Martin H. Trauth
MATLAB® Recipes for Earth Sciences
Martin H. Trauth

MATLAB® Recipes for Earth Sciences

With text contributions by Robin Gebbers and Norbert Marwan and illustrations by Elisabeth Sillmann

With 77 Figures and a CD-ROM
Various books on data analysis in earth sciences have been published during the last ten years, such as *Statistics and Data Analysis in Geology* by JC Davis, *Introduction to Geological Data Analysis* by ARH Swan and M Sandilands, *Data Analysis in the Earth Sciences Using MATLAB®* by GV Middleton or *Statistics of Earth Science Data* by G Borradaile. Moreover, a number of software packages have been designed for earth scientists such as the ESRI product suite ArcGIS or the freeware package GRASS for generating geographic information systems, ERDAS IMAGINE or RSINC ENVI for remote sensing and GOCAD and SURFER for 3D modeling of geologic features. In addition, more general software packages as IDL by RSINC and MATLAB® by The MathWorks Inc. or the freeware software OCTAVE provide powerful tools for the analysis and visualization of data in earth sciences.

Most books on geological data analysis contain excellent theoretical introductions, but no computer solutions to typical problems in earth sciences, such as the book by JC Davis. The book by ARH Swan and M Sandilands contains a number of examples, but without the use of computers. G Middleton’s book firstly introduces MATLAB as a tool for earth scientists, but the content of the book mainly reflects the personal interests of the author, rather than providing a complete introduction to geological data analysis. On the software side, earth scientists often encounter the problem that a certain piece of software is designed to solve a particular geologic problem, such as the design of a geoinformation system or the 3D visualization of a fault scarp. Therefore, earth scientists have to buy a large volume of software products, and even more important, they have to get used to it before being in the position to successfully use it.

This book on *MATLAB Recipes for Earth Sciences* is designed to help undergraduate and PhD students, postdocs and professionals to learn methods of data analysis in earth sciences and to get familiar with MATLAB, the leading software for numerical computations. The title of the book is an appreciation of the book *Numerical Recipes* by WH Press and others that is still very popular after initially being published in 1986. Similar to the book by Press and others, this book provides a minimum amount of
theoretical background, but then tries to teach the application of all methods by means of examples. The software MATLAB is used since it provides numerous ready-to-use algorithms for most methods of data analysis, but also gives the opportunity to modify and expand the existing routines and even develop new software. The book contains numerous MATLAB scripts to solve typical problems in earth sciences, such as simple statistics, time-series analysis, geostatistics and image processing. The book comes with a compact disk, which contains all MATLAB recipes and example data files. All MATLAB codes can be easily modified in order to be applied to the reader’s data and projects.

Whereas undergraduates participating in a course on data analysis might go through the entire book, the more experienced reader will use only one particular method to solve a specific problem. To facilitate the use of this book for the various readers, I outline the concept of the book and the contents of its chapters.

1. Chapter 1 – This chapter introduces some fundamental concepts of samples and populations, it links the various types of data and questions to be answered from these data to the methods described in the following chapters.

2. Chapter 2 – A tutorial-style introduction to MATLAB designed for earth scientists. Readers already familiar with the software are advised to proceed directly to the following chapters.

3. Chapter 3 and 4 – Fundamentals in univariate and bivariate statistics. These chapters contain very basic things how statistics works, but also introduce some more advanced topics such as the use of surrogates. The reader already familiar with basic statistics might skip these two chapters.

4. Chapter 5 and 6 – Readers who wish to work with time series are recommended to read both chapters. Time-series analysis and signal processing are tightly linked. A solid knowledge of statistics is required to successfully work with these methods. However, the two chapters are more or less independent from the previous chapters.

5. Chapter 7 and 8 – The second pair of chapters. From my experience, reading both chapters makes a lot of sense. Processing gridded spatial data and analyzing images has a number of similarities. Moreover, aerial
photographs and satellite images are often projected upon digital elevation models.

6. Chapter 9 – Data sets in earth sciences are tremendously increasing in the number of variables and data points. Multivariate methods are applied to a great variety of types of large data sets, including even satellite images. The reader particularly interested in multivariate methods is advised to read Chapters 3 and 4 before proceeding to this chapter.

I hope that the various readers will now find their way through the book. Experienced MATLAB users familiar with basic statistics are invited to proceed to Chapters 5 and 6 (the time series), Chapters 7 and 8 (spatial data and images) or Chapter 9 (multivariate analysis) immediately, which contain both an introduction to the subjects as well as very advanced and special procedures for analyzing data in earth sciences. It is recommended to the beginners, however, to read Chapters 1 to 4 carefully before getting into the advanced methods.

