Book Reviews

Published online: 03 May 2017.

Submit your article to this journal

Article views: 207

View Crossmark data

Taylor & Francis
Taylor & Francis Group

Exposure-Response Modeling: Methods and Practical Implementation.
Jixian Wang. Boca Raton, FL: Chapman & Hall/CRC Press, 2015, xix + 331 pp., $89.95(H), ISBN: 978-1-46-657320-8

The main thread of Jixian Wang's book is exposure-response (ER) modeling, with pharmacology, drug development, and epidemiological exposure as the main application areas. *Exposure-Response Modeling* surveys many key statistical models and methods, including mixed-effects models, measurement error models, time-to-event data models, dynamic and time series models, nonparametric and semiparametric regression models, Bayesian approaches, causal inference, sequential analysis, decision trees, and experimental design. Although most of these statistical methods are well known and widely used in many fields, the book also covers approaches unique to modeling ER relationships such as joint models for exposure and response, joint models for exposure and time-to-event, and differential equation models for pharmacokinetics/pharmacodynamics (PK/PD). The comprehensive survey of methods for ER relationships provided in this book will be interesting for applied researchers and for those wishing to quickly get an overview of ER modeling.

Chapter 1 provides a brief introduction to exposure-response (ER) models using three applied examples, which provide an overview of the purpose of the book. Chapter 2 introduces several standard models for exposure (PK models) and ER models—in particular, differential equation (DE) models for PK/PD. The chapter also discusses a variety of statistical models that can be applied in these settings, including linear, nonlinear, partial linear, semiparametric, and nonparametric regression models.

Each of Chapters 3–6 focuses on one type of data with corresponding ER models. Chapter 3 presents various mixed-effects models such as linear, nonlinear, generalized linear/nonlinear, and nonparametric/semiparametric mixed-effects models for longitudinal data (including both continuous and discrete longitudinal data). The inference issues for these mixed-effects models are discussed, although a more comprehensive survey on mixed-effects models for longitudinal data can be found in an earlier book by Wu and Zhang (2006). Chapter 4 discusses joint modeling of exposure and response. The chapter primarily adopts a measurement error perspective, with regression calibration used for parameter estimation and inference. Besides this simultaneous modeling approach, a two-step or sequential modeling (SM) strategy is discussed for fitting the joint model. Similar ideas appear in structural equation modeling (Kline 2016). In Chapter 5, a time-to-event response model is considered for ER modeling. Some basic time-to-event or survival models, including Cox models, accelerated failure model, competing risk models, models for recurrent events, frailty models, and models for interval-censored data are introduced for ER relationships. Time-varying covariates and semiparametric regression approaches are also discussed. Joint modeling of exposure (longitudinal data) and time-to-event response is briefly discussed, but the book would benefit from a more thorough coverage of this topic. Chapter 6 concerns dynamic ER relationships based on differential equation models, a topic that is not familiar to many statisticians. Simple PK

compartment models, indirect response models, and disease process (dynamic) models are discussed. These topics are also relatively new to many statisticians, but unfortunately, they are not covered thoroughly either.

Chapters 7–9 cover the topics of inference and decision-making approaches for ER models. In particular, Chapter 7 reviews Bayesian inference and decision analysis, discussing decision-making issues with respect to dose selection, multi-state models, and multiple objectives. Bayesian approaches for meta analyses of ER relationships are also introduced, along with applied examples. Causal inference for ER models is covered in Chapter 8 with a focus on adjustment of confounding bias. An important application of ER modeling is to determine the dose of drugs and optimal administration of drugs to achieve the best treatment effect. In Chapter 9, the author uses models for dose-exposure and exposure-response relationships to perform dose determination and adjustment. Dynamic treatment strategies, decision tree approaches, dose monitoring, and experimental design issues are carefully discussed in this chapter.

Chapter 10 introduces software tools for ER modeling, besides standard statistical software tools such as R and SAS, NONEM, a popular PK/PD modeling tool (implemented in Fortran), is briefly reviewed. Finally, some basic statistical concepts and approaches are briefly summarized in an Appendix.

In summary, the greatest strength of this book is that the models and methodologies are always motivated and explained by applications and examples, which effectively communicates to readers the basic ideas behind complex methodologies. Also, practical implementation and computer code are discussed alongside the methods, which will help readers to apply the methods to their own data. The complex methods are clearly presented without using sophisticated mathematical terminology, which has advantages and disadvantages. On the one hand, it is certainly useful for practitioners and applied researchers. However, young investigators interested in doing research in the field of ER modeling may wish that the book's coverage were more thorough in places, for example, when it comes to more recent developments.

