INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE, 2016
VOL. 30, NO. 6, 1250–1253

BOOK REVIEW

Spatial statistics and geostatistics: theory and applications for geographic information science and technology, by Yongwan Chun and Daniel A. Griffith, Thousand Oaks, CA, SAGE Publications, 2013, 181 pp., $47.70/£24.29 (paperback), ISBN 978-1-446-20174-9

This book is designed exclusively for a graduate-level, spatial statistics course at the University of Texas at Dallas (Griffith 2014). Thus, it is not an introductory textbook to spatial statistics and geostatistics usually full of reader-friendly step-by-step explanations and illustrations (e.g., Lloyd 2010, O'Sullivan and Unwin 2010). Rather it is targeted for those who seek to do something more requiring for an intermediate or advanced level of knowledge from the particular field. This 'uneasy' nature of the book is reinforced by the fact that the entire book revolves around a relatively new spatial statistical approach named the eigenvector spatial filtering (ESF) technique developed and disseminated by the authors.

This dictates its merits and, at the same time, its limitations. Its merits overwhelmingly come from its providing readers with a profound insight into the nature and implications of spatial autocorrelation and helping them scrutinize a practical framework to tackle some pitfalls endemic to it for their own research. Its limitations are rooted in a rather limited volume for such advanced topics: too brief, sometimes, reader-unfriendly descriptions and illustrations in some places; some missing popular topics (e.g., geographically weighted regression, GWR), and non-negligible variability between chapters in terms of difficulty to follow (for instance, some parts of Box–Cox power transformations in Chapter 4, the influences of different spatial resolutions on co-kriging in Chapter 6, and most topics in Chapter 8 and 9 are relatively difficult).

Prior to a chapter-by-chapter review, I would like to mention my somewhat uncomfortable feeling about the book title. It seems to me that spatial statistics and geostatistics are hardly juxtaposed with each other (for a similar view, see de Smith *et al*. 2015); rather, the latter is part of the former as seen in numerous textbooks (e.g., Ripley 1981, Cressie 1993, Bailey and Gatrell 1995, Schabenberger and Gotway 2005, Gaetan and Guyon 2010). The relationship is portrayed in the same way even in a book published by one of the authors (figure 1.1 in Griffith and Layne 1999, 6) where geostatistics is part of spatial statistics along with spatial autoregression, point pattern analysis, and image analysis. To my opinion, the title would better be 'spatial autoregression and geostatistics' built firmly on the paralleling relationships between them.

Each chapter has its pros and cons. Chapter 2 is concerned with the concept, indices, and impacts of spatial autocorrelation giving an exclusive emphasis on Moran's *I*. It contains several things that I really like: spatial autocorrelation indices including Moran's *I*, Geary's *c*, and join-count statistics, and more might be structurally identical; Moran's *I* can be defined as a product of two standardized vectors; the semi-variogram and Geary's *c* are conceptually parallel; the variance inflation resulting from the presence of spatial autocorrelation comes into play in various ways in accordance to different statistical distributions involved such as normal, Poisson, and binomial distributions. However, the chapter could have been benefited from the inclusion of further discussions on the distributional properties of at least Moran's *I*; the equations for expectation and variance under both the normality and randomization assumptions need to be included.

I really like a somewhat lengthy discussion on spatial sampling in Chapter 3 mainly because it offers an extremely important knowledge base for not only designing a new sampling scheme but also assessing and attempting to remedy pre-existing ones. In addition, its practical importance has increasingly been marked as individual- or institution-level data gathering practices have become more and more popular. In particular, the sampling simulation experiments with different spatial configurations covered in numerous pages in the chapter certainly provide readers with a new insight into designing a spatial sampling scheme. However, there is a mismatch between figure 3.1 and the corresponding descriptions in the text; this could have been fixed by switching the positions of (b) and (c).

