1 Introduction

1.1 Points of View

There are numerous groups of people interested in chemometrics. One of the problems over the past two decades is that each group has felt it is dominant or unique in the world. This is because scientists tend to be rather insular. An analytical chemist will publish in analytical chemistry journals and work in an analytical chemistry department, a statistician or chemical engineer or organic chemist will tend to gravitate towards their own colleagues. There are a few brave souls who try to cross disciplines but on the whole this is difficult. However, many of the latest advances in theoretical statistics are often too advanced for routine chemometrics applications, whereas many of the problems encountered by the practising analytical chemist such as calibrating pipettes and checking balances are often too mundane to the statistician. Cross-citation analysis of different groups of journals, where one looks at which journal cites which other journal, provides fascinating insights into the gap between the theoretical statistics and chemometrics literature and the applied analytical chemistry journals. The potential for chemometrics is huge, ranging from physical chemistry such as kinetics and equilibrium studies, to organic chemistry such as reaction optimisation and QSAR, theoretical chemistry, most areas of chromatography and spectroscopy on to applications as varied as environmental monitoring, scientific archaeology, biology, forensic science, industrial process monitoring, geochemistry, etc., but on the whole there is no focus, the ideas being dissipated in each discipline separately. The specialist chemometrics community tends to be mainly interested in industrial process control and monitoring plus certain aspects of analytical chemistry, mainly near-infrared spectroscopy, probably because these are areas where there is significant funding for pure chemometrics research. A small number of tutorial papers, reviews and books are known by the wider community, but on the whole there is quite a gap, especially between computer based statisticians and practising analytical chemists.

This division between disciplines spills over into industrial research. There are often quite separate data analysis and experimental sections in many organisations. A mass spectrometrist interested in principal components analysis is unlikely to be given time by his or her manager to spend a couple of days a week mastering the various ins and outs of modern chemometric techniques. If the problem is simple, that is fine; if more sophisticated, the statistician or specialist data analyst will muscle in, and try to take over the project. But the statistician may have no feeling for the experimental difficulties of mass spectrometry, and may not understand when it is most effective to continue with the interpretation and processing of data, or when to suggest changing some mass spectrometric parameters.

All these people have some interest in data analysis or chemometrics, but approach the subject in radically different ways. Writing a text that is supposed to appeal to a broad church of scientists must take this into account. The average statistician likes to build on concepts such as significance tests, matrix least squares and so on. A

statistician is unlikely to be satisfied if he or she cannot understand a method in algebraic terms. Most texts, even most introductory texts, aimed at statisticians contain a fair amount of algebra. Chemical engineers, whilst not always so keen to learn about distributions and significance tests, are often very keen on matrix algebra, and a chemometrics course taught by a chemical engineer will often start with matrix least squares and linear algebra.

Practical chemists, on the other hand, often think quite differently. Many laboratory based chemists are doing what they are doing precisely because at an early phase in their career they were put off by mathematicians. This is especially so with organic chemists. They do not like ideas expressed in terms of formal maths, and equations are 'turn offs'. So a lecture course aimed at organic chemists would contain a minimum of maths. Yet some of these people recognise later in their career that they do need data analytical tools, even if these are to design simple experiments or for linear regression, or in QSAR. They will not, however, be attracted to chemometrics if they are told they are required first to go on a course on advanced statistical significance testing and distributions, just to be able to perform a simple optimisation in the laboratory. I was told once by a very advanced mathematical student that it was necessary to understand Gallois field theory in order to perform multilevel calibration designs, and that everyone in chemometrics should know what Krilov space is. Coming from a discipline close to computing and physics, this may be true. In fact, the theoretical basis of some of the methods can be best understood by these means. However, tell this to an experimentalist in the laboratory that this understanding is required prior to performing these experiments and he or she, even if convinced that chemometrics has an important role, will shy away. In this book we do not try to introduce the concepts of Gallois field theory or Krilov space, although I would suspect not many readers would be disappointed by such omissions.

Analytical chemists are major users of chemometrics, but their approach to the subject often causes big dilemmas. Many analytical chemists are attracted to the discipline because they are good at instrumentation and practical laboratory work. The majority spend their days recording spectra or chromatograms. They know what to do if a chromatographic column needs changing, or if a mass spectrum is not fragmenting as expected. Few have opted to work in this area specifically because of their mathematical background, yet many are confronted with huge quantities of data. The majority of analytical chemists accept the need for statistics and a typical education would involve some small level of statistics, such as comparison of means and of errors and a little on significance tests, but the majority of analytical texts approach these subjects with a minimum of maths. A number then try to move on to more advanced data analysis methods, mainly chemometrics, but often do not recognise that a different knowledge base and skills are required. The majority of practising analytical chemists are not mathematicians, and find equations difficult; however, it is important to have some understanding of the background to the methods they use. Quite correctly, it is not necessary to understand the statistical theory of principal components analysis or singular value decomposition or even to write a program to perform this (although it is in fact very easy!). However, it is necessary to have a feel for methods for data scaling, variable selection and interpretation of the principal components, and if one has such knowledge it probably is not too difficult to expand one's understanding to the algorithms themselves. In fact, the algorithms are a relatively small part of the data

