Geoderma 267 (2016) 58–64

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Identifying sources of soil landscape complexity with spatial adjacency graphs

Jonathan D. Phillips

Tobacco Road Research Team, Department of Geography, University of Kentucky, Lexington, KY 40506-0027, USA

A R T I C L E   I N F O

Article history:
Received 10 September 2015
Received in revised form 15 December 2015
Accepted 22 December 2015
Available online 7 January 2016

Keywords:
Soil complexity
Soil variability
Spatial adjacency graph
Soil factor sequence
Spectral radius

A B S T R A C T

Soil landscapes often exhibit complex spatial patterns, with some aspects of soil variation apparently unrelated to measurable variations in environmental controls. However, these local, contingent complexities are not truly random or intrinsically unknowable. The purpose of this work is to develop and apply a method for identifying or teasing out causes of soil landscape complexity. Soil spatial adjacency graphs (SAG) represent the geography of soil landscapes as a network that can be analyzed using algebraic graph theory. These SAGs include linear sequential subgraphs that represent sequences of soil forming factors. The number and length of these soil factor sequences (SFS), and their associated spectral radius values, determine whether the SFS are sufficient to explain the spatial pattern of soil adjacency. SAGs and associated graph theory methods provide useful tools for guiding pedological investigations and identifying gaps in knowledge. The methods also allow sources of soil landscape complexity and variability to be determined in a way that can help assess the underlying deterministic sources of chaos and dynamical instability in pedology. The approach is applied to a soil landscape in central Kentucky, producing a SAG with 13 nodes (soil types) and 36 links indicating whether the soils occur contiguously. Five SFS were identified, the sum of whose spectral radius values is 6.35. The spectral radius of the SAG is 6.56, indicating that the SFS can explain most, but not all, of the complexity of the soil relationships. The analysis also points to potential environmental controls that could potentially enable full explanation.

1. Introduction

1.1. Soil variability and complexity

Spatial variability of soils and soil properties, and pedodiversity (related but not identical concepts) has often been found to encompass unexplained variation. Soil variability has long been attributed to a systematic, predictable component, and an apparently random noise component (Burrough, 1983). Some studies have found the apparent noise in many cases is actually attributable to deterministic complexity associated with dynamical instability and chaos (e.g. Culling, 1988; Ibáñez et al., 1990; Phillips, 1993, 2000, 2001b; Ibáñez, 1994; Phillips et al., 1996; Liebens and Schaetzl, 1997; Webster, 2000; Phillips and Marion, 2005; Toomanian et al., 2006; Milan et al., 2009; Borujeni et al., 2010). Several studies of pedodiversity have also found evidence of soil diversity arising from local, contingent, unobserved factors unrelated to measurable variations in soil forming factors such as topography, (micro)climate, parent material, vegetation, and geomorphic history (Phillips, 2001a, 2013b; Phillips and Marion, 2005; Ibáñez et al., 2009, 2013). Dynamical instabilities can magnify the effects of small variations in initial conditions or localized disturbances, and complex nonlinear interactions may obscure relationships between soils and soil-forming factors. Few, if any, would argue that these local, contingent, complexities are truly random or intrinsically unknowable. The purpose of this work is to develop and apply a method for identifying or teasing out pedological causes of complexity in soil geography at the landscape scale.

The pseudo-random or "noise" component of soil variability can be considered as contingent, based on the notion that instability, local disturbances, and other deviations from systematic patterns are geographically and/or historically contingent (Phillips, 2013a, 2013b). This recognizes (or assumes) that the unexplained variations are controlled by unobserved local geographical variations in environmental controls and/or specific local disturbances or events. In previous work I analyzed the soil spatial variability at the landscape scale (study sites of 20–70 ha) to determine the relative importance of variability linked to soil forming factors (considered as soil factor sequences or SFS) versus that associated with contingent factors such as local disturbances and dynamically unstable magnification of minor initial differences (Phillips, 2013a). In the present study that approach is expanded to determining how many SFS or factorial relationships are necessary to explain the spatial pattern of soils. The (implicit or explicit) assumption of most analyses of soil geography, pedodiversity, and soil spatial variability is that

E-mail address: jdp@uky.edu.

http://dx.doi.org/10.1016/j.geoderma.2015.12.019
0016-7061/

even complex, unexplained variations can potentially be explained, given enough local information. However, it is not practical to make detailed pedon-by-pedon examinations at a density sufficient to explain the high degree of soil variability often observed at the landscape scale. The approach developed here allows the determination of the minimum number of SFS that must be identified to explain the spatial pattern of soil adjacency.