References

Kline, R. B. (2016), *Principles and Practice of Structural Equation Modeling* (4th ed.), New York: Guilford Press. [458]

Wu, H., and Zhang, J. T. (2006), *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modeling Approaches*, New York: Wiley. [458]

Hulin Wu
University of Texas Health Science Center at Houston

Handbook of Design and Analysis of Experiments.
Angela Dean, Max Morris, John Stufken, and Derek Bingham (eds.). Boca Raton, FL: Chapman & Hall/CRC Press, 2015, xix+940 pp., $125.95(H), ISBN: 978-1-46-650433-2.

The *Handbook of Design and Analysis of Experiments* appears in the series entitled *Handbooks of Modern Statistical Methods*.

The aims and scope of the series are to cover the state of the art in the theory and applications of statistical methodology, and to present a good balance of theory and application through a synthesis of the key methodological developments and examples and case studies using real data. The volumes should be of primary interest to researchers and graduate students from (bio)statistics, but also appeal to scientists where the methodology is applied to real problems.

The *Handbook of Design and Analysis of Experiments* includes seven sections: (i) General principles; (ii) Designs for linear models; (iii) Designs accommodating multiple factors; (iv) Optimal design for nonlinear and spatial models; (v) Computer experiments; (vi) Cross-cutting issues; and (vii) Design for contemporary applications. Each section contains between two and five chapters. All chapters have been written by leading researchers. The quality and length of the chapters varies substantially.

Most of the chapters indeed provide a thorough overview of the state of the art in the theory of a specific subfield of design and analysis of experiments. However, virtually none of the chapters discuss applications to any appreciable depth. So, there is certainly not a good balance between theory and application. In addition, the mathematical level of most chapters is very high, due to which the handbook will, in my opinion, lack appeal to scientists other than researchers in design and analysis of experiments.

In terms of breadth, the *Handbook of Design and Analysis of Experiments* goes beyond well-known textbooks on the design and analysis of experiments, such as Atkinson, Donev, and Tobias (2007), Goos and Jones (2011), and Fedorov and Leonov (2013). Some of the handbook's chapters provide a broad overview of the existing work, and go beyond existing specialized textbooks by including recent developments and summarizing recently published series of papers. Nearly all authors stayed within their comfort zone and focused on design approaches they published on, thereby failing to integrate various streams of literature.

For instance, the handbook includes a chapter on regular screening designs and another on nonregular screening designs, but there is no comparison between the two kinds of screening designs, which may convince applied scientists that nonregular designs are worth considering even in the event regular alternatives are available. Moreover, increasingly so, practitioners are aware of the flexible optimal design methodology and consider nonorthogonal screening designs rather than the regular and nonregular screening designs discussed in the handbook. A comparison of the optimal design approach and the orthogonal design approach would have been a useful addition to the handbook. This would provide a true synthesis and integrate various streams of literature. A similar comment could be made concerning the chapter "Multistratum Fractional Factorial Designs," where the focus is exclusively on combinatorial constructions of regular designs. More flexible optimal experimental designs may, however, be more efficient and allow for more estimable effects.

Two chapters that are nice reads are the ones entitled "Designs for Selected Nonlinear Models" and "Designs for Generalized Linear Models." Despite the difficult nature of the topic, the authors (Biedermann and Yang and Atkinson and Woods) managed to write chapters that should be readable and usable by any motivated researcher. I do miss an in-depth discussion of Bayesian or pseudo-Bayesian optimal design for nonlinear and generalized linear models, however. Due to the ever-increasing computing speed and the availability of efficient numerical integration techniques, the (pseudo-)Bayesian approach is no longer computationally infeasible. The (pseudo-)Bayesian approach should therefore have received more attention in these chapters, and throughout the entire handbook. It would have been great, for instance, to see how (pseudo-)Bayesian designs (and maximin designs) compare to the locally optimal designs for the Michaelis–Menten model and the exponential model in the chapter "Designs for Selected Nonlinear Models." In that chapter, Figure 14.8 visualizes the drop in efficiency of locally optimal designs in the event the model parameters are misspecified. It would have been interesting to see similar pictures for (pseudo-)Bayesian (and maximin) designs.

