

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Publications, Agencies and Staff of the U.S.
Department of Commerce

U.S. Department of Commerce

6-2009

Review of *Applied Spatial Data Analysis with R* by R. S. Bivand, E. J. Pebesma, and V. Gomez-Rubio

Jay M. Ver Hoef
NOAA National Marine Mammal Lab

Follow this and additional works at: <https://digitalcommons.unl.edu/usdeptcommercepub>



Part of the [Environmental Sciences Commons](#)

Ver Hoef, Jay M., "Review of *Applied Spatial Data Analysis with R* by R. S. Bivand, E. J. Pebesma, and V. Gomez-Rubio" (2009). *Publications, Agencies and Staff of the U.S. Department of Commerce*. 188.
<https://digitalcommons.unl.edu/usdeptcommercepub/188>

This Article is brought to you for free and open access by the U.S. Department of Commerce at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Publications, Agencies and Staff of the U.S. Department of Commerce by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

BOOK REVIEWS

EDITOR:
THOMAS M. LOUGHIN

- | | | | |
|--|--------------------------|--|--------------------------|
| Applied Spatial Data Analysis with R
(R. S. Bivand, E. J. Pebesma,
and V. Gomez-Rubio) | <i>Jay M. Ver Hoef</i> | Adaptive Design Theory and Implementation Using SAS and R
(M. Chang) | <i>Feifang Hu</i> |
| Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology
(A. B. Lawson) | <i>J. Law</i> | Ecological Models and Data in R
(B. M. Bolker) | <i>Rachel M. Fewster</i> |
| Disease Surveillance: A Public Health Informatics Approach
(J. S. Lombardo and
D. L. Buckeridge, editors) | <i>Iris Pigeot</i> | Applied Multiway Data Analysis
(P. M. Kroonenberg) | <i>Renato Coppi</i> |
| Survival Analysis for Epidemiologic and Medical Research
(S. Selvin) | <i>M. G. Valsecchi</i> | <i>Brief Reports by the Editor</i> | |
| Survival and Event History Analysis: A Process Point of View
(O. O. Aalen, O. Borgan,
and H. K. Gjessing) | <i>Patricia Grambsch</i> | Sampling of Populations: Methods and Applications, 4th edition
(P. S. Levy and S. Lemeshow) | |
| Nonlinear Dimensionality Reduction
(J. A. Lee and M. Verleysen) | <i>Haonan Wang</i> | Applied Survival Analysis: Regression Modeling of Time-to-Event Data, 2nd edition
(D. W. Hosmer, S. Lemeshow, and S. May) | |
| Model Selection and Model Averaging
(G. Claeskens and N. L. Hjort) | <i>Thomas M. Loughin</i> | SAS for Data Analysis: Intermediate Statistical Methods
(M. G. Marasinghe and W. J. Kennedy) | |
| Meta-Analysis of Binary Data Using Profile Likelihood
(D. Böhning, R. Kuhnert,
and S. Rattanasiri) | <i>Eloise Kaizar</i> | Advances in Mathematical and Statistical Modeling
(B. C. Arnold, N. Balakrishnan, J. M. Sarabia,
and R. Mínguez, editors) | |
| Wavelet Methods in Statistics with R
(G. P. Nason) | <i>Jeffrey S. Morris</i> | An Introduction to Generalized Linear Models, 3rd edition
(A. J. Dobson and A. G. Barnett) | |
| | | Design and Analysis of Bioavailability and Bioequivalence Studies, 3rd edition
(S.-C. Chow and J.-P. Liu) | |

BIVAND, R. S., PEBESMA, E. J., and GOMEZ-RUBIO, V. **Applied Spatial Data Analysis with R**. Springer, New York, 2008. xiv+378 pp/\$59.95/€48.10/ ISBN 9780387781709.

Most authors struggle to pick a title that adequately conveys all of the material covered in a book. When I first saw *Applied Spatial Data Analysis with R*, I expected a review of spatial statistical models and their applications in packages (libraries) from the CRAN site of R. The authors' title is not misleading, but I was very pleasantly surprised by how deep the word "applied" is here. The first half of the book essentially covers how R handles spatial data. To some

statisticians this may be boring. Do you want, or *need*, to know the difference between S3 and S4 classes, how spatial objects in R are organized, and how various methods work on the spatial objects? A few years ago I would have said "no," especially to the "want" part. Just let me slap my EXCEL spreadsheet into R and run some spatial functions on it. Unfortunately, the world is not so simple, and ultimately we want to minimize effort to get all of our spatial analyses accomplished. The first half of this book certainly convinced me that some extra effort in organizing my data into certain spatial class structures makes the analysis easier and less subject to mistakes. I also admit that I found it very interesting and I learned a lot.

