetric dilutions of the stock solutions and adjusting nitric acid concentrations These solutions were prepared to cover the acid range Nitric acid concentrations were determined by a standard addition method [13].

We recorded spectra between 500 and 880 nm on each sample using a 0.2 cm path length flow cell. The spectrometer for these experiments was an LT Industries Quantum 1200. This instrument allows for the remote placement of sample cell and detector in an isolated glove box, with a fiber-optic bundle transporting the light between source, sample, and detector. The resolution obtained with this instrument is on the order of 1 nm with the scan for the visible region requiring 200 ms. For each sample, ten 200-ms scans were acquired and averaged.

Data analysis was performed using a PLS program developed at the University of Washington [14]. This code was implemented on a VAX 11-780.

Results

Visible spectra of the plutonium species appear in Figs. 1 and 2. Fig. 1 shows the sensitivity of several Pu(III) absorption bands in solutions containing 2.0 to 29.9 g/l of Pu(III). The nitric acid concentration in these four samples was approximately 1.3 *M*. In high-precision analytical measurements, the bands at 565 and 601 nm are commonly used to quantitate Pu(III) after adjustment of solution conditions. The effect of varying nitric acid concentration on these spectra is illustrated in Fig 2 where Pu(III) was held constant (6.0 g/l) and nitric acid was varied from 0.6 to 2.3 *M*. This effect is most readily observed at 565 nm, where the absorption peak tends to narrow or become more symmetrical with increasing nitric acid concentration, and between 750 and 825 nm, where a change in one or more underlying absorbance bands causes small changes in the spectra.

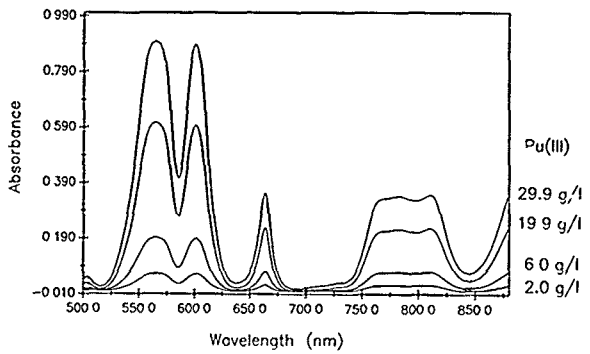


Fig 1 Absorbance spectra of Pu(III) from 2 0 to 29 9 g/l

Using the 25-sample calibration, separate PLS models were built for Pu(III) and nitric acid. All variables were mean centered and scaled by their standard deviation before the model was built. For both models the number of component vectors to use was determined by cross validation (alternating one-sample-removed method), and the final models included all 25 samples. Table 1 shows the percentage variance explained for these calibration samples by the PLS model for both Pu(III) and nitric acid and the spectra The first component explains 94 35% of the variance in the spectral responses Evidently Pu(III) changes are the cause of this because 98 80% of its variance is

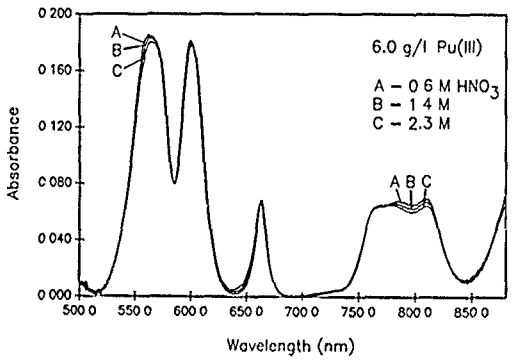


Fig. 2 Effect of nitric acid on Pu(III) absorbance spectra Nitric acid varies from 0 6 to 2.3 *M* with a constant 6 0 g/l Pu(III) concentration

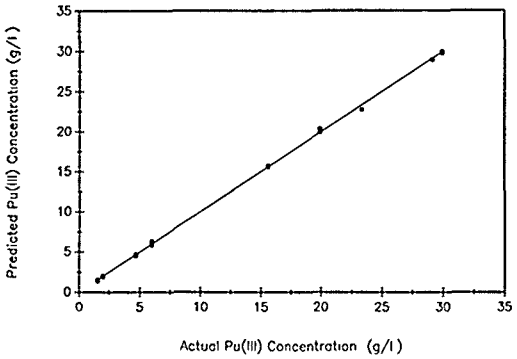


Fig 3 Actual Pu(III) concentration versus predicted Pu(III) concentration based on a two-latent-variable PLS model

explained by this component. This is as expected on the basis of Fig. 1. Nitric acid, however, has only 5 78% of its variance described by the first PLS component. For nitric acid more of the nitric acid variation is explained by components that explain lower amounts of spectral variance. Because very little of total spectral variance is used to model nitric acid molarity, we expect poorer results.

The accuracy of a multivariate model can be visually examined by plotting the actual calibration concentrations versus the predicted values for each sample. For Pu(III) the 25 sample concentrations are plotted versus their estimated concentra-

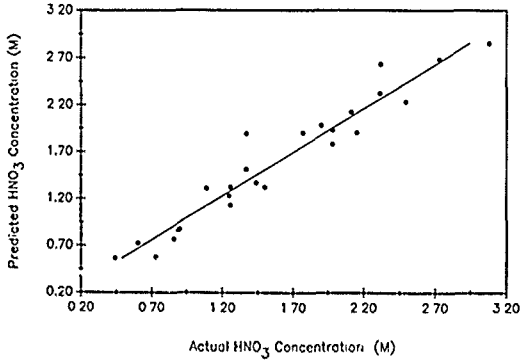


Fig. 4. Actual nitric acid concentration versus predicted nitric acid concentration from a six-latent-variable PLS model.

TABLE 1

Variance described by PLS models for Pu(III) and nitric acid

Latent variable	Spectral Response		Pu(III)		Nitric acid	
	Each (%)	Total (%)	Each (%)	Total (%)	Each (%)	Total (%)
1	94 53	94 53	98 80	98 80		
2	3 61	98 14	1.16	99 96		
1	94 35	94 35			5.78	5 78
2	1 81	96 17			29 55	35 34
3	3 51	99 68			1 17	36 51
4	0 14	99 82			28 12	64 63
5	0 05	99 87			16 91	81 55
6	0 02	99 90			11 52	93 07

tions using a two-latent-variable model shown in Fig 3. As expected, Pu(III) is well modeled with an r^2 statistic of 1.00 and a standard error of 0.20 g/l Fig 4 provides a similar plot of measured versus predicted concentrations for nitric acid using a six-latent-variable model. In this case the model describes the overall nitric acid effect on the spectra but with a greater degree of error than the Pu(III) model. The r^2 statistic for the nitric acid model was 0 93 with a standard error of 0.18 *M*.

A better measure of the validity of the calibration models is to examine that predictive capability using samples not included in the calibration sample set. To validate the constructed models, we analyzed a test set containing six samples with known Pu(III) and nitric acid concentrations in the same manner as the calibration set samples These samples were prepared using the same techniques as for the calibration samples. Table 2 compares the resulting predictions with known values. The calibration model is validated if the predicted values of unknowns are within the standard error range of the model, which is a calculation of the standard deviation of the model residuals. For example, approximately 95% of future samples should fall within twice the standard error if the unknowns come from the same population as the standards For Pu(III) with a standard error of 0.20 g/l, all of the predictions were within two standard errors, with four of the six predictions within one standard error. For nitric acid all predicted values are within the two standard error limit (0.18 *M* $CHNO_3$) estimated by the model, and half of these samples are within one standard error. The estimated standard errors of prediction were 0.25 g/l and 0 23 *M* for Pu(III) and nitric acid respectively, which is slightly greater than that of the calibration set for both analytes Although the number of samples was limited in both calibration and test sets, there was no statistical difference between the standard errors based on an *F*-test comparison The results of this test set provide confidence that both the Pu(III) and nitric acid models are valid over the range of concentrations normally encountered in the plutonium oxalate precipitation studies

We have demonstrated the use of the Pu(III)–nitric acid absorbance spectra coupled with PLS regression for the determination of Pu(III) and nitric acid concentrations over the analyte ranges of 1.99 to 29.9 g/l plutonium and 0 44 *M* and

TABLE 2

Prediction results for test set samples

Sample	Pu(III) (g/l)			Nitric acid (*M*)		
	True	Estimated	Difference	True	Estimated	Difference
1	1.99	2 00	0 01	1 98	1 96	0 02
2	5.97	5.99	0 02	1.15	1 47	0 32
3	29 9	30 3	0.4	1 07	0 92	0 15
4	19.9	19 7	0 2	2.13	2 47	0 34
5	4 67	4 62	0 05	2 08	1 97	0 11
6	15 6	15 2	0.4	0 94	1 16	0 22
Standard error of prediction			0 25			0 23

3.08 *M* nitric acid The precision of these predictions is suitable for studying the effects of oxalic acid and nitric acid concentrations during the precipitation of plutonium oxalate. Although greater precision could be obtained using other more complex methods, the information gained from these spectral measurements is adequate for real-time analyses The coupling of multivariate regression techniques with absorbance spectroscopy provides quantitation of both Pu(III) and nitric acid from a single, easy-to-perform spectral measurement, thereby simplifying the instrumentation used in studying the precipitation reaction.

