# BOOK REVIEWS

## EDITOR:
## GUILHERME J. M. ROSA

WANG, Y. **Smoothing Splines: Methods and Applications.** Chapman and Hall/CRC, Boca Raton, Florida, 2011. xxiv + 370 pp. \$89.95/€57.99, ISBN 9781420077551.

Due to its appealing properties, smoothing splines have always been one of the most popular smoothing methods for various problems. A well-known advantage of smoothing splines is that they are very convenient and reliable in estimating multivariate functions, benefited from the natural decomposition of the reproducing kernel Hilbert spaces corresponding to the functional ANOVA decomposition. A number of books have described smoothing splines alongside other smoothing methods; see Green and Silverman (1994), Eubank (1999), and Wood (2006), among others. As far as the reviewer knows, only two monographs have devoted solely to smoothing splines before the publication of this book. One is the classical and most cited Wahba (1990) that continues to provide motivation for researchers in this area. The other is Gu (2002) that gives a systematic introduction to smoothing splines and their applications to regression, density estimation, and hazard estimation. Gu's book also comes with an R package named gss. Similar to Gu (2002), this book by Wang provides a nice extension to Wahba (1990) as well as a really useful R package named `assist`. However, the focus of this book is different from its predecessors.

As a leading expert in smoothing splines, Wang has made numerous contributions to the generalization of smoothing spline regression models to incorporate heteroscedastic and/or correlated errors, nonlinear regression models, and/or parametric components. This book summarizes his work in a meticulously organized way. A distinguished strength of this book is the wide variety of real data sets used to illustrate models and methods. For each new method introduced in the book, the author always provides several real examples from different disciplines for illustration. For each of these real examples and synthetic demonstrations, the author always gives the complete step-by-step analysis code. All these are extremely helpful for practitioners to match their problems with appropriate methods and do their analysis in a sensible way.

There are some overlaps between this book and Gu (2002) and their R packages. Both Wang and Gu devote the first few chapters to smoothing spline regression models with Gaussian-type responses. Their respective implementations are the functions `ssr` in `assist` and `ssanova` in `gss`. The `ssr` function also implements generalized smoothing spline regression models with responses coming from a distribution in the exponential family. This is implemented as the `gssanova` function in `gss`. Although random effects were not considered in Gu's book, the `gss` package has been updated to include the random effects options for both ssanova and `gssanova`. Both packages can handle nonparametric and semiparametric regression models too. Note that the syntax of these functions are slightly different despite their similar functionalities.

Of course, there are major differences between this book and Gu (2002) and their R packages. Here are some highlights that can hopefully guide a user to select the appropriate tool. Gu's book and package offer tools for (conditional) probability density estimation, hazard function estimation, Cox models with nonparametric/semiparametric relative risk, and accelerated failure time models through location-scale families with nonparametric/semiparametric form of location function. On the other hand, Wang's book and package definitely provide more options for various regression models via the functions `ssr`, `nnr`, `snr`, `slm`, and `snm`. They allow the complications of random/mixed effects, nonlinear

regression models, semiparametric models, and any mix of them. The delicate tree diagram in Figure 1.5 of Chapter 1 displays how these functions extend some well-known R functions rooted from the function lm. The diagram is not only a must-visit for first-time readers to get an idea of the outline of the book, but also a place that should be consulted often to determine the right choice of model and R function.

For each method, the book provides all the necessary computational details including explicit formulae and detailed algorithms. The lavish style of the book is quite different from the conciseness in Wahba (1990) and Gu (2002). It is an ideal textbook for high-level graduate student course and an ideal reference for those who deal with complicated nonparametric or semiparametric regression models. What's lacking in the book is rigorous theoretical investigation of the methods, partly due to the extreme difficulty in such developments. The last chapter of Gu (2002) can be consulted for the theories on some smoothing spline models. Overall, I think this is a great book on smoothing splines that one should treasure like Wahba (1990) and Gu (2002).

### References

Eubank, R. (1999). *Nonparametric Regression and Spline Smoothing,* 2nd edition. New York: Dekker.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach.* London: Chapman and Hall.

