HOW TO USE EXCEL® IN ANALYTICAL CHEMISTRY AND IN GENERAL SCIENTIFIC DATA ANALYSIS

HOW TO USE EXCEL® IN ANALYTICAL CHEMISTRY

AND IN GENERAL SCIENTIFIC DATA ANALYSIS

Robert de Levie
Bowdoin College, Brunswick, ME

CAMBRIDGE UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2001

Printed in the United Kingdom at the University Press, Cambridge

Typeface Utopia 9.25/13.5pt and Meta Plus System QuarkXPress® [SE]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data
de Levie, Robert.
How to use Excel in analytical chemistry and in general scientific
data analysis / Robert de Levie.
p. cm.
ISBN 0 521 64282 5 (hbk.). – ISBN 0 521 64484 4 (pbk.)
1. Chemistry, Analytic–Statistical methods–Data processing.
2. Science–Statistical methods–Data processing. 3. Microsoft
Excel (Computer file) 4. Electronic spreadsheets. I. Title.
QD75.4.S8D4 1999
543′.00285′5369–dc21 98-50543 CIP

ISBN 0 521 64282 5 hardback
ISBN 0 521 64484 4 paperback

Disclaimer of warranty

The macros in this book are intended for instructional purposes, and are offered without any warranty, express or implied, of being free of errors and/or of being suitable for any particular application. They should not be used in any application that could possibly result in bodily harm and/or in damage to or loss of property. The author and his publisher disclaim any and all liability for direct, incidental, or consequential damages resulting from the use of any of the suggestions, programs, or parts of programs in this book.

This work is in copyright. Nevertheless, permission is granted for the macros to be used provided that such use does not result in computer code that is made available in connection with or conditioned upon a sale, rental, or other monetary transaction. However, insofar as some of these macros also contain material from the Numerical Recipes, and distribution of the latter, separately or as a macro, will require the explicit copyright permission from Numerical Recipes Software.

CONTENTS

Preface *[xi]*

PART I INTRODUCTION TO USING THE SPREADSHEET

1 How to use Excel *[1]*
1.1 Starting Windows *[2]*
1.2 A first look at the spreadsheet *[3]*
1.3 A simple spreadsheet and graph *[7]*
1.3a Making a graph in Excel 97 or a more recent version *[9]*
1.3b Making a graph in Excel 5 or Excel 95 *[11]*
1.4 Addressing a spreadsheet cell *[13]*
1.5 More on graphs *[15]*
1.6 Mathematical operations *[21]*
1.7 Error messages *[26]*
1.8 Naming and annotating cells *[27]*
1.9 Viewing the spreadsheet *[28]*
1.10 Printing *[29]*
1.11 Help! *[30]*
1.12 The case of the changing options *[31]*
1.13 Importing macros and data *[32]*
1.14 Differences between the various versions of Excel *[33]*
1.15 Some often-used spreadsheet commands *[35]*
1.16 Changing the default settings *[36]*
1.17 Summary *[37]*

PART II STATISTICS AND RELATED METHODS

2 Introduction to statistics *[39]*
2.1 Gaussian statistics *[39]*
2.2 Replicate measurements *[45]*

2.3 The propagation of imprecision from a single parameter *[51]*
2.4 The propagation of imprecision from multiple parameters *[54]*
2.5 The weighted average *[58]*
2.6 Least-squares fitting to a proportionality *[60]*
2.7 Least-squares fitting to a general straight line *[66]*
2.8 Looking at the data *[71]*
2.9 What is 'normal'? *[73]*
2.10 Poissonian statistics *[78]*
2.11 How likely is the improbable? *[79]*
2.12 Summary *[83]*

3 More on least squares *[90]*
3.1 Multi-parameter fitting *[90]*
3.2 Fitting data to a quadratic *[93]*
3.3 Least squares for equidistant data: smoothing and differentiation *[94]*
3.4 Weighted least squares *[99]*
3.5 Another example of weighted least squares: enzyme kinetics *[103]*
3.6 Non-linear data fitting *[105]*
3.6a Some kinetic data *[106]*
3.6b A double exponential *[109]*
3.6c False minima *[115]*
3.6d Enzyme kinetics revisited *[116]*
3.6e SolverAid *[117]*
3.7 Summary *[118]*