I thank the NASA/GSFC/METI/ERSDAC/JAROS and U.S./Japan ASTER Science Team and the director Mike Abrams for allowing me to include the ASTER images in the book. The book has benefitted from the comments of a large number of colleagues and students. I gratefully acknowledge my colleagues who commented earlier versions of the manuscript, namely Robin Gebbers, Norbert Marwan, Ira Ojala, Lydia Olaka, Jim Renwick, Jochen Rössler, Rolf Romer, and Annette Witt. Thanks also to the students Mathis Hein, Stefanie von Lonski and Matthias Gerber, who helped me to improve the book. I very much appreciate the expertise and patience of Elisabeth Sillmann who created the graphics and the complete page design of the book. I also acknowledge Courtney Esposito leading the author program at The MathWorks, Claudia Olrogge and Annegret Schumann at Mathworks Deutschland, Wolfgang Engel at Springer, Andreas Bohlen and Brunhilde Schulz at UP Transfer GmbH. I would like to thank Thomas Schulmeister who helped me to get a campus license for MATLAB at Potsdam University. The book is dedicated to Peter Koch, the late system administrator of the Department of Geosciences who died during the final writing stages of the manuscript and who helped me in all kinds of computer problems during the last few years.

Potsdam, September 2005

Martin Trauth
Contents

Preface V

1 Data Analysis in Earth Sciences 1
1.1 Introduction 1
1.2 Collecting Data 1
1.3 Types of Data 3
1.4 Methods of Data Analysis 7

2 Introduction to MATLAB 11
2.1 MATLAB in Earth Sciences 11
2.2 Getting Started 12
2.3 The Syntax 15
2.4 Data Storage 19
2.5 Data Handling 19
2.6 Scripts and Functions 21
2.7 Basic Visualization Tools 25

3 Univariate Statistics 29
3.1 Introduction 29
3.2 Empirical Distributions 29
3.3 Example of Empirical Distributions 36
3.4 Theoretical Distributions 41
3.5 Example of Theoretical Distributions 50
3.6 The t–Test 51
3.7 The F–Test 53
3.8 The $\chi^2$–Test 56
4 Bivariate Statistics

4.1 Introduction 61
4.2 Pearson’s Correlation Coefficient 61
4.3 Classical Linear Regression Analysis and Prediction 68
4.5 Analyzing the Residuals 72
4.6 Bootstrap Estimates of the Regression Coefficients 74
4.7 Jackknife Estimates of the Regression Coefficients 76
4.8 Cross Validation 77
4.9 Reduced Major Axis Regression 78
4.10 Curvilinear Regression 80

5 Time-Series Analysis

5.1 Introduction 85
5.2 Generating Signals 85
5.3 Autospectral Analysis 91
5.4 Crossspectral Analysis 97
5.5 Interpolating and Analyzing Unevenly-Spaced Data 101
5.6 Nonlinear Time-Series Analysis (by N. Marwan) 106

6 Signal Processing

6.1 Introduction 119
6.2 Generating Signals 120
6.3 Linear Time-Invariant Systems 121
6.4 Convolution and Filtering 124
6.5 Comparing Functions for Filtering Data Series 127
6.6 Recursive and Nonrecursive Filters 129
6.7 Impulse Response 131
6.8 Frequency Response 134
6.9 Filter Design 139
6.10 Adaptive Filtering 143

7 Spatial Data

7.1 Types of Spatial Data 151
7.2 The GSHHS Shoreline Data Set 152
7.3 The 2-Minute Gridded Global Elevation Data ETOPO2 154
7.4 The 30-Arc Seconds Elevation Model GTOPO30 157
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.5</td>
<td>The Shuttle Radar Topography Mission SRTM</td>
<td>158</td>
</tr>
<tr>
<td>7.6</td>
<td>Gridding and Contouring Background</td>
<td>161</td>
</tr>
<tr>
<td>7.7</td>
<td>Gridding Example</td>
<td>164</td>
</tr>
<tr>
<td>7.8</td>
<td>Comparison of Methods and Potential Artifacts</td>
<td>169</td>
</tr>
<tr>
<td>7.9</td>
<td>Geostatistics (by R. Gebbers)</td>
<td>173</td>
</tr>
<tr>
<td>8</td>
<td><strong>Image Processing</strong></td>
<td>193</td>
</tr>
<tr>
<td>8.1</td>
<td>Introduction</td>
<td>193</td>
</tr>
<tr>
<td>8.2</td>
<td>Data Storage</td>
<td>194</td>
</tr>
<tr>
<td>8.3</td>
<td>Importing, Processing and Exporting Images</td>
<td>199</td>
</tr>
<tr>
<td>8.4</td>
<td>Importing, Processing and Exporting Satellite Images</td>
<td>204</td>
</tr>
<tr>
<td>8.5</td>
<td>Georeferencing Satellite Images</td>
<td>207</td>
</tr>
<tr>
<td>8.6</td>
<td>Digitizing from the Screen</td>
<td>209</td>
</tr>
<tr>
<td>9</td>
<td><strong>Multivariate Statistics</strong></td>
<td>213</td>
</tr>
<tr>
<td>9.1</td>
<td>Introduction</td>
<td>213</td>
</tr>
<tr>
<td>9.2</td>
<td>Principal Component Analysis</td>
<td>214</td>
</tr>
<tr>
<td>9.3</td>
<td>Cluster Analysis</td>
<td>221</td>
</tr>
<tr>
<td>9.4</td>
<td>Independent Component Analysis (by N. Marwan)</td>
<td>225</td>
</tr>
</tbody>
</table>

**General Index**

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>231</td>
</tr>
</tbody>
</table>