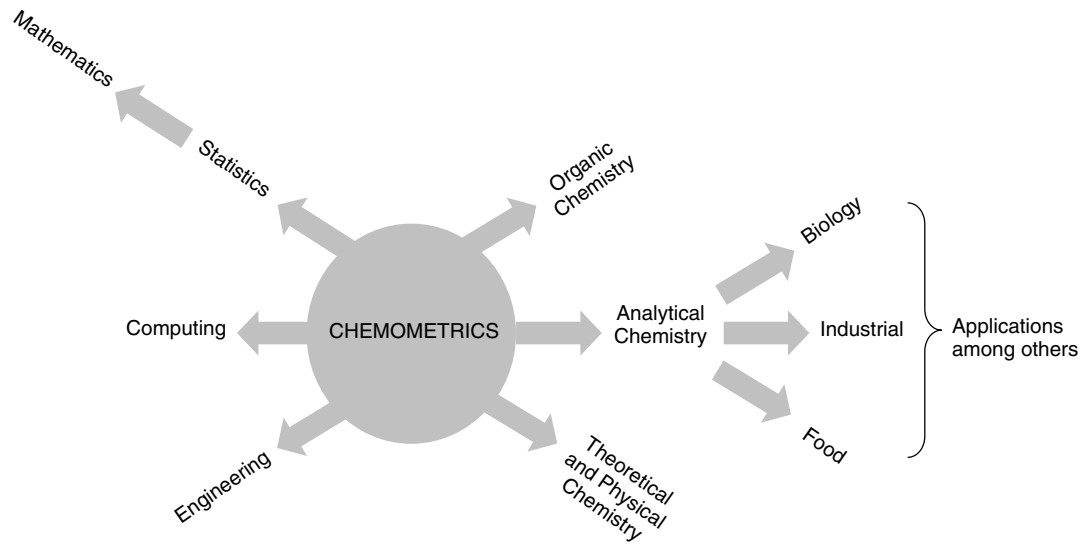

Figure 1.1
How chemometrics relates to other disciplines

analysis, and even in a commercial chemometric software package PCA or PLS (two popular approaches) may involve between 1 and 5 % of the code.

The relationship of chemometrics to different disciplines is indicated in Figure 1.1. On the left are the enabling sciences, mainly quite mathematical and not laboratory based. Statistics, of course, plays a major role in chemometrics, and many applied statisticians will be readers of this book. Statistical approaches are based on mathematical theory, so statistics falls between mathematics and chemometrics. Computing is important as much of chemometrics relies on software. However, chemometrics is not really computer science, and this book will not describe approaches such as neural networks or genetic programming, despite their potential importance in helping solve many complex problems in chemistry. Engineers, especially chemical and process engineers, have an important need for chemometric methods in many areas of their work, and have a quite different perspective from the mainstream chemist.

On the right are the main disciplines of chemistry that benefit from chemometrics. Analytical chemistry is probably the most significant area, although some analytical chemists make the mistake of claiming chemometrics uniquely as their own. Chemometrics has a major role to play and had many of its origins within analytical chemistry, but is not exclusively within this domain. Environmental chemists, biologists, food chemists as well as geochemists, chemical archaeologists, forensic scientists and so on depend on good analytical chemistry measurements and many routinely use multivariate approaches especially for pattern recognition, and so need chemometrics to help interpret their data. These scientists tend to identify with analytical chemists. The organic chemist has a somewhat different need for chemometrics, primarily in the areas of experimental design (e.g. optimising reaction conditions) and QSAR (quantitative structure–analysis relationships) for drug design. Finally, physical chemists such as spectroscopists, kineticists and materials scientists often come across methods for signal deconvolution and multivariate data analysis.