In the chapter "Design for Discrete Choice Experiments," Grossmann and Schwabe spend a substantial amount of space summarizing their own results. However, unlike other chapter authors, they do provide a nonnegligible discussion of alternative approaches. While they do not go into much technical detail, they spend several pages discussing computational approaches. This kind of discussion is missing entirely, for instance in the chapter "Response Surface Experiments and Designs" (Khuri and Mukhopadhyay), which merely lists standard designs and models for quantitative experimental factors and for mixtures, ignoring the common complications that qualitative factors often need to be studied too, multi-factor or multi-ingredient constraints are often present and restricted randomization issues render standard designs infeasible.

In conclusion, the handbook fails to provide the synthesis and the balance between theory and application promised in the aims and scope of the series "Handbooks of Modern Statistical Methods." Most chapters are far too theoretical to be useful and accessible to applied scientists. Moreover, the way in which most topics are covered is strongly biased by the chapters' authors. Of course, as the handbook's editors write in the preface, *Not every topic can be covered in great detail in a book such as this one. Indeed, much more could be said on the important topics of screening designs, Bayesian designs, and clinical trials, for example, ...* I agree with this, as well as with the editors' statement that the handbook *gives a taste of the broad range of uses of experimental design, the current knowledge in these areas, and some indications for further reading on related topics.*

References

Atkinson, A. C., Donev, A. N., and Tobias, R. (2007), *Optimum Experimental Designs, With SAS*, Oxford: Oxford University Press. [459]

Fedorov, V. V., and Leonov, S. L. (2013), *Optimal Design for Nonlinear Response Models*, New York: Chapman & Hall. [459]

Goos, P., and Jones, B. (2011), *Optimal Design of Experiments: A Case Study Approach*, New York: Wiley. [459]

Peter Goos
KU Leuven University of Antwerp

The Seven Pillars of Statistical Wisdom. Stephen M. Stigler. Cambridge, MA, and London, England: Harvard University Press, 2016, 230 pp., $22.95(P), ISBN: 978-0-67-408891-7.

Professor Stigler has been informing, educating, and entertaining statisticians from a historical perspective for well over 40 years. While many of his publications (e.g., Stigler 1986) are History of Statistics, with capital H and S, he has also given us many gems (e.g., Stigler 1999), which are history of statistics with small h and s: highly readable dissertations with a historico-statistical approach to a topic, an idea, a time, a place, or a person. He sees statistics everywhere, and so his writings intertwine our field with the wider societies and cultures in which it is embedded. Invariably he excites our curiosity, gives us new insights, and reveals surprises.

This charming book is in his best tradition. It presents an indirect answer to the question every generation of statisticians must consider "What is Statistics?" His approach is to identify and then discuss seven principles, "seven pillars that have supported our field in different ways in the past and promise to do so into the indefinite future" (p. 2). His view is that each was "revolutionary when introduced, and each remains a deep and important conceptual advance" (p. 2). He also believes firmly that "These ideas are not part of Mathematics, nor are they part of Computer Science. They are centrally of Statistics..." (p. 11). I am sure all readers of this review have reflected on these issues at one time or another: what do we know and do, and how does that differ from what is known and done by others, who do not call themselves statisticians? Stigler's book can be viewed as the distillation of a career-long reflection on these questions.

I was strongly tempted to list the seven pillars, and discuss them here, but I have resisted. It seems to me that would be a serious spoiler, as you, the reader, should first have the pleasure of deciding upon your own seven pillars of statistical wisdom. Then you can purchase, borrow, or steal the book and compare your list with Stigler's. At that point you will also have the opportunity to read the fascinating historical highways and byways that he shows around each pillar.

I had the privilege of being in the audience of Stigler's Presidential Invited Address at the 2014 JSM, where he presented his seven pillars, and discussed some of the material in this book. It was an unforgettable event. If you are impatient, you can experience something similar by viewing his presentation at https://ww2.amstat.org/meetings/jsm/2014/webcasts/index.cfm

Two important questions remain, the first being: Who should read this book? As I have already implied, all readers of *JASA* reviews, and all attendees at JSM Presidential Invited Addresses, but who else? The book's level of sophistication varies from very basic to quite high. If you gave this book to a high school student interested in statistics, they would be able to follow some of it, perhaps 20%, though I surmise that quite a lot of the other 80% would be inspirational. For graduate students, much more would be accessible, and for many, there would be "aha moments," where, by virtue of the stories Stigler tells, and his references to contemporary treatments, they might suddenly "get" something that had previously been presented to them as an obscure theoretical notion. For the rest of us, a concise

account of the way one of our most respected colleagues sees our field will be of great interest, all the more so in the case of Stephen Stigler, as his views are so well informed by his historical perspective.

