Geographical & Environmental Modelling, Vol. 6, No. 1, 2002, 41–57

Carfax Publishing
Taylor & Francis Group

Spatial Structures and Scale in Categorical Maps

JONATHAN D. PHILLIPS

ABSTRACT *Spatial variability is typically represented by categorical maps or by samples taken at specific locations (point data). The regions or patches represented in categorical maps are intuitive and consistent with much geographical (and ecological, pedological, etc.) theory, but point data are more amenable to geostatistics and other approaches which allow examination of scales of variability and of spatial structure. The state probability function (SPF) is introduced as a tool for evaluating the spatial structure of landscapes represented as categorical maps. The SPF is based on the degree of similarity between mapped units or classificatory categories a given distance apart. A plot of the SPF vs. distance produces a graph which can be interpreted similarly to a variogram. The method is tested/illustrated using detailed soil map data from eastern North Carolina. The SPF is shown to be appropriately sensitive to major landscape boundaries across which the suite of soil types changes, and to the degree of clustering of similar units.*

Introduction

Several research traditions in physical geography are based on the premise that there are strong links between landscape patterns and landscape processes, function, and evolution. This is particularly explicit in landscape ecology (e.g. Forman, 1995; Gustafson, 1998; Kupfer, 1995) and in soil geography (Fridland, 1976; Hole & Campbell, 1985; Ibáñez *et al.*, 1995). In both subfields a number of methods have been developed to quantify and characterize spatial patterns, with the ultimate goal of identifying and elucidating landscape processes. The pattern–process link is perhaps less explicit in other subfields of physical geography, but seems clearly related to the form–process connections often addressed in geomorphology. Walsh *et al.* (1998) suggested that a more explicit application of the 'pattern–process–scale paradigm' of landscape ecology to geomorphology would be fruitful.

Spatial patterns are generally represented either as categorical maps or as spatial point samples. Maps offer many advantages, but point data allow the use of spatial statistical tools to address important questions such as appropriate scales of analysis, and the quantification of spatial structure. The purpose of this paper is to introduce a method for identifying scales of variability and quantifying spatial structure of categorical maps, in a manner analogous to that of the geostatistical variogram.

Jonathan D. Phillips, Tobacco Road Research Team, Department of Geography, University of Kentucky, Lexington, KY 40506, USA. E-mail: jdp@pop.uky.edu

1361-5939 print/1469-8323 online/02/010041-17
DOI: 10.1080/13615930220127288

The quantitative tools for evaluating spatial patterns of landscapes may be grouped into two general types, associated with categorical maps and point data. Indices essentially generate a single number to characterize some aspect of the spatial pattern of a categorical map. Examples include entropy statistics and indices of fragmentation and contagion. The second type involves functions, which generally involve the depiction (commonly by a plot or graph) of some measure of spatial variability over size, scale, or distance. Examples include autocorrelation, spectral, and semivariance functions. Reviews are given by several authors (Gustafson, 1998; Haines-Young & Chopping, 1996; Ibáñez *et al.*, 1995; McBratney, 1998). There are examples of hybrids or combinations of the two, as when indices of spatial structure are derived from functions, or when indices are calculated at different resolutions to produce a function. However, Gustafson's (1998) review of the state of the art in quantifying spatial pattern identified a clear gap between the type of information generated by categorical maps and the tools available for characterizing their spatial patterns, and geostatistical functions for quantifying the spatial structure of point data.

There is therefore a need for a spatial function applicable to categorical data. More explicitly, the idea is to develop a method whereby a graph comparable to, and interpreted in the same way as, a variogram may be constructed for categorical maps and data. The variogram or semivariogram is a common tool for use with quantitative data in geostatistics and spatial statistics (McBratney, 1998; Oliver *et al.*, 1989).

Spatial Functions for Categorical Maps

Variograms and related geostatistical functions require interval or ratio-level quantitative data. It is therefore difficult to apply these methods to categorical maps. In some cases this can be overcome by deriving quantitative measures. For example, soil or community types on a soil or vegetation map might be assigned quantitative values for particular parameters via lookup tables. Alternatively, a cartometric approach might be used—for example, at sample points on the map the distance to the nearest soil or vegetation boundary could be measured. Another approach is that of Bregt *et al.* (1992), who devised the state probability difference function (spdf), described in more detail below. The spdf is based on categorical data and produces a plot which can elucidate the structure of spatial variability. The weaknesses of the spdf are that it does not account for varying degrees of similarity among categorical units, and does not produce a graphic resembling a variogram.

A more familiar approach is the adaptation of geostatistics to categorical attributes. This is described briefly here to provide a frame of reference before presenting the state probability function. Goovaerts (1999) discusses categorical applications of geostatistics in the context of examining the spatial variability of measured soil parameters grouped into several categories—for example, pH values grouped into reaction classes, or contamination levels classified according to risk or toxicity thresholds. Let a quantitative, point-measured variable x be grouped into a categorical attribute S with k possible states S_k, $k = 1, 2, \ldots, k$ and u_a denote a unique location. Then the function

$$i(u_a; z_k) = 1 \qquad \text{if } z(k_a) = s_k$$

and

$$i(u_a; z_k) = 0 \qquad \text{otherwise}$$

where z_k is the value or state of the categorical variable.

Goovaerts (1999, p. 56) then presents a version of the semivariance function:

$$\hat{\gamma}(h;z_k) = 1/[2N(h)] \sum_{a=1}^{N(h)} [i(u_a;s_k) - i(u_a + h;s_k)]^2 \tag{1}$$

where $N(h)$ is the number of points h units apart. Semivariance is plotted against the spatial lag distance h to produce what Goovaerts (1999) calls an indicator variogram, which is the basis for indicator kriging.

There is no reason equation (1) could not be applied to data which are categorical to begin with. Like the spdf, the categorical semivariance function does not recognize varying degrees of similarity—two locations are either identical (1) or different (0). Also like the spdf, but unlike the regular variogram, theoretical limits for the indicator variogram are easily established: $1/[2N(h)]$ where all points are the same, 0.5 where no two points are the same, and $(1/k)^2/2$ for a random sequence where any of the k states is equally likely at any location.

The goal here is to produce a function which recognizes varying degrees of similarity in categorical data, which produces a visual tool comparable to the variogram, and which can be interpreted on an intuitively appealing 0 to 1 scale. The state probability function described below is applied here as a new and separate method, but can readily be generalized into modifications of indicator geostatisics or the spdf that account for varying degrees of difference between categorical variables.

The State Probability Function

The state probability function (SPF) is designed to determine how the spatial structure and variability of a categorical phenomenon is related to spatial scale or distance. The SPF is intended to produce information broadly similar to that of equation (1), with two major exceptions. The first is that degrees of difference between map categories can vary according to degrees of difference rather than in a binary (same or different) fashion. Second, the SPF is scaled to an intuitively appealing 0 (no difference) to 1 (maximum difference) scale.