Gu, C. (2002). *Smoothing Spline ANOVA Models.* New York: Springer-Verlag.

Wahba, G. (1990). *Spline Models for Observational Data,* Volume 59: *CBMS-NSF Regional Conference Series in Applied Mathematics.* Philadelphia: SIAM.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R.* Boca Raton, Florida: Chapman and Hall/CRC.

Pang Du
Department of Statistics
Virginia Tech
Blacksburg, Virginia, USA

CRESSIE, N. and WIKLE, C. K. **Statistics for Spatio-Temporal Data.** John Wiley & Sons, Inc., Hoboken, New Jersey, 2011. xvii + 588 pp. $89.95/€68.40. ISBN 9780471692744.

This book is about the statistical analysis of data observed in space and time. It is both a thoughtful book and a book that covers a large amount of material. The book would be a good reference for a researcher in the field.

Before reading this book, I naively expected that the bulk of it would contain a chapter on time series analysis, a chapter on spatial statistics, and finally a chapter combining these two topics into spatio-temporal statistical models. The book does contain these chapters, and in addition a chapter about exploratory methods for spatio-temporal data. The chapter on spatio-temporal statistical models has subchapters on spatio-temporal covariance functions, spatio-temporal kriging, time-series of spatial processes, spatio-temporal point processes, and spatio-temporal components-of-variation models—more or less as I expected. However, this chapter does not form the main focus of the book, which instead is "hierarchical dynamic spatio-temporal models" (hierarchical DSTMs) presented in the subsequent three chapters. Actually, the field of spatio-temporal statistical modeling has developed much from the space-time separable models that were considered 15 years ago, and now also incorporates literature from other fields like the atmospheric sciences. I have been delighted to learn about all this.

Hierarchical DSTMs have been developed within the last 15 years (see for example Wikle, Berliner, and Cressie, 1998; Berliner, Wikle, and Cressie, 2000; Wikle et al. 2001; and Wikle and Hooten, 2010). In the book the presentation of hierarchical DSTMs is divided into three chapters about models, implementation and inference, and data examples, respectively. In the first chapter, the authors strongly advocate the use of the hierarchical model framework since the conditional thinking allows us to actually model the dynamics of the latent spatio-temporal process as it evolves in time. The latent process is typically described by a Markovian-dependence structure in time. The authors argue that knowledge about physical/chemical/biological/ecological dynamic relationships should be utilized to parameterize this dependence structure, for example by using stochastic versions of differential or difference equations that describe the system. The topic of the second chapter is the wide range of different types of algorithms (Kalman filtering, McMC, INLA, etc.) that have been used for statistical inference. Statistical analysis of spatio-temporal data has to a large degree been determined by the available computing technology and statistical computing methodology, making this a very important part of the field. The third chapter contains four data examples about: forecasting of sea surface temperatures, analysis of remote sensing data, invasive species modeling, and modeling of near-surface wind fields, which illustrates the wide range of applications.

The amount of material covered in the book is impressive. The book contains many references, each chapter has a section at the end with bibliographical notes, and the book summarizes and discusses a large number of different approaches within each chapter. This makes the book very useful as a reference on the topic. On the other hand, it also makes the book more difficult to read and digest, and probably a less ambitious book would have been preferable for use in a course. In my view, the book is not well suited for use as a textbook in a course, but is a very good companion for a researcher in the field.

To summarize, this book has the clear potential to become the standard book reference in spatio-temporal statistics, similarly to the way Cressie (1991) became the standard reference in spatial statistics for a whole decade. I recommend readers with interest in spatio-temporal statistics to buy the book.

REFERENCES

Berliner, L. M., Wikle, C. K., and Cressie, N. (2000). Long-lead prediction of Pacific SSTs via Bayesian dynamic models. *Journal of Climate* **13,** 3953–3968.

Cressie, N. (1991). *Statistics for Spatial Data.* New York: John Wiley & Sons.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics* **5,** 117–154.