QUALITATIVE DETERMINATION OF Pu(IV) COMPLEX COMPOSITION

The Vis–NIR absorption spectra of Pu(IV) in nitric acid have several intense bands [15] The number, position, and intensity of these bands depend on the total nitric acid ($CHNO_3$) molarity and the plutonium oxidation state The spectra may also be influenced by the presence of other cations and anions Thus, it was hypothesized that Vis–NIR absorption spectroscopy could provide information important for the chemical characterization of acidic plutonium solutions Such information could be used to chemically adjust such solutions before their treatment by ion exchange. This study was designed to determine the effect of fluoride and oxalate on the chemistry of Pu(IV)–nitric acid solutions as evidenced by changes in their spectra. Fluoride and oxalate complexes of plutonium do not adsorb to the ion exchange resins being used in this study.

The research questions posed were

- How many different absorbing species are present in the plutonium solutions ranging from 4 *M* to 10 *M* $CHNO_3$ in the presence of either fluoride or oxalate?
- What spectral changes result from the addition of fluoride or oxalate to Pu(IV)–nitric acid solution?
- Can the distribution ratios (R_ds) and initial concentrations of nitric acid, plutonium, fluoride, and oxalate be predicted from the Vis–NIR spectra of the solutions?
- Can we develop a classification procedure using Vis–NIR spectra that will separate good solutions from bad ones with respect to ion exchange behavior (as defined by R_ds)?

Experimental

Solutions and spectroscopy

The data sets used in this study consisted of spectra collected from two different experiments, which were identical except for the substitution of oxalate for fluoride in the second experiment. The solutions used are described in Table 3. Nitric acid molarities ranged from 4 0 to 10 0. Fluoride and oxalate concentrations ranged from $8\,37 \times 10^{-3}$ *M* to $3\,35 \times 10^{-2}$ *M* plus a zero value For all fluoride and oxalate concentrations, two different concentrations of Pu(IV), 8.37×10^{-3} *M* and 4.18×10^{-2} *M*, were used The spectra from solutions containing no fluoride or oxalate are common to both data sets.

All the spectra were recorded after sufficient time for the solutions to equilibrate with a Quantum 1200 Vis–NIR spectrometer from L T Industries The wavelength region recorded was from 400 to 880 nm in 0 4-nm increments The solutions were contacted with anion exchange resin (40–70 mesh Lewatit MP-500-FK) after their spectra were recorded. The R_d values were calculated by using initial and final plutonium concentrations for the fluoride data. The R_d analyses are not presented for oxalate data.

Data reduction, analysis, and interpretation

Preprocessing the spectral data consisted of several steps that were not always performed, de-

TABLE 3

Composition of solutions used for effect of fluoride or oxalate on spectra of Pu(IV)–nitric acid solutions *

Nitric acid	4 *M*, 5 *M*, 6 *M*, 7 *M*, 8 *M*, 9 *M*, 10 *M*
Plutonium	$8\,37 \times 10^{-3}$ *M*, $4\,18 \times 10^{-2}$ *M*
Fluoride or oxalate	0 00, $8\,37 \times 10^{-2}$ *M*, $1\,67 \times 10^{-2}$ *M*, $2\,51 \times 10^{-2}$ *M*, $3\,35 \times 10^{-2}$ *M*

* At each combination of nitric acid molarity and plutonium concentration, solutions containing either fluoride or oxalate at the indicated concentrations were prepared

pending on the particular study objectives. To reduce the number of variables that the computer programs must handle, all the spectra were reduced from 1200 to 600 absorbance values per spectrum by performing a two-point average of successive absorbance values. Occasional baseline shifts were corrected by a simple baseline subtraction method. For each spectrum this involved determining the minimum absorbance value, a_k, in that spectrum; computing the average of a_{k-1}, a_k, and a_{k+1}; and subtracting this average from every absorbance value in each spectrum. More sophisticated methods of baseline correction for these spectra would be difficult to implement because the spectra are so complex. Absorbance values approached baseline in only one or two spectral intervals. To adjust for different concentrations of plutonium in different data sets, we normalized the spectra to a sum of 1.0, $a_k^* = a_k/(\text{Sum } a_k)$, $k = 1$ to 600 for each spectrum. However, this normalization is not done when the best model for predicting plutonium concentrations is desired.

Data analysis methods consisted mainly of variations of the mathematical-statistical procedures most commonly referred to as principal components analysis. All of these methods involve decomposition and analysis of a spectral data matrix whose individual rows consist of the Vis–NIR spectrum of one of the experimental solutions under study. The specific methods used were pattern recognition based on principal components modeling (SIMCA) [16], pattern recognition based on nearest neighbor classification [17], pattern recognition based on other methods contained in the ADAPT package [18], and principal components regression [19,20].

Results

For each data set, there are 70 spectra corresponding to seven $CHNO_3$ molarities, five fluoride or oxalate concentrations, and two plutonium concentrations ($2 \times 5 \times 7 = 70$). Thus, we have a large number of spectra that are quite complex and that vary considerably with changing concentrations. Fig. 5 demonstrates this complexity and the changes caused by fluoride at 8 M $CHNO_3$ for a 8.37×10^{-3} M plutonium solution. The highest fluoride concentration is a 4:1 fluoride-to-plutonium molar ratio. The peaks with 0.0 M fluoride at 420 and 850 nm are absent in the high-fluoride spectrum. There are numerous changes in relative peak heights. The band at 475

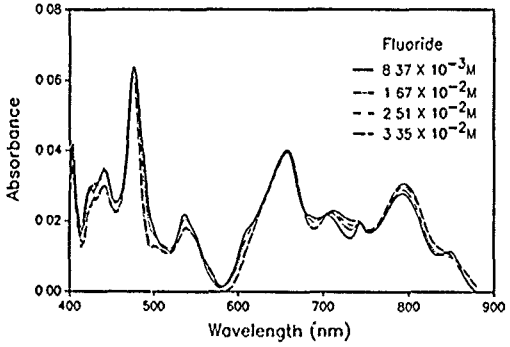


Fig. 5. Spectra of 8.37×10^{-3} M Pu(IV) with fluoride.

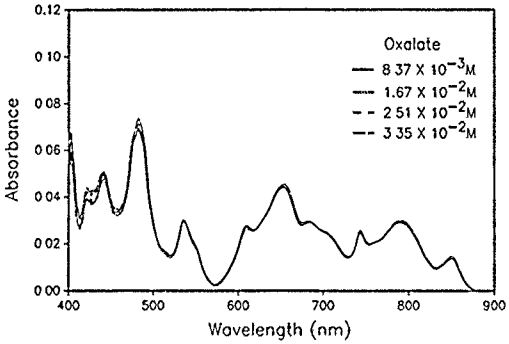


Fig. 6 Spectra of 8.37×10^{-3} *M* Pu(IV) with oxalate

nm is less intense in the high-fluoride spectrum. Oxalate does not have as great an effect on the spectra of Pu(IV)–nitric acid solutions as does fluoride (Fig. 6).

Number of absorbing species

Matrix rank determination has become a fairly common procedure in spectroscopy for estimating the number of absorbing species in a series of mixtures [21]. This procedure is valid provided Beer's model is obeyed, that is, if the total absorbance is a linear summation of the absorbances of the individual species. The major difficulty with the procedure is determining the chemically meaningful rank. Because of noise, the mathematical rank will usually be the lesser of I and K for a data set composed of I spectra. The ith row of the matrix contains the spectrum for the ith mixture, and K is the number of wavelengths at which there are absorbance values. Various methods for determining the number of absorbing species have been proposed. In this paper, we will discuss only the method based on cross validation. The spectral data matrix used for this analysis consisted of either the fluoride or oxalate spectral data set. In each case, there are 70 spectra with 600 absorbance values, i.e., I by K = 70 by 600.

Cross validation

The cross validation for principal components analysis contained in the set of pattern recognition computer programs, SIMCA, was used for the present problem. SIMCA's program module CPRIN was used with the cross validation option. In cross validation, a subset of the data is excluded from the data set. Then a model is developed, and the excluded data values are estimated (predicted) by using the model. The sum of the squared differences between each true value and each predicted value is the predicted residual error sum of squares (PRESS). Next, the excluded data subset is returned to the modeled data set, and a different subset of the data is excluded. Again, a model is developed and used to predict the excluded subset. This process continues until all data have been excluded and predicted one time for each value of J (number of components). If, after allowing for degrees of freedom, PRESS continues to decrease upon addition of component J, component J is assumed to model nonrandom variation in the data. However, if PRESS for component J is greater than PRESS for component $J - 1$, component J is assumed to be modeling only random noise in the data. In this case component J should not be used, and we assume

TABLE 4

Cross validation results for determining the number of linear independent components in the fluoride and oxalate spectral data matrices

J	Fluoride		Oxalate	
	Variance explained	PRESS * $J/(J-1)$	Variance explained	PRESS * $J/(J-1)$
1	75 15	0 50	90 59	0 31
2	21 75	0 36	6 67	0.55
3	1 42	0 75	1 48	0 69
4	0 49	0 86	0 53	0 78
5	0 42	0 83	0.30	0 79
6	0 26	0 86	0 11	0 88
7	0 14	0 87	0 05	0.97
8	0 06	0 99	0 04	0 96
9	0 05	0 96	0 03 **	1 00
10	0 03 **	1 00	0 02	1 00

* For $J=1$, PRESS for $J-1$ is based on the variance explained by using the average values.

** A strict interpretation of cross validation results shows that there are nine and eight components in the fluoride and oxalate data sets

there are $J-1$ linearly independent components in the entire data set. If the spectra of the individual chemical species add linearly, i.e. if Beer's model is obeyed, this number is the same as that of absorbing species in the solutions from which the spectra were obtained.