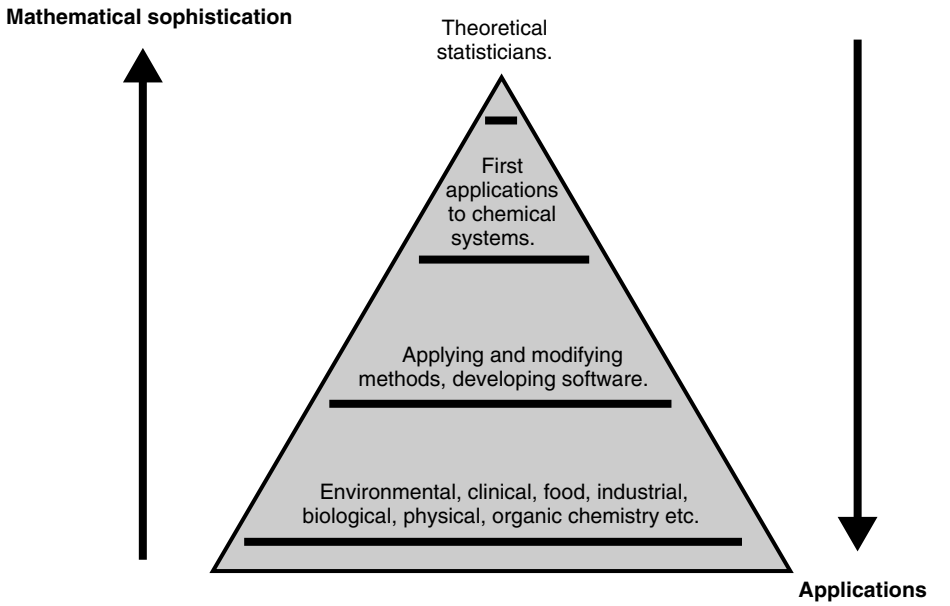

Figure 1.2
People interested in chemometrics

Different types of people will be interested in chemometrics, as illustrated in Figure 1.2. The largest numbers are application scientists. Many of these will not have a very strong mathematical background, and their main interest is to define the need for data analysis, to design experiments and to interpret results. This group may consist of some tens of thousands of people worldwide, and is quite large. A smaller number of people will apply methods in new ways, some of them developing software. These may well be consultants that interface with the users: many specialist academic research groups are at this level. They are not doing anything astoundingly novel as far as theoretical statisticians are concerned, but they will take problems that are too tough and complex for an applications scientist and produce new solutions, often tinkering with the existing methods. Industrial data analysis sections and dedicated software houses usually fit into this category too. There will be a few thousand people in such categories worldwide, often organised into diverse disciplines. A rather smaller number of people will be involved in implementing the first applications of computational and statistical methods to chemometrics. There is a huge theoretical statistical and computational literature of which only a small portion will eventually be useful to chemists. In-vogue approaches such as multimode data analysis, Bayesian statistics, and wavelet transforms are as yet not in common currency in mainstream chemistry, but fascinate the more theoretical chemometrician and over the years some will make their way into the chemists' toolbox. Perhaps in this group there are a few hundred or so people around the world, often organised into very tightly knit communities. At the top of the heap are a very small number of theoreticians. Not much of chemical data analysis is truly original from the point of view of the mathematician – many of the 'new' methods might have been reported in the mathematical literature 10, 20 or even 50 years ago; maybe the number of mathematically truly original chemometricians

is 10 or less. However, mathematical novelty is not the only sign of innovation. In fact, much of science involves connecting ideas. A good chemometrician may have the mathematical ability to understand the ideas of the theoretician and then translate these into potential applications. He or she needs to be a good listener and to be able to link the various levels of the triangle. Chemical data analysis differs from more unitary disciplines such as organic chemistry, where most scientists have a similar training base, and above a certain professional level the difference is mainly in the knowledge base.

Readers of this book are likely to be of two kinds, as illustrated in Figure 1.3. The first are those who wish to ascend the triangle, either from outside or from a low level. Many of these might be analytical chemists, for example an NIR spectroscopist who has seen the need to process his or her data and may wish some further insight into the methods being used. Or an organic chemist might wish to have the skills to optimise a synthesis, or a food chemist may wish to be able to interpret the tools for relating the results of a taste panel to chemical constituents. Possibly you have read a paper, attended a conference or a course or seen some software demonstrated. Or perhaps in the next-door laboratory, someone is already doing some chemometrics, perhaps you have heard about experimental design or principal components analysis and need some insight into the methods. Maybe you have some results but have little idea how to interpret them and perhaps by changing parameters using a commercial package you are deluged with graphs and not really certain whether they are meaningful. Some readers might be MSc or PhD students wishing to delve a little deeper into chemometrics.

