ECOGRAPHY 25: 553–557, 2002

Integrating the statistical analysis of spatial data in ecology

A. M. Liebhold and J. Gurevitch

Liebhold, A. M. and Gurevitch, J. 2002. Integrating the statistical analysis of spatial data in ecology. – Ecography 25: 553–557.

In many areas of ecology there is an increasing emphasis on spatial relationships. Often ecologists are interested in new ways of analyzing data with the objective of quantifying spatial patterns, and in designing surveys and experiments in light of the recognition that there may be underlying spatial pattern in biotic responses. In doing so, ecologists have adopted a number of widely different techniques and approaches derived from different schools of thought, and from other scientific disciplines. While the adaptation of a diverse array of statistical approaches and methodologies for the analysis of spatial data has yielded considerable insight into various ecological problems, this diversity of approaches has sometimes impeded communication and retarded more rapid progress in this emergent area. Many of these different statistical methods provide similar information about spatial characteristics, but the differences among these methods make it difficult to compare the results of studies that employ contrasting approaches. The papers in this mini-series explore possible areas of agreement and synthesis between a diversity of approaches to spatial analysis in ecology.

A. M. Liebhold (aliebhold@fs.fed.us), Northeastern Research Station, USDA Forest Service, Morgantown, WV 26505, USA. – J. Gurevitch, Dept of Ecology and Evolution, State Univ. of New York, Stony Brook, NY 11794-5245, USA.

The lack of spatial independence in ecological data has typically been viewed as a problem that can obscure one's ability to understand the biology of the organisms being studied. Various methods have been devised for eliminating or avoiding the effects of spatial dependence in measuring biotic responses. For example, sampling of ecological data has been typically carried out by stratifying across space and averaging to infer underlying processes and mechanisms. However, over the last 20 yr, ecologists have begun to realize that there is a lot of important biology in the spatial dependence of biotic responses, and have become increasingly interested in examining spatial relationships directly. Where previous research ignored or sought to remove the effects of space, new research has sought to explicitly understand, measure and model spatial patterns in biotic responses as a critical aspect of the ecology of many organisms and systems.

Quantitative examination of spatially explicit data in ecology is broadly categorized as "spatial analysis" (e.g., Legendre and Fortin 1989). Besides this emerging area of spatial analysis, other developments have advanced our understanding of spatial relationships in ecology, including the construction of spatial ecology theory and spatially explicit models of ecological processes (e.g., Hassell et al. 1991, Dunning et al. 1995, Tilman and Karieva 1997). Perhaps equally important to conceptual advances, the availability of modern computer hardware and software (e.g., geographical information systems, increased computer speed and memory) has expanded our abilities to address many of the most interesting and critical problems in spatial ecology. Prior to the availability of these tools, analysis of many of the most important problems regarding spatial data was impossible because of the sheer magnitude of data and the complexity of their analysis.

Motivations for spatial analysis are diverse but the common thread is the quantification of spatial patterns. Not surprisingly, there is considerable variety in the types of statistical methods that have been selected to

Accepted 24 January 2002

ISSN 0906-7590

analyze and model spatial variation in ecological data. Some of these methods have a long history in ecology, such as the early variance-mean analyses of Bliss (1941), Greig-Smith (1952), Taylor (1961), and Iwao (1972), the distribution work of Patil (e.g., Patil and Stiteler 1974), and "spatial pattern analysis" of grid and transect data (e.g., Hill 1973, Greig-Smith 1979, and recently summarized by Dale 1999). In some cases, specific models were built (e.g., the "functional response" of predators to host densities, Holling 1959) that in effect attempted to capture spatially-dependent processes without explicit reference to spatial location. The development of certain new, specialized statistical metrics has been motivated by the emerging field of landscape ecology, which focuses on spatial processes operating over various spatial extents (Turner 1989, Wiens 1989). However, many methods currently being applied for spatial analysis in ecology were originally developed in other scientific disciplines such as geography and mining geology.