The final question is: Only seven? Stigler is well aware that all seven of his principles "date back at least to the first half of the twentieth century, some to antiquity" (p. 195). He goes on to say that "None of them is out of date, but we may still ask if more is needed in the modern age" (p. 195). He hints at an eighth pillar, and remarks "History suggests that this will not appear easily or in one step" (p. 203). This wonderful book closes on a note of optimism regarding our future.

References

Stigler, S. M. (1986), *The History of Statistics: The Measurement of Uncertainty before 1900*, Cambridge, MA: Harvard University Press. [460]
—— (1999), *Statistics on the Table: The History of Statistical Concepts and Methods*, Cambridge, MA: Harvard University Press. [460]

Terry Speed
The Walter and Eliza Hall Institute of Medical Research

Spatial and Spatio-Temporal Geostatistical Modeling and Kriging. José-María Montero, Gema Fernández-Avilés, and Jorge Mateu. New York: Wiley, 2015, xxii+357 pp., $105.00(H), ISBN: 978-1-11-841318-0.

The book begins with a brief description of its scope, which is modeling and analysis of real-valued data indexed by locations in space and/or time. Admirably, the authors make an early attempt to motivate study of the topic by considering estimation and inference for the mean of a random field from the sample mean of a set of spatial observations. It is demonstrated that failing to account for correlation in the observations produces confidence intervals for the mean that are too narrow—an important point indeed.

Chapter 2 contains definitions and discussions of regionalized variables, random functions, and the concepts of stationarity, intrinsic stationarity, and isotropy. In a 300-page book, it is expected that there will be a few typos and some imperfect notational choices, but the authors should have been more careful in their treatment of these basic principles. For example, Definitions 2.2.2, 2.2.4, 2.2.5, and 2.2.6—introducing random fields, variances, covariances, and variograms—all contain either a typo or undefined notation. Perhaps worse, the discussion and figure caption following Definition 2.3.1 confuses stationarity and isotropy.

Chapter 3 covers covariance functions, variograms and empirical variograms, nugget effects, anisotropy, and various methods for fitting models to data. There is a thorough discussion of how aspects of the covariance function or variogram relate to properties of the associated random field, including the importance of the functions' behavior at the origin. In the list of covariance functions, however, the authors curiously leave out the Matérn model, which is popular precisely because it has a parameter that flexibly controls its behavior at the origin. To their credit, the authors dismiss the Gaussian covariance as

unrealistic. Two detailed examples with data are presented; one on the process of computing an empirical variogram from spatial data, and another on fitting a parametric variogram function to data using various fitting methods. The authors unconvincingly try to dismiss maximum likelihood (ML) methods, using an old criticism that the likelihood function could be multimodal, a criticism that could equally apply to the weighted least-square methods for which they advocate. Even if the likelihood were multimodal, this would merely reflect uncertainties about the parameters contained in the data. The authors also claim without evidence that ML estimators suffer from "frequent severe downward bias," even though ML estimators are known to be consistent in some scenarios (Mardia and Marshall 1984). There is a discussion of composite likelihood methods at the end of the chapter, but more attention could have been paid to methods for estimating covariance parameters when the datasets are too large for the covariance matrix to be stored in computer memory. This is a common issue in modern spatial data analysis, and there are approaches that have good statistical properties—for example, the independent blocks composite likelihood method—that could be explained in a page or two.

Chapter 4 introduces kriging as best linear unbiased prediction, complete with many worked examples showing how the kriging weights depend on the relative configuration of the points and the assumed variogram model. The authors also carefully distinguish between estimating integrals of a random field over a block versus making predictions of a random field at points, discussing point prediction and the estimation of a mean parameter as limiting cases of block prediction. The material included is well thought out and helps build intuition for how kriging procedures behave. A brief mention is made of the connection between kriging and conditional expectation in the multivariate normal distribution. A more thorough discussion of the link would have been helpful for readers with a background in multivariate statistics. Although there are many examples of simulated data, no details about how one might simulate data are presented, and conditional simulations are noticeably absent. The chapter concludes with several sections on more exotic interpolation procedures; median polish kriging, disjunctive kriging, and indicator kriging, which may be of interest to some.