PART III CHEMICAL EQUILIBRIUM

4 Acids, bases, and salts *[121]*
4.1 The mass action law and its graphical representations *[121]*
4.2 Conservation laws, proton balance, and pH calculations *[127]*
4.3 Titrations of monoprotic acids and bases *[130]*
4.4 Schwartz and Gran plots *[133]*
4.5 The first derivative *[136]*
4.6 A more general approach to data fitting *[142]*
4.7 Buffer action *[146]*
4.8 Diprotic acids and bases, and their salts *[148]*
4.9 Polyprotic acids and bases, and their salts *[152]*
4.10 Activity corrections *[155]*
4.11 A practical example *[161]*
4.12 Summary *[172]*

5 Other ionic equilibria *[175]*
5.1 Complex formation *[175]*

5.2 Chelation *[180]*
5.3 Extraction *[182]*
5.4 Solubility *[185]*
5.5 Precipitation and dissolution *[190]*
5.6 Precipitation titrations *[194]*
5.7 The von Liebig titration *[200]*
5.8 The graphical representation of electrochemical equilibria *[204]*
5.9 Redox titrations *[211]*
5.10 Redox buffer action *[217]*
5.11 Summary *[220]*

PART IV INSTRUMENTAL METHODS

6 Spectrometry, chromatography, and voltammetry *[223]*
6.1 Spectrometric pK_a determination *[223]*
6.2 Multi-component spectrometric analysis 1 *[225]*
6.3 Multi-component spectrometric analysis 2 *[227]*
6.4 The absorbance–absorbance diagram *[231]*
6.5 Chromatographic plate theory 1 *[234]*
6.6 Chromatographic plate theory 2 *[239]*
6.7 Peak area, position, and width *[243]*
6.8 Determining the number of theoretical plates *[245]*
6.9 Optimizing the mobile phase velocity *[248]*
6.10 Polarography *[251]*
6.11 Linear sweep and cyclic voltammetry 1 *[257]*
6.12 Linear sweep and cyclic voltammetry 2 *[261]*
6.13 Summary *[263]*

PART V MATHEMATICAL METHODS

7 Fourier transformation *[265]*
7.1 Introduction to Fourier transformation *[265]*
7.2 Interpolation and filtering *[277]*
7.3 Differentiation *[285]*
7.4 Aliasing and leakage *[288]*
7.5 Convolution *[295]*
7.6 Deconvolution *[304]*
7.7 Summary *[309]*

8 Standard mathematical operations *[311]*
8.1 The Newton–Raphson method *[311]*
8.2 Non-linear least squares *[313]*
8.3 Signal averaging *[314]*

8.4 Lock-in amplification *[316]*
8.5 Data smoothing *[318]*
8.6 Peak fitting *[324]*
8.7 Integration *[328]*
8.8 Differentiation *[331]*
8.9 Semi-integration and semi-differentiation *[335]*
8.10 Interpolation *[336]*
8.11 Matrix manipulation *[337]*
8.12 Overflow *[341]*
8.13 Summary *[343]*

9 Numerical simulation of chemical kinetics *[345]*
9.1 Introduction *[345]*
9.2 The explicit method *[346]*
9.2a First-order kinetics *[346]*
9.2b Numerical accuracy *[348]*
9.2c Dimerization kinetics *[350]*
9.2d A user-defined function to make the spreadsheet more efficient *[351]*
9.2e Trimerization kinetics *[353]*
9.2f Monomer–dimer kinetics *[354]*
9.2g Polymerization kinetics *[355]*
9.3 Implicit numerical simulation *[359]*
9.3a First-order kinetics *[359]*
9.3b Dimerization kinetics *[360]*
9.3c Trimerization kinetics *[361]*
9.3d Monomer–dimer kinetics *[362]*
9.3e Polymerization kinetics *[363]*
9.4 Some applications *[365]*
9.4a Autocatalysis *[365]*
9.4b Heterogeneous catalysis *[367]*
9.4c The steady-state approximation *[369]*
9.4d Oscillating reactions: the Lotka model *[372]*
9.5 Summary *[374]*