While the adaptation of such a diverse array of statistical approaches and methodologies for analysis of spatial data has yielded considerable insight into various ecological problems, this diversity of approaches has sometimes impeded communication and retarded more rapid progress in this emergent area (but see ver Hoef et al. 1993). While in some cases these statistical methods provide similar information about spatial characteristics, differences among them make it difficult to compare studies that employ different approaches. Each approach has strengths and weaknesses for certain problems and for specific types of data. However, the existence of parallel, non cross-referenced literatures, each with its own terminology has caused many ecologists to only apply what ever approach they learned first. In order to counter this understandable tendency, and to open a potentially valuable toolbox of diverse approaches to ecologists, we organized a workshop on statistical methods for spatial analysis in ecology at the National Center for Ecological Analysis and Synthesis (NCEAS). In order to obtain a broad perspective, we invited participants from a variety of backgrounds and with expertise in a diversity of approaches. This mini-series represents a synthesis of the studies by this working group.

How has information about spatial pattern been used to answer ecological questions and what are some of the ways in which ecologists have approached spatial analysis? In some studies, analysis of spatial pattern is used to evaluate hypotheses about the processes responsible for the observed patterns. Though several different ecological processes may be capable of generating the same spatial pattern, quantification of spatial pattern may provide clues as to the identity of processes. For example, the existence of certain spatial patterns may rule out specific ecological processes. In addition to answering or generating causal questions about ecological processes, quantification of spatial pattern may be used to correct for spatial dependence of the data in other analyses. Most parametric statistical analyses, for example, are based on assumptions of independence among samples, but the ubiquity of spatial dependence in ecological data often leads to violations of this assumption (e.g., Legendre 1993, 2000, Méot et al. 1998). While this has been known for many years, the importance of this problem is just beginning to be appreciated fully by many ecologists and there is a need to identify appropriate methods for accounting for spatial pattern in the adjustment of statistical tests (Lennon 2000).

Several different major schools of spatial analysis from other disciplines have been adopted by ecologists. The first of these comes from geography, and its methods include the use of statistics (e.g., Moran's I, Geary's C) to measure spatial autocorrelation (Moran 1950, Sokal and Oden 1978, Cliff and Ord 1981), as well as the employment of Mantel tests (e.g., Fortin and Gurevitch 1993, Legendre 2000). This school of analysis generally incorporates hypothesis testing as an important part of the analysis. These methods have also been adapted for correcting for spatial autocorrelation in data when testing hypotheses about relationships between two or more variables, particularly by using Mantel tests (e.g., Smouse et al. 1986, Legendre 1993, Manly 1997, Méot et al. 1998).

A different tradition for the statistical analysis of spatial data adopted by ecologists, known as geostatistics, was originally developed for applied geological problems. These methods use various statistical methods (e.g., construction of variograms and determination of spatial covariance) to quantify spatial autocorrelation (or "spatial dependence"; e.g., Isaaks and Srivastava 1989, Rossi et al. 1992, Liebhold et al. 1993). While equivalencies exist for many of the statistics employed in this tradition and those used in geography, these schools of spatial statistical analysis developed independently, have different terminology, and have different perspectives. For example, geostatistics emphasizes spatial estimation, rather than hypothesis testing (e.g., Liebhold et al. 1991). Methods (including kriging and spatial simulation) have been developed in geostatistics to estimate values at unsampled locations via interpolation from nearby locations, and to provide confidence intervals for interpolated values. These methods are used for estimating entire surfaces or mapping spatial data. Kriging and related spatial estimation procedures increasingly are being used in both basic and applied ecology (Robertson 1987, Liebhold et al. 1993).

