Quantitative Methods of Data Analysis for the Physical Sciences and Engineering

This book provides thorough and comprehensive coverage of most of the new and important quantitative methods of data analysis for college and graduate students and practitioners. In recent years, data analysis methods have exploded alongside advanced computing power, and an understanding of such methods is critical to getting the most out of data and for extracting signal from noise. The book excels in explaining difficult concepts through simple explanations and detailed explanatory illustrations. Most unique is the focus on confidence limits for power spectra and their proper interpretation, something rare or completely missing in other books. Likewise, there is a thorough discussion of how to assess uncertainty via use of Expectancy, and easy-to-apply and -understand Bootstrap method. The book is written so that descriptions of each method are as self-contained as possible. Many examples are presented to clarify interpretations, as are user tips in highlighted boxes.

Douglas G. Martinson is a Lamont Research Professor in the Division of Ocean and Climate Physics at Columbia University’s Lamont–Doherty Earth Observatory. As a physical oceanographer who researches the role of polar oceans in global climate, his research involves the collection of a large amount of data and considerable quantitative analysis. He developed the course on Quantitative Methods of Data Analysis as an Adjunct Professor for the Department of Earth and Environmental Sciences at Columbia University, and received an Outstanding Teacher Award in 2004.
Quantitative Methods of Data Analysis for the Physical Sciences and Engineering

DOUGLAS G. MARTINSON

Columbia University
To the love of my life, my wife Rhonda
Contents

Preface  xiii
How to Use This Book  xiii
Acknowledgments xv

Part I     Fundamentals

1 The Nature of Data and Analysis  3
1.1 Analysis  3
1.2 Data Nomenclature  3
1.3 Representing Discrete Data and Functions as Vectors  5
1.4 Data Limits  6
1.5 Data Errors  8
1.6 Practical Issues  12

2 Probability Theory  15
2.1 Overview  15
2.2 Definitions  16
2.3 Probability  18
2.4 Univariate Distributions  19
2.5 Multivariate Distributions  27
2.6 Moments of Random Variables  31
2.7 Common Distributions and Their Moments  50
2.8 Take-Home Points  59
2.9 Questions  60

3 Statistics  62
3.1 Overview  62
3.2 Estimation  62
3.3 Estimating the Distribution  66
3.4 Point Estimates  69
3.5 Principle of Maximum Likelihood (An Important Principle)  76
## Contents

3.6 Interval Estimates  
3.7 Hypothesis Testing  
3.8 Sample-Based Distributions  
3.9 Take-Home Points  
3.10 Questions  

### Part II Fitting Curves to Data

4 Interpolation  
4.1 Overview  
4.2 Piecewise Continuous Interpolants  
4.3 Continuous Interpolants  
4.4 Take-Home Points  
4.5 Questions  

5 Smoothed Curve Fitting  
5.1 Overview  
5.2 Introduction  
5.3 Functional Form of the Curve  
5.4 Defining “Best” Fit  
5.5 Determining Parameter Values for a Best-Fit Curve  
5.6 Orthogonal Fitting of a Straight Line  
5.7 Assessing Uncertainty in Optimal Parameter Values  
5.8 Assessing the Fit of the Best-Fit Curve  
5.9 Take-Home Points  
5.10 Questions  

6 Special Curve Fitting  
6.1 Overview  
6.2 Weighted Curve Fits  
6.3 Constrained Fits  
6.4 Robust Curve Fits  
6.5 Regression/Calibration  
6.6 Correlation Coefficient  
6.7 Take-Home Points  
6.8 Questions
## Contents