The data variables were not scaled. Two different SIMCA runs were made, one for the fluoride and one for the oxalate data set with each spectrum normalized to a sum of 1.0. The results of these two analyses are listed in Table 4 in terms of the ratio of PRESS for J components to the PRESS for $J-1$ components The variance explained by each component is also tabulated. These PRESS ratios indicate nine components for the fluoride spectra and eight components for the oxalate spectral data set. In the absence of fluoride or oxalate, studies indicated five or six components. Thus, the addition of fluoride or oxalate to solutions of Pu(IV)-nitric acid (4 M-10 M) add about three or four observable components.

In this study we applied SIMCA, nearest neighbor, Bayes quadratic classifier, and the linear learning machine from ADAPT [18] to investigate their usefulness for classifying the fluoride or oxalate Pu(IV)-nitric acid solutions. For the ADAPT analysis, the R_d values were used to divide the fluoride spectral data set into 'good' and 'bad' categories For the SIMCA pattern recognition approach, data were not divided into separate categories before analysis because it is possible to visually see the separation when plotting certain of the sample scores.

ADAPT results on fluoride spectra

The classification results appear in Table 5. The input data to these pattern recognition methods consisted of the principal component scores of the spectra rather than the spectra themselves All the methods were able to separate spectra representing good and bad R_d values reasonably well. The linear learning machine correctly categorized all 70 spectra, and the Bayes quadratic classifier only missed 1 out of 70. The nearest neighbor results vary a little depending on the number of voting neighbors. Apparently three, five, or seven voting neighbors give equivalent results, but none are as good as the Bayes or learning machine methods.

SIMCA Plots for fluoride and oxalate spectra

We developed a six-component model using SIMCA and the principal components of Vis-NIR spectra obtained from 39 solutions. The 39 solutions contained only nitric acid ranging from about 1 M to 14 M and Pu(IV). No fluoride or oxalate

TABLE 5

Pattern recognition summary results for fluoride spectra using the Bayes quadratic classifier, linear learning machine, and nearest neighbor algorithm in ADAPT

	Good samples (High R_d)		Bad samples (Low R_d)		No. of neighbors
	No correct	No incorrect	No correct	No incorrect	
Bayes	26	3	43	1	
Learning machine	26	0	44	0	
Nearest neighbor	22	4	36	8	1
	23	3	39	5	3
	24	2	38	6	5
	23	3	39	5	7

was in the data set. (SIMCA mean centers and autoscales the spectral intensities before calculating the principal components.) Here we use the first two principal components derived by this model to compare how the spectra of fluoride and oxalate data plot as compared with Pu(IV)–nitric acid.

The scores of the first two components for the 39 training samples are plotted in Fig 7, which shows a nice semicircular trend of increasing $CHNO_3$ molarity from the top left to the middle right of the graph The numbers in the figure with an appended H designate total nitric acid molarity $CHNO_3$. The numbers with a prefixed T were all between 6 5 and 8.5 *M* nitric acid, with nitric acid molarity increasing from left to right. The desirable samples, from an ion exchange perspective, plot at the bottom of the figure as 7H Clearly, given the location of a solution containing only Pu(IV) and nitric acid on this figure, the ap-

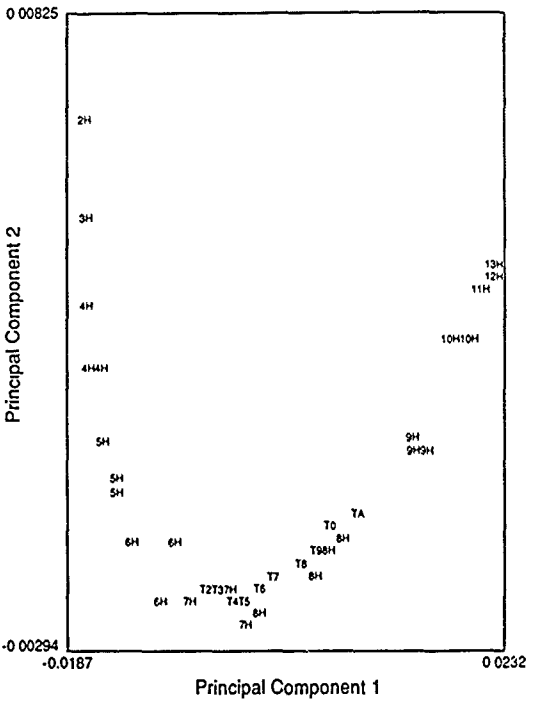


Fig. 7. Plot of first two principal components of 39 Pu(IV)–nitric acid samples making up the training set

proximate quantity of acid or base to add for adjusting the solution chemistry in a desired direction could be specified.

This same principal components model was used to calculate scores for all samples of the fluoride and oxalate data sets. The scores of the first two principal components are plotted together with those of some of the training samples in Figs. 8 and 9 for fluoride and oxalate samples, respectively. (Training samples are in bold print.) Fig. 8 shows the fluoride spectra plot in the plane above the semicircle defined by the training set. Again, the numbers refer to nitric acid molarity and the Fs to fluoride samples. For a constant nitric acid molarity, greater fluoride-to-plutonium concentration ratios plot higher in the graph. If aluminum were added to complex the fluoride in an unknown solution that plotted in the middle of Fig. 8, its position in this graph would move down and to the right. Upon arriving at the semicircle representing the training set, a base, such as sodium hydroxide, could be added to the solution

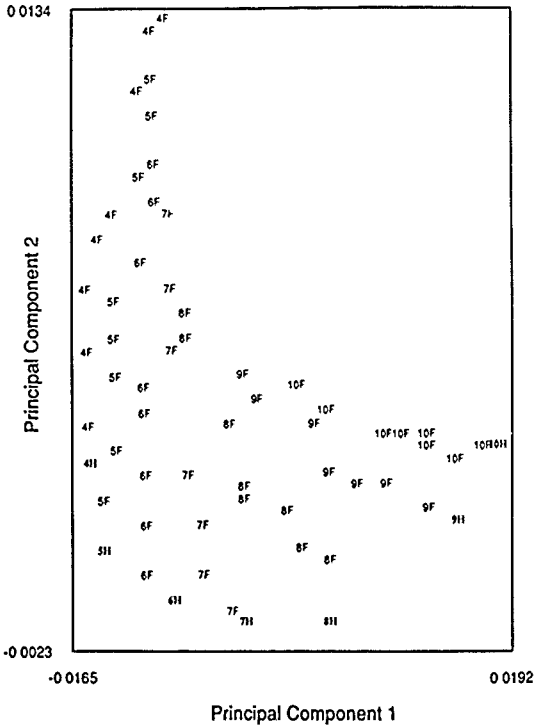


Fig. 8 Plot of first two principal components for fluoride samples and some Pu(IV)–nitric acid training samples

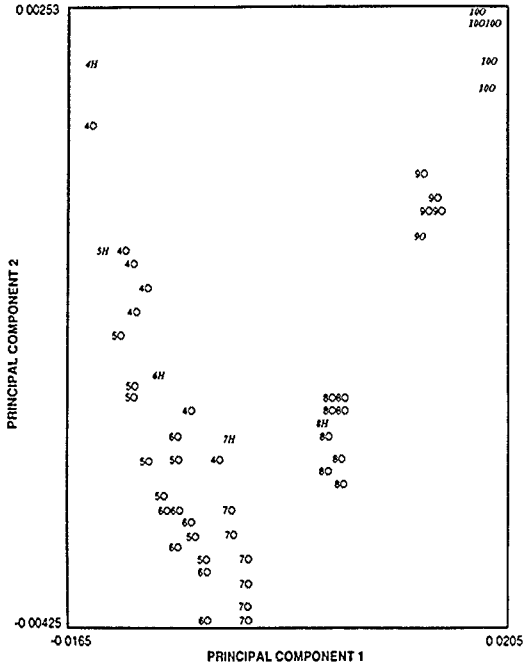


Fig 9 Plot of first two principal components for oxalate samples and some Pu(IV)–nitric acid training samples

until the measured spectrum's principal components plotted between 7 *M* and 8 *M* $CHNO_3$, i.e., 7H and 8H. In this way, Vis–NIR spectrometry could be used for real-time adjustment of the solution chemistry to arrive at a system desirable for ion exchange.