PART VI SPREADSHEET PROGRAMMING

10 Some useful macros *[375]*
10.1 What is a macro? *[375]*
10.1a The macro module *[376]*
10.1b Reading and modifying the contents of a single cell *[378]*
10.1c Reading and modifying the contents of a block of cells *[382]*
10.1d Two different approaches to modifying a block of cells *[384]*
10.1e Numerical precision *[387]*

10.1f Communication via boxes *[389]*
10.1g Subroutines *[393]*
10.2 A case study: interpolating in a set of equidistant data *[395]*
10.2a Step-by-step *[396]*
10.2b The finished product *[401]*
10.3 Propagation of imprecision *[405]*
10.4 Installing and customizing a macro *[410]*
10.4a Installing external macros *[410]*
10.4b Assigning a shortcut key *[411]*
10.4c Embedding in a menu *[412]*
10.4d Miscellany *[414]*
10.5 Fourier transformation *[415]*
10.5a Forward Fourier transformation *[416]*
10.5b Descriptive notes *[420]*
10.5c A bidirectional Fourier transformation macro *[421]*
10.6 Convolution and deconvolution *[426]*
10.7 Weighted least squares *[432]*
10.7a The algorithm *[432]*
10.7b Implementation *[433]*
10.8 More about Solver *[442]*
10.8a Adding uncertainty estimates to Solver *[442]*
10.8b Incorporating Solver into your macro *[448]*
10.9 Smoothing and differentiating equidistant data *[449]*
10.10 Semi-integration and semi-differentiation *[460]*
10.11 Reducing data density *[463]*
10.12 An overview of VBA *[466]*
10.12a Objects *[467]*
10.12b Properties and methods *[467]*
10.12c Data types *[468]*
10.12d Expressions *[469]*
10.12e Statements *[470]*
10.12f Active regions *[470]*
10.12g Cells *[471]*
10.12h Ranges *[472]*
10.12i Subroutines *[473]*
10.12j Macros *[474]*
10.12k Functions *[474]*
10.12l Message boxes *[475]*
10.12m Input boxes *[475]*
10.12n Dialog boxes *[476]*
10.12o Collective statements *[476]*
10.12p For…Next loops *[477]*
10.12q Do loops *[478]*

10.12r Conditional statements *[479]*
10.12s Exit statements *[480]*
10.13 Summary *[481]*

Index *[483]*

PREFACE

Chemistry is an experimental science, and primarily lives in the laboratory. No book on spreadsheets will change that. However, many aspects of chemical analysis have significant quantitative, mathematical components, and many of these can be illustrated effectively using spreadsheets. At the same time, the spreadsheet is a very accessible tool for data analysis, an activity common to all of the physical sciences. This book emphasizes the use of spreadsheets in data analysis, while at the same time illustrating some of the underlying principles. The basic strength of spreadsheets was summarized by the name of the very first spreadsheet, VisiCalc, in that it facilitates the *visualization of calculations*, and thereby can help to make theory and data analysis come to life.

Spreadsheets are well-recognized for their near-immediate response to changes in their input parameters, for their ease in making graphs, for their open format and intuitive layout, and for their forgiving error-handling. For these reasons they are usually considered to be the most easily learned computer tools for numerical data analysis. Moreover, they are widely available, as they are often bundled with standard word processors.

Spreadsheets used to be far inferior to the so-called higher-level computer languages in terms of the mathematical manipulations they would support. In particular, numerical methods requiring iterations used to be awkward on a spreadsheet. Fortunately, this has changed with the introduction, in version 5 of Excel, of a macro language (Visual BASIC for Applications, or VBA) that allows the inclusion of standard computer code. Now the immediacy of the spreadsheet and the convenience of its graphical representations can be combined with the wide availability in the literature of sophisticated higher-level programs to make the spreadsheet a powerful scientific as well as didactic tool.

Of course, spreadsheets cannot do everything. While they make quite competent graphs, they lack some of the stunning three-dimensional representations of more specialized, graphics-oriented packages. Moreover, spreadsheets cannot handle symbolic mathematics, and they are unsuitable for highly specialized, computation-intensive tasks such as molecular modeling. However, they are unmatched for ease of learning, and for general availability and price.

