Legendre, P. 2019. Foreword.

Pp. xi–xv *in*: Griffith, D. A., Y. Chun & B. Li, Foreword by P. Legendre. 2019. *Spatial regression analysis using eigenvector spatial filtering*. Elsevier Science BV / Academic Press, Amsterdam. ISBN 978-0-12-815043-6.

Foreword

Geographers have long understood that natural phenomena contain spatial structures. Tobler's First Law of Geography provocatively characterizes this behavior of nature: "everything is related to everything else, but near things are more related than distant things" Tobler (1970, p. 236). This statement is a vivid representation of the phenomenon known as spatial (auto)correlation. Without it, natural phenomena would be disorganized, and physical, geological, and ecological processes, among others, could not take place. On the pragmatic side, if nature did not display spatial structures, there would be no geography or geographers.

Statisticians have long been interested in the description and quantification of spatial structures displayed by data observed in geographic space. These scientists started with a method derived from regression analysis, known as trend surface modeling. Krumbein (1956, 1959) and Grant (1957) first used this method in the earth sciences, following an earlier proposal by "Student" to describe temporal variation using a polynomial function of time ("Student" [Gosset], 1914).

In the early days of spatial analysis by practitioners in various fields, spatial correlation in variables was viewed merely as a nuisance: its presence in data made the usual tests of significance invalid when applied without corrections, whereas modified tests were difficult to implement and had not been fully worked out by statisticians. In my own field of specialty, community ecology, key papers appeared in Levin (1992) and Legendre (1993), arguing that spatial structures were a most important characteristic of the distribution of organisms and natural populations in ecosystems, and were worthy of study for their own sake.

Trend surface analysis, which uses a polynomial function of the geographic coordinates of study sites, was used in early attempts at spatial modeling. This is a rather crude method: to model fine spatial structures would require a polynomial equation with more monomials than observations, which in turn would render the method useless in practice in regression. Researchers then started looking for an applicable method that would produce fine-resolution spatial models with a reasonable number of parameters, one that could be applied to irregularly spaced study sites and be used to model univariate or multivariate response data.

Moran eigenvector spatial filtering: Multiple origins and convergence

A method to model multiscale spatial patterns based on spatial eigenvectors is known as Moran eigenvector spatial filtering (MESF). This book by Griffith, Chun, and Li is about that methodology. Interestingly, this method was developed independently and nearly simultaneously in two different fields, statistical geography (Griffith, 1996, 2000) and quantitative community ecology (Borcard & Legendre, 2002[1]). Following their first paper, Borcard, Legendre, Avois-Jacquet, and Tuomisto (2004) published a series of real-world ecological applications of this method. A few years later, Dray, Legendre, and Peres-Neto (2006) formalized the theory of Moran's eigenvector maps (MEM).

Griffith's original goal was to filter the effect of spatial autocorrelation out of model residuals, transferring this component to a model's conditional mean (i.e., intercept), whereas that of Legendre and his coauthors was to explicitly model the multiscale nature of univariate or multivariate response data[2]. The MESF method of analysis was based on earlier developments by geographers to analyze binary spatial connection (i.e., spatial weights) matrices (SWM; Garrison & Marble, 1964; Gould, 1967; Tinkler, 1972; Griffith, 1996). The two groups quickly realized that their methods had the same algebraic bases and that their objectives were interchangeable. Researchers from these two groups jointly published a paper unifying the terminology and defining the field of *spatial eigenfunction analysis* (Griffith & Peres-Neto, 2006), which encompasses all methods based on eigenvectors describing the spatial relationships among study sites.

Subsequently, Griffith and Legendre had an opportunity to exchange notes in August 2007 during a conference organized by Academia Sinica in Taipei, where they had been invited separately and independently to present their methods. They explained to the audience that the two methods, although formally presented in different ways, were actually one and the same.

---

[1] I had presented this method two years earlier in a keynote address delivered at the Modelling Complex Systems conference in Montréal in July 2000.

[2] In the fields of community ecology and biogeography, and contrary to many problems in geography, most response datasets are (highly) multivariate and nonnormal.