The first two principal component scores of the oxalate spectra, together with those of some of the training samples, are plotted in Fig. 9. In contrast to the fluoride samples, these samples generally plot on or below the semicircle defined by the training samples. This plot verifies our earlier observation that oxalate does not have as much effect on the Pu(IV)–nitric acid spectra as does fluoride (Fig 6), which agrees with the known chemistry of these systems

SUMMARY

We have shown that Vis–NIR spectra of Pu(IV)–nitric acid solutions containing either fluoride or oxalate provide information concerning the solution chemistry Pattern recognition methods based on the spectra can be used to determine chemical character of the solutions. Plots of prin-

cipal component scores provide information about the solution chemistry that could be used to adjust solution conditions to desired states

REFERENCES

1 G A Burney and J A Porter, Solubilities of Pu(III), Am(III), and Cm(III) oxalates, *Nuclear Chemistry Letters*, 3 (1967) 79–85

2 J M Cleveland, *The Chemistry of Plutonium*, American Nuclear Society, La Grange Park, IL, 1979, pp 401–403.

3 L.E. Wangen, M V. Phillips and L.F Walker, *Use of Multivariate Calibration for Plutonium Quantitation by the Pu(III) Spectrophotometric Method*, U S Department of Energy Report LA-11297, 1988

4 P G Hagan and F.J Miner, *Spectrophotometric Determination of Plutonium III, IV, and VI in Nitric Acid Solutions*, Atomic Energy Report RFP-1391, 1969

5 D R Van Hare, *Analysis of Special Recovery Samples by Pu(III) Spectrophotometry*, U S Department of Energy Report, DP-1713

6 D L Baldwin and R W Stromatt, *Plutonium, Uranium, Nitrate Measurements in Purex Process Stream by Remote Fiber Optic Diode Array Spectrophotometry*, U S Department of Energy Report PNL-SA-15318, 1987

7 A Lorber, L E. Wangen and B R. Kowalski, A theoretical foundation for the PLS algorithm, *Journal of Chemometrics*, 1 (1987) 19–31

8 P Geladi and B R. Kowalski, Partial least squares regression a tutorial, *Analytica Chimica Acta*, 185 (1986) 1–17

9 D M Haaland and E.V Thomas, Partial least squares methods for spectral analyses 1 Relation to other quantitative calibration methods and the extraction of qualitative information, *Analytical Chemistry*, 60 (1988) 1193–1208

10 M Otto and W Wegscheider, Spectrophotometric multicomponent analysis applied to trace metal determinations, *Analytical Chemistry*, 57 (1985) 63–69

11 M Martens and H Martens, Near-infrared reflectance determination of sensory quality of peas, *Applied Spectroscopy*, 40 (1986) 303–310

12 J L. Parker, *A Plutonium Solution Assay System Based on High-Resolution Gamma Ray Spectroscopy*, U S Department of Energy Report LA-8146-MS, 1980

13 E.W. Baumann and B H Torrey, Determination of free acid by standard addition with potassium thiocyanate as a complexant, *Analytical Chemistry*, 56 (1984) 682–685

14 D. Veltkamp and B R Kowalski, *PLS 2-Block Modeling, Version 3 0*, Center for Process Analytical Chemistry, BG-10, University of Washington, Seattle, WA, 1988

15 J L Ryan, Species involved in the anion-exchange absorption of quadrivalent actinide nitrates, *Journal of Physical Chemistry*, 64 (1960) 1375–1385

16 S Wold and M Sjostrom, SIMCA a method for analyzing chemical data in terms of similarity and analogy, in B Kowalski (Editor), *Chemometrics Theory and Application*, ACS Symposium Series No 52, American Chemical Society, Washington, DC, 1977, pp 243–282

17 M A Sharaf, D L Illman and B R. Kowalski, *Chemometrics*, Wiley-Interscience, New York, 1986, pp 234–239

18 A.J Stuper, W E Brugger and P C Jurs, A computer system for structure–activity studies using chemical structure information handling and pattern recognition techniques, in B Kowalski (Editor), *Chemometrics Theory and Applications*, ACS Symposium Series No 52, American Chemical Society, Washington, DC, 1977, pp 165–191

19 M A Sharaf, D L Illman and B R Kowalski, *Chemometrics*, Wiley-Interscience, New York, 1986, pp 281–292

20 D L Massart, B G M Vandeginste, S.N Deming, Y Michotte and L Kaufman, *Chemometrics A Textbook*, Elsevier, Amsterdam, 1988

21 E.R Malinowski and D G Howery, *Factor Analysis in Chemistry*, Wiley-Interscience, New York, 1980

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 259
Elsevier Science Publishers B V., Amsterdam

Discussion of "Determining chemical characteristics of plutonium solutions using visible spectrometry and multivariate chemometric methods" by W.P. Carey and L.E. Wangen

Mark E. Johnson
Department of Statistics, University of Central Florida, Orlando, FL 32816 (U S A)

This paper is an ideal contribution to the Statistics in Chemistry conference held in College Station The authors present several challenging problems which they address in an intelligent fashion using PLS regression and a variety of pattern recognition techniques.

Their most successful application is in the determination of Pu(III) and nitric acid using PLS regression The results on test samples given in Table 2 provide strong evidence that the authors can predict unknown concentrations. Perhaps the authors might comment on any operator or technician effects Obviously, they are adept at using the LT Industries Quantum 1200 device. In routine operations by lesser skilled technicians, would the performance be so good?

The questions related to qualitative determination of Pu(IV) complex composition are clearly more difficult and the results not so clear-cut. I am a little unclear on the results in Figs. 5 and 6 for the zero fluoride and oxalate concentration. If there were five concentrations used, where is the fifth curve?

Figs. 7–9 are curious. Many times statisticians neglect the very useful technique of designating points on plots as a value-added characteristic Fig 7 seems (unfortunately) to set a standard by which we view Figs. 8 and 9. The scatter in Figs. 8 and 9 is much more than in Fig. 7 What fraction of the variation is explained by the first two principal components?

One final comment concerns Table 5. Although Bayes and learning machine dominate nearest neighbor procedures here, I am unwilling as yet to dismiss nearest neighbor (the authors do not suggest this but a reader might inadvertently conclude as much). I suspect that if the data were a tad more 'noisy', nearest neighbor might make a comeback. What is it about the authors' application that favors Bayes and learning machine?

Chemometrics and Intelligent Laboratory Systems, 10 (1991) 261–270
Elsevier Science Publishers B.V., Amsterdam

Transformation robust experimental design with application to some problems in chemistry

Young-Il Kim
Department of Management Science, Chung-Ang University, Seoul (Korea)

Christopher J. Nachtsheim *
Curtis L. Carlson School of Management, University of Minnesota, Minneapolis, MN 55455 (U.S.A.)

(Received 30 May 1990, accepted 5 November 1990)

Abstract

Kim, Y.-I. and Nachtsheim, C.J., 1991. Transformation robust experimental design with application to some problems in chemistry. *Chemometrics and Intelligent Laboratory Systems*, 10: 261–270.

In this paper we consider the selection of an appropriate experimental design when the exact form of the error distribution is unknown. The goal of error-robust design is to design an experiment so that the 'ill-effects' resulting from a lack of knowledge of the error structure will be minimal. Numerical algorithms for computer construction of error-robust designs are developed and the method is illustrated in connection with the design of experiments for nonlinear modeling of chemical reactions.

1. INTRODUCTION

The examination of standard statistical techniques in order to determine their sensitivity to assumptions and development of new techniques that are insensitive to assumptions have been major areas of statistical research in the last two decades. Experimental design is an area in which it is particularly important to investigate questions of robustness because an experimenter's assumptions about the experimental process are critical in determining the design. Furthermore, the design must be chosen before the data are collected and so cannot be discarded if the data indicate that the assumptions are seriously violated. Thus it is important to examine experimental designs for their sensitivity to assumptions.

Generally, we observe that the design chosen will explicitly depend on the experimenter's
(1) design criterion;
(2) definition of the design space;
(3) a priori specification of the model.
By 'model' we mean the distribution of a response $y(x)$, at some point x in the q-dimensional design space χ. Unfortunately, precise a prior specification of points (1)–(3) is often difficult in practice. This fact has led statisticians to search for ways of constructing designs where one or more of the items listed cannot be so explicitly stated.

For example, with regard to (1) above, Box [1] stressed the need to design experiments with many, sometimes conflicting, goals in mind, not just one implied by a single design criterion. Kiefer [2] examined the robustness of optimal designs to

changes in criteria. Welch [3] presented a method for cataloguing designs that are optimal by one criterion, so that further comparisons among these optimal designs could be made on the basis of other criteria

The question of robustness to assumptions concerning the true model η has been widely studied Two different, but complementary, approaches have been taken. The first approach has sought designs that will yield reasonable results for the proposed model even though it is known to be inexact. Steinberg and Hunter [4] call these 'model-robust designs'. For examples of work in this realm, see refs. 5–9. The second approach has focused on developing designs that facilitate improvement of the proposed model by trying to highlight suspected inadequacies. Steinberg and Hunter call these designs 'model-sensitive designs'. Examples are given in refs. 10–17, among others.

Special 'model-robustness' problems arise in the design of experiments for nonlinear models This is because the best design depends, in general, on the unknown parameter values. Investigators are thus placed in a paradoxical position of having to known at design stage (at least approximately) the very quantities that they are conducting the experiment to estimate. Little has been done to assess the robustness of nonlinear designs to misspecification of θ. (Chaloner and Larntz [18] develop a Bayesian approach in which only a prior distribution for θ is required.) For reviews of nonlinear designs, see refs. 19 and 20, among others.

A final area of robustness concerns the sensitivity of designs to the specification of error structure. The occurrence of outliers and missing observations represent two ways in which these assumptions may be violated. A number of authors have studied design in such circumstances. See, for example, refs. 8 and 21–23 regarding design in the presence of outliers. Also see refs. 24 and 25 concerning design when missing data might be a problem. Concerning lack of independence in the error terms, see refs. 26 and 27.

Surprisingly little has been done, however, with regard to the designs that are robust to the general misspecification of the error structure. In what follows, we consider the construction of such designs The only relevant paper on this issue was found to be ref. 28 They applied a 'power transformation weigthing' technique to develop sequential experimental designs for precise parameter estimation of the model and transformation parameters together.

