Compared to What? The Failure of Spatial Analysis to Embrace Geographic Concepts to Aid the Analysis of Event Information

Marc P. Armstrong
Department of Geography and
Program in Applied Mathematical and Computational Sciences
The University of Iowa
Iowa City, IA 52242
marc-armstrong@uiowa.edu

0.0 Preamble

An event is defined here as an observation of a particular class of activity with an associated space-time coordinate. Thus, a report of an infectious disease might have a location (address) and a date associated with it. The address can be geocoded to provide a coordinate value for mapping. Events can be represented in different dimensions (0, 1,2, *n*), though for the sake of discussion I refer mainly to events represented as points. Events can be instantaneous, or they may have a specific duration or periodicity.

1.0 Introduction

Methods of spatial analysis can be placed into several categories. For example, methods have been developed to describe various aspects of point patterns (e.g., nearest neighbor statistic), while other methods are used to determine whether points are randomly distributed (e.g., quadrat analysis). However, methods that are intended to determine if a distribution of points is geographically distributed according to some theoretical expectation of location require further development. By theoretical expectation, in this case, I am referring to distributions that would be influenced by forces that give rise to concepts such as central place theory. In addition, little work has been done to establish how specific types of geographic information and knowledge can be introduced into the analysis of spatial data. If points are analyzed to determine if they are significantly different than, say, a Poisson distribution, then we are missing a determination of whether they are located significantly close to Interstate highways or whether they are located at the ends of spokes emanating from a particular airline hub. These and other types of patterns are difficult (or impossible) to detect using the conventional suite of spatial analysis tools, yet patterns such as these could easily result from aerosol dispersal of microbes from the back of a truck or from contamination originating in the air-handling system of an airport. We need to upgrade our tools to detect events and place them in real contexts for analysis.

2.0 Hot Spots – The Geographical Thermometer

Conventional methods of analysis use statistical computations to help determine whether events are unusual with respect to other events. We can define methods based on case-control (Cuzick-Edwards), frequency (quadrat analysis), or deviation from neighbors (*Gi**). In each case there is some expectation associated with what we are observing but aside from distance and contiguity little other geographical information is normally employed. For example, though anisotropy is well used in kriging, directionally-differentiated analyses are rarely used in other forms of spatial analysis. A richer set of geographical concepts (Nystuen, 1963) need to be introduced into the assessment of spatial pattern and process.

3.0 Event Rates—At this rate we'll never get there…

A chronic problem when analyzing event data is the computation of rates of occurrence. It is possible to define rates on the basis of pre-defined geographical units such as counties, but such units may have nothing to do with the substance of an analysis. Consequently, researchers have begun to develop approaches that flexibly aggregate data to support rate computations (Rushton, 2003). Often a lattice structure is used. A so-called spatial filter searches for points within some radius of a lattice point and assigns the count of events observed to the lattice point. Other methods do not rely on a fixed radius, but instead adjust its size to include a minimum count of events. When the events are summed for one variable, a similar approach is used for another variable and a rate is computed. For example, all births in the region surrounding a lattice point can be located to form a denominator and all birth defect incidents in the same region can be analyzed to form a numerator in order to calculate a birth defect rate. If rates are computed, they can be interpolated and mapped as a continuous variable or analyzed as discrete points on the lattice. Of course, the specification of the lattice spacing is a key issue, and decisions about spacing will clearly play a role in what is observed in any subsequent analysis.

When a surface is created several new techniques have been developed to determine whether important topographic features are present.

4.0 Piqued Peak—Topographic Features on Statistical Surfaces

If a statistical surface is created, it can be visualized and analyzed by experts (visual inspection) to determine high and low areas. However, additional tools can be created to enable users to delimit the "topology of topography": peaks, pits, passes, ridges, and valleys (Warntz, 1966; Pfaltz, 1976). Though such features are difficult to define in an ontological sense (Mark and Smith, in press), they can play important roles in furthering understanding about pattern and process, especially if they correspond with other geographical features such as rivers and interstate highways. These topographic features can be extracted using concepts developed to analyze real topography. One way to pre-process the information is to subject the surface to what is sometimes referred to as "echelon analysis". In a metaphorical sense, the surface could be placed in a tub and slowly flooded to reveal isolated peaks as islands. If the water were released, ridges would emerge and closed depressions would remain filled with water, while valleys would channel it away toward an outlet.

When this procedure is applied using either continuous or discrete time sampling the resulting surface can be analyzed for emerging peaks and thresholds can be specified to determine if special attention is warranted. Noisy surfaces can be visualized in such a way that the visual assessment of regions is made easy for users. One way is to use optimization routines to combine the cells that are by-products of interpolation into clusters such that a local measure of spatial autocorrelation in the resulting classes is maximized.