Chapter 4 deals with the concept and aspects of spatial heterogeneity and how to investigate it. It is delighting to see that the authors succeed in explaining and illustrating the fact that spatial heterogeneity refers to the spatial variability not only in mean and variance but also in spatial autocorrelation itself which effectively evokes the necessity of local indices of spatial autocorrelation discussed in Chapter 6. Spatial heterogeneity might be a better title for the chapter and its second section (4.2 spatial weights matrices) is somewhat *heterogeneous*; it would be much better if the topics relating to spatial weights matrices are addressed in Chapter 2 along with spatial autocorrelation indices. There might be another issue worthwhile to note: the notation for spatial weights matrices. The authors make a clear distinction between **C** and **W**; the former binary, contiguity-based ones and the latter row-standardized ones. We might need more characters referring to other types of spatial weights matrices; e.g., binary, but non-contiguity-based ones, distance-based ones, contiguity-distance hybrid ones, and more. Furthermore, we might need to make another clear distinction between zero diagonal and non-zero diagonal ones (see Lee 2004, 2009). In conclusion, we need to establish a general, notational rule for spatial weights matrices which is expected to encompass all kinds of defining spatial relationships among observational units.

Chapter 5 deals with various types of spatial regression models which are designed to alleviate the problem of spatial autocorrelation in regression residuals. It begins with the usefulness of simultaneous autoregressive (SAR) specification in revealing and edifying the pitfalls of related standard statistical procedures such as ordinary least squares (OLS) regression, ANOVA, product moment correlation coefficients, and two-group discriminant function analysis. It furthers the mission to the generalized linear model (GLM) involving various types of non-normal probability models by means of the ESF technique. The inclusion of ESF might be what makes this book really unique and extremely useful, and the accompanying *R* codes will provide readers with better chance to exploit the technique in their research. However, it seems to me that the ESF technique in the text comes out of nowhere and the description on it is somewhat too brief and reader-unfriendly. There should have been a section in Chapter 5 titled 'Eigenvector spatial filtering' and a step-by-step explanation under the title leading to a better understanding about the rationale and procedure of the technique.

Chapter 6 is concerned with spatial heterogeneity in spatial autocorrelation focusing on local spatial autocorrelation indices and exploratory techniques associated with them. First of all, I really like their addressing to the multiple testing issue for local statistics at first, because it is rarely acknowledged or largely ignored by plain users usually working with off-the-shelf software. It would be much better if the equation for the sampling variance of local Moran's I_i is given in a more general way because it only applies to situations in which a binary spatial weights matrix is involved (see Lee 2009). The authors provide a way of obtaining vectors of spatially varying coefficients by means of the ESF technique. However, it would be much better if they provide a comparison between their approach and a more *prominent* approach, GWR.

Chapter 7 is divided into two parts; the former is about kriging whereas the latter is mostly about the usefulness of spatial linear operators for multivariate spatial statistics such as principal component analysis (PCA), factor analysis (FA), multivariate analysis of variance (MANOVA), and canonical correlation analysis (CCA). The latter is what I think really makes this book special, since the notions of spatial autocorrelation should not confined to multiple regression and should be extended to spatialize any forms of multivariate spatial analyses. This book clarifies that when spatial liner operator filtered data are entered rather than original data, multivariate analyses could furnish new, more spatially intelligent findings. The notions and procedure of correlation coefficient decomposition provides an invaluable insight into bivariate association and verify the usefulness of ESF approach. Now a correlation coefficient can be seen as a conceptual amalgam of five components: pure correlation (correlation between spatially filtered variables), correlation induced by common spatial autocorrelations, two cross-correlations induced by unique spatial autocorrelations, and uncorrelation between unique spatial autocorrelations.