Other approaches to spatial analysis, such as dispersion indices, were developed specifically for ecological applications (see review in Dale 1999). In the late 1960s and 1970s a variety of methods were developed to infer spatial pattern from the frequency distribution of sample counts (e.g., Taylor 1961, Iwao 1972). Although these methods can be used to differentiate among some types of spatial patterns, they do not use information about the spatial location of samples and fall short in their abilities to differentiate between types of patterns. Other methods developed for ecological applications were designed for specific types of data. Nearest neighbor and related statistics, for example, were developed for analysis of data consisting of exhaustive maps depicting the locations of all individuals (e.g., trees in a forest; Ripley 1979, 1981). Another group of methods were developed for applications in landscape ecology for analysis of polygonal data (e.g., polygons representing isolated forest stands in a largely agricultural land use; McGarigal and Marks 1995). In addition to the methods broadly described above, other approaches, such as spectral analysis (Ford and Renshaw 1984), wavelet analysis (e.g. Bradshaw and Spies 1992, Dale and Mah 1998), and fractal analysis (Milne 1992) have been applied to the analysis of spatial data in ecology.

At first glance, these ecological methods appear to be fundamentally different from methods derived from geological or geographical applications, but as Ripley (1981) showed, for example, point locations can be aggregated into counts in quadrats which can then be analyzed for spatial autocorrelation. Similar transformations could be applied to polygonal data as well. Thus, relationships may exist among these methods, but these relationships generally remain obscure to most ecologists.

The papers presented in this mini-series were written with the intention of bridging the disjunction between these methods, schools and terminology for spatial analysis in ecology. We present an overview and comparison of the various statistical approaches to analysis of spatial data in ecology and explore the similarities, differences, strengths and weaknesses of the varying methods. These comparisons are not limited solely to the spectrum of methods for quantifying spatial patterns, but also include exploration of methods for incorporating information about spatial dependence in hypothesis testing, modeling relationships between variables, and sampling and experimental design. We also discuss issues that pervade virtually all of the approaches to statistical handling of spatial data – the issues of scale, including the unit size, shape, lag and extent used for observation and analysis.

The mini-series consists of five research papers. The first paper (Dale et al. 2002) examines the mathematical and conceptual relationships among various spatial statistics methods. These methods are viewed from the perspective of their mathematical equivalencies (or lack thereof) and are evaluated both informally and technically. By highlighting the similarities and differences among the methods, this paper provides a unified framework for comparing the descriptive information provided by the varying approaches to spatial analysis.

In the second paper, the authors (Perry et al. 2002) continue the comparison of statistical methods for quantifying spatial patterns in ecological data. While the paper by Dale et al. (2002) focuses on comparison of the mathematical relationships among the various statistics, the comparison in Perry et al. (2002) emphasizes their applied characteristics. They accomplish this first, by presenting a "taxonomy" of the various types of spatial data and then describing which methods may be applied to each data type and what descriptive information about spatial pattern is provided by each method. Finally, the paper illustrates the use of the various methods by applying the different methods to the same example data sets. This paper is intended to provide information that ecologists can use to identify which statistical method is appropriate for their data and the type of descriptive information they need.

The third paper (Legendre et al. 2002) examines the consequences of spatial structure for the design and analysis of field surveys. Spatial autocorrelation and other kinds of spatial structure can mask relationships between biological response variables (e.g., abundance) and explanatory (habitat) variables or they may lead to false conclusions that there are relationships when in fact none exist. Different survey designs vary in their ability to mitigate such problems. In this paper, the authors simulate a series of data sets with known spatial autocorrelation and compare the use of different survey designs for the detection of biotic responses. They also evaluate a statistical method for eliminating or controlling the effect of spatial autocorrelation during analysis.

The fourth paper (Keitt et al. 2002) also addresses the problem of assessing relationships between biotic responses and environmental factors in the presence of spatial autocorrelation. While the paper by Legendre et al. (2002) focuses on adjusting survey techniques and analytical methods to account for spatial autocorrelation when testing the significance of relationships, Keitt et al. (2002) focus on methods for obtaining accurate models (estimation) of species-environment relationships where spatial autocorrelation is present. They illustrate these methods for accounting for spatial autocorrelation in models using three different ecological datasets.