Chapters 5 and 6 extend the material on spatial analyses to the spatial-temporal case, and Chapter 7 has an extensive 88-page summary of various spatial-temporal covariance functions. This material should be useful to a graduate or higher level researcher working on problems in space-time covariance functions. A nonstatistician practitioner could get some benefit from reading the first few sections of Chapter 7 but would be overwhelmed by the later sections, especially since little guidance on covariance function selection is given. Although the list is extensive, there is too much emphasis on sum, product, and product-sum models, none of which are smoother away from the origin than they are at the origin, which leads to suboptimal predictions (Stein 2005a). Welcome attention is paid to models that are not fully symmetric and to nonstationary models, which are important for modeling environmental processes. This book does not mention models that are defined in the spectral domain in time but not space, an underused framework that is incredibly useful for modeling and computation for space-time data

collected by fixed monitoring stations recording information at regular time intervals (Stein 2005b). In fact, all of the space-time data examples in this book are collected in this way.

I teach a course in applied spatial statistics at NC State. While in theory, a book with this title could be appropriate as a text for a portion of the course, this book has too many errors, notational problems, unfounded criticisms, and omissions of important content for me to recommend it for that course. We also have a more advanced course in spatial statistics, but I would prefer to use (Banerjee, Carlin, and Gelfand 2014) over this book for that course. The depth of material in Motero et al.'s book on spatial-temporal covariance functions is not available in other books, so it does serve as a useful reference in that regard. The authors should be commended for including numerous worked examples and extensive online material, complete with R code for reproducing most of the figures. The book concludes with a chapter in functional geostatistics. This chapter provides an interesting introduction for readers not familiar with functional modeling and estimation.

References

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: CRC Press. [461]

Mardia, K. V., and Marshall, R. J. (1984), "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression," *Biometrika*, 71, 135–146. [461]

Stein, M. L. (2005a), "Space-Time Covariance Functions," *Journal of the American Statistical Association*, 100, 310–321. [461]

—— (2005b), "Statistical Methods for Regular Monitoring Data," *Journal of the Royal Statistical Society*, Series B, 67, 667–687. [461]

Joseph Guinness
North Carolina State University

Spatial Microsimulation with R. Robin Lovelace and Morgane Dumont. Boca Raton, FL: Chapman & Hall/CRC Press, 2016, xxi + 259 pp., $89.95(P), ISBN: 978-1-49-871154-8.

Spatial Microsimulation with R, written by Robin Lovelace and Morgane Dumont details the method of spatial microsimulation in general, and its implementation using R. The book comprises 13 chapters, providing a thorough and comprehensive treatment of the entire spatial microsimulation process, from defining concepts through developing and executing a spatial microsimulation model.

The field of microsimulation deals with simulating the behaviors and interactions of individuals to assess the effect of individual level rules on a broader system of interest. Spatial microsimulation represents a specific field of microsimulation focused on analyzing individual level data allocated to specific geographical regions. Spatial microsimulation has been used primarily in health, estimating the smoking rates in various communities (Tomitz, Clarke, and Rigby 2008), transportation, notably the RAMBLAS model that simulated daily patterns in the Netherlands (Veldhuisen, Timmermas, and Kapoen 2000), and economic policy in the EUROMOD project that

investigates the implications of changes in economic policy (Sutherland and Figari 2013). This growing field has great potential for expansion and this book provides insight into approaching broader applications.

The book is divided into three sections: (I) Introducing spatial microsimulation in R, (II) Generating spatial microdata, and (III) Modeling spatial microdata. The sections and chapters follow a logical progression, beginning with foundational topics to establish an understanding of core concepts and generally growing more complex throughout.

The authors clearly state in Chapter 1 that the book should be considered a "general-purpose introduction to the field and its implementation" in R. A broad range of individuals can benefit from this text, regardless of levels of familiarity with Spatial Microsimulation or R software.

Spatial Microsimulation: A Reference Guide for Users by Tanton and Edwards (2012) is the most comparable text to Lovelace and Dumont's work. *Spatial Microsimulation with R* presents many of the same concepts, but in a more cohesive manner and with the added benefit of walk-throughs with R code examples.