### Part III  Sequential Data Fundamentals

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Serial Products</td>
<td>205</td>
</tr>
<tr>
<td>7.1</td>
<td>Overview</td>
<td>207</td>
</tr>
<tr>
<td>7.2</td>
<td>Statistical Considerations</td>
<td>209</td>
</tr>
<tr>
<td>7.3</td>
<td>Convolution</td>
<td>222</td>
</tr>
<tr>
<td>7.4</td>
<td>Serial Correlation</td>
<td>234</td>
</tr>
<tr>
<td>7.5</td>
<td>Take-Home Points</td>
<td>249</td>
</tr>
<tr>
<td>7.6</td>
<td>Questions</td>
<td>250</td>
</tr>
<tr>
<td>8</td>
<td>Fourier Series</td>
<td>252</td>
</tr>
<tr>
<td>8.1</td>
<td>Overview</td>
<td>252</td>
</tr>
<tr>
<td>8.2</td>
<td>Introduction</td>
<td>253</td>
</tr>
<tr>
<td>8.3</td>
<td>Periodic Functions</td>
<td>265</td>
</tr>
<tr>
<td>8.4</td>
<td>Fourier Series</td>
<td>270</td>
</tr>
<tr>
<td>8.5</td>
<td>Take-Home Points</td>
<td>270</td>
</tr>
<tr>
<td>8.6</td>
<td>Questions</td>
<td>270</td>
</tr>
<tr>
<td>9</td>
<td>Fourier Transform</td>
<td>271</td>
</tr>
<tr>
<td>9.1</td>
<td>Overview</td>
<td>271</td>
</tr>
<tr>
<td>9.2</td>
<td>Discrete Periodic Data</td>
<td>271</td>
</tr>
<tr>
<td>9.3</td>
<td>Discrete Sine and Cosine Transforms</td>
<td>282</td>
</tr>
<tr>
<td>9.4</td>
<td>Continuous Sine and Cosine Transforms</td>
<td>288</td>
</tr>
<tr>
<td>9.5</td>
<td>The Fourier Transform</td>
<td>289</td>
</tr>
<tr>
<td>9.6</td>
<td>Fourier Transform of Non-Periodic Data</td>
<td>296</td>
</tr>
<tr>
<td>9.7</td>
<td>Fourier Transform Properties</td>
<td>301</td>
</tr>
<tr>
<td>9.8</td>
<td>Fourier Transform Theorems</td>
<td>311</td>
</tr>
<tr>
<td>9.9</td>
<td>Fast Fourier Transform</td>
<td>319</td>
</tr>
<tr>
<td>9.10</td>
<td>Take-Home Points</td>
<td>320</td>
</tr>
<tr>
<td>9.11</td>
<td>Questions</td>
<td>321</td>
</tr>
<tr>
<td>10</td>
<td>Fourier Sampling Theory</td>
<td>322</td>
</tr>
<tr>
<td>10.1</td>
<td>Overview</td>
<td>322</td>
</tr>
<tr>
<td>10.2</td>
<td>Sampling Theorem</td>
<td>323</td>
</tr>
<tr>
<td>10.3</td>
<td>Relationship between Discrete and Continuous Transform</td>
<td>338</td>
</tr>
<tr>
<td>10.4</td>
<td>Other Sampling Considerations</td>
<td>347</td>
</tr>
<tr>
<td>10.5</td>
<td>Take-Home Points</td>
<td>348</td>
</tr>
<tr>
<td>10.6</td>
<td>Questions</td>
<td>348</td>
</tr>
</tbody>
</table>
11 Spectral Analysis
11.1 Overview
11.2 Noise in the Spectrum
11.3 More Stable Estimates of the Fourier Coefficients
11.4 Spectral Estimation in Practice
11.5 Bootstrap Testing with Time Series
11.6 Take-Home Points
11.7 Questions

12 Cross-Spectral Analysis
12.1 Overview
12.2 Joint PDF Moments in the Time Domain
12.3 Frequency Domain Estimation of the ccf
12.4 Statistical Considerations
12.5 Take-Home Points
12.6 Questions

13 Filtering and Deconvolution
13.1 Overview
13.2 Frequency Domain Representation
13.3 Special Types of Filters
13.4 Practical Considerations
13.5 Inverse Filtering (Deconvolution)
13.6 Exact (Deterministic) Deconvolution
13.7 Best-Fit Deconvolution
13.8 Take-Home Points
13.9 Questions

14 Linear Parametric Modeling
14.1 Overview
14.2 Discrete Linear Stochastic Process Models
14.3 Model Identification and Solution
14.4 Parameter Estimation
14.5 Forecasting
14.6 Parametric Spectral Estimation
14.7 Take-Home Points
14.8 Questions
14.9 Time Series References
<table>
<thead>
<tr>
<th>15</th>
<th>Empirical Orthogonal Function (EOF) Analysis</th>
<th>495</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.1</td>
<td>Overview</td>
<td>495</td>
</tr>
<tr>
<td>15.2</td>
<td>Introduction</td>
<td>495</td>
</tr>
<tr>
<td>15.3</td>
<td>Eigenvector Analysis</td>
<td>499</td>
</tr>
<tr>
<td>15.4</td>
<td>Principal Components (PC)</td>
<td>510</td>
</tr>
<tr>
<td>15.5</td>
<td>Singular Spectrum Analysis (SSA)</td>
<td>524</td>
</tr>
<tr>
<td>15.6</td>
<td>Take-Home Points</td>
<td>533</td>
</tr>
<tr>
<td>15.7</td>
<td>Questions</td>
<td>533</td>
</tr>
</tbody>
</table>