A word about the theoretical background for MESF in ecology

By asking me to write this "Foreword" for their book, Griffith, Chun, and Li offered me an opportunity to explain the theoretical bases that make ecologists interested in spatial correlation and its modeling by spatial eigenfunctions. For that, I have to go back a little in the history of community ecology. This is the branch of ecology devoted to the scientific study of relationships among the species forming natural communities, as well as relationships between these species and their environmental conditions. In the 1990s, ecologists became aware that different kinds of generating processes could produce spatial correlation in data. The main mechanisms are the following: (1) *induced spatial dependence*: the functional dependence of given response data (e.g., species) on a set of explanatory variables; this process is in action when species forming natural communities are dependent upon the environmental conditions in which they are found; (2) *true autocorrelation*: spatial correlation that may occur in multivariate data because of functional interactions among the species in a multivariate data matrix; and (3) *historical dynamics:* manifestations of past natural events, such as isolation by geographic barriers and disturbances of various kinds (e.g., storms, forest fires, volcanic eruptions, and landslides), and anthropogenic causes, such as agriculture, logging, mining, and constructions of various sizes; these past processes may have caused spatial structures to emerge and may have left traces in present-day data that can be identified and modeled as spatial structures. Researchers in other fields could apply these hypotheses, conceptualizations, or theory elements to the explanation of spatial structures they find in their data.

The methodological developments of MESF to date by quantitative geographers are described in detail in the 10 chapters of this book. Meanwhile, methodological developments of MESF continue in ecological research. Blanchet and his coauthors developed asymmetric spatial eigenvector maps in the late 2000s (Blanchet, Legendre, & Borcard, 2008; Blanchet, Legendre, Maranger, Monti, & Pepin, 2011), a method designed to model the effects of directional physical processes, such as marine and river currents, on ecological communities. Guénard, Legendre, Boisclair, and Bilodeau (2010) decomposed the correlation between variables into spatial scales and then extended their method to multivariate response data matrices (Guénard & Legendre, 2018). Following another research path, Guénard, Legendre, and Peres-Neto (2013) extended the spatial eigenvector framework to the modeling of phylogenetic trees and used MESF eigenvectors to predict different types of traits and properties unobserved in rare or endangered species (Guénard, Boisclair, & Legendre, 2015; Guénard, von der Ohe, Walker, Lek, and Legendre (2014). Spatial eigenvectors also were used to develop a test for space–time interaction in repeated surveys (through time) of sets of sites without replication (Legendre, De Cáceres, & Borcard, 2010).

Simultaneous development of a methodology by researchers in different disciplines is an indication of its strength. The MESF method was developed independently by two groups of researchers, at about the same time, which may provide users of the method more confidence in it. Reviewing the variety of ways MESF analysis has been applied to real-world data by geographers, on the one hand, and by ecologists, on the other hand, may give users in different fields ideas for applications that they initially had not considered.

In addition, ecologists are interested in software and, in particular, in R. They developed an R package devoted to spatial and temporal analysis called adespatial (Dray et al., 2016, the first version released on CRAN), with a strong emphasis on spatial eigenfunction analysis. This software and that presented in especially Chapter 4 of this book provide a powerful toolbox for practitioners.

Scientists who apply MESF methods to analyze data recognize that spatial eigenfunctions, which describe unmeasured spatial relationships among the sites constituting study units, may be a proxy for unmeasured explanatory variables. This perspective means, in practice, that one can use these eigenfunctions to model and predict spatial structures without detailed quantitative knowledge of all explanatory variables affecting a set of observed data.

Extensions and the future of MESF analysis

Extensions of MESF to the analysis of temporal data (time series) and to space–time data appeared in the 2010s. In the Preface of this book, Griffith, Chun, and Li mention these developments, which were the result of the work of several groups of researchers in statistical geography. Ecologists also extended MESF to the analysis of space–time data (Legendre & Gauthier, 2014).

I would like to express my highest appreciation to Professors Griffith, Chun, and Li for the immense amount of work they have completed to produce the 10 chapters of this book. Developers and users of spatial eigenfunction methods in all fields of the natural and social sciences will read, study, and refer to this book, which constitutes the first comprehensive compendium of spatial eigenfunction research. This book represents a necessary and effective step toward future development and diffusion of the method. Furthermore, seeing what new understanding will be achieved by researchers about the role of spatial structures in natural and man-made systems—researchers who now have a firm basis of knowledge upon which to base their applications of MESF and its future developments—will be exciting.