Let a landscape or spatial pattern take one of q distinct states S_k, $k = 1, 2, 3, \ldots, q$. These states could be, for instance, soil mapping units, morphostratigraphic units, or vegetation community types. Associated with each k is some indicator of similarity assigned to each map category. These are arranged such that the states are increasingly dissimilar from 1 to q. Thus there is a similarity gradient, so that the similarity of system states S_i and S_j is given by the absolute value of $(S_i - S_j)$. If $S_i - S_j = 0$ the system states are identical. The maximum degree of dissimilarity would be associated with $|S_i - S_j| = S_q - S_1$.

In this study the states are soil types. Other examples might include successional stages, or ecosystem or community types as they occur along gradients of climate or disturbance. Early successional states, for instance, would be more similar to each other than to late successional states. Or, for soils or landforms in a catenary sequence, those units in comparable catenary positions would be more similar than those at different points on a catena. For example, suppose we are analyzing a map of vegetation communities that includes categories of salt marsh, brackish marsh, freshwater marsh, freshwater swamp, moist oak flats, and mesic upland forests. Comparing two points that are both oak flats would yield a value of zero. Comparing salt marsh with mesic forests would constitute the maximum degree of difference. The comparison of, say, freshwater marsh and swamp would yield a positive value less than the maximum. These varying degrees of similarity are reflected in the similarity indicators, which may be based on the arrangement of the states or categories along some gradient, a ranking, or a quantitative index assigned to each state.

A state dissimilarity index ($\varnothing$) for any two locations or map units S_i, S_j is defined by

$$\varnothing = |S_i - S_j| / (S_q - S_1). \tag{2}$$

This index would be zero where the states are identical, and $\varnothing = 1$ where there is the maximum possible degree of dissimilarity. The similarity index is therefore context sensitive.

A state probability function can be computed, which represents the probability that the same system state will be found at some specified separation distance:

$$\mathrm{SPF}_h = \left(\sum^{n-h} \varnothing_h \right) \Big/ (n - h) \tag{3}$$

where h is the lag or separation distance. For a spatial transect with samples at 10 m intervals, for instance, SPF_h for $h = 3$ would be based on all pairs of sites 30 m apart. The emphasis here is on spatial structure, but the SPF could also be applied to temporal sequences, such as identified states or stages in an historical timeline, or horizons, facies, etc. in a dated stratigraphic sequence. The application of the SPF to a vertical feature such as a soil or weathering profile or sedimentary sequence might be useful in determining the extent to which the sequence may be time-transgressive. This type of application will be illustrated later.

Plotting SPF_h against h allows the structure of spatial variation in system states to be assessed. This state probability plot is broadly analogous to a variogram, particularly the indicator variogram associated with indicator kriging. One important difference between the SPF and traditional indicator geostatistics is that the variogram is derived from quantitative measurements (interval or ratio data), while the SPF is based on a qualitative categorization of system state (ordinal data). Another is that the SPF must range from 0 to 1, while the upper limit to the indicator variogram is 0.5. Some simple modifications to indicator kriging designed for use with categorical data would give information similar to that of the SPF function.

In a random sequence, where all states are equally likely at all locations, the SPF would be the same at all lags:

$$\mathrm{SPF}_{\mathrm{random}} = \left\{ \sum_{i=1}^{q} \left[\left(\sum_{j=1}^{q} |S_i - S_j| \right) \Big/ q \right] \right\} \Big/ (q - 1). \tag{4}$$

Lower SPF values indicate a higher degree of system state similarity. This could be interpreted as spatial clustering or persistence. Low SPF values would reflect not only large contiguous areas of the same type, but also gradual transitions between similar types. Higher SPF values signify a lower degree of similarity. This suggests more fragmentation or smaller areas of categorical units, and more abrupt transitions between dissimilar states. Quantitatively, persistence and clustering is indicated where

$SPF < SPF_{random}$. Where the SPF exceeds the random value, this suggests a situation whereby the existence of a given state decreases the probability of finding a similar state at the associated lag distance.

The interpretation of the SPF and the plots of SPF vs. lag are in some ways analogous to interpretations of the geostatistical variogram. The plot indicates whether variability is bounded or unbounded. Bounded plots reach a maximum value (called a sill, to borrow a term from geostatistics). Beyond this point, the SPF may fluctuate narrowly around the sill value, but no further upward trend is evident. The lag or distance associated with the point at which the sill is reached represents the range of spatial dependence. The possibility of periodicity is indicated by dips or hole effects in the plot.

The SPF is an adaptation of the spatial probability difference function (spdf) applied to ordinal soil data by Bregt *et al.* (1992). The spdf is conceptually similar to the SPF, but is based strictly on whether a pair S_i, S_j are the same or different. In the SPF each individual comparison $\varnothing$ can range from 0 to 1, but in the spdf the comparable measure must be either 0 or 1. Further, the spdf is designed so that the value of unity is associated with similarity rather than difference, while the SPF takes the opposite approach. This is so that, like the familiar variogram, higher values are associated with greater difference and variance.

The SPF is advantageous for some problems in that it accommodates variable degrees of similarity. The spdf and the categorical semivariance are based on a binary classification—compared samples are either the same or different. Using examples of soils in this study, a soil of the Wagram series might be compared with soils of the Troup or Bayboro series. In a binary comparison system, Wagram vs. Bayboro and Wagram vs. Troup would both yield values of unity (or zero in the spdf or categorical semivariance). However, the Wagram and Troup soils are very similar, differing only with respect to the thickness of the sandy A and F horizons. The Bayboro series, however, is a poorly drained, finer grained soil formed in different parent material. The SPF, as applied to the study transects, yields $\varnothing = 0.83$ for Wagram vs. Bayboro and $\varnothing = 0.17$ for Wagram vs. Troup.

Random and Chaotic SPF Functions

A sequence of 170 random integers from one to nine was produced using a random number generator and rounding the results to the nearest integer. The SPF function was then calculated to lag $h = 42$, based on the rule of thumb commonly applied in spatial statistics whereby the maximum lag should be about 1/4 the total number of points in a series. At longer lags there are too few data points available. The SPF plot (Figure 1) shows the function oscillating around the expected SPF_{ran} value.

A chaotic sequence of the same length was also generated using the logistic map. Starting with a seed value of 0.5, the successive values were produced using this equation: $X_a = rX_{a-1}(1 - X_{a-1})$, with $r = 3.75$. This is known to produce a chaotic sequence of numbers between 0 and 1. The SPF was applied directly to this sequence; and to a sequence whereby the results were converted to integers from 1 to 9, with similar results. The SPF plot for the chaotic data is shown in Figure 2.