This paper has the following structure. We first review the design of experiments in the presence of known, non-constant variance in Section 2. In Section 3, a general definition of error-robustness is developed and a number of examples are considered. Carroll and Ruppert [29] recently advocated a new method (power transformation on both sides—PTBS) for simultaneous estimation of the regression parameters and index of the 'best' power transformation, λ We show in Section 4 that designs that are robust (in a sense to be described) to the eventual specification of λ are related to error-robust designs Two important examples from the literature are studied in Section 5. Some closing remarks are given in Section 6

2 OPTIMAL DESIGN IN THE PRESENCE OF NON-CONSTANT VARIANCE

2.1 Notation

In what follows, we assume that responses are independent having mean $E(y(x)) = \eta(x, \theta)$ and variance Var $(y(x)) = \sigma^2(x, \lambda)$ where θ and λ are unknown parameter vectors of dimensions p and q respectively We use the term error function in connection with $\sigma^2(x, \lambda)$, its inverse, $\sigma^{-2}(x, \lambda)$, is termed the efficiency function, where we shall assume $0 < \sigma^2(x, \lambda) < \infty$. For brevity we will often the abbreviated form $\sigma^2(x)$

Consider an N-point experiment in which n, observations are taken at the points $x_i \in \chi$ for $i = 1, 2, \ldots, n$ such that $\sum_{i=1}^{n} n_i = N$. Such an experiment can be described by a measure $\xi[N]$ as follows:

$$\xi[N](x) = \begin{cases} n_i; & \text{if } x = x_i \in \{x_1, \ldots, x_n\} \\ 0; & \text{otherwise} \end{cases}$$

Let $S(\xi[N]) = \{x_1, \ldots, x_n\}$ denote the support of $\xi[N]$. Note that if $\xi_N = \xi[N]/N$, then ξ_N is a discrete probability measure on ξ Thus, an exact

or discrete experimental design is a probability measure ξN on the design space χ subject to the restriction that $N\xi_N(x)$ is an integer.

Removing the restriction that $\xi_N(x)$ be a multiple of $1/N$, the set of approximate experimental designs on χ is denoted by

$$\Xi = \left\{ \xi \,|\, \int_{\chi} d\xi(x) = 1,\ \xi(x) \geq 0, \quad \text{for every} \quad x \in \chi \right\}$$

An (approximate) design problem, specified by the triplet (η, σ^2, χ), is solved by selection of an approximate design $\xi \in \Xi$ for the model η, the design space χ and the error function σ^2. Note that in many design problems an exact design χ_N can be approximated by an approximate design ξ.

2.2 Measures of optimality

We assume that least squares estimates $\hat{\theta}$ of the parameter θ are to be obtained. Let $f(x, \theta) = \partial\eta(x, \theta)/\partial\theta$ and

$$F(\theta) = \begin{bmatrix} f^T(x_1, \theta) \\ \vdots \\ f^T(x_n, \theta) \end{bmatrix}$$

Then for these estimates (with $n_i = 1$), the asymptotic covariance is given by

$$\mathrm{Var}(\hat{\theta}) = \left[F(\hat{\theta})^T V^{-1} F(\hat{\theta}) \right]^{-1}$$

where $V = \mathrm{diag}\ \{\sigma^2(x_1, \lambda), \ldots, \sigma^2(x_n, \lambda)\}$. For linear models, the so-called design matrix, $X = F(\theta)$, is independent of θ. For any N-point discrete design ξ_N, we have

$$F(\hat{\theta})^T V^{-1} F(\hat{\theta}) = N \sum_{x \in S(\xi_N)} \sigma^{-2}(x, \lambda) f(x, \hat{\theta}) f^T(x, \hat{\theta}) \xi_N(x)$$

$$= N \int_{\chi} \sigma^{-2}(x, \lambda) f(x, \hat{\theta}) f^T(x, \hat{\theta})\, d\xi_N(x)$$

and hence the i, jth element of $F(\hat{\theta})^T V^{-1} F(\hat{\theta})/N$ is $\sigma^{-2}(x, \lambda) f_i(x, \theta) f_j(x, \theta)$, averaged with respect to the discrete probability measure ξ_N. In general, the normalized information matrix of an experimental design ξ is

$$M(\xi, \theta) = \int_{\chi} \sigma^{-2}(x, \lambda) f(x, \theta) f^T(x, \theta)\, d\xi(x)$$

The dispersion matrix $M^{-1}(\xi, \theta)$ is sometimes written $D(\xi, \theta)$.

Many criteria have been proposed for optimizing the selection of a design ξ for the design problem (η, σ^2, χ). Generally, the criteria are based on some functional of the information matrix, $M(\xi, \theta)$. Motivation for such criteria is often based on the properties of the resulting least squares estimate $\hat{\theta}$. For example, a design ξ_D is defined to be D-optimal for (η, σ^2, χ) and prior estimate θ_0 if

$$\max_{\xi \in \Xi} |M(\xi, \theta_0)| = |M(\xi_D, \theta_0)|$$

By definition, D-optimal designs minimize the (asymptotic) generalized variance of the least squares estimate of θ.

Alternatively, suppose that an experimenter is concerned with prediction. The least squares estimate of the mean response at a point x is

$$\hat{y}(x) = \eta(x, \hat{\theta})$$

$$\mathrm{Var}(\hat{y}(x)) = \mathrm{Var}(\eta(x, \hat{\theta})) \approx f^T(x, \hat{\theta}) D(\xi, \hat{\theta}) f(x, \hat{\theta}) = d(x, \hat{\theta}, \xi)$$

G-optimal designs minimize the maximum normalized variance of prediction $\sigma^{-2}(x, \lambda)\, d(x, \xi, \hat{\theta})$. Formally, a design ξ^* is G-optimal if

$$\min_{\xi \in \Xi} \max_{x \in \chi} \sigma^{-2}(x, \lambda)\, d(x, \xi, \theta_0) = \min_{x \in \chi} \sigma^{-2}(x, \lambda)\, d(x, \xi^*, \theta_0)$$

The D-efficiency of a design ξ for (η, σ^2, χ) and prior estimate θ_0, with respect to ξ_1, is

$$D(\xi, \xi_1, (\eta, \sigma^2, \chi)) = \left\{ \det M^{-1}(\xi_1, \theta_0) \det M(\xi, \theta_0) \right\}^{1/p}$$

Similarly, the G-efficiency of a design ξ for (η, σ^2, χ) and prior value θ_0, with respect to ξ_1, is

$$G(\xi, \xi_1, (\eta, \sigma^2, \chi)) = \max_{x \in \chi} d(x, \xi_1, \theta_0) / \max_{x \in \chi} d(x, \xi, \theta_0)$$

The following result, given by Kiefer and Wolfowitz [30] in the context linear models and later [31] in the context of nonlinear models, shows that D- and G-optimal designs are equivalent.

THEOREM 1. The following conditions are equivalent:
(a) ξ^* is D-optimal
(b) ξ^* is G-optimal
(c) $\max_{x \in \chi} \sigma^{-2}(x, \lambda)\, d(x, \epsilon, \theta_0) = p$.
The set of all designs satisfying these conditions is convex, and the corresponding information matrices are identical.

The equivalence of conditions (a) and (c) yields a simple method for checking the optimality of a candidate design ξ. If the maximum normalized prediction variance is greater than p, then ξ is not D-(G-)optimal. Numerical algorithms [32] for constructing D-(G-)optimal designs make direct use of this condition.

We note that in practice $\sigma^2(x, \lambda)$ is usually assumed constant. The impact of this assumption can be illustrated by the following example. Suppose $\eta(x, \theta) = f^T(x)\theta$, where $f^T(x) = (1, x, x^2)$ and $\chi = [-1,1]$. Suppose also that $\sigma^2(x, \lambda) = \frac{1}{2}[(\lambda - 1)x + \lambda + 1)]$ for $\lambda \geq 1$. Thus, the error variance increases linearly with slope $(\lambda - 1)$ over the design space χ and if $\lambda = 1$, $\sigma^2(x, \lambda) = 1$. Table 1 shows D-(G-)optimal designs for various λs. Note that as the value of λ increases, the design shifts the middle support point toward the left side of the design space. Surprisingly, the D-optimal design shifts mass toward lower variance (high efficiency) region of the design space. This pattern has consistently appeared in worked examples. For further results see refs. 32 and 33. We note from the table that the G-efficiences are monotonically decreasing in λ. For example, with $\lambda = 9$, the G-efficiency of ξ_1 is 0.888. This very simple example illustrates the nonrobustness of the usual optimal design and, we think, motivates the need for the study of designs which are robust to misspecification of σ^2. In the following section we introduce the concept of error-robustness and develop methods for constructing robust designs.

TABLE 1

Location of interior points (x_λ) of G-optimal designs (ξ_λ) for quadratic regression for various λ, $\chi = [-1,1]$, $\sigma^2(x, \lambda) = \frac{1}{2}[(\lambda - 1)x + \lambda + 1]$, $\xi_\lambda(\pm 1) = \xi_\lambda(x_\lambda) = \frac{1}{3}$

λ	Interior point x_λ	G-Efficiency of ξ_1
1	0	1.000
3	−0.141191	0.958
5	−0.183268	0.924
7	−0.221089	0.902
9	−0.241081	0.888

3 ERROR-ROBUST DESIGN

The concept of an error function is critical in both design and analysis. In data analysis contexts, graphical examination of scatterplots of residuals versus predictors or fitted values is used to detect nonconstant variance. A systematic megaphone shape in the plot would indicate that the variance of the response depends on the quantity plotted on the x-axis. Cook and Weisberg [34] suggested an alternative approach for diagnosing non-constant error terms. It involves expansion of the regression model by assuming a particular, though widely applicable, functional form for the variance:

$$\mathrm{var}(y(x)) \propto \exp(\lambda^T x)$$

where λ is an unknown parameter vector. Cook and Weisberg utilized this difinition to propose the score test and the equivalent graphical method for testing the assumption of constant error terms in linear regression. Many of the error functions commonly encountered in data analysis arise as special cases of this important, general form. Suppose we expand $\mathrm{var}(y(x)) = \exp(\lambda^T x)$ about $x = 0$ in a single dimension. Then

$$\mathrm{var}(y(x)) \propto 1 + \lambda x + \lambda^2 x^2/2$$

and we specify $\sigma^2(x)$ as proportional to a quadratic function of x. Specifying only the first term implies that $\sigma^2(x)$ is proportional to x, which

may be a very natural assumption in a comparatively narrow range.