Spreadsheets can be used as glorified calculators. There is nothing wrong with that, but there is no need to write about such rather obvious applications here, since there are already

a sufficient number of books devoted to this topic. Instead I have tried to illustrate some of the more subtle aspects of data analysis, some of the more specialized features of chemical equilibrium, some of the more abstract underpinnings of modern chemical instrumentation, and some of the finer points of numerical simulation. The choice and sequence of topics closely follows the order in which these are typically encountered in textbooks in analytical chemistry, so that this book can readily be used in courses in quantitative or instrumental chemical analysis. Since the choice of topics is rather wide, the reader is welcome to pick and choose from among these according to his or her own preference and need.

Most chapters start with a brief summary of the theory in order to put the spreadsheet exercises in perspective, and to define the nomenclature used. The standard versions of Excel 95 through Excel 2000 for Windows 95 or Windows 98 are used. Many exercises use the Solver and the Analysis ToolPak, both of which are available in the standard Excel packages but may have to be loaded separately, as add-ins, in case this was not done initially. When use of chapter 10 is contemplated, the VBA help file should also be loaded.

While the specific spreadsheet instructions in this book are for Excel 97 on IBM-compatible computers, they can all be implemented readily (i.e., with no or very minor modifications) in Excel 5 (for Windows 3.1), Excel 95 (for Windows 95), Excel 98 (for the Mac), or Excel 2000 (for Windows 98 or Windows 2000). Moreover, I have indicated where Excel 5 and Excel 95 require different procedures from those in Excel 97, 98, or 2000, namely in their handling of graphs and macros. There are some minor differences between the Excel versions for IBM-compatible and MacIntosh computers. The most important of these are listed in chapter 1; none of them are serious.

Many exercises also work in the earlier versions (1 through 4) of Excel. However, these earlier versions cannot handle VBA macros, so that those spreadsheet exercises that use macros for weighted least squares, fast Fourier transformation, and convolution, cannot be run with versions preceding Excel 5. (Specifically, these are exercises 3.4 and beyond in chapter 3, and all exercises in chapter 7.) Moreover, the macros described in chapter 10 cannot be used in these earlier versions.

Many of the exercises in this book can also be run on spreadsheets other than Excel. In that case, however, apart from the impossibility to import higher-level computer programs into the spreadsheet, the user may also lack the convenience of a powerful multi-parameter non-linear least squares routine such as Solver. Given the choice of writing a book to fit all spreadsheets, or one that exploits the extra power of modern Excel, I have opted for the latter.

The purpose of this book is not to provide its readers with a set of prepackaged routines, into which they merely enter some constants. Instead, the emphasis is on letting the readers gain enough familiarity and experience to enable them to use spreadsheets independently, and in other scientific contexts, while at the same time illustrating a number of interesting features of analytical chemistry. In most cases, no theory is derived, and the reader should consult standard texts on statistics and on quantitative and instrumental chemical analysis for the necessary background information, as well as for a perspective on the strengths and weaknesses of the various methods.

The reader may discover some unavoidable parallelism between the material in this book and that in my undergraduate textbook, *Principles of Quantitative Chemical Analysis*,

McGraw-Hill, 1997, and even some remnants of my *Spreadsheet Workbook for Quantitative Chemical Analysis*, McGraw-Hill, 1992. This is partially because I have retained some of the didactic innovations introduced in these earlier texts, such as an emphasis on the progress of a titration rather than on the traditional titration curve, the use of buffer strength rather than buffer value, and the use of the abbreviations h and k in the description of electrochemical equilibria. However, the present text exploits the power of Excel to go far beyond what was possible in those earlier books.

For a few problems that would require the reader to write some rather complex macros, these have been provided. They are fully documented and explained in chapter 10, and can be downloaded from http://uk.cambridge.org/chemistry/resources/delevie Note that their code is readily accessible, and that the reader is not only encouraged to modify them, but is given the tools to do so. Again, the idea is to empower the reader to incorporate existing higher-language code into macros, in order to increase the reach and usefulness of Excel.

The first chapter introduces the reader to the software; it can be speed-read or skipped by those already familiar with Windows- or Mac-based spreadsheets. The last chapter discusses macros, which can convert a spreadsheet into a powerful computing tool. Sandwiched between these are the four main parts of this book: statistics and related methods, chemical equilibrium, instrumental methods, and mathematical analysis. These parts can be used independently, although some aspects introduced in chapters 2 and 3 are used in subsequent chapters, and the spreadsheet instructions tend to become somewhat less detailed as the text progresses.