The chaotic sequence produces a plot that is subtly but distinctly different from that of the random sequence. The mean value of the SPF averaged over all lags is in both cases equal to the expected value for a random sequence (0.42). The three major differences in the plots are that, as compared with the random sequence, the chaotic plot exhibits a high SPF value at lag 1, with rapid decay, a greater variation

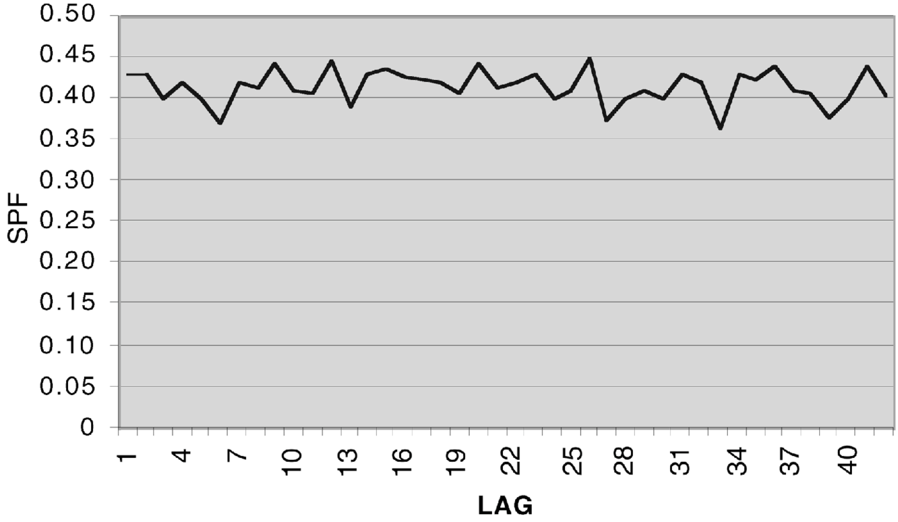

Figure 1. SPF plot for a synthetic random sequence.

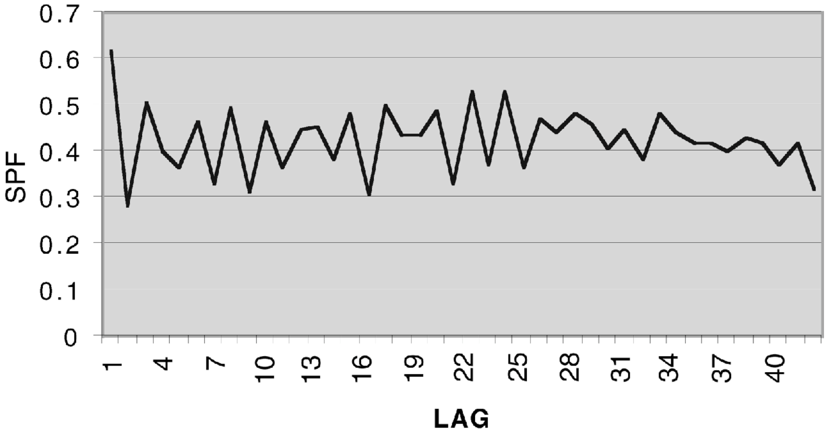

Figure 2. SPF plot for a synthetic chaotic sequence.

of SPF values, and an apparent trailing trend at higher lags. There is no obvious explanation for the latter, which could be a numerical artifact. The other features can be interpreted, however.

The high SPF value at lag $= 1$ is readily explained by the nature of the chaotic logistic equation. At lag 1, any given value of X_a cannot produce values close to X_a. High or low SPF values imply greater predictability—that is, given a particular state, one may predict either a high or low likelihood of finding a similar state. Thus a chaotic sequence would be expected to show either high or low (relative to the expected random value) SPF at lag 1, with a decay toward less predictability.

The random plot generally varies between 0.35 and 0.45, while the chaotic plot varies at lag > 1 between about 0.3 and 0.5. This is also explicable by the deterministic nature of the chaotic logistic equation. Extensive analysis of this equation shows that there are periods or cycles embedded within the apparent chaos (May, 1976). The high and low values of SPF are related to the cyclic elements.

These results suggest that the SPF function is capable of detecting subtly different signatures of spatial structures, and may be useful in distinguishing between deterministic and stochastic sources of uncertainty. This possibility deserves further investigation.

Difference Values

In the soil map example given below, the difference values are based on integer rankings from 1 to q. However, the SPF does not require integer similarity values, nor does it require equal intervals between adjacent system states or map categories. In some situations the assignment of similarity values is clearly a judgement call. Thus the spdf or categorical geostatistics may have some appeal in situations where objectivity is a concern. In some types of categorical data, such as some hierarchical classification schemes, it may be possible to objectively assign difference values based on the level of hierarchy at which two states or categories differ. Unfortunately, this is not the case with US soil taxonomy, where soils of different orders, for instance (such as a Paleudult and a Paleustalf) may be more similar than soils of the same order (for example a Paleustalf and a Cryoboralf).

In other cases indices of similarity may be applied. In the example application to a bisequal soil profile presented below, an index of pedogenic development is applied to soil horizons for the purpose of calculating difference values.

Study Area and Field Methods

The SPF function was tested, and its use illustrated, via applications to a soil landscape in eastern North Carolina. The North Carolina (NC) coastal plain is a flat to gently rolling region underlain by unconsolidated marine, coastal, and fluvial sediments of Pliocene to Holocene age. The climate is humid subtropical, with mean annual precipitation in the study areas of about 1200 mm yr^{-1}. The dominant soil order on uplands is Ultisols, which are highly weathered, acidic soils with argillic horizons and characterized by low base saturation.

The two study sites, the Clayroot and Littlefield sites (Figure 3), are in Pitt County, NC, near the villages of Clayroot and Littlefield, Both are small agricultural watersheds used for a multiyear-study of soil erosion, transport, deposition, and redistribution studies (Slattery *et al*., 1998). Soils were mapped in detail at both sites in connection with the latter study. Data from both the Clayroot and Littlefield sites are based on detailed soil maps produced at a scale of 1:3700 and 1:3550, respectively.

Published soil maps for the sites are not consistent with contemporary taxonomy, were found to be inaccurate, and in any case lack the resolution required for the geomorphic and hydrological studies at the site (see Slattery *et al*., 1998). Thus the soils at the Clayroot and Littlefield sites were field-mapped at a detailed scale. The mapping is based on a combination of soil pits and trenches, soil auger and probe sampling, and soil exposures in drainage ditches and canals. At Clayroot three pits were excavated, and a total of 274 auger or probe samples were taken at 10 m intervals along one or more transects across each field at the 19.4 ha study area. In the 71 ha Littlefield site there were two pits, two trenches, 33 shallow excavations at apparent sites of soil deposition or redistribution, and 260 auger or probe samples.