The second group already has some mathematical background but wishes to enter the triangle from the side. Some readers of this book will be applied statisticians, often

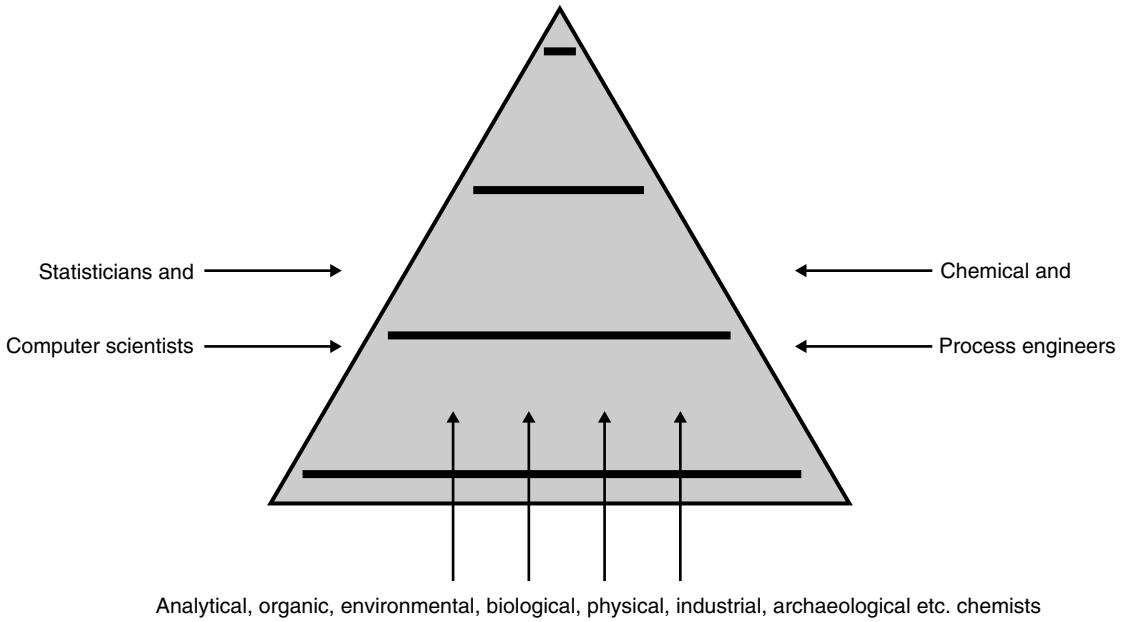

Figure 1.3
Groups of people with potential interest in this text

working in industry. Matrix algebra, significance tests and distributions are well known, but what is needed is to brush up on techniques as applied specifically to chemical problems. In some organisations there are specific data processing sections and this book is aimed as a particularly useful reference for professionals working in such an environment. Because there are not a large number of intensive courses in chemical data analysis, especially leading to degrees, someone with a general background in statistical data analysis who has moved job or is taking on extra responsibilities will find this book a valuable reference. Chemical engineers have a special interest in chemometrics and many are encountering the ideas when used to monitor processes.

1.2 Software and Calculations

The key to chemometrics is to understand how to perform meaningful calculations on data. In most cases these calculations are too complex to do by hand or using a calculator, so it is necessary to use some software.

The approach taken in this text, which differs from many books on chemometrics, is to understand the methods using numerical examples. Some excellent texts and reviews are more descriptive, listing the methods available together with literature references and possibly some examples. Others have a big emphasis on equations and output from packages. This book, however, is based primarily on how I personally learn and understand new methods, and how I have found it most effective to help students working with me. Data analysis is not really a knowledge based subject, but more a skill based subject. A good organic chemist may have an encyclopaedic knowledge of reactions in their own area. The best supervisor will be able to list to his or her students thousands of reactions, or papers or conditions that will aid their students, and with experience this knowledge base grows. In chemometrics, although there are quite a number of named methods, the key is not to learn hundreds of equations by heart, but to understand a few basic principles. These ideas, such as multiple linear regression, occur again and again but in different contexts. To become skilled in chemometric data analysis, what is required is practice in manipulating numbers, not an enormous knowledge base. Although equations are necessary for the formal description of methods, and cannot easily be avoided, it is easiest to understand the methods in this book by looking at numbers. So the methods described in this text are illustrated using numerical examples which are available for the reader to reproduce. The datasets employed in this book are available on the publisher's Website. In addition to the main text there are extensive problems at the end of each main chapter. All numerical examples are fairly small, designed so that you can check all the numbers yourselves. Some are reduced versions of larger datasets, such as spectra recorded at 5 nm rather than 1 nm intervals. Many real examples, especially in chromatography and spectroscopy, simply differ in size to those in this book. Also, the examples are chosen so that they are feasible to analyse fairly simply.