Wikle, C. K., Milliff, D., Nyscka, D., and Berliner, L. M. (2001). Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association* **96,** 382–297.

Wikle, C. K. and Hooten, M. B. (2010). A general science-based framework for dynamical spatio-temporal models (with discussion). *Test* **19,** 417–468.

Ole F. Christensen
Center for Quantitative Genetics and Genomics
Aarhus University
Tjele, Denmark

DEY, D. K., GHOSH, S. and MALLICK, B. K. (eds) **Bayesian Modeling in Bioinformatics.** Chapman and Hall/CRC, Boca Raton, Florida, 2010. xxvi + 440 pp. $89.95/£59.99. ISBN 9781420070170.

During the last decade, there have been enormous advances in genomics and molecular biology, which enable innovative insights into biology. The challenge of analyzing and interpreting the vast amounts of data from microarrays and other high-throughput technologies has led to collaborations amongst statistics, computer sciences, biology, and other fields. This book discusses the development and application of Bayesian statistical methods for the analysis of high-throughput genomic and proteomic data arising in order to address questions in molecular and structural biology and medical research.

This book consists of 17 review articles, which discuss a variety of recently developed Bayesian methods for analyzing genomic and proteomic data. The editors aim to provide a broad overview of relevant topics rather than to present comprehensive reviews in all areas of bioinformatics. All chapters are independently written and could be read as independent review articles. New topics are explicitly and carefully introduced and the articles would be easy-to-read for researchers and graduate students.

In Chapter 1, methods for processing and analyzing microarray time-course experiments are provided. This chapter discusses a variety of recently developed Bayesian techniques for detection of differentially expressed genes in time series microarray experiments. In Chapter 2, classification methodology for differential gene expression using hierarchical Bayesian models is presented. Specifically, optimal Bayesian rules and multiple testing methods using false discovery rate are discussed. In Chapter 3, applications of mode-oriented stochastic search algorithms to discrete multi-way data from genome-wide studies are discussed. Chapter 4 discusses nonparametric Bayesian approaches in bioinformatics. Nonparametric modeling using Dirichlet process mixture models is discussed for multiple testing, high-dimensional regression, clustering, and functional data analysis. Chapter 5 discusses measurement error problems in analyzing cDNA microarray data, especially for survival analysis settings. In Chapter 6, robust Bayesian inference of differential gene expression is discussed. The author develops an R package `bridge` for implementing these methods. In Chapter 7, the Bayesian hidden Markov modeling approach for analyzing CGH array data is reviewed. In Chapter 8, the authors discuss Bayesian approaches to phylogenetic analysis. Chapter 9 is concerned with identification of biomarkers in high-throughput microarray data. Chapter 10 reviews Bayesian methods using sparsity priors for protein–protein interaction predictions. In Chapter 11, Bayesian network learning of microarray gene expression data are discussed. Chapter 12 reviews in vitro to in vivo factor profiling in genomic expression. In Chapter 13, proportional hazards regression using Bayesian kernel machines is discussed. Chapter 14 discusses model-based inference based on a hierarchical Beta mixture model for protein biomarker discovery using mass spectrometry. Chapter 15 focuses on Bayesian methodology for detecting differentially expressed genes. Simulation experiments for comparing some standard methods are implemented. In Chapter 16, Bayes and empirical Bayes methods for spotted microarray data analysis are discussed. The author develops MATLAB$^®$ packages `MicroBayes` and `GEBCauchy`. Finally, Chapter 17 discusses the Bayesian classification method for QTL mapping and develops a MATLAB$^®$ package `QTLBayes`.

Readers would be required to have some knowledge about Bayesian statistics and computational techniques such as Markov Chain Monte Carlo, but the all chapters feature thorough gentle introductions. This book would be an excellent reference for researchers and graduate students interested in learning about recently developed Bayesian approaches to genomic and proteomic data.

Hisashi Noma
Department of Data Science
The Institute of Statistical Mathematics
Tachikawa, Tokyo, Japan

CHAUDHURI, A. **Randomized Response and Indirect Questioning Techniques in Surveys.** Chapman & Hall/CRC, Boca Raton, Florida, 2011. vx + 212 pp., $99.95/€66.99. ISBN 9781439836576.