Although this book should by no means be considered a stand-alone tutorial on using R, ample time is given throughout to help readers familiarize themselves with the software. The authors give steps on installing R, downloading data, and setting up projects in Chapter 2, in addition to providing an entire appendix ("Getting up-to-speed with R") with further assistance in understanding some of the functionality of R and providing supplemental resources on the software. The authors also provide a helpful online resource in the form of a GitHub repository (*https://github.com/Robinlovelace/spatial-microsim-book*) to help readers learn by doing and promote reproducibility.

Chapter 3 serves primarily to ensure that the audience is on the same page, outlining the meaning of various terms used throughout the book as well as clarifying exactly what is meant by spatial microsimulation (the overall modeling approach as opposed to a narrow methodology). For additional clarity, the authors identify what spatial microsimulation is not, differentiating the method from closely related concepts. The chapter concludes by concisely stating the main assumptions underlying microsimulation models.

Part II (Chapters 4–10) address generating spatial microdata and represents the core of the material. Chapter 4 deals with data preparation, using SimpleWorld (a dataset containing 33 individuals with various attributes spread across three geographic zones) as a motivating example. In discussing concepts of the identification/selection of target and constraint variables and the categorization and structure of the microdata, the authors avoid the assumption many texts make that all data satisfy all theoretical assumptions and come from clean, comprehensive, coherent sources. The authors discuss considerations of combining variables from different sources, arrangements of constraints, and data cleaning, which are all crucial topics for modeling and data analysis in general but often are not given sufficient discussion.

Chapter 5 is the longest and most detailed in the book; some less technical readers may begin to find the material difficult to consume during this chapter. The chapter represents a shift from conceptual to practical in which the authors provide thorough coverage of population synthesis, the process of converting input microdata and constraint variables into spatial microdata. The discussion of weighting algorithms as a means of allocating individuals to geographic zones includes a comparison of methods (deterministic vs. stochastic and the broader reweighting vs. combinatorial optimization methods). While the authors go beyond simply presenting a method and moving forward, the comparison of weighting algorithms can be a dense topic and the discussion would have been better served with some references for additional detail. A comprehensive treatment of the commonly used iterative proportional fitting (IPF) reweighting algorithm follows the discussion of weighting algorithms. Considering the significance allocating individuals to geographic zones in spatial microsimulation and the role IPF plays in this allocation, IPF merits an extensive examination, which the authors provide. Both the theory of IPF as well as its specific implementation in R via the `ipfp` and `mipfp` packages (as well as a comparison of the two packages) are presented in the chapter, again using SimpleWorld as an illustration. Chapter 7 introduces the CakeMap example, aimed to estimate cake consumption in Leeds. Through CakeMap, the authors demonstrate the steps and considerations presented in Chapter 4 (including rationalizing the selection of constraint variables, data preparation) as well as providing another comparison of the `ipfp` and `mipfp` packages as presented in Chapter 5.

Continuing the pattern of presenting a range of methods and approaches, Chapter 6 is devoted to presenting alternative approaches to IPF for population synthesis, including Generalized Regression Weighting (GREGWT) procedure, combinatorial optimization, the `simPop` R package, and the Urban Data Science Toolkit (UDST). GREGWT and combinatorial optimization are explored in more detail than simPop and UDST, but the general inclusion of alternatives to consider is helpful should the reader desire to do additional research.

Chapter 8 addresses model checking and gives approaches for ensuring spatial microsimulation models operate as intended. Both internal and external validation are discussed, although there is minimal information on external validation of spatial microsimulation models; this is a challenging topic due to the lack of necessary data. The authors discuss in-sample testing approaches as well as providing potential causes of poor model fit, providing an internal validation of CakeMap example from the previous chapter.

Certain situations arise in which no individual level data are available for population synthesis. Chapter 9 discusses options to perform population synthesis in these cases: *global cross-tables and local marginal distributions* and *two-level aggregated data*. The former approach contains ample code to demonstrate using the SimpleWorld example, but the later does not have any code examples.

Chapter 10 covers the topic of household allocation, a process for assigning individuals generated as a result of spatial microdata and assigning them to household units. The authors cover allocation methods based on having either (1) separate datasets containing information on individuals and households, respectively, but lacking linkage between the two or (2) no access to household-related data but individual level variables related to characteristics of their households. Both situations are explored in detail, with relevant examples and illustrations; however, there are no R code examples to help guide the reader through this process using the software.