**Appendix 1** Overview of Matrix Algebra

**Appendix 2** Uncertainty Analysis

References

Index
Preface

This book is the outcome of a one-semester graduate class taught in the Department of Earth and Environmental Sciences at Columbia University, although the book could be used over two or even three semesters, if desired. I have taught this class since 1985, having taken over from a departing marine seismologist who had taught the course as one on Fourier analysis, the only topic that computers of the day were capable of performing, because of the development of the Fast Fourier Transform. However, at that time computers were rapidly becoming powerful enough to allow application of methods requiring more power and memory. New methods were sprouting yearly, and as the computers grew faster, previously sluggish methodologies were becoming realizable. At the time I started teaching the course, there were no textbooks (none!) that gave a thorough introduction to the primary methods. Numerical Recipes – published in the early 1980s – did present a brief overview and the computer code necessary to run nearly every method, and it was a godsend. It occurred to me that my class notes should be converted to a book to fill this void. Over the last 30 years many other books have been published, but in my opinion there is still a need for an introductory-level book that spans a broad number of the most useful techniques. Regardless of its introductory nature, I have tried to give the reader a complete enough understanding to allow him or her to properly apply the methods while avoiding common pitfalls and misunderstandings.

I try to present the methods following a few fundamental themes: e.g., Principle of Maximum Likelihood for deriving optimal methods, and Expectancy for estimating uncertainty. I hope this makes these important themes better understood and the material easier to grasp.

How to Use This Book

This book is designed to fill many needs, according to the level of the student. Some, like myself, see the methods clearly if they understand their complete derivation, while others don’t require that detailed understanding. In an effort to satisfy both readerships, I have placed complete derivations in boxes highlighted with 25 percent grayscale: these boxes are optional, and the reader is free, if preferred, to skip the box and go straight to the answer (all equations in derivation boxes are prefaced by “D” – for example, “D5.1”). There are student exercises at the end of each chapter, some of which require
computing. I do not present code because it changes so quickly, but I do show some MATLAB code in the solution manual. Data for exercises requiring such can be found at www.cambridge.org/martinson. Most of the examples in the book are taken from the natural sciences, although they are presented so as to be understandable to anyone. Special user tips are included in boxes highlighted with 15 percent grayscale. I have attempted to make each chapter stand on its own (as far as possible), so the reader doesn’t need to have read the entire book in order to understand material from previous chapters. This should make the book good for easy use and reference for a particular method.

Read it, practice, and when not sure what road to take, take all possible roads and then determine which is the most appropriate for your particular analysis. Then, maybe present several results explaining the differences, and why you favor the method you choose.
Acknowledgments

As with all books evolving from a course, one must acknowledge the considerable input from students and teaching assistants. As any teacher knows, it is usually the one teaching who learns more than anyone – when first teaching this course there were numerous derivations that only were partly developed, then a “miracle occurred” that skipped some “intuitively obvious” steps to the final result. No such skipped steps occur within this book. Over the years, excellent questions from students that I could not answer on the spot forced me to fill in many aspects of the material. So I offer a heartfelt thanks to those who stumped me in class. In that same vein, I would appreciate hearing about any errors still present in the book. The class has benefited from some incredibly smart and motivated teaching assistants, and many of the exercises appearing at the end of chapters originated from them (special thanks go to Sharon Stammerjohn, Chen Chen, and Darren McKee, among many others). Unfortunately, as I transformed my class notes into a textbook, my wife, Rhonda, became a writer’s widow for nearly a year – I can’t thank her enough for all the support she has given me. And finally, but not least, thanks to my editor, Matt Lloyd, at Cambridge University Press, who was a constant source of improvements and encouragement!

Finally, that ubiquitous message accompanying all such books: any errors in the book are strictly mine. Oh, and the other statement: any views expressed in this book (and there are many) are entirely mine. Enjoy!