One of the difficulties is to decide what software to employ in order to analyse the data. This book is not restrictive and you can use any approach you like. Some readers like to program their own methods, for example in C or Visual Basic. Others may like to use a statistical packages such as SAS or SPSS. Some groups use ready packaged chemometrics software such as Pirouette, Simca, Unscrambler and several others on the market. One problem with using packages is that they are often very

focused in their facilities. What they do they do excellently, but if they cannot do what you want you may be stuck, even for relatively simple calculations. If you have an excellent multivariate package but want to use a Kalman filter, where do you turn? Perhaps you have the budget to buy another package, but if you just want to explore the method, the simplest implementation takes only an hour or less for an experienced Matlab programmer to implement. In addition, there are no universally agreed definitions, so a 'factor' or 'eigenvector' might denote something quite different according to the software used. Some software has limitations, making it unsuitable for many applications of chemometrics, a very simple example being the automatic use of column centring in PCA in most general statistical packages, whereas many chemometric methods involve using uncentred PCA.

Nevertheless, many of the results from the examples in this book can successfully be obtained using commercial packages, but be aware of the limitations, and also understand the output of any software you use. It is important to recognise that the definitions used in this book may differ from those employed by any specific package. Because there are a huge number of often incompatible definitions available, even for fairly common parameters, in order not to confuse the reader we have had to adopt one single definition for each parameter, so it is important to check carefully with your favourite package or book or paper if the results appear to differ from those presented in this book. It is not the aim of this text to replace an international committee that defines chemometrics terms. Indeed, it is unlikely that such a committee would be formed because of the very diverse backgrounds of those interested in chemical data analysis.

However, in this text we recommend that readers use one of two main environments.

The first is Excel. Almost everyone has some familiarity with Excel, and in Appendix A.4 specific features that might be useful for chemometrics are described. Most calculations can be performed quite simply using normal spreadsheet functions. The exception is principal components analysis (PCA), for which a small program must be written. For instructors and users of VBA (a programming language associated with Excel), a small macro that can be edited is available, downloadable from the publisher's Website. However, some calculations such as cross-validation and partial least squares (PLS), whilst possible to set up using Excel, can be tedious. It is strongly recommended that readers do reproduce these methods step by step when first encountered, but after a few times, one does not learn much from setting up the spreadsheet each time. Hence we also provide a package that contains Excel add-ins for VBA to perform PCA, PLS, MLR (multiple linear regression) and PCR (principal components regression), that can be installed on PCs which have at least Office 97, Windows 98 and 64 Mbyte memory. The software also contains facilities for validation. Readers of this book should choose what approach they wish to take.

A second environment, that many chemical engineers and statisticians enjoy, is Matlab, described in Appendix A.5. Historically the first significant libraries of programs in chemometrics first became available in the late 1980s. Quantum chemistry, originating in the 1960s, is still very much Fortran based because this was the major scientific programming environment of the time, and over the years large libraries have been developed and maintained, so a modern quantum chemist will probably learn Fortran. The vintage of chemometrics is such that a more recent environment to scientific programming has been adopted by the majority, and many chemometricians swap software using Matlab. The advantage is that Matlab is very matrix oriented and it is most

convenient to think in terms of matrices, especially since most data are multivariate. Also, there are special facilities for performing singular value decomposition (or PCA) and the pseudoinverse used in regression, meaning that it is not necessary to program these basic functions. The user interface of Matlab is not quite as user-friendly as Excel and is more suited to the programmer or statistician rather than the laboratory based chemist. However, there have been a number of recent enhancements, including links to Excel, that allow easy interchange of data, which enables simple programs to be written that transfer data to and from Excel. Also, there is no doubt at all that matrix manipulation, especially for complex algorithms, is quite hard in VBA and Excel. Matlab is an excellent environment for learning the nuts and bolts of chemometrics. A slight conceptual problem with Matlab is that it is possible to avoid looking at the raw numbers, whereas most users of Excel will be forced to look at the raw numerical data in detail, and I have come across experienced Matlab users who are otherwise very good at chemometrics but who sometimes miss quite basic information because they are not constantly examining the numbers – so if you are a dedicated Matlab user, look at the numerical information from time to time!

An ideal situation would probably involve using both Excel and Matlab simultaneously. Excel provides a good interface and allows flexible examination of the data, whereas Matlab is best for developing matrix based algorithms. The problems in this book have been tested both in Matlab and Excel and identical answers obtained. Where there are quirks of either package, the reader is guided. If you are approaching the triangle of Figure 1.3 from the sides you will probably prefer Matlab, whereas if you approach it from the bottom, it is more likely that Excel will be your choice.