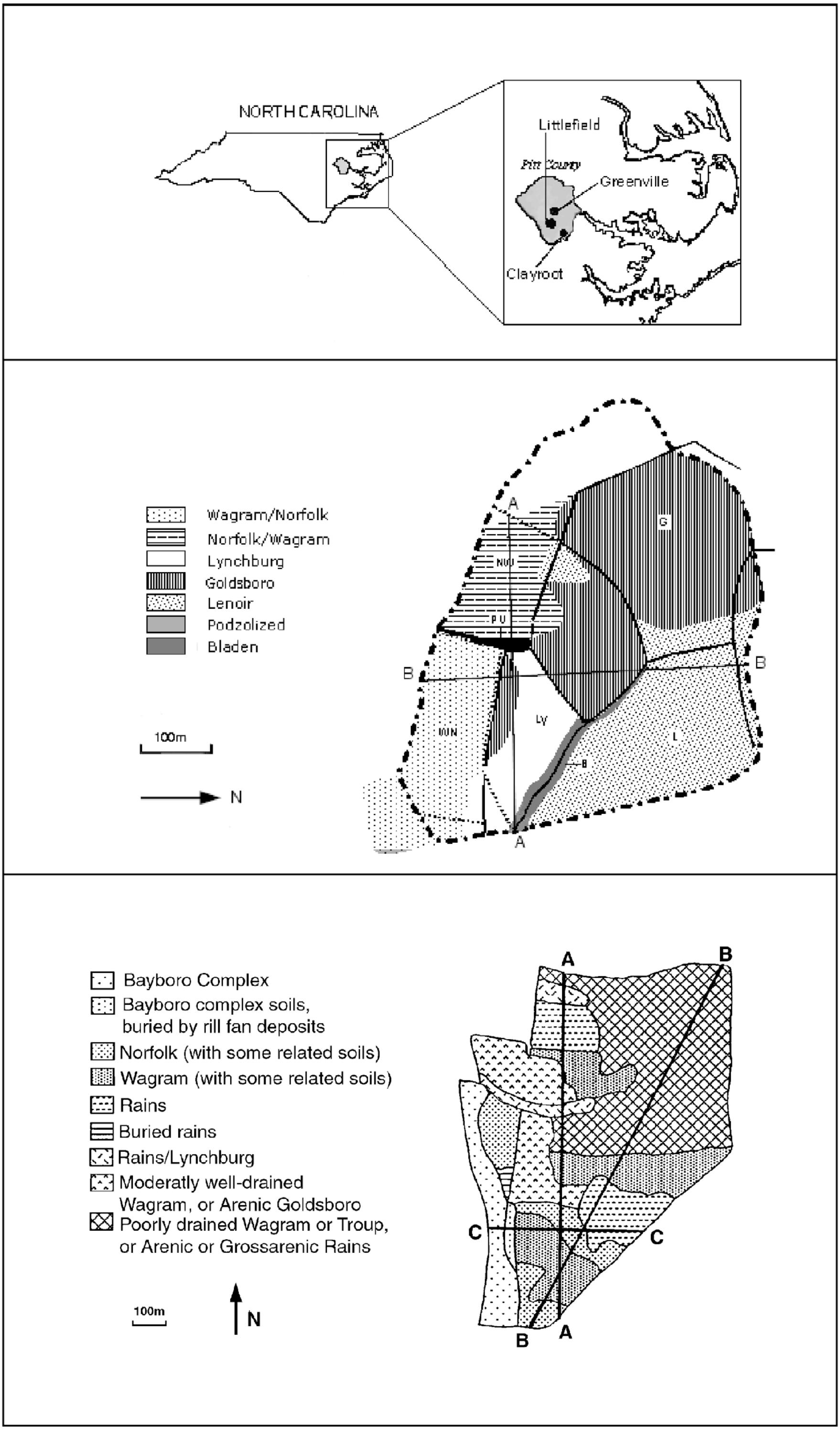

Figure 3. Soil maps and location map for the Clayroot and Littlefield study sites. Sample transects are shown on the soil maps.

At both sites, additional excavations and probe samples were taken on an *ad hoc* basis to confirm soil mapping and in conjunction with goals of the geomorphic studies (Phillips *et al*., 1999a, b). Soils were described and classified based on standard methods (Soil Survey Staff, 1993), and their morphology and landscape relationships are described in detail elsewhere (Phillips *et al*., 1999a, b).

Soil maps of Clayroot and Littlefield (Figure 3) were produced based on the fieldwork. Linear transects were drawn on the soil maps for sampling at 1 mm intervals. This sampling interval represents 3.7 and 3.55 m on the ground, respectively, for the two study areas. While the precise starting and ending points of these transects were randomly selected, their general locations were arbitrarily chosen to examine different soil landscape patterns. At Littlefield, for instance, transect A was chosen to reflect the full range of soil variability mapped at the site. Transect B was selected to examine a situation where a large portion of the transect is occupied by a single soil type. Transect C was picked to reflect a situation where two distinct geomorphic settings are encountered.

The soil mapping units in each sample were simply ranked on an integer scale according to similarity. For example, a sample of five soils would be ranked from 1 to 5, with soils 1 and 5 the most dissimilar, and soils one ranking apart the most similar. The rankings are unique to each sample so that the same soil series may have different rankings in different transects. This is the simplest possible way to assign similarity values, which need not be integers and which could have variable intervals between ranked soils. More precise and detailed assignments of similarity values could work as well. Experiments with more complex ranking schemes did not materially change the outcome of this particular analysis, so the simple system is retained here.

The ranking was based on genetic and morphological properties (Table 1). The Norfolk, Goldsboro, Lynchburg and Rains soils are formed in the same parent material and are part of a drainage catena (well to poorly drained). They differ substantially only in properties related to drainage and water table depth. The Wagram series differs from the Norfolk only in having thicker A and E horizons. The Arenic and Grossarenic Paleaquults are the poorly drained parts of a local drainage catena where Wagram is the well-drained member. The Podzolized Ultisols (see Phillips *et al*., 1999a, b) represent Norfolk, Wagram, and Arenic/Grossarenic Paleaquult soils buried by sandy aeolian deposits, and thus occupy one extreme of the rankings. The Lenoir, Bladen, and Bayboro series are generally finer than the other soils at the sites and are poorly drained. They occupy the opposite end of the rankings, most closely related to the more poorly drained members of the Norfolk catena. The relationships among these soils at the regional scale are outlined in detail by Daniels *et al*. (1984) and Phillips (1990). The local soil geography and geomorphology at the study sites, and detailed characteristics of the soils, are discussed by Phillips *et al*. (1999a, b).