Part III (Chapters 11 and 12) covers modeling spatial microdata. Richard Ellison and David Hensher contributed Chapter 11, which specifically covers the Transport and Environment Strategy Impact Simulator (TRESIS) approach to spatial microsimulation. As the name suggests, the approach is more focused on transportation and comprises a set of demand models that are used together; the focus from a spatial microsimulations perspective in the context of this book is the residential location model. Following the demand model concept of TRESIS, the authors discuss an approach for using demand models to allocate synthetic households (households sampled based on a set of demographic variables and the incidence of combinations of these variables in the population of interest) to zones in R.

The final chapter, contributed by Maja Založnik, discusses spatial microsimulation for agent-based modeling (ABM), itself a rapidly growing field and a natural connection to make with spatial microsimulation. Following a brief discussion of various ABM software, the author demonstrates spatial microsimulation, again leveraging the SimpleWorld example, in NetLogo. NetLogo is a natural choice as it is user-friendly and interacts well with R via the `RNetLogo` package. Between the details presented in the text and the additional references provided in the chapter, one could get up to speed with NetLogo relatively quickly, although the text is by no means a stand-alone NetLogo how-to, nor is it intended to be.

Overall, *Spatial Microsimulation with R* is a well-written and concise work on a topic of broad appeal. The book is structured in a logical way, which makes it straightforward for the reader to pull out pertinent information. It is a useful reference that provides value for individuals of diverse backgrounds and can be a valuable resource for individuals seeking new applications for spatial microsimulation.

References

Sutherland, H., and Figari, F. (2013), "EUROMOD: The European Union Tax-Benefit Microsimulation Model," *International Journal of Microsimulation*, 6, 4–26. [462]

Tanton, R., and Edwards, K. (2012), *Spatial Microsimulation: A Reference Guide for Users*, Dordrecht: Springer. [462]

Tomitz, M. N. M., Clarke, G. P., and Rigby, J. E. J. (2008), "The Geography of Smoking in Leeds: Estimating Individual Smoking Rates and the Implications for the Location of Stop Smoking Services," *Area*, 40, 341–353. [461]

Veldhuisen, J., and Timmermas, H., and Kapoen, L. (2000), "RAMBLAS: A Regional Planning Model Based on the Microsimulation of Daily Activity travel Patterns," *Environment and Planning A*, 32, 427–443. [461]

Daniel P. Heard
United Services Automobile Association

Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. Tailen Hsing and Randall Eubank. New York: Wiley, 2015, xiii+333 pp., $95.00(H), ISBN: 978-0-47-001691-6.

The steadily increasing volume of methodological research in statistics in the area of functional data analysis (FDA) reflects the prevalence of such data in various scientific disciplines. Previous texts in the area, most notably the landmark text by Ramsay and Silverman (2005), highlight the breadth of application fields for FDA and rely mostly on the intuition supplied by an understanding of multivariate data analysis techniques, providing an easily accessible and attractive introduction to FDA. In contrast, the text *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators* by Tailen Hsing and Randall Eubank seeks to provide mathematically precise formulations of data objects in FDA, allowing the reader to move past the point where multivariate intuition fails and to grasp the inherent difficulties in dealing with infinite-dimensional functional objects. With this new book, Hsing and Eubank have filled a long-existing void in the area by gathering scattered theoretical results from the literature into one central repository for use by methodological researchers.

The overlap with existing texts is minimal, as real data examples are entirely avoided, allowing the earlier published works to retain their relevance in the field. In fact, without previous exposure to such texts that present an array of case studies and applications, this new book would prove insufficient to gain an appreciation for the value of FDA. It is clear that the intent of the authors is to provide researchers who have an interest in deriving theoretical justifications for their own methods with the basic toolkit to do so. The book is accessible to advanced graduate students with a solid background in linear algebra, real analysis, multivariate statistics, and measure-theoretic probability. With such an audience in mind, the authors mention that it may be appropriate for a special topics course. To be successful, such a course would likely need to supplement the text with motivating examples, except in the most specialized settings where theory is the sole interest of the student. One alternative is to study the text in a weekly reading group. Having done this with an FDA research group at UC Davis, I can attest to the efficacy of this approach.

