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

Biometrics 65, 660–673
June 2009

BOOK REVIEWS

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.