Two final words of caution are needed. The first is that some answers in this book have been rounded to a few significant figures. Where intermediate results of a calculation have been presented, putting these intermediate results back may not necessarily result in exactly the same numerical results as retaining them to higher accuracy and continuing the calculations. A second issue that often perplexes new users of multivariate methods is that it is impossible to control the sign of a principal component (see Chapter 4 for a description of PCA). This is because PCs involve calculating square roots which may give negative as well as positive answers. Therefore, using different packages, or even the same package but with different starting points, can result in reflected graphs, with scores and loadings that are opposite in sign. It is therefore unlikely to be a mistake if you obtain PCs that are opposite in sign to those in this book.

1.3 Further Reading

There have been a large number of texts and review articles covering differing aspects of chemometrics, often aimed at a variety of audiences. This chapter summarises some of the most widespread. In most cases these texts will allow the reader to delve further into the methods introduced within this book. In each category only a few main books will be mentioned, but most have extensive bibliographies allowing the reader to access information especially from the primary literature.

1.3.1 General

The largest text in chemometrics is published by Massart and co-workers, part of two volumes [1,2]. These volumes provide an in-depth summary of many modern

chemometric methods, involving a wide range of techniques, and many references to the literature. The first volume, though, is strongly oriented towards analytical chemists, but contains an excellent grounding in basic statistics for measurement science. The books are especially useful as springboards for the primary literature. This text is a complete rewrite of the original book published in 1988 [3], which is still cited as a classic in the analytical chemistry literature.

Otto's book on chemometrics [4] is a welcome recent text, that covers quite a range of topics but at a fairly introductory level. The book looks at computing in general in analytical chemistry including databases, and instrumental data acquisition. It does not deal with the multivariate or experimental design aspects in a great deal of detail but is a very clearly written introduction for the analytical chemist, by an outstanding educator.

Beebe and co-workers at Dow Chemical have recently produced a book [5] which is useful for many practitioners, and contains very clear descriptions especially of multivariate calibration in spectroscopy. However there is a strong 'American School' originating in part from the pioneering work of Kowalski in NIR spectroscopy and process control, and whilst covering the techniques required in this area in an outstanding way, and is well recommended as a next step for readers of this text working in this application area, it lacks a little in generality, probably because of the very close association between NIR and chemometrics in the minds of some. Kramer has produced a somewhat more elementary book [6]. He is well known for his consultancy company and highly regarded courses, and his approach is less mathematical. This will suit some people very well, but may not be presented in a way that suits statisticians and chemical engineers.

One of the first ever texts in the area of chemometrics was co-authored by Kowalski [7]. The book is somewhat mathematical and condensed, but provided a good manual for the mathematically minded chemometrician of the mid-1980s, and is a useful reference. Kowalski also edited a number of symposium volumes in the early days of the subject. An important meeting, the NATO Advanced Study School in Cosenza, Italy, in 1983, brought together many of the best international workers in this area and the edited volume from this is a good snapshot of the state-of-the-art of the time [8], although probably the interest is more for the historians of chemometrics.

The present author published a book on chemometrics about a decade ago [9], which has an emphasis on signal resolution and minimises matrix algebra, and is an introductory tutorial book especially for the laboratory based chemist. The journal *Chemometrics and Intelligent Laboratory Systems* published regular tutorial review articles over its first decade or more of existence. Some of the earlier articles are good introductions to general subjects such as principal components analysis, Fourier transforms and Matlab. They are collected together as two volumes [10,11]. They also contain some valuable articles on expert systems.

Meloun and co-workers published a two volume text in the early 1990s [12,13]. These are very thorough texts aimed primarily at the analytical chemist. The first volume contains detailed descriptions of a large number of graphical methods for handling analytical data, and a good discussion of error analysis, and the second volume is a very detailed discussion of linear and polynomial regression.

Martens and Martens produced a recent text which gives quite a detailed discussion on how multivariate methods can be used in quality control [14], but covers several aspects of modern chemometrics, and so should be classed as a general text on chemometrics.

1.3.2 Specific Areas

There are a large number of texts and review articles dealing with specific aspects of chemometrics, interesting as a next step up from this book, and for a comprehensive chemometrics library.

1.3.2.1 Experimental Design

In the area of experimental design there are innumerable texts, many written by statisticians. Specifically aimed at chemists, Deming and Morgan produced a highly regarded book [15] which is well recommended as a next step after this text. Bayne and Rubin have written a clear and thorough text [16]. An introductory book discussing mainly factorial designs was written by Morgan as part of the Analytical Chemistry by Open Learning Series [17]. For mixture designs, involving compositional data, the classical statistical text by Cornell is much cited and recommended [18], but is quite mathematical.