Chapter 1 introduces functional data as observed sample paths of a stochastic process. In practice, these are observed only on a discrete grid, giving what appear to be high-dimensional multivariate data, regarding which the authors warn, "Rote application of [multivariate data analysis] technology is simply not the avenue one should follow..."(Page 2). Chapter 2 provides the essential background in functional analysis for Banach and Hilbert spaces, which, in and of itself, is distinctive among statistical texts. The treatment of these topics is rigorous and entirely self-contained, with detailed proofs given. More importantly, the chosen topics are well-suited for a statistical audience as prominent, but unnecessary, elements in typical functional analysis texts are omitted, such as the Hahn-Banach and Baire Category theorems. Even more interesting are the thorough considerations of specialized Hilbert spaces relevant in FDA, namely, reproducing kernel Hilbert spaces (RKHS) and Sobolev spaces. An understanding of these special cases, far from being superfluous, allows one to exploit the unique nature of functional data through concepts of smoothness. Indeed, the dynamics of functional data are often the focus of study in applied settings.

Chapters 3–4 deal with linear functionals and operators, including compact and integral operators that lie at the heart of

FDA. In particular, the properties of compact operators give key insights into the unique challenges posed by functional data that are not found in multivariate data. Since these operators are to be estimated, Chapter 5 develops perturbation theory as a prerequisite for Functional Principal Component Analysis in Chapter 9. This theory is presented in the classical way, relying heavily on complex analysis and contour integrals. As such, the results are quite precise, generally yielding equalities rather than bounds on eigenvalue or eigenfunction differences, but this comes at the cost of geometric intuition. For a more basic and intuitive approach avoiding complex analysis, the nonasymptotic bounds provided in Bosq (2000) on the eigenvalue and eigenfunction discrepancies in terms of covariance operator differences should be sufficient and likely preferable.

A measure-theoretic definition of random Hilbert elements, as a generalization of random variables, is given in Chapter 7, with the corresponding distributional elements of a mean function and covariance operator. When the Hilbert space is a function space, this approach is compared to the competing view, mentioned previously, in which the observed functions are sample paths of a mean-square continuous stochastic process. The advantages and disadvantages of each view are discussed, and the now famous Karhunen–Loève decomposition is presented, hinting at its use for data exploration and dimensionality reduction.

While other texts, such as the aforementioned book by Bosq (2000), do provide statistical theory applicable to FDA, this book is unique in its realistic treatment of the data actually available to the analyst, where the functional data trajectories are available on a sparse or dense grid, but never fully observed. The authors review nonparametric smoothing and regularization in Chapter 6 in the context of penalized least squares, including sections on smoothing parameter selection and connections to splines. These results, developed for the approximation of a single curve, are later incorporated into the results on mean and covariance estimation in Chapter 8, where dependencies are present in the data. A particularly useful contribution are unified results for dense and sparse observation designs in Theorems 8.2.1 and 8.2.4 along with their corollaries, echoing the work of Li and Hsing (2010). In Chapter 9, these results are combined with the perturbation theory from Chapter 5 to obtain limit distributions and rates of convergence for the eigenvalue and eigenfunction estimates, although the prediction of functional principal component scores is not discussed. The book finishes by exploring canonical correlation and linear regression for functional data. Here, by carefully defining canonical correlation to avoid the nonexistence of the inverse covariance operator for random elements of L^2, the authors demonstrate the links between canonical correlation and various other methods that can be seen as extensions of multivariate techniques, including regression, factor analysis, and discriminant analysis, and MANOVA.

In summary, the new text by Hsing and Eubank is an important addition to the researcher's FDA library. As is the case in any statistical field of study, the benefits of a firm understanding of the mathematical foundations of FDA go beyond proof-writing abilities, extending to deeper insight and creative thinking when it comes to real data analysis. I would recommend this text to any academic wishing to conduct methodological research in FDA, in addition to any practitioner with the question, "What makes functional data unique?"

References

Bosq, D. (2000), *Linear Processes in Function Spaces*, New York: Springer. [464]

Li, Y., and Hsing, T. (2010), "Uniform Convergence Rates for Nonparametric Regression and Principal Component Analysis in Functional/Longitudinal Data," *The Annals of Statistics*, 38, 3321–3351. [464]

Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis*, New York: Springer. [463]

Alexander M. Petersen
University of California, Santa Barbara