The authors of the final paper of the mini-series (Dungan et al. 2002) examine many aspects of scale, which is an issue of general importance for the analysis of spatially structured data in ecology. First, they address the confused and variable use of the word "scale" in the ecological literature. They argue that ecologists need to clarify the use of this term in the future because of its many possible meanings. Finally they pay special attention to sample unit size, shape, lag and extent. They discuss how virtually all spatial statistics are affected by these four characteristics of a set of observations and their analysis and how they may alter conclusions about ecological phenomena.

Our intention in these papers was to synthesize the varying approaches to the statistical analysis of spatial data in ecology. These papers advance us toward a more unified overview of these methods and should serve both as a resource for ecologists in their selection of appropriate approaches for data analysis and allow ecologists to compare results from different studies that may have utilized different statistical approaches. Despite these advances, more work will certainly be needed to address a host of related issues. For example, there are certain uses of spatial statistics, such as spatial interpolation and simulation, that we did not address. Another problem the group identified was the disparate nature of software available for calculating spatial statistics. Ultimately it would be desirable to develop software package(s) (or components of other statistical packages) that are capable of computing a full array of spatial statistics. We feel that over time, many of these deficiencies will be addressed as spatial statistics play an increasingly central role in ecological research.

Acknowledgements – This work was conducted as part of the "Integrating the Statistical Modeling of Spatial Data in Ecology" Working Group supported by the National Center for Ecological Analysis and Synthesis, a center funded by NSF (Grant #DEB-94-21535), the Univ. of California-Santa Barbara, the California Resources Agency, and the California Environmental Protection Agency.

References

Bliss, C. I. 1941. Statistical problems in estimating populations of Japanese beetle larvae. – J. Econ. Entomol. 34: 221–232.

Bradshaw, G. A. and Spies, T. A. 1992. Characterizing canopy gap structure in forests using wavelet analysis. – J. Ecol. 80: 205–215.

Cliff, A. D. and Ord, J. K. 1981. Spatial processes: models and applications. – Pion Limited.

Dale, M. R. T. 1999. Spatial pattern analysis in plant ecology. – Cambridge Univ. Press.

Dale, M. R. T. and Mah, M. 1998. The use if wavelets for spatial pattern analysis in ecology. – J. Veg. Sci. 9: 805–814.

Dale, M. R. T. et al. 2002. Conceptual and mathematical relationships among methods for spatial analysis. – Ecography 25: 558–577.

Dungan, J. L. et al. 2002. A balanced view of scale in spatial statistical analysis. – Ecography 25: 626–640.

Dunning, J. B. Jr et al. 1995. Spatially explicit population models: current forms and future uses. – Ecol. Appl. 5: 3–11.

Ford, E. D. and Renshaw, E. 1984. The interpretation of process from pattern using two-dimensional spectral analysis: modelling single species patterns in vegetation. – Vegetatio 56: 113–123.

Fortin, M. J. and Gurevitch, J. 1993. Mantel tests: spatial structure in field experiments. – In: Scheiner, S. M. and Gurevitch, J. (eds), Design and analysis of ecological experiments. Chapman and Hall, pp. 342–359.

Greig-Smith, P. 1952. The use of random and contiguous quadrats in the study of structure in plant communities. – Ann. Bot. (Lond.) 16: 293–316.

Greig-Smith, P. 1979. Pattern in vegetation. – J. Ecol. 67: 755–779.

Hassell, M., Comins, H. and May, R. M. 1991. Spatial structures and chaos in insect population dynamics. – Nature 353: 255–258.

Hill, M. O. 1973. The intensity of spatial pattern in plant communities. – J. Ecol. 61: 225–235.