Straight-line transects were used due to a deliberate effort to produce SPF plots for qualitatively different spatial patterns of soils. However, the SPF does not require a linear transect, and any regularly spaced set of sampling points including grids, triangular lattices, and non-linear transects would suffice.

Application to a Vertical Sequence

The SPF function can be applied to a vertical feature to address questions in the temporal domain. This is illustrated here with respect to a bisequal soil. The origin

Table 1. Taxonomy[a] and similarity rankings of soils at the study sites

Clayroot transects A and B		
Podzolized Ultisols[b]	Spodic Paleudult	1
Wagram	Arenic Kandiudult	2
Norfolk	Typic Kandiudult	3
Goldsboro	Aquic Paleudult	4
Lynchburg	Aeric Paleaquult	5
Lenoir	Aeric Paleaquult	6
Bladen	Typic Albaquult	7
Littlefield transect A		
Podzolized Ultisols[b]	Spodic Paleudult	1
Arenic & Grossarenic Paleaquults[c]		2
Wagram	Arenic Kandiudult	3
Goldsboro/Wagram[d]	Arenic Kandiudult/Aquic Arenic Paleudult	4
Norfolk	Typic Kandiudult	5
Rains/Lynchburg	Typic Paleaquult/Aeric Paleaquult	6
Rains	Typic Paleaquult	7
Littlefield transect B		
Podzolized Ultisols[b]	Spodic Paleudult	1
Arenic & Grossarenic Paleaquults[c]		2
Wagram	Arenic Kandiudult	3
Goldsboro/Wagram[d]	Arenic Kandiudult/Aquic Arenic Paleudult	4
Norfolk	Typic Kandiudult	5
Buried Bayboro	Thapto-Umbric Paleaquult	6
Bayboro	Umbric Paleaquult	7
Littlefield transect C		
Wagram	Arenic Kandiudult	1
Norfolk	Typic Kandiudult	2
Rains	Typic Paleaquult	3
Buried Bayboro	Thapto-Umbric Paleaquult	4
Bayboro	Umbric Paleaquult	5

[a] Classification to subgroup level, US Soil Taxonomy.

[b] These are soils, generally Norfolk and Wagram, buried by sandy aeolian sediments, in which secondary podzolization is occurring.

[c] This is a soil complex composed of soils similar to the Rains series, but with substantially thicker sand and loamy sand surficial horizons.

[d] This is a soil complex consisting of the Wagram series and a soil that is similar to the Wagram, but somewhat poorly drained, and similar to the Goldsboro, but arenic.

of bisequal soils (soils with two vertical horizon sequences) is uncertain. Though the definition of bisequal soils holds that the sequa should have formed contemporaneously, this situation is difficult to distinguish from features inherited from parent material, or from the effects of secondary pedogenesis in deposits overlying buried soil profiles.

The Littlefield and Clayroot sites contain soils with multiple horizon sequences, but these are known to result from secondary pedogenesis in sandy aeolian deposits (Phillips *et al.*, 1999a, b). The SPF was applied to the Mandarin soil series (sandy, siliceous, thermic Oxyaquic Alorthods), a bisequal soil mapped in eastern North Carolina and elsewhere in the southern Atlantic Coastal Plain (NCSS, 2000). The official series description for the Mandarin was taken from the US Department of Agriculture's official series descriptions database (Table 2; NCSS, 2000). The Langley-Turnbaugh and Evans (1994) index of soil development was applied to each horizon

Table 2. Profile description of the Mandarin series at the type location, taken from NCSS (2000). Values of the Langley-Turnbaugh and Evans (1994) development index are shown in brackets for each horizon

TYPICAL PEDON: Mandarin fine sand, on a smooth convex 0.5% slope, in forest. (Colors are for moist soil.)

A—0 to 4 inches; dark gray (10YR 4/1) fine sand; weak fine granular structure; very friable; extremely acid; clear wavy boundary (2 to 8 inches thick). [150]

E1—4 to 8 inches; light brownish gray (10YR 6/2) fine sand; single grained; loose; extremely acid; clear wavy boundary. [110]

E2—8 to 26 inches; light gray (10YR 7/1) fine sand; single grained; loose; strongly acid; abrupt wavy boundary. (Combined thickness of the E horizon is 10 to 24 inches.) [95]

Bh1—26 to 30 inches; very dark grayish brown (10YR 3/2) fine sand; weak fine subangular blocky structure; friable; in places sand grains well coated with organic matter; very strongly acid; gradual wavy boundary. [200]

Bh2—30 to 35 inches; very dark brown (10YR 2/2) fine sand; weak fine subangular blocky structure; friable; in places sand grains well coated with organic matter; few medium faint dark brown (10YR 3/3) soft masses of iron accumulation; very strongly acid; clear wavy boundary. [235]

Bh3—35 to 40 inches; black (5YR 2/1) fine sand; moderate medium subangular blocky structure; friable; in places sand grains well coated with organic matter; few fine prominent yellowish brown (10YR 5/4) soft masses of iron accumulation; very strongly acid; gradual wavy boundary. (Combined thickness of the Bh horizon is 5 to 35 inches.) [260]

BE—40 to 46 inches; brown (10YR 5/3) fine sand; single grained; loose; moderately acid; gradual smooth boundary (0 to 22 inches thick). [125]

E′1—46 to 56 inches; light gray (10YR 7/2) fine sand; single grained; loose; slightly acid; gradual wavy boundary. [100]

E′2—56 to 62 inches; white (10YR 8/1) fine sand; single grained; loose; few medium faint very pale brown (10YR 7/3) soft masses of iron accumulation; neutral; gradual wavy boundary. [110]

E′3—62 to 73 inches; grayish brown (10YR 5/2) fine sand; single grained; loose; neutral; gradual wavy boundary. (Combined thickness of the E′ horizon is 0 to 30 inches.) [120]

B′h—73 to 80 inches; black (10YR 2/1) fine sand; few fine distinct white (10YR 8/1) bodies; weak fine subangular blocky structure; friable; in places sand grains coated with organic matter; moderately acid. [205]

TYPE LOCATION: Duval County, Florida; 3000 feet north of Atlantic Boulevard, 0.7 mile west of Girvin Road in NE1/4NW1/4, Sec. 22, T. 2 S., R. 28 E.

described in the database, producing index values of 95 to 260. Higher values indicate a greater degree, and lower values a lesser degree, of pedogenic development. The application and calculation of the index is described in Table 3.

The system states and degrees of difference between sample points are defined by the profile development index. The top, A horizon, for each example, has an index value of 150 and is 4 in. thick. Thus the first four values in the sequence have a value of 150. This produced a sequence of 80 values for the 80 in. profile description, with 10 different index values (there are 11 horizons but two, the E1 and E′2, have the same value).

