BOOK REVIEWS

SPATIAL STATISTICS AND GEOSTATISTICS: THEORY AND APPLICATIONS FOR GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY, By Chun, Yongwan, and Daniel A. Griffith. London: Sage, 2013.

Spatial statistics and geostatistics are vital research tools for geographers. Almost every geography student learns at one point or another about the well-known spatial dependency phenomenon of autocorrelation, whereby nearby spatial units have more in common (hence positive autocorrelation) than spatial units further away. This book doesn't break with this tradition. In fact autocorrelation is the common factor that continues throughout the book in various analytical forms and statistical applications. Indeed the stated key aim of this book is to reveal the impacts of spatial autocorrelation on georeferenced data analysis.

The book itself is not a large volume – net 167 pages. It contains a lot of relevant and important information. Every chapter always starts with a set of clearly defined learning objectives which help the reader to get orientated in the general structure and flow of the text. Numerous application examples including maps, figures and tables are set within each chapter. However, most of the examples shown in the book are from data related to the island of Puerto Rico. Specifically the authors use extensively the same set of Digital Elevation Model (DEM) data. The examples are almost entirely based on physical geographic attributes like temperature, precipitation, land coverage etc. The only social example is a very basic transportation application of traffic flows (Ch. 4) and a spatially dependent work trip model (Ch. 5). "R" statistics environment is used extensively throughout the book and each chapter ends with an "R" code for implementation. Readers familiar with using "R" will be able to reconstruct these examples and learn more.

The book takes somewhat for granted that the reader is well acquainted with statistical theory and spatial analysis methods. Some definitions are used but not always clearly explained or explained after initially introduced. A glossary could have helped unfamiliar readers without interfering with the text flow. The material also requires the reader to be well acquainted in reading matrix form equations and relevant notations. This is not always easy and it might have been useful also to show the more simple bivariate cases in parallel.

Overall I found there is a schism-like allure to the writing style. At times it feels indeed like two writers contributed selectively to each chapter. Writer 1 focuses strongly on theory and mathematics; Writer 2, in contrast, mainly works on describing the examples, but it is not always clear to the reader how the theory that Writer 1 develops in much detail is in fact implemented by Writer 2. While results in the form of maps, figures and tables are shown in abundance, not always is it entirely apparent how the equations or models were actually applied and how these results were essentially obtained.

Chapter 1 is actually an Introduction and describes very plainly the structure of the book and a very basic introduction to "R". The informative sections start with Chapter 2 which describes the fundamental statistical properties of spatial dependency and autocorrelation. It discusses the two well-known indices for measuring autocorrelation: the Moran I coefficient and the Geary Ratio (GR). This chapter is a review of what most geographers are well acquainted with but includes more information on the variance inflation attributes of autocorrelation in familiar statistical distributions. Readers acquainted with the Moran I properties will be able to learn more on the properties of the GR and its relation to the semi-variogram and the correlogram plots.

If Chapter 2 dealt with spatial statistical measures, Chapter 3 complements with a discussion of the important aspects of statistical random sampling under conditions of spatial dependency. Sampling designs for spatial data are reviewed (mainly the hexagonal tessellation). Using examples of the Puerto Rico DEM data simulation experiments are conducted on three different synthetic spatial surfaces. One main contribution of this chapter is the discussion of resampling techniques – the Bootstrap and the Jackknife - and how these can be used to determine an effective statistical design for spatial data. A second important contribution of this chapter is the discussion on sample size and why effective samples in the presence of spatial dependency can be substantially smaller than for independent observations because of redundancy issues.

Chapter 4 describes spatial data composition and testing of homogeneity and heterogeneity in georeferenced observations. Box-Cox power transformations to normality and back transformations are embedded throughout this chapter. Hypotheses testing of spatial heterogeneity (homogeneity of variance and equality of means) both regional and directional are developed using ANOVA inference methods. Spatial weighting is introduced using eigenvector spatial filtering to capture autocorrelation. The chapter concludes with a discussion of anisotropy (elevation directional dependency).

Chapter 5 is probably the core of the entire book. It deals with translation of familiar statistical linear and nonlinear regression models to the case of spatially dependent data. It commences with the comparison of Ordinary Linear Regression (OLS) to the familiar spatial autoregressive (SAR) model and the less familiar eigenvector spatial filtering model in the presence of spatially autocorrelated data. Next, spatial filtering techniques are applied to treat nonlinear models like the Poisson and Binomial/Logistic regression models. The key example here is a work trip generation model where both geographic distributions and spatial interaction are accounted for.

While previous chapters dealt with global statistics, Chapter 6 examines local autocorrelation statistics and hot and cold spots. Its key feature is an innovative method to conduct multiple testing for positively correlated data with effective sample sizes. Next, local indices of spatial association (LISA) and the Getis-Ord statistics are discussed. Spatially varying coefficients conclude this chapter relating local to global measures of spatial autocorrelation and their variation across space.

Chapter 7 presents a relatively advanced discussion on quantifying spatial variance and covariance and the use of semi-variogram models (kriging and co-kriging). Readers who are new to this topic should first read more elementary resources as the discussion assumes prior knowledge. Kriging and co-kriging techniques are first discussed followed by a comparison between analysis results obtained with different spatial resolutions as model covariates. Next, spatial filtering methods (e.g. Cochrane-Orcutt) are presented for differentiating between spatial and a-spatial covariance. The reader will find important comparisons for familiar multivariate data analysis with georeferenced data including Principle Components (Factor Analysis), MANOVA, and Canonical Correlation. Correlation coefficient decomposition (Product moment correlation coefficients) with spatial filtering is a novel advanced topic that concludes the chapter. Overall this chapter will help researchers answer if autocorrelation effects in the data significantly matters and if so how.

No statistical method review can be complete without a thorough discussion of missing or incomplete data. Here, Chapter 8 shows how to interpolate spatial (two dimensional) data. The discussion shows that krigging is the best linear unbiased estimator (BLUP) equivalent to an expectation-maximization (EM) solution. EM-based data imputation methods for response variable missing value are further developed using spatially dependent variants of ANCOVA and linear regression. The SAR and eigenvector spatial filtering models for data imputation and the associated level of uncertainty in the estimates are further discussed.

Chapter 9 breaks away from traditional frequentist statistics that form the theoretical basis for all the previous chapters. It introduces the reader to Bayesian map analysis (using the freeware WINBUGS and GeoBugs) and Monte-Carlo spatial simulations. Bayesian methods for postulating prior/posterior probability distributions are presented followed by a discussion on mixed models with random parameters. The use of Markov Chain Monte-Carlo simulations is shown to estimate parameter values based on likelihood maximization techniques. Another topic discussed is designing Monte Carlo spatial simulation experiments when there are no analytical solutions and complementing previous discussion on the bootstrap and jackknife inference methods. The final section raises awareness of the important topic of spatial error and uncertainty (random variable variation and location error and interaction of both with autocorrelation). This section also serves as an informal summary of the pervious chapters.

The book does not provide the reader with an overall summary and conclusions Chapter. Rather throughout the book in several of the more advanced chapters, the authors provide recaps of previous information from the last chapters and put this into the correct context. This is not always reader-friendly. A summary chapter that connects the eight chapters together can be an important contribution to add in a future revised edition. In sum, this book is an important additional reference source for an important methodological topic that will be beneficial to advanced students and researchers.

Eran Ben-Elia
Ben-Gurion University of the Negev