Randomized response (RR) is a statistical method that is used to collect truthful answers in a survey interview from a sample of a finite population when a direct stigmatizing question could conceivably provide evasive answer bias. Hence, when

a survey has direct questions that could potentially not allow for reliable study due to the incorporation of false or evasive answers, RR or indirect questioning in general creates conditions that avoid these evasive answers. Consequently, randomized responses can be used in a good analysis to estimate the population parameters of interest. RR techniques are based usually on the objective statistical inference with other new complexities for obtaining information with two or more questions selected and with a procedure of randomization for the selection of the question. The resulting question of this randomization is known to respondents but not to the interviewer. Consequently, the interviewer is unaware of the question being answered, but the interviewee knows it and can answer it without error. In this way, the respondent's privacy is partially protected.

In this book, the simplicity of the original Warner's RR technique for a simple random sample with replacement of a finite population is developed with other techniques along 10 chapters, especially in Chapters 2, 3, 4, 5, 7, 8, 9, and 10. Chapter 6 is devoted specifically to the issue of protection of privacy.

In the process of RR, there is a loss of precision in the estimation of the finite population parameters with respect to a direct question with truthful answers, but its use is justified because of the resulting elimination of evasive answers bias that invalidate in practice the work of collecting answers.

As other authors pointed out in practical cases, the RR loss in precision increases when the protection of privacy of surveyed or interviewed people increases.

The book compiles the important mathematical contributions in the area of RR, but it does not present simple methods for applications of the techniques in the practice of surveying. The book is a compendium of theories and seems to forget the interest of these techniques in origin is to deal with a problem in the practice of surveys, such as in medicine and in sociology. However, readers interested in a summary of the statistical advances on RR, will certainly find this book to be a good resource.

In my viewpoint, the main interest of a book on RR techniques is to provide simple practical and efficient use of these techniques for applied purposes, while mathematical curiosities should be a second priority. The book, written by a renowned mathematical statistician, places more emphasis on this mathematical development, which does not offer additional clarity for practical researchers.

Throughout the book, the author revises the application of RR techniques in the estimation of attribute or quantitative variables and parameters, and provides different ways to treat maximum likelihood estimation, and optional RR devices and techniques that encompass quantitative variables. He also gives his personal viewpoint on different related alternative to RR techniques such as the item count and nominative techniques, and the three-card method. The book is mainly technical and provides a catalogue of references for summarizing up to date contributions in this area.

Finally, I think that this book shows the scientific level of the author, Professor Arijit Chaudhuri, who has contributed substantially to the areas of RR and indirect questioning. It is

a well-written and edited book and I congratulate the author for his work.

Mariano Ruiz Espejo
Departamento de Matemáticas
Fundamentales
Universidad Nacional de Educación a Distancia
Madrid, Spain

HINKELMANN, K. (ed.) **Designs and Analysis of Experiments, Volume 3: Special Designs and Applications.** Wiley, Hoboken, 2012. xxvii + 555 pp. $125.00/€100.20. ISBN 9780470530689.

*Designs and Analysis of Experiments, Volume 3: Special Designs and Applications* edited by Klaus Hinkelmann is the third of Wiley's three-volume set of books on designs and analysis of experiments. The first two volumes, authored by K. Hinkelmann and O. Kempthorne, introduce and develop principles and methods of design and analysis of experiments that are approached through the general linear model supported by the randomization theory. However, experimentation has spread throughout very special fields whose particularities imply the use of much more intricate planning and modeling. This seems to be the motivation of this third volume that encompasses a collection of different field of applications, raging from biological areas (genetics, medical, and agriculture) to engineering, business and others.

Producing a single book integrating potential subjects for so wide types of applications is, in the least, an enormous challenge. A group of 30 leading experts on different topics produced the 15 chapters of this book. Their job was to give an account of historical development of the subject, to describe the state of art on design and data analysis, and to provide some account of software use. It is clear that many authors have put the most effort to produce an entertaining and didactic coverage of material in such short space. However, it is possible to spot a quite large number of editing imperfections/mistakes, some small but others quite unbelievable like the reproduction of the same table twice in pages 490 and 492. Each chapter is self-contained with some relationships between a few of them.