1.3.2.2 Pattern Recognition

There are several books on pattern recognition and multivariate analysis. An introduction to several of the main techniques is provided in an edited book [19]. For more statistical in-depth descriptions of principal components analysis, books by Joliffe [20] and Mardia and co-authors [21] should be read. An early but still valuable book by Massart and Kaufmann covers more than just its title theme 'cluster analysis' [22] and provides clear introductory material.

1.3.2.3 Multivariate Curve Resolution

Multivariate curve resolution is the main topic of Malinowski's book [23]. The author is a physical chemist and so the book is oriented towards that particular audience, and especially relates to the spectroscopy of mixtures. It is well known because the first edition (in 1980) was one of the first major texts in chemometrics to contain formal descriptions of many common algorithms such as principal components analysis.

1.3.2.4 Multivariate Calibration

Multivariate calibration is a very popular area, and the much reprinted classic by Martens and Næs [24] is possibly the most cited book in chemometrics. Much of the text is based around NIR spectroscopy which was one of the major success stories in applied chemometrics in the 1980s and 1990s, but the clear mathematical descriptions of algorithms are particularly useful for a wider audience.

1.3.2.5 Statistical Methods

There are numerous books on general statistical methods in chemistry, mainly oriented towards analytical and physical chemists. Miller and Miller wrote a good introduction [25] that takes the reader through many of the basic significance tests, distributions, etc. There is a small amount on chemometrics in the final chapter. The Royal Society

of Chemistry publish a nice introductory tutorial text by Gardiner [26]. Caulcutt and Boddy's book [27] is also a much reprinted and useful reference. There are several other competing texts, most of which are very thorough, for example, in describing applications of the t-test, F-test and analysis of variance (ANOVA) but which do not progress much into modern chemometrics. If you are a physical chemist, Gans' viewpoint on deconvolution and curve fitting may suit you more [28], covering many regression methods, but remember that physical chemists like equations more than analytical chemists and so approach the topic in a different manner.

1.3.2.6 Digital Signal Processing and Time Series

There are numerous books on digital signal processing (DSP) and Fourier transforms. Unfortunately, many of the chemically based books are fairly technical in nature and oriented towards specific techniques such as NMR; however, books written primarily by and for engineers and statisticians are often quite understandable. A recommended reference to DSP contains many of the main principles [29], but there are several similar books available. For nonlinear deconvolution, Jansson's book is well known [30]. Methods for time series analysis are described in more depth in an outstanding and much reprinted book written by Chatfield [31].

1.3.2.7 Articles

We will not make very great reference to the primary literature in this text. Many of the authors of well regarded texts first published material in the form of research, review and tutorial articles, which then evolved into books. However, it is worth mentioning a very small number of exceptionally well regarded tutorial papers. A tutorial by Wold and co-workers on principal components analysis [32] in the 1980s is a citation classic in the annals of chemometrics. Geladi and Kowalski's tutorial [33] on partial least squares is also highly cited and a good introduction. In the area of moving average filters, Savitsky and Golay's paper [34] is an important original source.

1.3.3 Internet Resources

Another important source of information is via the Internet. Because the Internet changes very rapidly, it is not practicable in this text to produce a very comprehensive list of Websites; however, some of the best resources provide regularly updated links to other sites, and are likely to be maintained over many years.

A good proportion of the material in this book is based on an expanded version of articles originally presented in ChemWeb's e-zine the *Alchemist*. Registration is free [35] and past articles are in the chemometrics archive. There are several topics that are not covered in this book. Interested readers are also referred to an article which provides a more comprehensive list of Web resources [36].

Wiley's *Chemometrics World* is a comprehensive source of information freely available to registered users via their SpectroscopyNOW Website [37], and the datasets and software from this book are available via this Website.

There are one or two excellent on-line textbooks, mainly oriented towards statisticians. Statsoft have a very comprehensive textbook [38] that would allow readers to delve into certain topics introduced in this text in more detail. Hyperstat also produce an

on-line statistics textbook, mainly dealing with traditional statistical methods, but their Website also provides references to other electronic tutorial material [39], including Stockburger's book on multivariate statistics [40].