5.0 Compared to what?

Though normally stated in "null" form, in a general sense statistical hypothesis tests check for the presence of an effect by comparing an observed sample with an expected value of a parameter that would obtain in the absence of the effect. What constitutes an expectation, however, should be made contingent on theoretical knowledge about geography. For example, certain regularities in settlement size and spacing can be observed in the real world. In addition the range of goods available at each of these places also follows theoretical expectations. In simplest form, one can observe such patterning in a single dimension, such as along a highway or in a river corridor but more commonly we observe events in some type of geographical space.

Getis and Boots (1978: 3) develop a useful typology of events that is represented in a 3D space. The *y*-axis in their model represents the relative number of events and whether they are being spawned (births) or diminished (deaths). The *x*-axis represents the relative concentration of events as they either diffuse or agglomerate. The *z*-axis represents time. This conceptual model is, however, somewhat incompletely developed. In particular, it fails to address processes that are more topological than geometrical. For example, while the model handles contagious diffusion (birth, diffuse, time) it does not include hierarchical diffusion.

A related problem concerns the specification of an expected pattern against which an observed pattern is compared. Many traditional methods of spatial analysis have the investigator formulate a null hypothesis under which an observed pattern is compared to one that is assumed to exhibit complete spatial randomness (CSR). In practice, the CSR pattern is generated using a Poisson process model in which *n* points are placed into a region, where each location has an equal likelihood of selection, and where each point is independent of the location of other points. This, however, is only a crude starting point since well-established geographical principles would suggest that CSR is a rather poor null model. In simple terms, either the known distribution of background events can be used, or a theoretical one can be substituted. In a pure form a Lösch or Christaller model can be specified to approximate a central place hierarchy. However, even that adjustment may be inadequate since observed distributions are often different than theoretical expectations because of variability in the environment. Indeed, this geographic variability led Smailes (1946) to call Christaller's work into question. Such variability can be accommodated, however. For example, Rushton (1971) has shown how central place distributions can be deformed to account for variation in population densities.

Tests to evaluate the observed and expected distributions can be derived based on principles of Monte Carlo simulation.

Other (some rather obvious) factors to consider when formulating hypothesis tests:

- Water runs downhill. The characteristics of hydrological networks need to be considered in any informed spatial analysis of water-borne hazards.
- Wind speed and direction are not distributed in a circular normal fashion; they are seasonally biased. The characteristics of the prevailing winds need to be considered in any informed spatial analysis of air-borne hazards.
- Highways are railroads are not uniformly distributed. The characteristics of the transportation network needs to be considered in any informed spatial analysis of transit-borne hazards.
- Airline routes tend to be organized in a "hub and spoke" pattern. The characteristics of airline connections need to be considered in any informed spatial analysis of airline-borne hazards.
- Stuff diffuses. Analyses of events need to consider both contagious and hierarchical diffusion processes.

The introduction of different types of knowledge can aid in the analysis of spatial event data sets. Clearly there is a need for knowledge about the *geometrical* characteristics of the analysis—this may refer to scale, and related notions of dimension, distance, and topology. *Structural* knowledge refers to the type of knowledge that would be applied to specify the kinds of factors that would be included in any evaluation of patterns that have some expected values that are different than CSR. *Procedural* knowledge is used to select the algorithms that are used to implement the analyses.

6.0 Exploration and Visualization

Though it is clear that much work remains in the area of spatial analysis, it is also clear that new approaches to the exploratory visualization of the results of such analyses also must be developed. One way to glean a maximum amount of information from graphically-represented information is to define a set of criteria that would guide search for pattern and process. These criteria are then optimized and the results presented to the data explorer. This approach could be patterned after a process described by Armstrong, Xiao and Bennett (2003). In that approach choropleth classification criteria are placed into a multi-objective framework and optimal solutions that satisfy the criteria are calculated and reported in textual, graph, and map formats (Figure 1). The location of a solution in a decision space is indicated by the graph in the lower center of the figure. Class breaks are shown in the lower left window and the choropleth map that is the result of selecting one of the solutions is shown in the top left. Similar environments can be created for dot value and size trade-offs, and for graduated symbol maps in which symbol size trade-offs exist.