The results of Section 2 indicate that optimal designs depend on the model specification η as well as the variance function σ^2. As stated previously, it is typically the case in practice that the variance function $\sigma^2(x, \lambda)$ cannot be determined before experimentation. Given that the true error function $\sigma^2(x, \lambda)$ is unknown we will consider a design ξ to be robust to specification of $\sigma^2(x, \lambda)$ if ξ is highly efficient for error functions likely to be encountered in practice. More specifically we shall assume that σ^2 is an unknown element of some known space of error functions, E. We will then attempt to characterize designs that are efficient, in a sense to be described, for all possible $\sigma^2 \in \mathrm{E}$. To do so, we shall require the following result, due to Atwood [35], which relates the D and G efficiencies of a design

THEOREM 2 Let ξ_{σ^2} be the D-optimal design for (η, σ^2, χ). Then for any design ξ in Ξ,

$$\mathrm{D}(\xi, \xi_{\sigma^2}, (\eta, \sigma^2, \chi)) \geq \mathrm{G}(\xi, \xi_{\sigma^2}, (\eta, \sigma^2, \chi))$$

G-efficiency provides a lower bound for the D-efficiency of a design ξ with respect to the D-optimal design ξ_{σ^2}. Following Thibodeau [8], in context of model robustness, we attempt to construct designs having high D-efficiency for each $\sigma^2 \in \mathrm{E}$ by maximizing the lower bound. Loosely speaking, we will consider a design error-robust if its G-efficiency is high for every $\sigma^2 \in \mathrm{E}$. Thus no matter what the subsequent analysis indicates regarding choices of σ^2, the D-efficiency of the design will be relatively high. Formally, we have

Definition 1. The design $\xi^* \in \Xi$ is error-robust if and only if

$$\max_{\xi \in \Xi} \min_{\sigma^2 \in \mathrm{E}} \mathrm{G}(\xi, \xi_{\sigma^2}, (\eta, \sigma^2, \chi)) = \min_{\sigma^2 \in \mathrm{E}} \mathrm{G}(\xi^*, \xi_{\sigma^2}, (\eta, \sigma^2, \chi))$$

Notice that because the number of parameters in the model, p, does not change with σ^2, Definition 1 indicates that a design is error-robust design if and only if

$$\min_{\xi \in \Xi} \max_{\sigma^2 \in \mathrm{E}} \max_{x \in \mathrm{X}} \sigma^{-2}(x)\, \mathrm{d}(x, \xi, \theta_0) = \max_{\sigma^2 \in \mathrm{E}} \max_{x \in \mathrm{X}} \sigma^{-2}(x)\, \mathrm{d}(x, \xi^*, \theta_0)$$

Thus the error-robust design minimizes the 'worst case' normalized maximum variance of fitted values.

In most instances, analytic characterization of the error-robust design is impossible, and numerical methods are required See Kim [33] for some notable exceptions. The following algorithm, which is a simple modification of one by Fedorov [32], can be used for computer construction of error-robust designs.

Algorithm 1

1. Specify nonsingular starting design ξ_1 Set $i = 1$.
2. Find x_i such that $\max_{\sigma^2 \in \mathrm{E}} \max_{x \in \mathrm{X}} \sigma^{-2}(x)\, \mathrm{d}(x, \xi_i, \theta_0) = \sigma^{-2}(x_i)\, \mathrm{d}(x_i, \xi_i, \theta_0)$.
3. Let $\alpha_i = 1/(i + s)$, $s \geq 0$, and form $\xi_{i+1} = (1 - \alpha_i)\xi_i + \alpha_i \xi_{x_i}$, where ξ_{x_i} places unit mass at x_i Update D.
4. Check for convergence. One simple approach is as follows. Assume $k \geq 2$ is a user defined integer. Typically, $k \approx 5$. Let

   $$\delta_j = \sigma^{-2}(x_{i-j+1})\, \mathrm{d}(x_{i-j+1}, \xi_{i-j+1}, \theta_0) \qquad 1 \leq j < \min[i, k]$$

   Let s_δ^2 be the sample variance of the $\{\delta_i\}$. If $i \geq k$ and s_δ^2 is sufficiently small, stop. Otherwise, set $i = i + 1$ and go to 2.

Note that the sequence $\{\alpha_i\}$, as specified above, will not, in general, lead to monotonically decreasing $\sigma^{-2}(x_i)\, \mathrm{d}(x_i, \xi_i, \theta_0)$

As a simple illustration, consider again the quadratic regression model $f^T(x) = (1, x, x^2)$ with $\mathrm{E} = \{\sigma^2(x) \mid \sigma^2(x) \propto (\lambda - 1)x + (\lambda + 1),\ \lambda = 1, 3, 5, 7, 9\}$. The following design was found to be error-robust using the algorithm described above: $\xi(\pm 1) = 0.325$, $\xi(0.039609) = 0.182$, $\xi(-0.260323) = 0.167$. Table 2 presents G-ef-

TABLE 2

G-efficiencies of various designs ξ_λ^* for quadratic regression on $\chi = [-1,1]$ $\sigma^2(x) \propto (\lambda - 1)x + (\lambda + 1)$

Design	Actual γ				
	1	3	5	7	9
ξ_1	1 0	0.958	0 924	0 902	0 888
ξ_3	0 948	1 0	0 996	0 998	0.994
ξ_5	0 914	0 993	1 0	0 998	0 994
ξ_7	0 876	0 979	0 997	1 0	0 999
ξ_9	0 854	0 969	0 992	0 998	1.0
Robust	0 974	0 981	0 979	0 978	0 974

ficiencies of designs constructed under varying assumptions about λ For example, the first row summarizes the performance of the optimal design under assumption $\lambda = 1$, for various alternative 'true' efficiency functions As noted previously, if λ turns out to be 9 by subsequent analysis, the design will be 88 8% G-efficient. The worst case occurs in the lower left-hand corner of the table. Here the experimenter has assumed $\lambda = 9$, when λ turns out to be 1, in which case the G-efficiency of the D-optimal design is 85.4%. In contrast, the worst-case G-efficiency of the error-robust design is 97.4% Interestingly, the error-robust design consists of 4 support points. Intuitively, mass at $x = 0$ 039609 was required to protect against $\lambda = 1$ where mass $x = -0.260323$ was required for protection against $\lambda = 9$. This intuitive explanation is supported by the fact that during execution of the computer algorithm, maximization of $\sigma^{-2}(x)\, d(x, \xi, \theta_0)$ occurred only at $\lambda = 1$ or $\lambda = 9$. This example suggests that for $\lambda = [a,b]$, in some cases a reasonable approximation to the error-robust design will be obtained by mixing the D-optimal designs ξ_a and ξ_b appropriately.

4 POWER-TRANSFORMATION ROBUST DESIGN

Recently Carroll and Ruppert [29] introduced a method, power transformation on both sides, PTBS, for simultaneous estimation of regression parameters and an appropriate power transformation index. They discussed its use with known, nonlinear regression models. Suppose the known mean structure, which may be derived, for example, from a physical system, is $E(y(x)) = \eta(x, \theta)$ and that $\eta(x, \theta) > 0$ for $x \in \chi$. Errors $\{\epsilon\}$ are not necessarily additive (or constant over χ) implying

$$y(x) = g(\eta(x, \theta), \epsilon)$$

For example, if the errors are log normal and $g(a, b) = ab$ (i e., errors are multiplicative), taking logs yields

$$\log(y(x)) = \log \eta(x, \theta) + \epsilon$$

Where $\{\epsilon\}$ are normally distributed. This type of situation led Carroll and Ruppert to consider a family of strictly monotonic transformations $h(y, \lambda)$, indexed by the q-vector λ, and to assume that for some value of λ, say λ_0,

$$h(y, \lambda_0) = h(\eta(x, \theta), \lambda_0) + \epsilon$$

This approach is in the spirit of Box and Cox [36], who suggested the well known power transformation family:

$$h(y, \lambda) = y^{(\lambda)} = (y^\lambda - 1)/\lambda \quad \text{if } \lambda = 0$$
$$= \log(y) \quad \text{if } \lambda \neq 0$$

Box and Cox sought a transformation that achieves (a) a simple additive or linear model, (b) homoscedastic errors, and (c) normally distributed errors In PTBS regression, both the response and the model are transformed via h. An important advantage of PTBS regression is that the original meaning of the parameters is preserved Estimation of θ and λ in PTBS regression is typically carried out via normal theory maximum likelihood.

For the above model, we have

$$\frac{\partial \eta^{(\lambda)}(x, \theta)}{\partial \theta} = f_\lambda(x, \theta)$$
$$= \eta(x, \theta)^{\lambda - 1} f(x, \theta)$$

where $f(x, \theta)$ is as defined previously: $f(x, \theta) = \partial \eta(x, \theta)/\partial \theta$ Given λ, information matrix and variance functions are defined as

$$M_\lambda(\xi, \theta)$$
$$= \int_\chi \eta(x, \theta)^{2(\lambda - 1)} f(x, \theta) f^T(x, \theta)\, d\xi(x)$$

and

$$d_\lambda(x, \xi, \theta) = \eta(x, \theta)^{2(\lambda-1)} f^T(x, \theta) \times M_\lambda^{-1}(\xi, \theta) f(x, \theta)$$

respectively.