The treatment of statistics is focused on explicit applications of both linear and nonlinear least-squares methods, rather than on the alphabet soup (F, Q, R, T, etc.) of available tests. However, within that rather narrow framework, many practical aspects of error analysis and curve fitting are considered. They are chosen to illustrate the now almost two centuries old dictum of de Laplace that the theory of probability is merely common sense confirmed by calculation.

Since the spreadsheet is eminently capable of doing tedious numerical work, exact mathematical expressions are used as much as possible in the examples involving chemical equilibria. Similarly, the treatment of titrations emphasizes the use of exact mathematical relations, which can then be fitted to experimental data. In some of the exercises, the student first computes, say, a make-believe titration curve, complete with simulated noise, and is then asked to extract from that curve the relevant parameters. The make-believe curve is clearly a stand-in for using experimental data, which can be subjected to the very same analysis.

For the more instrumental methods of quantitative chemical analysis, I have taken a rather eclectic approach, merely illustrating some aspects that are especially suitable for spreadsheet exploration, such as Beer's law and its applications to the analysis of multicomponent mixtures, chromatographic plate theory, polarography, and cyclic voltammetry.

Because of its important place in modern chemical instrumentation, an entire chapter is devoted to Fourier transformation and its applications, including convolution and deconvolution. The chapter on mathematical analysis illustrates several aspects of signal handling traditionally included in courses in instrumental analysis, such as signal averaging and synchronous detection, that deal with the relation between signal and noise. Its main focus,

however, is on numerical analysis, and it covers such aspects as finding roots and fitting curves, integrating, differentiating, smoothing, and interpolating data. Numerical solution of differential equations is the focus of chapter 9, where we discuss a number of kinetic schemes, partially to counterbalance the earlier emphasis on equilibrium behavior.

The final chapter describes the nitty-gritty of macros, and illustrates how they can be used to make the spreadsheet do many amazing things in exchange for relatively little effort on the part of the user, who can simply incorporate pre-existing, well-documented, widely available algorithms.

The aim of this book, then, is to illustrate numerical applications rather than to explain fundamental concepts. Theory is mentioned only insofar as it is needed to define the nomenclature used, or to explain the approach taken. This book can therefore be used in conjunction with a regular textbook in analytical chemistry, in courses on quantitative or instrumental chemical analysis. It can also serve as a stand-alone introduction to modern spreadsheet use for students of chemistry and related scientific disciplines, provided they are already familiar with some of the underlying scientific concepts. Because of its emphasis on exercises, this book is also suitable for individual, home use.

I am grateful to Drs. T. Moisio and M. Heikonen of Valio Ltd, Helsinki, for permission to use their unpublished experimental data in chapter 4, to Professor Phillip Barak of the University of Minnesota for permission to include his adaptive-degree least-squares algorithm in chapter 10, and to Numerical Recipes Software of Cambridge Massachusetts for permission to use some subroutines from the Numerical Recipes.

I am indebted to Professors Nancy Gordon and Gale Rhodes of the University of Southern Maine, Professor Barry Lavine of Clarkson University, Professors Panos Nikitas and Nanna Papa-Louisi of Aristotle University, as well as to Mr. William H. Craig and Professors Andrew Vogt, George Benke, and Daniel E. Martire of Georgetown University, for their many helpful and constructive comments and suggestions. I am especially indebted to Professor Joseph T. Maloy of Seton Hall University for his extensive advice.

I am grateful to Georgetown University for a sabbatical leave of absence, which gave me the unbroken time to work on this book, and to Professor Nancy Gordon of the University of Southern Maine in Portland, Maine, and Professor Panos Nikitas of Aristotle University of Thessaloniki, Greece, for their gracious hospitality during the writing of it. Finally I thank my son, Mark, for his invaluable help in getting me started on this project, and my wife, Jolanda, for letting me finish it.

User comments, including corrections of errors, and suggestions for additional topics and/or exercises, are most welcome. I can be reached at RDELEVIE@BOWDOIN.EDU Corrections will be posted in the web site

http://uk.cambridge.org/chemistry/resources/ delevie

From this web site you can also download the data set used in section 4.11, and the macros of chapter 10.