If the dual horizon sequences are inherited from parent material layering or from multiple depositional episodes the SPF plot should exceed the expected random value at some lag. This is because at some distance the comparisons will be dominantly between horizons formed from different materials or at different times and there should be ‘repulsion’; i.e. a higher likelihood of finding different degrees of development than would be expected in a random sequence. If the sequa formed contemporaneously (i.e. are acquired during pedogenesis rather than inherited), the SPF plot should stay below the expected random value.

Table 3. Development index (modified slightly and simplified from Langley-Turnbaugh and Evans (1994). The index is the sum of the point totals assigned below

A. Moist Munsell Color (rubification and melanization)
  1. 10 points for each increase in hue redness
     5Y = 0 points to 10R = 80 points
  2. 5 points for each increase in chroma for chroma 0 to 4; 10 points for each increase 4 to 8
     0 = 10 points; 4 = 30 points; 8 = 70 points
  3. 10 points for each decrease in value (increase in darkness)
     2 = 90 points; 10 = 10 points

B. Clay films. Points assigned for each increase in abundance, thickness, and location of clay films
  1. very few = 10 points; few = 20; common = 30; many = 40; continuous = 50
  2. thin = 10 points; moderately thick = 20; thick = 30
  3. 10 points for pore linings; 20 for ped faces or bridging grains

C. Dry or moist consistence. 10 points for each increase in hardness
  1. loose = 10 points to extremely firm = 60 points (moist)
     loose = 10 points to extremely hard = 60 points (dry)

D. Structure (grade and type)
  1. weak = 10 points; moderate = 20; strong = 30
  2. massive or single grained = 0 points; granular = 5; platy = 10; blocky (angular or subangular) = 30; prismatic or columnar = 40

E. Cutans or cementing agents common: 50 points

Results

Clayroot and Littlefield Soil Maps

The soil geography, pedology, and geomorphology of the sites is described in detail elsewhere (Phillips *et al.*, 1999a, b). The SPF_h values were plotted against h for each transect, up to a value of h equal to approximately 25% of the total length of the sample transect.

Transect A at Clayroot (Figure 4) does not reach a sill value, indicating that spatial dependence persists up to at least $h = 25$ (92.5 m), and that the farther apart one

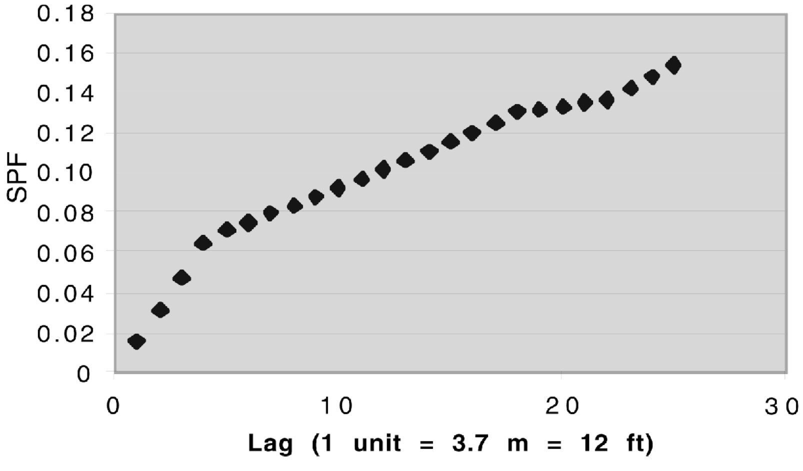

Figure 4. SPF plot for Clayroot transect A.

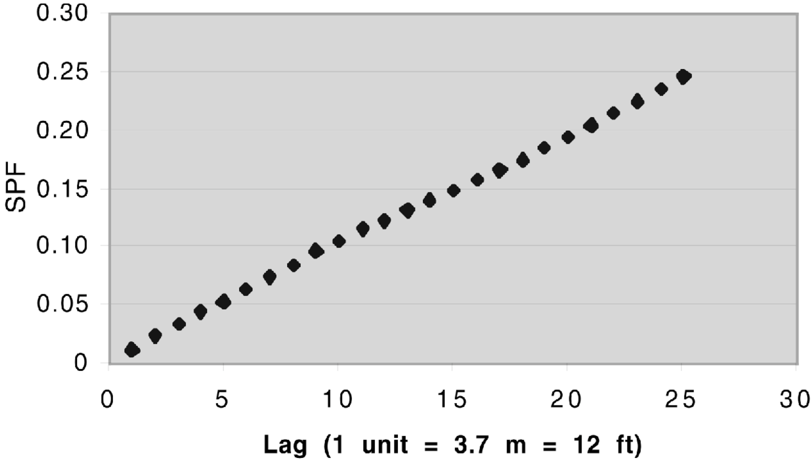

Figure 5. SPF plot for Clayroot transect B.

gets up to at least that distance, the more different the soils are likely to be. This reflects the fact that one end of the transect is dominated by the poorly and somewhat poorly drained Bayboro and Lynchburg soils, while the other is the well-drained Norfolk/Wagram complex. Likewise, Clayroot B (Figure 5) has an unbounded SPF plot which accurately reflects different soil suites characterizing either end of the transect. Both samples, particularly transect A, have values well below that expected for a random sequence. The spatial structure is thus characterized by persistence and contiguity over distances of about 100 m or less. The higher SPF values in transect B reflect a more irregular pattern, which includes some small soil bodies which do not show up on the publication-size map.

The transects at Littlefield were purposefully chosen to reflect three different soil map trends, and the SPF plots reflect this. Transect A (Figure 6), chosen to reflect the full variety of soils mapped at the site, reaches a sill at $h = 15$ and a maximum

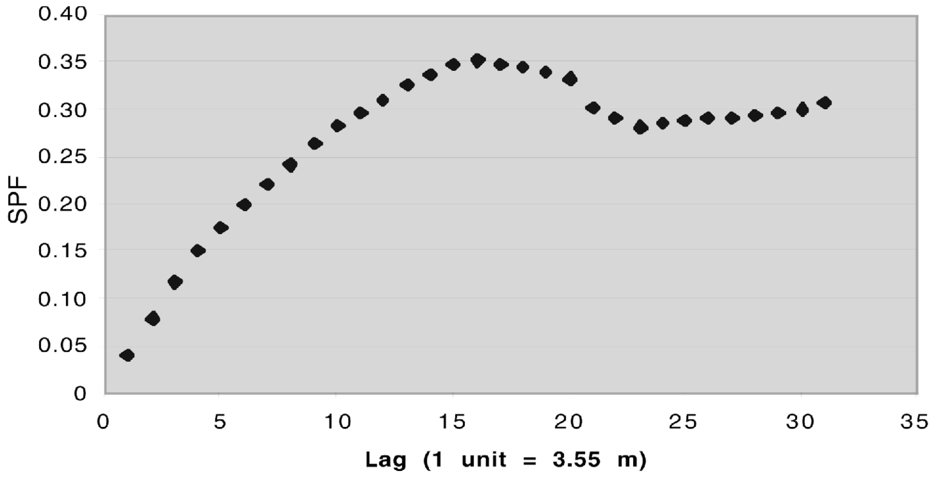

Figure 6. SPF plot for Littlefield transect A.