The above expressions indicate that the design problem may be viewed as standard, with induced efficiency function $\sigma^{-2}(x, \lambda) = \eta(x, \theta)^{2(\lambda-1)}$. It is also apparent that choice of design will depend on an experimenter's a priori suspicions concerning λ. Typically, one takes $\lambda = 1$ and hopes for the best, although consequences can be dire. For example, suppose that the underlying theoretical model is quadratic and errors are multiplicative and log normal. That is, $\eta(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2$ and $\lambda = 0$ gives the appropriate transformation. For $\theta_0^T = (1, 1, 1)$ and $\chi = [0,1]$, the design $\xi_0(\pm 1) = \frac{1}{2}$, $\xi_0(0.373) = \frac{1}{2}$ is D-optimal. On the other hand, if the experimenter assumes $\lambda = 1$, and obvious choice might be the usual D-optimal design ξ_1, which places $\frac{1}{3}$ mass at the points ± 1 and 0. Since $\max_{x \in \chi} d_0(\chi, \xi_1, \theta_0) = 3.56$, ξ_1 is 84% G-efficient. If the appropriate λ is -1 or 2, the G-efficiency of ξ_1 drops to 47% in both cases.

The above discussion motivates the need for designs for PTBS regression that are robust to specification of λ for λ in a specified set L. We offer the following.

Definition 2 The design $\xi^* \in \Xi$ is power-transformation (PT) robust if and only if

$$\min_{\xi \in \Xi} \max_{\lambda \in L} \max_{x \in \xi} d_\lambda(x, \xi) = \max_{\lambda \in L} \max_{x \in \chi} d_\lambda(x, \xi^*)$$

where $d_\lambda(x, \xi) = f_\lambda^T(x) M_\lambda^{-1} f_\lambda(x)$.

As noted, for a specified regression function $\eta(x, \theta)$, the Carroll and Ruppert family of transformations indexed by $\lambda \in L$ induces a corresponding family of induced error function $E_\lambda = \{\eta(x, \theta)^{-2(\lambda-1)} | \lambda \in L\}$.

Thus Definition 2 may be restated in the following way.

Definition 3. The design $\xi^* \in E$ is PT-robust if and only if ξ^* is error-robust for E_λ.

Since PT-robustness is a special case of error-robustness, the algorithm previously developed for computer construction of robust designs is applicable.

5 TRANSFORMATION ROBUSTNESS APPLICATIONS

The following two examples are taken from literature and are frequently cited in papers on nonlinear design. These examples illustrate how inefficient the usual D-optimal designs can be in the presence of uncertainty about the error structure, and the efficacy of the robust approach.

Example 1. The following experiment was reported by Box and Hunter [20] and has been discussed by numerous authors. The purpose of the experiment is to model some catalytic reactions of the type $R \rightarrow P_1 + P$, in which the reagent R is some quaternary or primary alcohol from a log chain, the product P_1 is an olefin and the product p is water. The theoretical model for such a reaction is

$$\eta(x, \theta) = \frac{\theta_1 \theta_3 x_1}{1 + \theta_1 x_1 + \theta_2 x_2}$$

where η is the speed of the chemical reaction, x_1 is the partial pressure of the product P, x_2 is the partial pressure of the product P_1, θ_1 is a reaction parameter, θ_2 is the absorption equilibrium constant for the product P_1, and θ_3 is the effective constant of the reagent R.

For purposes of design construction, following Box and Hunter [20], the prior values of the parameters were fixed at $\theta_0^T = [2.9, 12.2, 6.9]$. It was assumed that observations are possible in the region $\chi = \{x_1, x_2 | 0 \le x_1 \le 3,\ 0 \le x_2 \le 3\}$, which leads to the locally D-optimal design, $\xi_D(0.3, 0.0) = \xi_D(3.0, 0.0) = \xi_D(3.0, 0.8) = 1/3$. ξ_D and χ are pictured in Fig. 1a. The fact that the design does not cover the design space leads one to question the logic of the design, unless the experimenter has particularly strong faith in his assumptions. The D-optimal designs for $\lambda = -1$ and $\lambda = 0$ are pictured in Figs. 1b and 1c, respectively. Note that for $\lambda < 1$ the efficiency function is undefined at

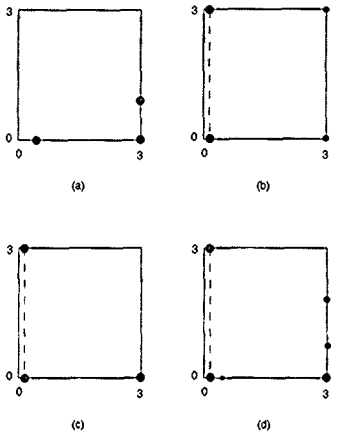


Fig 1 Optimal designs for Example 1 $\chi = [0,3]^2$ (a) Optimal design for $\lambda = 1$ $\xi(0\,3,0) = \xi(3,0) = \xi(3,0\,8) = 1/3$ (b) Optimal design for $\lambda = -1$. $\xi(0\,1,0) = \xi(0\,1,3) = 1/3$, $\xi(3,0) = \xi(3,3) = 1/6$ (c) Optimal design for $\lambda = 0$ $\xi(0\,1,0) = \xi(3,0) = \xi(0\,1,3) = 1/3$ (d) Robust design $\xi(3,0) = 0\,216$; $\xi(3,0\,8) = 0\,17$, $\xi(0\,3,0) = 0\,101$, $\xi(0\,1,3) = 0\,227$, $\xi(0\,1,0) = 0\,212$, $\xi(3,1\,7) = 0\,073$

$x_1 = 0$. Thus it was necessary to truncate the design space such that $\chi = \{(x_1, x_2) \mid \Delta < x_1 \leq 3,\ 0 \leq x_2 \leq 3\}$ for some $\Delta > 0$. We chose $\Delta = 0.1$. The truncation is indicated in Figs. 1b and 1c.

The PT-robust design is pictured in Fig. 1d. G-efficiencies of the robust design for various true λ are summarized in Table 3. Notice that the worst case G-efficiencies result for $\lambda = \pm 1$ with both values being about 66%. As was indicated in the previous example the error-robust design seems to represent the best trade-off possible between D-optimal d igns for $\lambda = \pm 1$. Table 3 also shows that this 66% G-efficiency is high in comparison to the $\approx 0\%$ G-efficiency resulting from the case when we assume $\lambda = 0$, and λ turns out to be 1.

TABLE 3

G-efficiencies for designs in Example 1

λ Assumed	λ		
	-1	0	1
-1	1 000	0 520	0 220
0	0 997	1.000	0 000
1	0 000	0 214	1 000
Robust	0 656	0 754	0 662

Example 2 The following model was studied by Carr [37]

$$\eta(x, \theta) = \frac{\theta_1\theta_3(x_2 - x_3/1.632)}{1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_3}$$

where η is the rate of disappearance of *n*-pentane, x_1, x_2, x_3 are the partial pressures of hydrogen, *n*-pentane and *i*-pentane respectively, θ_1 is a reaction parameter and θ_2, θ_3, θ_4 are equilibrium constants (psia^{-1}). For this problem, $\chi = \{(x_1, x_2, x_3) \mid 107 \leq x_1 \leq 471,\ 69 \leq x_2 \leq 294;\ 11 \leq x_3 \leq 121\}$. Box and Hill [38] later used power transformation weighting to fit the model to Carr's

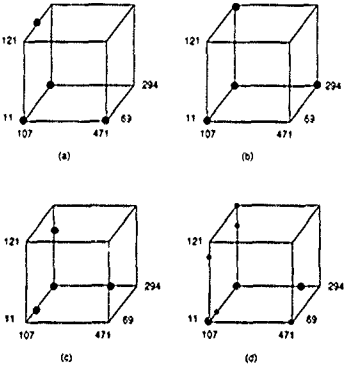


Fig. 2 Optimal designs for Example 2 (a) Optimal design for $\lambda = 0$ $\xi(107, 294, 11) = \xi(471, 69, 11) = \xi(107, 69, 11) = \xi(107, 125\,5, 121) = 0\,25$ (b) Optimal design for $\lambda = 0\,5$ $\xi(107, 294, 11) = \xi(471, 294, 11) = \xi(107, 69, 11) = \xi(107, 294, 121) = 0\,25$ (c) Optimal design for $\lambda = 1$: $\xi(107, 294, 11) = \xi(380, 294, 11) = \xi(107, 125\,25, 11) = \xi(107, 294, 93\,5) = 0\,25$ (d) Robust design: $\xi(107, 294, 11) = 0\,189$, $\xi(107, 125\,25, 11) = 0\,074$, $\xi(107, 294, 93\,5) = 0\,116$, $\xi(380, 294, 11) = 0\,181$, $\xi(107, 69, 93\,5) = 0\,099$, $\xi(471, 69, 11) = 0\,118$; $\xi(107, 69, 11) = 0\,142$, $\xi(107, 294, 121) = 0\,082$.

TABLE 4

G-efficiencies for designs in Example 2

λ Assumed	λ		
	0	1/2	1
0	1 000	0 350	0 094
1/2	0 536	1 000	0 822
1	0 304	0 663	1 000
Robust	0 768	0 767	0 768

24 observations. We consider the construction of a PT-robust design.