Chapter 8 deals with the paralleling relationships between kriging as a geostatistical interpolation method and spatial autoregression and ESF as a spatial statistical imputation method. This chapter successively explains the superiority of ESF by showing its extensibility to the generalized linear model for spatial data which encompasses binomial, Poisson, negative binomial regression models. Chapter 9 covers additional, advanced topics including some basics of Bayesian spatial statistics, Markov chain Monte Carlo (MCMC) techniques, and Monte Carlo simulation experiments. The last one is linked to the eigenvector selection for ESF approach.

Another advantage of the book is that it is on free computation environment provided with R. Each chapter ends with R codes used for the statistical results, graphs, and maps displayed in it. For self-exercise purposes for such advanced spatial statistical techniques, the codes seem much more useful than those provided by other R-based introductory textbooks (Kalkhan 2011, Plant 2012, Bivand *et al*. 2013, Dorman 2014, Brunsdon and Comber 2015). However, those who intend to learn both the R language and spatial statistics in a single book might hope to have an entire chapter dedicated solely to the R basics.

In conclusion, this book is an excellent invitation to the amazing world of spatial statistics. The entry barrier of the field is not low at all and this book is not spoon-feeding at all. However, a Greek saying is always correct: ‘The roots of education are bitter, but the fruit is sweet.’

References

Bailey, T.C. and Gatrell, A.C., 1995. *Interactive spatial data analysis*. New York, NY: Longman, 413.

Bivand, R.S., Pebesma, E.J., and Gómez-Rubio, V., 2013. *Applied spatial data analysis with R*. 2nd ed. New York, NY: Springer, 405.

Brunsdon, C. and Comber, L., 2015. *An introduction to R for spatial analysis & mapping*. Thousand Oaks, CA: SAGE Publications, 343.

Cressie, N., 1993. *Statistics for spatial data*. Revised ed. New York, NY: John Wiley & Sons, 900.

de Smith, M.J., Goodchild, M.F., and Longley, P.A., 2015. *Geospatial analysis: A comprehensive guide to principles, techniques and software tools*. 5th ed. Leicester: Winchelsea Press, 750.

Dorman, M., 2014. *Learning R for geospatial analysis*. Birmingham: Packt Publishing, 345.

Gaetan, C. and Guyon, X., 2010. *Spatial statistics and modeling*. New York, NY: Springer, 297.

Griffith, D.A., 2014. Reflections on the current state of spatial statistics education in the United States: 2014. *Geo-spatial Information Science*, 17 (4), 229–235. doi:10.1080/10095020.2014.986834

Griffith, D.A. and Lane, L.J., 1999. *A casebook for spatial statistical data analysis: A compilation of analyses of different thematic data sets*. New York, NY: Oxford University Press, 506.

Kalkhan, M.A., 2011. *Spatial statistics: GeoSpatial information modeling and thematic mapping*. Boca Raton, FL: CRC Press, 166.

Lee, S.-I., 2004. A generalized significance testing method for global measures of spatial association: an extension of the Mantel test. *Environment and Planning A*, 36 (9), 1687–1703. doi:10.1068/a34143
Lee, S.-I., 2009. A generalized randomization approach to local measures of spatial association. *Geographical Analysis*, 41 (2), 221–248. doi:10.1111/j.1538-4632.2009.00749.x
Lloyd, C.D., 2010. *Spatial data analysis: An introduction for GIS users*. New York, NY: Oxford University Press.
Oxford Uni, D. and Unwin, D.J., 2010. *Geographic information analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons.
Plant, R.E., 2012. *Spatial data analysis in ecology and agriculture using R*. Boca Raton, FL: CRC Press, 631.
Ripley, B.D., 1981. *Spatial statistics*. New York, NY: Wiley & Sons, 252.
Schabenberger, O. and Gotway, C.A., 2005. *Statistical methods for spatial data analysis*. Boca Raton, FL: Chapman & Hall/CRC, 488.

Sang-Il Lee
Department of Geography Education, Seoul National University, Seoul 08826, South Korea
si_lee@snu.ac.kr

http://dx.doi.org/10.1080/13658816.2015.1107074