As a summary I would group the chapters as:

1) Genetic/field experiments: Three chapters, Chapter 1, on genetic crosses, is a quite long piece of work on mating designs as well as *environmental* designs for offsprings, in which complete and incomplete block designs are largely considered, optimality and efficiency of designs are discussed, quite detailed account of variance component models is given, computation code in R for data analysis is referred to the Wiley URL (ftp://ftp.wiley.com/public/sci_tech_med/special_designs—that I had temporary difficulty to access due to slow response from the server); Chapter

2, on microarray experiments, describes phase 1 and phase 2 designs and their linking considering some optimality, and a data analysis example with elementary SAS/PROC MIXED code is included; and Chapter 3, on field experiments in agriculture in which methods for accounting for spatial trend through the spatial linear mixed model are very briefly discussed, and illustrative data analyses for two experiments using ASReml R are presented without code (readers are pointed to the general ASReml R URL for details);

2) *Special* statistical models: Three chapters, Chapter 4 is concerned with optimum designs for generalized linear models, i.e., binary and count responses, in which finding approximate locally optimum designs is the emphasis; Chapter 10 deals with optimum designs for discrete choice models (multinomial logit model) that are extensively used for studying preferences on products or services, including SAS code for data analysis, and Chapter 15, on directional responses where emphasis is given to methods for modeling circular/spherical data arising mainly in studies of animal or plant behavior, ANOVA type analysis are described including a section on resampling techniques;

3) Randomized clinical trials: Three chapters, Chapter 5 establishes the particularities, principles and concerns governing experiments using humans, in particular phase 3 type trials, methods for sample size calculations, and a short description of analysis for time-to-event data (with SAS and Splus codes) are presented; Chapter 6 summarizes important aspects on trial monitoring justified by ethical considerations, and an overview on methods for establishing stopping rules and preserving error rates and power (due to the problem of performing several interim analysis in sequential trials) is provided; Chapter 7 addresses methods caring for the validity of inferential procedures while changing prespecified probabilities to assign new recruited patients to treatments, and strong argument in favor of adapting probabilities is discussed based on ethical grounds—in the light of responses from interim analysis and/or patient covariates some information about the best treatment may be available;

4) Computer/simulation experiments: Two chapters (11 and 12), that I do not feel able to comment except that I found Chapter 11 enlightening to someone that does not have knowledge on computer experiments as myself; and, at last;

5) Designs involving factorial treatments: four chapters, two of them (Chapters 8 and 9) related to studying fractions of factorials with respect to desired properties for fitting many sets of possible models, in which design properties as estimation capacity, minimum aberration and extensions are discussed; Chapter 13, on robust parameter designs, exploits designs and analysis approaches for experiments dealing with the problem of mean and variance modeling of product/process responses depending on several experimental as well as noise factors; and Chapter 14, on split-plot response surface designs, in which the main focus is on equivalent estimation, formulae for exact tests are given and some sort of residual analysis for both strata are outlined.

Without surprise, I ended up with the feeling of superficiality/briefness of some parts of the book although I understand that more background means many more pages. Perhaps I would prefer that the space on software code and long output had been saved since, in fact, the ones provided are not essential, even for beginners. To my knowledge this is the unique book covering such a wide range of specialties. My impression is that the material presented in this book is useful for researchers desiring to have a general account of the state of art of planning experiments going beyond the classics and/or analyzing data yielded from complex designed experiments. The applied statistician whose profile is to give support to experimenters from a wide range of fields would also benefit from having this book in the shelf to be consulted as needed, preferably after the revision is performed. Each chapter also provides a sort of up-to-date literature review to start with although any reference list of a book can always be dependent on the taste of the authors.

Luzia Trinca
Departamento de Bioestatística
Universidade Estadual Paulista
Botucatu, Brazil