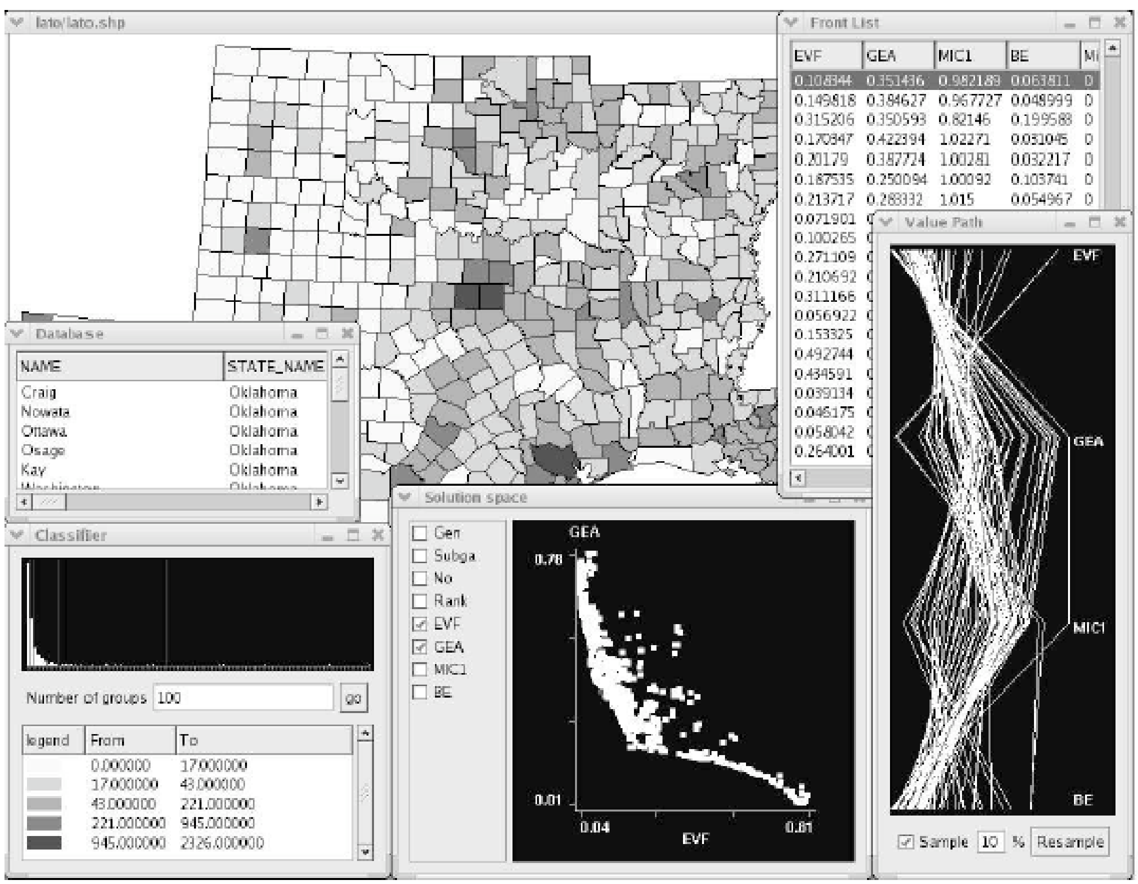

Figure 1. The multi-criteria choropleth map classification environment described by Armstrong, Xiao and Bennett (2003).

References

Armstrong, M.P., Xiao, N. and Bennett, D.A. 2003. Using genetic algorithms to create multicriteria class intervals for choropleth maps. *Annals of the Association of American Geographers*, **93** (3) 595-623.

Boots, B. and Getis, A. 1988. *Point Pattern Analysis*. Beverly Hills, CA: Sage Publications.

Cuzick, J. and Edwards, R. 1990. Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society*, *Series B*, **52**, 73–104.

Getis, A. and Boots, B. 1978. *Models of Spatial Processes: An Approach to the Study of Point, Line and Area Patterns*. New York, NY: Cambridge University Press.

Getis, A. and Ord, J.K. 1992. The analysis of spatial association by distance statistics. *Geographical Analysis* 24 189-206.

Kulldorff M. 1997. A spatial scan statistic. *Communications in Statistics: Theory and Methods*, 26:1481-1496.

Ord, J. K. and Getis, A. 1995. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical Analysis,* 27, 286-306Patil

Nystuen, J, 1963. Identification of some fundamental spatial concepts. *Papers of the Michigan Academy of Sciences* **48**, 373-384. (Reprinted in B. Berry and D. Marble, 1968 *Spatial Analysis: A Reader in Statistical Geography*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.)

Mark D. M. and Smith, B. In press. A Science of Topography: Bridging the Qualitative-Quantitative Divide In Michael P. Bishop and Jack Shroder, editors *Geographic Information Science and Mountain Geomorphology* Chichester, England: Springer-Praxis.

Pfaltz, J. 1976, Surface networks. *Geographical Analysis*, **8**, 77-93.

Rogerson , P. 2001. Monitoring point patterns for the development of space-time clusters. *Journal of the Royal Statistical Society, Series A* , 164: 87-96.

Rushton, G. 1971. Map transformations of point patterns: central place patterns in areas of variable population density. Discussion Paper No. 17. Iowa City, IA: Department of Geography, The University of Iowa.

Rushton, G. 2003. Public health, GIS, and spatial analytic tools. *Annual Review of Public Health* 24:43-56.

Smailes, A. 1946. The urban mesh of England and Wales. *Transactions, Institute of British Geographers*, 11:

Warntz, W. 1966, The topology of a socioeconomic terrain and spatial flows. *Papers of the Regional Science Association*, **17**, 47-61.