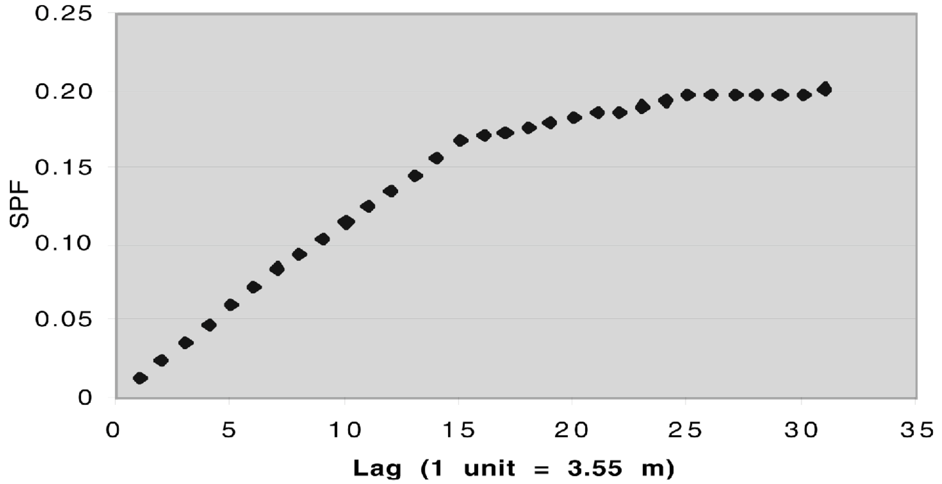

Figure 7. SPF plot for Littlefield transect B.

at $h = 16$ (about 55 m). The maximum value is above the expected value for a random sequence (0.326), but the slight downturn in the graph suggests that the sill value is roughly equal to $\mathrm{SPF}_{\mathrm{random}}$. The Littlefield B plot (Figure 7) clearly shows the influence of having roughly half the sample occupied by a single contiguous soil type, as the SPF reaches a sill value of only 0.2. There are three regions of the plot. The SPF rises relatively steeply from $h = 1$ to 16 (4 to 57 m), steadily but less steeply from $h = 16$ to 25 (60 to 89 m), and shows no spatial dependence beyond $h = 25$ (90 m).

Littlefield transect C (Figure 8) reflects a different situation. Not only does it follow a lowland-to-upland gradient, but also incorporates two distinct environments. At the lower end of the transect, soil parent materials are poorly drained alluvial terrace sediments of the Bayboro and buried Bayboro types. The upper end is typical

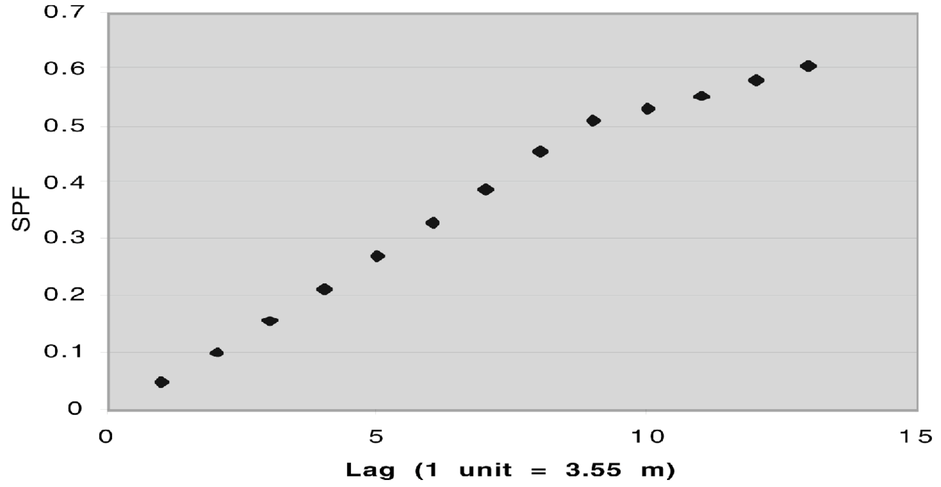

Figure 8. SPF plot for Littlefield transect C.

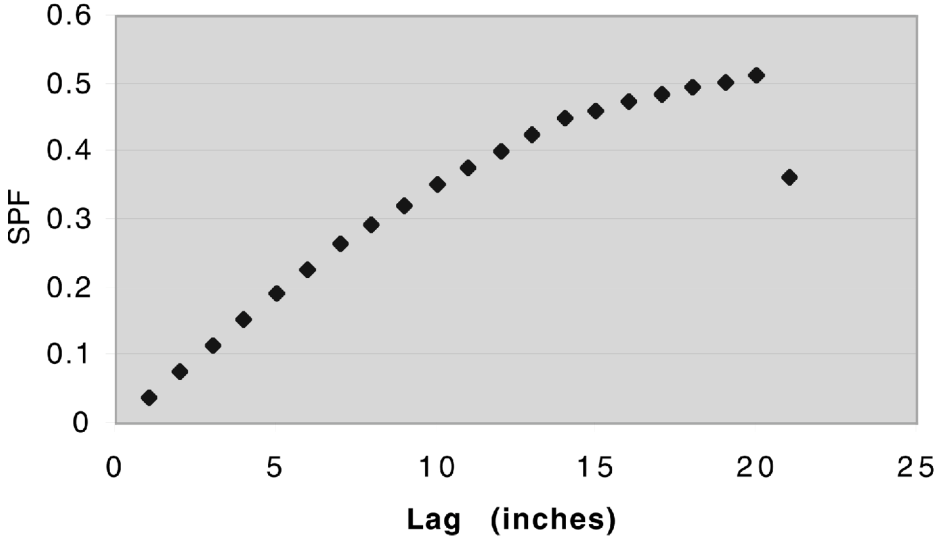

Figure 9. SPF plot for Mandarin soil profile (Table 2). The final data point represents the expected random value.

coastal plain upland parent material, and may be further subdivided into a well-drained portion (Norfolk and Wagram) and a poorly drained area (Rains). It may also be noted that, unlike any other transect in this study, all individual soil types occur in single contiguous groupings along the sample. The result is that the SPF plot not only displays unbounded variation, but reaches a very high SPF value of about 0.6. From $h = 6$ (21 m) onward, the SPF exceeds the expected value for a random sequence when $q = 5$ (0.32). This means that over distances greater than about 20 m, the spatial dependence is in the form of repulsion or dispersal—pairs of sites are more likely to be different than similar.

