

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.

Cambridge University Press

978-1-107-02976-7 — Quantitative Methods of Data Analysis for the Physical Sciences and Engineering

Douglas G. Martinson

Frontmatter

[More Information](#)

Quantitative Methods of Data Analysis for the Physical Sciences and Engineering

DOUGLAS G. MARTINSON

Columbia University



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press

978-1-107-02976-7 — Quantitative Methods of Data Analysis for the Physical Sciences and Engineering

Douglas G. Martinson

Frontmatter

[More Information](#)

CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781107029767

DOI: 10.1017/9781139342568

© Douglas G. Martinson 2018

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2018

Printed in the United Kingdom by TJ International Ltd. Padstow Cornwall

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data

Names: Martinson, Douglas G.

Title: Quantitative methods of data analysis for the physical sciences and engineering /

Douglas G. Martinson (Columbia University, New York)

Description: Cambridge, United Kingdom ; New York, NY : Cambridge University Press, 2018.

Identifiers: LCCN 2017055413 | ISBN 9781107029767 (hbk.) | ISBN 1107029767 (hbk.)

Subjects: LCSH: Statistics. | Physical sciences – Statistical methods. | Engineering – Statistical methods.

Classification: LCC QA276 .M3375 2018 | DDC 519.5–dc23

LC record available at <https://lcn.loc.gov/2017055413>

ISBN 978-1-107-02976-7 Hardback

Additional resources for this publication at www.cambridge.org/martinson

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

To the love of my life, my wife Rhonda

Cambridge University Press
978-1-107-02976-7 — Quantitative Methods of Data Analysis for the Physical Sciences and Engineering
Douglas G. Martinson
Frontmatter
[More Information](#)

Contents

| | | |
|---------------|--|------------------|
| | <i>Preface</i> | <i>page</i> xiii |
| | How to Use This Book | xiii |
| | <i>Acknowledgments</i> | xv |
| Part I | Fundamentals | 1 |
| 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 |

| | | |
|----------------|---|------------|
| viii | Contents | |
| | 3.6 Interval Estimates | 80 |
| | 3.7 Hypothesis Testing | 86 |
| | 3.8 Sample-Based Distributions | 96 |
| | 3.9 Take-Home Points | 100 |
| | 3.10 Questions | 101 |
| Part II | Fitting Curves to Data | 103 |
| 4 | Interpolation | 105 |
| | 4.1 Overview | 105 |
| | 4.2 Piecewise Continuous Interpolants | 110 |
| | 4.3 Continuous Interpolants | 124 |
| | 4.4 Take-Home Points | 126 |
| | 4.5 Questions | 127 |
| 5 | Smoothed Curve Fitting | 128 |
| | 5.1 Overview | 128 |
| | 5.2 Introduction | 128 |
| | 5.3 Functional Form of the Curve | 129 |
| | 5.4 Defining “Best” Fit | 130 |
| | 5.5 Determining Parameter Values for a Best-Fit Curve | 138 |
| | 5.6 Orthogonal Fitting of a Straight Line | 158 |
| | 5.7 Assessing Uncertainty in Optimal Parameter Values | 159 |
| | 5.8 Assessing the Fit of the Best-Fit Curve | 172 |
| | 5.9 Take-Home Points | 176 |
| | 5.10 Questions | 176 |
| 6 | Special Curve Fitting | 178 |
| | 6.1 Overview | 178 |
| | 6.2 Weighted Curve Fits | 178 |
| | 6.3 Constrained Fits | 185 |
| | 6.4 Robust Curve Fits | 193 |
| | 6.5 Regression/Calibration | 193 |
| | 6.6 Correlation Coefficient | 195 |
| | 6.7 Take-Home Points | 200 |
| | 6.8 Questions | 201 |