Carroll and Ruppert obtained the PTBS parameters estimates $(\hat{\theta}, \hat{\lambda}) = (39.2, 0.043, 0\,021, 0\,104, 0.72)$. These point estimates for $\hat{\theta}$ are used as prior values in what follows. Reasonable values of λ were thought to be in the interval $L = [0,1]$. Computational constraints forced us to rather severly discretize both L and χ. We took $L = \{0,0.5,1.0\}$ and to approximate χ, we used as a candidate set corresponding to the 5^3 factorial region The error-robust design is pictured in Fig 2d For reference, the D-optimal designs for $\lambda = 0$, 1/2, and 1 are pictured in Figs 2a–c. Table 4 gives G-efficiencies for varying designs and assumptions about λ For example, a G-efficiency of 30.4% occurs when $\lambda = 0$ and constant variance is assumed. In contrast, the minimum G-efficiency for the error-robust design is 76%.

6 CONCLUSIONS

In this paper we have summarized research directed toward the characterization of designs that are insensitive to the specification of error structure. We have developed the related concepts of error and transformation robustness and examined a number of designs that were approximately optimal by our stated criterion. Some obvious extensions, however, are still needed. While the designs calculated are reasonably robust to the specification of error structure in the nonlinear case they suffer from the need to specify θ a prior. One way of alleviating this difficulty may be to combine the maximum approach suggested herein with the methods of Bayesian nonlinear design as described in ref 18. Such methods are currently under investigation

REFERENCES

1 G E.P Box, Choice of response surface design and alphabetical optimality, *Utilitas Mathematica*, 21B (1982) 11–55

2 J Kiefer, Optimal design variation in structure and performance under change of criterion, *Biometrika*, 62 (1975) 277–288

3 W.J Welch, Branch and bound search for experimental designs based on D-optimality and other criteria, *Technometrics*, 24 (1982) 41–48

4 D M Steinberg and W G Hunter, Experimental design review and comment, *Technometrics*, 26 (1984) 71–96

5 G E.P Box and N R Draper, A basis for the selection of a response surface design, *Journal of the American Statistical Association*, 54 (1959) 622–653

6 K. Kussmaul, Protection against assuming the wrong degree in polynomial regression, *Technometrics*, 11 (1969) 677–682

7 E Läuter, Experimental design in a class of models, *Mathematische Operations Forschung und Statistik*, 5 (1974) 379–398

8 L A Thibodeau, Robust design for regression problem, *Ph.D Dissertation*, University of Minnesota, Dept of Statistics, 1977

9 R D Cook and C.J Nachtsheim, Model robust, linear-optimal designs, *Technometrics*, 24 (1982) 49–54

10 W G Hunter and A M Reiner, Designs for discriminating between two rival models, *Technometrics*, 7 (1965) 307–323

11 G E.P Box and W.J Hill, Discrimination among mechanistic models, *Technometrics*, 9 (1967) 57–71

12 S M Stigler, Optimal experimental design for polynomial regression, *Journal of the American Statistical Association*, 66 (1971) 311–318

13 A C Atkinson, Planning experiments to detect inadequacies in regression models, *Biometrika*, 59 (1972) 275–293

14 A C Atkinson and V V Fedorov, The design of experiments for discriminating between two rival models, *Biometrika*, 62 (1975) 57–70

15 A C Atkinson and V V Fedorov, Optimal design experiments for discriminating between several models, *Biometrika*, 62 (1975) 289–303

16 L.R. Jones and T.J Mitchell, Design criteria for detecting model inadequacy, *Biometrika*, 65 (1978) 541–551

17 M D Morris and T.J Mitchell, Two level multifactor designs for detecting the presence of interactions, *Technometrics*, 25 (1983) 345–355

18 K. Chaloner and K Larntz, Optimal Bayesian design applied to logistic regression experiments, *Journal of Statistical Planning and Inference*, 21 (1989) 191–208.

19 G E.P. Box and H L. Lucas, Design of experiments in nonlinear situations, *Biometrika*, 46 (1959) 77–90
20 G E.P. Box and W.G Hunter, The experimental study of physical mechanism, *Technometrics*, 7 (1965) 23–42
21 G E.P. Box and N.R. Draper, Robust design, *Biometrika*, 62 (1975) 347–352
22 N R. Draper and A M Herzberg, Designs to guard against outliers in the presence of model bias, *Canadian Journal of Statistics*, 7 (1979) 127–135.
23 C.J. Nachtsheim, Contributions to optimal experimental design, *Ph.D Dissertation*, Dept. of Statistics, University of Minnesota, 1979.
24 Z Galil and J. Kiefer, Comparison of design for quadratic regression on cubes, *Journal of Statistical Planning and Inference*, 1 (1977) 121–132.
25 D F Andrews and A M Herzberg, The robustness and optimality of response surface designs, *Journal of Statistical Planning and Inferences*, 3 (1979) 249–257
26 J Sacks and D Ylvisaker, Designs for regression problem with correlated errors, *Annals of Mathematical Statistics*, 37 (1966) 66–89
27 P G Bickel and A.M Herzberg, Robustness of design against auto correlation in time I asymptotic theory, optimality for location and linear regression, *Annals of Statistics*, 7 (1979) 77–95
28 D J Pritchard and D W. Bacon, Accounting for heteroscedasticity in experimental design, *Technometrics*, 19 (1977)
29 R.J Carroll and D Ruppert, Power transformation when fitting theoretical models to data, *Journal of the American Statistical Association*, 68 (1984) 771–781
30 J Kiefer and J Wolfowitz, The equivalence to two extreme problems, *Canadian Journal of Mathematics*, 12 (1960) 363–366.
31 L.V White, An extension of the general equivalence theorem to nonlinear models, *Biometrika*, 60 (1973) 345–348
32 V V Fedorov, *Theory of Optimal Experiments*, Academic Press, New York, 1972
33 Y Kim, Error-robust statistical experimental design, with application to model-based sampling in auditing, *Ph D Dissertation*, Dept of Management Sciences, University of Minnesota, 1987
34 R D Cook and S Weisberg, Diagnostics for heteroscedasticity in regression, *Biometrika*, 70 (1983) 1–10
35 C L Atwood, Optimal and efficient design of experiments, *Annals of Mathematical Statistics*, 40 (1969) 1570–1602
36 G E.P Box and D R Cox, Analysis of transformations (with discussion), *Journal of the Royal Statistical Society, Series B*, 26 (1964) 211–246
37 N L Carr, Kinetics of catalytic isomerisation of *n* pentane, *Industrial Engineering Chemistry*, 52 (1960) 391–396
38 G E P Box and W.J Hill, Correcting inhomogeneity of variance with power transformation weighting, *Technometrics*, 16 (1974) 385–389

GENERAL INFORMATION

A detailed leaflet *Information for Authors* is available from the publisher, Elsevier Science Publishers B V , P O Box 330, 1000 AH Amsterdam, The Netherlands, upon request, and has also been published in Vol 9, No. 3 The most important items are given below

Types of contributions *Chemometrics and Intelligent Laboratory Systems* publishes original research papers, short communications, software descriptions and tutorial reports

The journal publishes papers from all areas of mathematics, including computer science, numerical methods, operations research, probability and statistics The motivation and results of the papers must be understandable to chemists and chemometricians

The journal also participates actively in software dissemination through articles on software developments, software descriptions and reviews of software

Short communications are usually complete descriptions of limited investigations, and should generally not exceed four printed pages

Editors responsible for tutorial reports are R G Brereton, R E Dessy and D R Scott Tutorial reports are published by invitation of the Editors, but may also submitted They will be refereed in the usual manner

Submission of papers Papers should be written in English Manuscripts (three copies are required) should be submitted to one of the Editors whose addresses are mentioned on page 2 of the cover Illustrations must also be submitted in triplicate One set should be in a form ready for reproduction, the other two may be of lower quality

Manuscript preparation The manuscript should be typed in double spacing on consecutively numbered pages of uniform size The title page should include the title, name(s) of author(s) and their affiliations The author to whom correspondence should be addressed must be indicated All papers begin with an abstract (50–250 words) which comprises a brief factual account of the contents of the paper As a rule, papers should be divided into sections, headed by a caption (e g , Abstract, Introduction, Experimental, etc)

References References should be numbered in the order in which they are cited in the text, and listed in numerical sequence on a separate sheet at the end of the article The numbers should appear in the text at the appropriate places in square brackets, on the line The reference list must contain the names of all authors, full titles of articles and full journal names, and first and last pages of each cited paper

Tables Tables should be compiled on separate sheets They must be numbered with Arabic numerals and have brief descriptive headings They should be referred to in the text

Illustrations All illustrations are to be numbered consecutively using Arabic numerals and should be referred to in the text Legends must be typed together on a separate sheet Line drawings should either be drawn in Indian ink or submitted as sharp glossy prints suitable for immediate reproduction Any lettering must be large enough to stand photographic reduction Half-tone illustrations should be black and white prints on glossy paper and have as much contrast as possible Colour illustrations may occasionally be published

Proofs One set of proofs will be sent to the author to be checked for printer s errors

Reprints Fifty reprints will be supplied free of charge Additional reprints (minimum 100) may be ordered by the authors The order form containing price quotations will be sent together with the galley proof of the article

Advertisements Advertisement rates are available from the Advertising Manager, Elsevier Science Publishers B V , P O Box 211, 1000 AE Amsterdam, The Netherlands, tel (20) 5803714, telex 18582 ESPA NL, telefax (20) 5803769

0169-7439/91/$03 50

This issue is printed on acid-free paper

Printed in The Netherlands