Mandarin Soil Profile

The SPF plot for the vertical sequence of the Mandarin soil profile is shown in Figure 9. The plot exceeds the expected random value at all lags greater than about 12 in. (30 cm). On average, points more than about a foot apart in the vertical sequence differ more in their degree of pedogenic development than would be anticipated in a random sequence. The interpretation is similar to that of Figure 8, and in this case suggests that the dual horizon sequences are inherited. Because the Mandarin is formed in sandy marine sediments such as coastal dunes, it is likely that multiple depositional episodes account for the vertical variations.

Discussion

The SPF function provides a means for quantifying spatial scale and structure in categorical maps. For the testing and illustrative purposes of this study, the SPF was applied to transects where the qualitative nature of the spatial pattern was already known, so that the SPF plots could be examined for fidelity to phenomena observed in the field. However, the SPF could as easily be applied to a grid pattern or any other regularly spaced spatial lattice. Difference values were established on an integer scale, but non-integer metrics could as readily be used.

The SPF provides at least some of the advantages associated with analysis of categorical maps and with geostatistical analysis of point data. The SPF recognizes varying degrees of similarity rather than depending on a binary same:different comparison. It also produces values lying along an intuitive 0 to 1 continuum.

Comparison of the SPF plots for the synthetic random and chaotic sequences suggests that the SPF may be sensitive to deterministic complexity signals. This needs further investigation, particularly in view of the fact that real geographic data are unlikely to reflect pure deterministic chaos or white noise. There is also the potential to apply the SPF in the temporal domain—for example, in a dated stratigraphic sequence. This, too, deserves further research.

The application of the SPF to the detailed soil maps shows that the function is sensitive to known spatial properties. When the sample transects cross major soil boundaries where distinctly different suites of soils exist on either side, the SPF plot is unbounded, reflecting the fact that greater dissimilarity accompanies greater separation distances. Likewise, the plots are bounded when no such boundaries are crossed. The SPF values accurately reflect the degree of spatial variability, rising for more complex and varied patterns.

The application to the vertical sequence illustrates the potential to use the SPF function in the temporal domain to address questions such as whether vertical variations are inherited or acquired. In the case of the bisequal Mandarin soil, the repulsion evident in the SPF plot suggests multiple depositional episodes. Note that the information here is too limited to settle the question of how the Mandarin horizon sequences formed; the point is to illustrate the use of the SPF in the vertical dimension to address a temporal question.

Conclusions

The state probability function is a useful tool for evaluating the spatial structure of landscapes represented as categorical maps. It allows the construction of a plot which can be interpreted similarly to a variogram, and has the additional advantage of having a predefined, finite range and a reference value associated with randomness. SPF analyses of data from detailed soil maps from eastern North Carolina show the function to be appropriately sensitive to major landscape boundaries across which the suite of soil types changes, and to the degree of clustering of similar units.

Acknowledgements

The fieldwork at the Littlefield and Clayroot sites was supported by US Department of Agriculture grant number 95–37107–2180. Paul Gares and Mike Slattery did most of the work on that project, leaving me free to dig holes.

References

Bregt, A.K., Stoorvogel, J.J., Bouma, J. & Stein, A. (1992) Mapping ordinal data in soil survey: a Costa Rican example. *Soil Science Society of America Journal*, 56, 525–531.

Daniels, R.B., Kleiss, H.J., Buol, S.W., Byrd, H.J. & Phillips, J.A. (1984) *Soil Systems of North Carolina*. Raleigh: NC Agricultural Research Service Bulletin 467.

Forman, R.T.T. (1995) *Land Mosaics*. New York: Cambridge University Press.

Fridland, V.M. (1976) *The Pattern of the Soil Cover* (translated from Russian). Jerusalem: Israel Program for Scientific Translations.

Goovaerts, P. (1999) Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma*, 89, 1–45.

Gustafson, E.J. (1998) Quantifying landscape spatial pattern: what is the state of the art? *Ecosystems*, 1, 143–156.

Haines-Young, R. & Chopping, M. (1996) Quantifying landscape structure: a review of landscape indices and their application to forested landscapes. *Progress in Physical Geography*, 20, 418–445.

Hole, F.D. & Campbell, J.B. (1985) *Soil Landscape Analysis*. Totowa, NJ: Rowman & Allanheld.

Ibáñez, J.J., De-Alba, S., Bermudez, F.F. & Garcia-Alvarez, A. (1995) Pedodiversity: concepts and measures. *Catena*, 24, 215–232.

Kupfer, J.A. (1995) Landscape ecology and biogeography. *Progress in Physical Geography*, 19, 18–34.

Langley-Turnbaugh, S.J. & Evans, C.V. (1994) A determinative soil development index for pedo-stratigraphic studies. *Geoderma*, 61, 39–59.

May, R.M. (1976) Simple mathematical models with very complicated dynamics. *Science*, 261, 459–467.

McBratney, A.B. (1998) Some considerations on methods for spatially aggregating and disaggregating soil information. In: P.A. Finke, J. Bouma & M.R. Hoosbeek, Eds, *Soil and Water Quality at Different Scales*. Dordrecht: Kluwer, pp. 51–64.

NCSS (National Cooperative Soil Survey) (2000) Official Series Descriptions. US Department of Agriculture database, URL http://www.statlab.iastate.edu/cgi-bin/osd/.

Oliver, M., Webster, R. & Gerrard, A.J. (1989) Geostatistics in physical geography. Part I: theory. *Transactions of the Institute of British Geographers*, 14, 259–269.

Phillips, J.D. (1990) Relative ages of wetland and upland surfaces as indicated by pedogenic development. *Physical Geography*, 11, 363–378.

Phillips, J.D., Gares, P.A. & Slattery, M.C. (1999a) Truncation and accretion of soil profiles on coastal plain croplands: implications for sediment redistribution. *Geomorphology*, 28, 119–140.

Phillips, J.D., Golden, H., Capiella, K., Andrews, B., Middleton, T., Downer, D., Kelli, D. & Padrick, L. (1999b) Soil redistribution and pedologic transformations in coastal plain croplands. *Earth Surface Processes and Landforms*, 24, 23–39.

Slattery, M.C., Gares, P.A. & Phillips, J.D. (1998) Quantifying soil erosion and sediment delivery on North Carolina coastal plain croplands. *Conservation Voices*, 1, 20–25.

Soil Survey Staff (1993) *Soil Survey Manual*. Washington: US Department of Agriculture, Handbook 18.

Walsh, S.J., Butler, D.R. & Malanson, G.P. (1998) An overview of scale, pattern, process relationships in geomorphology: a remote sensing and GIS perspective. *Geomorphology*, 21, 183–205.