The book begins with an introductory chapter that leans toward computing and data. It lists four main types of spatial data: points, lines, polygons, and grids (essentially very small polygons that emulate a continuous surface). Chapter 2 begins the description of classes and methods for spatial data in R. Not surprisingly, there is a class for each type of spatial data, and an extra sparse matrix representation of a grid. All of these classes inherit a certain underlying structure from a foundational spatial class. The payoff for taking the extra energy and time to structure your data as an object in Chapter 2 is immediately apparent from Chapter 3, which covers visualizing spatial data. Chapter 4 covers spatial data import and export. Several years ago, I struggled on a project that required managing various spatial data with different projections and support using the `rgdal` package in R; I really wish I had this book at that time! Chapters 4 and 5 go into greater detail on handling spatial data, including checking and manipulating topologies, and even creating your own spatial data classes.

The second half of the book covers statistical analysis, divided into the three main areas of spatial statistics; spatial point patterns (Chapter 7), geostatistics (Chapter 8), and lattice models (called areal data modeling and disease mapping in this book, Chapters 9–11). As the authors admit, these are fairly thin and eclectic treatments of these subjects, which are designed to “showcase” the use of spatial models using some of the packages in R; they are not designed as an introduction to spatial statistics. For example, Chapter 7 uses sophisticated examples of epidemiology from point locations of disease. The geostatistics chapter jumps rather quickly to multivariable models. Modeling areal data includes topics such as Moran eigenvectors and Marshall’s global EB estimator.

In summary, this is an excellent book that should be on the shelf of any applied statistician who is analyzing spatial data using R. It is not appropriate as a text on spatial statistics, but it would be a valuable companion to any course that uses spatial packages in R.

JAY M. VER HOEF
NOAA National Marine Mammal Lab
University of Alaska, Fairbanks
Fairbanks, Alaska, U.S.A.

LAWSON, A. B. **Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology**. CRC Press, Boca Raton, Florida, 2008. xvii + 344 pp. US\$79.95/£44.99. ISBN: 9781584888406.

The author of this book has contributed a series of books on methodologies in spatial analysis of disease data for various levels of readers. His co-authored book *An Introductory Guide to Disease Mapping* (Lawson and Williams, 2001) is still a good choice for novice readers of disease mapping. This introductory book, however, is limited on the practical aspects of disease mapping (McNally, 2002). It lacks technical details and is far from comprehensive for readers who are learning to apply statistical methodologies in disease mapping. His other book, *Statistical Methods in Spatial Epidemiology* (Lawson, 2006), which covers comprehensively statistical techniques in the field of spatial epidemiology (Bailey, 2007), is a valuable reference for statisticians. Spatial epidemiology, also referred

to as disease mapping, analyzes the spatial distribution of diseases taking into account spatial autocorrelation of disease outcomes. Nowadays, a Bayesian approach is becoming more and more popular, not just among biostatisticians, but many other disciplines including medical geographers, public health professionals, environmental health workers, planners, and social scientists. A book is needed for nonstatisticians in these fields who want to apply Bayesian disease mapping but find it challenging. In this regard, the author should be congratulated for writing this new book that is most welcome and timely. This book fills the gap between an introductory and advanced publication on spatial epidemiology with a modern focus on Bayesian methodologies.

This book is an excellent reference for intermediate learners of Bayesian disease mapping. It is divided into two main parts: background and themes. The part on background provides a clear introduction to basic ideas and concepts of Bayesian methods, including Bayesian inference and modeling, computational issues, residuals, and goodness-of-fit. This part contains exercises at the end of the chapters, except for Chapter 1 where datasets are introduced. However, there are no answers provided to the exercises. The second part of this book covers a wide range of applications of Bayesian methodologies in disease mapping. It includes common applications of disease map construction, relative risk estimation, disease cluster detection, ecological analysis, as well as some advanced applications that have received increased attention in the past few years. These relatively advanced applications include multiple scale analysis (also referred to as the problem of modifiable areal unit, change of support, or spatial misalignment, where spatial data at different aggregation levels are used within an analysis), multivariate disease analysis (where multiple diseases are jointly modeled spatially), spatial survival and longitudinal analysis (where both temporal variation and spatial effects are considered in survival and longitudinal analysis), and spatiotemporal disease mapping (where disease maps are analyzed with an associated temporal dimension as applied in the modeling of spatiotemporal spread of infectious disease). Discussions on applications are separated into those for case event (point) and count (aggregated) data. Examples and datasets for illustration are mostly from the United States and United Kingdom. Materials in the book are clearly presented and well laid out in general making it more attractive to read than many other books on Bayesian methodologies, in part because it uses short thick lines and plenty of spacing in between sections. That said, definitions or explanations of abbreviations used (e.g., CH, UH, AKA) are often not given on the same page. This might create difficulties for some readers in following the associated discussions, especially for those who do not read chapter by chapter from the beginning of the book.

Like many other textbooks on Bayesian methodologies, this book provides readers the datasets and `WinBUGS` codes for some of the examples discussed. It would be better if `WinBUGS` codes for all models fitted were provided to guide readers to implement any of the models. For example, in the analysis of Georgia low birth weight on pages 262–264, seven models are compared but the `WinBUGS` code for only one of the models is provided. At the time when this review was written, those datasets, `WinBUGS` file, and R codes from the book that are supposed to be acquirable from the author’s website