Data analysis comprises more than chemometrics, statistics or any single named discipline in seeking out the information contained in any given set of data. Happily, despite its title, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant* recognizes that simple fact and makes serious efforts to include examples of many of the various ways in which information about the underlying behavior of a system can be extracted from data obtained from the system. This includes both standard (for example, chemometrics and statistics) and nonstandard (for example, “Manhattan Distance”) ways to view and examine data. A short summary of elementary statistics is included in an appendix. While welcome, the necessary brevity of the treatment probably makes it too compressed for a reader to actually learn statistics from that presentation.

The book is organized according to application, rather than by technique. Thus, principal components largely are described in the chapter on pattern recognition, and partial least squares (PLS) is dealt with in the chapter on calibration. Every chapter includes several problems at the end to give readers practice in applying, and to promote thought about, the principles developed during the course of the chapter.

In the introductory chapter the author discusses the various ways different disciplines (physics, biology, and engineering) approach the examination of data, and the variety of backgrounds and mathematical sophistication these disciplines bring to this study. He then goes on to introduce his own philosophy of data analysis and how to learn it, including the importance of practical application. It is peppered with good advice (“Scale your variables appropriately before comparing them”, for example). To that end the author uses programs written in MATLAB and Excel, and through their use illustrates the concepts with numerous examples, both graphical and numerical. He also makes the programs and several of the data sets available to his readers on the Internet, an innovation in publishing that expands the book beyond what is contained between its covers. He concludes this chapter with a discussion and list of other resources, including both print and electronic (Internet-based) resources.

The second chapter, dealing with experimental designs, is a wonderful merging of statistical principles with more-sophisticated chemometric designs for experiments. The principles of applying distributional and numerical statistical tests to the data are explained, with numerical examples. The use of standard statistical hypothesis tests (t- and F-tests, for example) is a good example of the application of rigorous statistical principles to the analysis of chemometric algorithms and is extremely welcome in a book about chemometrics. The fact that so much material is packed into a book of finite size means, however, that some subjects are not explained in the detail needed to prevent the novice from falling into error. An example of this is the author’s discussion of cumulative normal distribution (CND) plots. The presentation and illustration of the application of this tool is virtually unique in the chemometric literature; however, guidelines for its use in distinguishing real effects from the variability of random data are absent. Cuthbert Daniel has shown that, especially for small numbers of samples, the variability due to randomness can mislead the novice into mistaking them for real effects (see C. Daniel and F. Wood, *Fitting Equations to Data* [Wiley, 1971] and also C. Daniel, *Applications of Statistics to Industrial Experimentation* [Wiley, 1976]). In the same books, Daniel presented illustrative examples based upon the CND of normal random numbers for comparison as a reference to which actual CND plots can be compared. As Brereton points out, before computers were widely available, compilations such as Daniel’s were the only way in which such information could be disseminated. Nowadays, however, because computers and appropriate software are readily available, users of these tools are well advised to create their own comparison plots by applying the tool (the CND plot, in this case) to known-random, known-normal synthetic data. Brereton initially applies these tests to simple, statistically based designs. Then more complex and sophisticated experimental designs (for example, mixture designs, response-surface designs, and so forth) are introduced, and the analysis based upon statistical principles continued as long as...
reasonably possible. Other characteristics of designs, such as the leverage of the design points and how the overall design affects them, are also treated.

The third chapter, “Signal Processing,” deals with methods of analyzing sequential signals. After a brief introductory passage presenting some of the common application areas, the chapter again starts with elementary methods (smoothing, curve fitting), and expands its coverage to discussions of the effects of noise and digitization. It then goes into descriptions of some of the more complex and sophisticated techniques (Savitzky-Golay methods, Fourier transforms, and so forth) and ends with discussions of the latest cutting-edge methodologies, which still are being examined largely by the research community. Examples are wavelet transforms, maximum entropy, and Bayesian methods.

The fourth chapter, “Pattern Recognition,” again follows the overall architecture of the book by starting with descriptive material and elementary methods (in this case “naked-eye” methods such as dendrograms). The chapter then goes into a fairly extensive examination of principal components, illustrating it with an application to spectra of chromatographic effluent, having overlapping peaks and using the application of principal components to the chromatographic data to separate the peaks.

The fifth chapter, “Calibration,” continues the pattern by starting with a one-variable example, presenting both the “classical” and “inverse” methods. Here again, as in previous chapters, secondary considerations are discussed as well as the main subject. The author describes and discusses effects of intercept, centering, and the locations of errors in the two cases. The concepts then are expanded to use multiple-regression and principal component analysis (PCA) as multivariate calibration methods. Following this is an extensive discussion of PLS in several of its varieties including the extension of the discussion to more-advanced topics, particularly trilinear PLS1 (the application of the PLS algorithm to a single dependent variable at a time). This extension requires use of the next mathematical step beyond vectors and matrices: tensors. Tensor analysis requires some industrial-strength math, far beyond what chemists are used to dealing with. For this reason it is questionable whether its inclusion in this book is appropriate, even though some of the more advanced researchers are investigating the topic. The chapter winds up with a discussion of methods for validating a calibration model once it is produced.

The sixth chapter, “Evolutionary Signals,” deals with what basically is a specialized application of time-series analysis; it goes into depth about the analysis of spectroscopic detection of chromatographic effluent. Examples of the applications examined are the use of PCA and PLS to separate the various components of a sample, when their chromatographic peaks are overlapped, and having separated them, to quantitate one or more of the components. The general problem is divided into multiple possible situations based upon the degree of overlap, from the easiest (simple partial overlap) to the most difficult (a small peak completely embedded under a large peak). Various auxiliary techniques are also discussed, and recommendations are made.

Appendix: There is only a single appendix, but the topics covered vary so widely that each one could have merited a separate appendix of its own. The appendix discusses the basics of vector and matrix math, the description of PCA and PLS in terms of matrices, basic statistics including hypothesis testing (and some tables of common statistics), and a discussion of how to use Excel and MATLAB for performing chemometric and other forms of data analysis. Again, inclusion of these topics is welcome, although the treatment probably makes the discussion too compressed for a reader to actually learn from the presentation. On the other hand, the topics can serve as introductions to more-extensive treatments, so that readers will be familiar with the terminology and approaches used when delving more fully into a topic of interest. They could also be valuable reviews for those readers who once knew the topics but whose knowledge has since gotten rusty.