Pierre Legendre
Université de Montréal, Montreal, QC, Canada
April, 2019

References

"Student" [W. S. Gosset]. (1914). The elimination and spurious correlation due to position in time or space. *Biometrika*, *10*, 179–180.

Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Modelling directional spatial processes in ecological data. *Ecological Modelling*, *215*, 325–336.

Blanchet, F. G., Legendre, P., Maranger, R., Monti, D., & Pepin, P. (2011). Modelling the effect of directional spatial ecological processes at different scales. *Oecologia*, *166*, 357–368.

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, *153*, 51–68.

Borcard, D., Legendre, P., Avois-Jacquet, C., & Tuomisto, H. (2004). Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, *85*, 1826–1832.

Dray, S., Blanchet, G., Guénard, G., Jombart, T., Legendre, P., & Wagner, H. H. (2016). *adespatial: multivariate multiscale spatial analysis. R package version 0.0-2. http://cran.r-project.org/package=adespatial.*

Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, *196*, 483–493.

Garrison, W. L., & Marble, D. F. (1964). Factor analytic study of the connectivity of a transportation matrix. *In vol. 12. Papers and Proceedings, Regional Science Association*, (pp. 231–238).

Gould, P. R. (1967). On the geographical interpretation of eigenvalues. *Transactions of the Institute of British Geographers*, (42), 53–92.

Grant, F. A. (1957). A problem in the analysis of geophysical data. *Geophysics*, *22*, 309–344.

Griffith, D. (1996). Spatial autocorrelation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. *Canadian Geographer*, *40*, 351–367.

Griffith, D. (2000). A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems*, *2*, 141–156.

Griffith, D., & Peres-Neto, P. R. (2006). Spatial modeling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*, *87*, 2603–2613.

Guénard, G., Boisclair, D., & Legendre, P. (2015). Phylogenetics to help predict active metabolism. *Ecosphere*, *6*, art62.

Guénard, G., & Legendre, P. (2018). Bringing multivariate support to multiscale codependence analysis: assessing the drivers of community structure across spatial scales. *Methods in Ecology and Evolution*, *9*, 292–304.

Guénard, G., Legendre, P., Boisclair, D., & Bilodeau, M. (2010). Multiscale codependence analysis: an integrated approach to analyze relationships across scales. *Ecology*, *91*, 2952–2964.

Guénard, G., Legendre, P., & Peres-Neto, P. R. (2013). Phylogenetic eigenvector maps: a framework to model and predict species traits. *Methods in Ecology and Evolution*, *4*, 1120–1131.

Guénard, G., von der Ohe, P. C., Walker, S. C., Lek, S., & Legendre, P. (2014). Using phylogenetic information and chemical properties to predict species tolerances to pesticides. *Proceedings of the Royal Society B*, *281*, 20133239.

Krumbein, W. C. (1956). Regional and local components in facies maps. *Bulletin of the American Association of Petroleum Geologists*, *40*, 2163–2194.

Krumbein, W. C. (1959). Trend-surface analysis of contour type maps with irregular control-point spacing. *Journal of Geophysical Research*, *64*, 823–834.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology*, *74*, 1659–1673.

Legendre, P., De Cáceres, M., & Borcard, D. (2010). Community surveys through space and time: testing the space–time interaction in the absence of replication. *Ecology*, *91*, 262–272.

Legendre, P., & Gauthier, O. (2014). Statistical methods for temporal and space–time analysis of community composition data. *Proceedings of the Royal Society B*, *281*, 20132728.

Levin, S. A. (1992). The problem of pattern and scale in ecology. *Ecology*, *73*, 1943–1967.

Tinkler, K. J. (1972). The physical interpretation of eigenfunctions of dichotomous matrices. *Transactions of the Institute of British Geographers*, (55), 17–46.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit Region. *Economic Geography*, *46*, 234–240. Supplement.

Spatial Econometrics and Spatial Statistics

SPATIAL REGRESSION ANALYSIS USING EIGENVECTOR SPATIAL FILTERING

DANIEL A. GRIFFITH
YONGWAN CHUN
BIN LI

Foreword by
PIERRE LEGENDRE

Series Editor
GIUSEPPE ARBIA

ELSEVIER
ACADEMIC PRESS
An imprint of Elsevier