Holling, C. S. 1959. Some characteristics of simple types of predation and parasitism. – Can. Entomol. 91: 385–398.

Isaaks, E. H. and Srivastava, R. M. 1989. An introduction to applied geostatistics. – Oxford Univ. Press.

Iwao, S. 1972. Application of the m-m^* method to the analysis of spatial patterns by changing the quadrad size. – Res. Popul. Ecol. 14: 97.

Keitt, T. H. et al. 2002. Accounting for spatial pattern when modeling organism-environmental interactions. – Ecography 25: 616–615.

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

Legendre, P. 2000. Comparison of permutation methods for the partial correlation and partial Mantel tests. – J. Stat. Comput. Simul. 67: 37–74.

Legendre, P. and Fortin, M.-J. 1989. Spatial pattern and ecological analysis. – Vegetatio 80: 107–138.

Legendre, P. et al. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. – Ecography 25: 601–615.

Lennon, J. J. 2000. Red-shifts and red herrings in geographical ecology. – Ecography 23: 101–113.

Liebhold, A. M. et al. 1991. Geostatistical analysis of gypsy moth, *Lymantria dispar*, egg mass populations. – Environ. Entomol. 20: 1407–1417.

Liebhold, A. M., Rossi, R. E. and Kemp, W. P. 1993. Geostatistics and geographic information systems in applied insect ecology. – Annu. Rev. Entomol. 38: 303–327.

Manly, B. 1997. Randomization, bootstrap and Monte Carlo methods in biology, 2nd ed. – Chapman and Hall.

McGarigal, K. and Marks, B. J. 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. – U.S. Dept of Agriculture, Forest Service, Pacific Northwest Research Station, Gen. Tech. Rep. PNW-GTR-351.

Méot, A., Legendre, P. and Borcard, D. 1998. Partialling out the spatial component of ecological variation: questions and propositions in the linear modeling framework. – Environ. Ecol. Stat. 5: 1–27.

Milne, B. T. 1992. Spatial aggregation and neutral models in fractal landscapes. – Am. Nat. 139: 32–57.

Moran, P. A. P. 1950. Notes on continuous stochastic phenomena. – Biometrika 37: 17–23.

Patil, G. P. and Stiteler, W. M. 1974. Concepts of aggregation and their quantification: a critical review with some new results and applications. – Res. Popul. Ecol. 15: 238–254.

Perry, J. N. et al. 2002. Illustrations and guidelines for selecting statistical methods for quantifying spatial pattern in ecological data. – Ecography 25: 578–600.

Ripley, B. D. 1979. Tests of "randomeness" for spatial point patterns. – J. R. Stat. Soc. B 41: 368–374.

Ripley, B. D. 1981. Spatial statistics. – Wiley.

Robertson, G. P. 1987. Geostatistics in ecology: interpolating with known variance. – Ecology 68: 744–748.

Rossi, R. E. et al. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. – Ecol. Monogr. 62: 277–314.

Smouse, P. E., Long, J. C. and Sokal, R. R. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. – Syst. Zool. 35: 627–632.

Sokal, R. R. and Oden, N. L. 1978. Spatial autocorrelation in biology. 1. Methodology. – Biol. J. Linn. Soc. 16: 199–228.

Taylor, L. R. 1961. Aggregation, variance and the mean. – Nature 189: 732.

Tilman, D. and Karieva, P. (eds) 1997. Spatial ecology: the role of space in population dynamics and interspecific interactions. – Princeton Univ. Press.
Turner, M. G. 1989. Landscape ecology: the effect of pattern on process. – Annu. Rev. Ecol. Syst. 20: 171–197.
ver Hoef, J. M., Cressie, N. A. C. and Glenn-Lewin, D. C. 1993. Spatial models for spatial statistics: some unification. – J. Veg. Sci. 4: 441–452.
Wiens, J. A. 1989. Spatial scaling in ecology. – Funct. Ecol. 3: 385–397.