| | Contents | ix |
|-----------------|---|-----|
| Part III | Sequential Data Fundamentals | 205 |
| 7 | Serial Products | 207 |
| | 7.1 Overview | 207 |
| | 7.2 Statistical Considerations | 209 |
| | 7.3 Convolution | 222 |
| | 7.4 Serial Correlation | 234 |
| | 7.5 Take-Home Points | 249 |
| | 7.6 Questions | 250 |
| 8 | Fourier Series | 252 |
| | 8.1 Overview | 252 |
| | 8.2 Introduction | 253 |
| | 8.3 Periodic Functions | 253 |
| | 8.4 Fourier Series | 265 |
| | 8.5 Take-Home Points | 270 |
| | 8.6 Questions | 270 |
| 9 | Fourier Transform | 271 |
| | 9.1 Overview | 271 |
| | 9.2 Discrete Periodic Data | 271 |
| | 9.3 Discrete Sine and Cosine Transforms | 282 |
| | 9.4 Continuous Sine and Cosine Transforms | 288 |
| | 9.5 The Fourier Transform | 289 |
| | 9.6 Fourier Transform of Non-Periodic Data | 296 |
| | 9.7 Fourier Transform Properties | 301 |
| | 9.8 Fourier Transform Theorems | 311 |
| | 9.9 Fast Fourier Transform | 319 |
| | 9.10 Take-Home Points | 320 |
| | 9.11 Questions | 321 |
| 10 | Fourier Sampling Theory | 322 |
| | 10.1 Overview | 322 |
| | 10.2 Sampling Theorem | 323 |
| | 10.3 Relationship between Discrete and Continuous Transform | 338 |
| | 10.4 Other Sampling Considerations | 347 |
| | 10.5 Take-Home Points | 348 |
| | 10.6 Questions | 348 |

| | | |
|-----------|---|-----|
| x | Contents | |
| 11 | Spectral Analysis | 350 |
| 11.1 | Overview | 350 |
| 11.2 | Noise in the Spectrum | 351 |
| 11.3 | More Stable Estimates of the Fourier Coefficients | 357 |
| 11.4 | Spectral Estimation in Practice | 391 |
| 11.5 | Bootstrap Testing with Time Series | 400 |
| 11.6 | Take-Home Points | 403 |
| 11.7 | Questions | 404 |
| 12 | Cross-Spectral Analysis | 406 |
| 12.1 | Overview | 406 |
| 12.2 | Joint PDF Moments in the Time Domain | 406 |
| 12.3 | Frequency Domain Estimation of the ccf | 414 |
| 12.4 | Statistical Considerations | 419 |
| 12.5 | Take-Home Points | 423 |
| 12.6 | Questions | 424 |
| 13 | Filtering and Deconvolution | 425 |
| 13.1 | Overview | 425 |
| 13.2 | Frequency Domain Representation | 427 |
| 13.3 | Special Types of Filters | 430 |
| 13.4 | Practical Considerations | 436 |
| 13.5 | Inverse Filtering (Deconvolution) | 437 |
| 13.6 | Exact (Deterministic) Deconvolution | 437 |
| 13.7 | Best-Fit Deconvolution | 446 |
| 13.8 | Take-Home Points | 454 |
| 13.9 | Questions | 455 |
| 14 | Linear Parametric Modeling | 456 |
| 14.1 | Overview | 456 |
| 14.2 | Discrete Linear Stochastic Process Models | 458 |
| 14.3 | Model Identification and Solution | 469 |
| 14.4 | Parameter Estimation | 476 |
| 14.5 | Forecasting | 479 |
| 14.6 | Parametric Spectral Estimation | 479 |
| 14.7 | Take-Home Points | 491 |
| 14.8 | Questions | 491 |
| 14.9 | Time Series References | 492 |

| | Contents | xi |
|------|---|-----|
| 15 | Empirical Orthogonal Function (EOF) Analysis | 495 |
| 15.1 | Overview | 495 |
| 15.2 | Introduction | 495 |
| 15.3 | Eigenvector Analysis | 499 |
| 15.4 | Principal Components (PC) | 510 |
| 15.5 | Singular Spectrum Analysis (SSA) | 524 |
| 15.6 | Take-Home Points | 533 |
| 15.7 | Questions | 533 |
| | <i>Appendix 1 Overview of Matrix Algebra</i> | 535 |
| | <i>Appendix 2 Uncertainty Analysis</i> | 572 |
| | <i>References</i> | 596 |
| | <i>Index</i> | 599 |

Cambridge University Press
978-1-107-02976-7 — Quantitative Methods of Data Analysis for the Physical Sciences and Engineering
Douglas G. Martinson
Frontmatter
[More Information](#)

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!

Cambridge University Press
978-1-107-02976-7 — Quantitative Methods of Data Analysis for the Physical Sciences and Engineering
Douglas G. Martinson
Frontmatter
[More Information](#)