1.4 References

1. D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics Part A*, Elsevier, Amsterdam, 1997.
2. B. M. G. Vandeginste, D. L. Massart, L. M. C. Buydens, S. de Jong, P. J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics Part B*, Elsevier, Amsterdam, 1998.
3. D. L. Massart, B. G. M. Vandeginste, S. N. Deming, Y. Michotte, and L. Kaufman, *Chemometrics: a Textbook*, Elsevier, Amsterdam, 1988.
4. M. Otto, *Chemometrics: Statistics and Computer Applications in Analytical Chemistry*, Wiley-VCH, Weinheim, 1998.
5. K. R. Beebe, R. J. Pell and M. B. Seasholtz, *Chemometrics: a Practical Guide*, Wiley, New York, 1998.
6. R. Kramer, *Chemometrics Techniques for Quantitative Analysis*, Marcel Dekker, New York, 1998.
7. M. A. Sharaf, D. L. Illman and B. R. Kowalski, *Chemometrics*, Wiley, New York, 1996.
8. B. R. Kowalski (Editor), *Chemometrics: Mathematics and Statistics in Chemistry*, Reidel, Dordrecht, 1984.
9. R. G. Brereton, *Chemometrics: Applications of Mathematics and Statistics to Laboratory Systems*, Ellis Horwood, Chichester, 1990.
10. D. L. Massart, R. G. Brereton, R. E. Dessy, P. K. Hopke, C. H. Spiegelman and W. Wegscheider (Editors), *Chemometrics Tutorials*, Elsevier, Amsterdam, 1990.
11. R. G. Brereton, D. R. Scott, D. L. Massart, R. E. Dessy, P. K. Hopke, C. H. Spiegelman and W. Wegscheider (Editors), *Chemometrics Tutorials II*, Elsevier, Amsterdam, 1992.
12. M. Meloun, J. Militky and M. Forina, *Chemometrics for Analytical Chemistry*, Vol. 1, Ellis Horwood, Chichester, 1992.
13. M. Meloun, J. Militky and M. Forina, *Chemometrics for Analytical Chemistry*, Vol. 2, Ellis Horwood, Chichester, 1994.
14. H. Martens and M. Martens, *Multivariate Analysis of Quality*, Wiley, Chichester, 2000.
15. S. N. Deming and S. L. Morgan, *Experimental Design: a Chemometric Approach*, Elsevier, Amsterdam, 1994.
16. C. K. Bayne and I. B. Rubin, *Practical Experimental Designs and Optimisation Methods for Chemists*, VCH, Deerfield Beach, FL, 1986.
17. E. Morgan, *Chemometrics: Experimental Design*, Wiley, Chichester, 1995.
18. J. A. Cornell, *Experiments with Mixtures: Design, Models, and the Analysis of Mixture Data*, Wiley, New York, 2nd edn, 1990.
19. R. G. Brereton (Editor), *Multivariate Pattern Recognition in Chemometrics, Illustrated by Case Studies*, Elsevier, Amsterdam, 1992.
20. I. T. Joliffe, *Principal Components Analysis*, Springer-Verlag, New York, 1987.
21. K. V. Mardia, J. T. Kent and J. M. Bibby, *Multivariate Analysis*, Academic Press, London, 1979.
22. D. L. Massart and L. Kaufmann, *The Interpretation of Analytical Chemical Data by the Use of Cluster Analysis*, Wiley, New York, 1983.
23. E. R. Malinowski, *Factor Analysis in Chemistry*, Wiley, New York, 2nd edn, 1991.
24. H. Martens and T. Næs, *Multivariate Calibration*, Wiley, Chichester, 1989.
25. J. N. Miller and J. Miller, *Statistics for Analytical Chemistry*, Prentice-Hall, Hemel Hempstead, 1993.
26. W. P. Gardiner, *Statistical Analysis Methods for Chemists: a Software-based Approach*, Royal Society of Chemistry, Cambridge, 1997.
27. R. Caulcutt and R. Boddy, *Statistics for Analytical Chemists*, Chapman and Hall, London, 1983.

28. P. Gans, *Data Fitting in the Chemical Sciences: by the Method of Least Squares*, Wiley, Chichester, 1992.
29. P. A. Lynn and W. Fuerst, *Introductory Digital Signal Processing with Computer Applications*, Wiley, Chichester, 2nd edn, 1998.
30. P. A. Jansson (Editor), *Deconvolution: with Applications in Spectroscopy*, Academic Press, New York, 1984.
31. C. Chatfield, *Analysis of Time Series: an Introduction*, Chapman and Hall, London, 1989.
32. S. Wold, K. Esbensen and P. Geladi, *Chemom. Intell. Lab. Syst.*, **2**, 37 (1987).
33. P. Geladi and B. R. Kowalski, *Anal. Chim. Acta*, **185**, 1 (1986).
34. A. Savitsky and M. J. E. Golay, *Anal. Chem.*, **36**, 1627 (1964).
35. *www.chemweb.com*.
36. R. G. Brereton, Chemometrics on the Net, *Alchemist*, 2 April 2001 *(www.chemweb.com)*.
37. *www.spectroscopynow.com*.
38. *www.statsoft.com/textbook/stathome.html*.
39. *http://davidmlane.com/hyperstat/*.
40. *www.psychstat.smsu.edu/MultiBook/mlt00.htm*.