Soil forming factors are environmental controls that control or influence variations in soil properties and soil types. The classic factors are climate, biota, topography, parent material, and time or surface age. In addition, there may be other locally or regionally significant factors such as non-topographically driven (e.g., aeolian, tillage) soil redistribution, disturbance regimes, land use, etc. A number of specific criteria or variables are used to represent these factors. Soil mapping, as well as pedology and soil geography, are based on the concept of sequences of soil variation based on soil forming factors. For instance, climo- and lithosequences represent systematic variation of soil properties along gradients of climate, and lithology or parent material, respectively. Catenas, repeated sequences of soils formed from similar parent material under similar climatic conditions but whose characteristics differ because of variations in relief and drainage, are another example. SFS may represent spatial gradients, such as a climosequence along gradients of moisture or temperature. Other SFS may be associated with categorical variations that may or may not occur along spatial gradients – for instance, a SFS based on coarser to finer textural differentiations.

Note here a key difference between soil forming factors and distinguishing characteristics used to identify, classify, or map soils. Distinguishing characteristics are used in taxonomy and soil keys, and may include soil forming factors, such as parent material or landform position. However, distinguishing characteristics are often features acquired during pedogenesis are reflect soil forming factors. For example, diagnostic horizons, chemical properties such as pH or cation exchange capacity, and horizonation are distinguishing characteristics that reflect, but are not, soil forming factors. In general, soil forming factors are those that could be measured or identified independently of soil properties.

1.2. Spatial adjacency graphs

In a spatial adjacency graph (SAG) the graph nodes or vertices are nominal or categorical spatial entities—in this case soil types, but landform types, geological formations, or vegetation communities and other entities could be similarly treated. Any pair of nodes is connected (i.e., there exists an edge or link between them) if they are spatially contiguous. Thus, if soil types *A* and *B* at least sometimes occur adjacent to each other, they are connected. If they are never spatially adjacent, there is no edge connecting *A*, *B*. In the SAGs used by Phillips (2013a), for instance, connectivity or spatial adjacency was based on whether soil types (taxa) occurred within the same mapping unit, or in mapping units with shared boundaries. Heckmann et al. (2015) consider SAGs to be intermediate between spatially explicit graphs, where nodes represent specific locations or areas, and structural graphs, where nodes represent system components (for instance, most state-and-transition models can be represented as structural graphs).

The method in this paper is based on algebraic graph theory, which focuses on analysis of graph adjacency matrices. An adjacency matrix for a network with *N* nodes is an *N x N* matrix with cell entries of zero if the row and column nodes are unconnected, and nonzero otherwise. For the case of a SAG, cell values are *1* if the row and column nodes are spatially contiguous (e.g., share boundaries on a map or occur within the same mapping unit), and *0* otherwise (by convention, diagonal entries are zero).

Eigenvalues of the adjacency matrix may be simple or complex. The largest eigenvalue (λ_1) (or its real part if complex) of the adjacency matrix is the graph spectral radius. The spectral radius has a maximum value of *N-1* for a fully connected SAG (e.g., any soil type can occur adjacent to any other), and a minimum for linear sequential or chain-type graphs. Spectral radius is a key indicator of many network properties. λ_1 is sensitive to the number of cycles in the graph, defined as sequences of edges that begin and end at the same node. Spectral radius is also inversely related to critical coupling strength, a threshold at which a graph transitions from incoherent to coherent behavior (Restrepo et al., 2006, 2007). Coherence is not directly relevant to SAGs, but this property reflects the fact that spectral radius is an indicator of the complexity of the graph (see, e.g., Fath, 2007; Phillips, 2011a, 2011b).

Denoting the SAG as *G*, *G'* is a subgraph of *G* (i.e., a graph whose nodes and edges are subsets of *G*). Standard principles from algebraic graph theory show that

$$\lambda_1\ (G) > \lambda_1 (G') \tag{1}$$

$$\lambda_1 (G) < \sum^{n} \lambda_1 (G'_k) \tag{2}$$

where there are $k = 1, 2, \ldots, n$ subgraphs of *G*.

Phillips (2013a) used this in the analysis of soil SAGs, based on the idea that SFS are linear sequential subgraphs of the SAG. The sum of the spectral radii associated with each identified SFS (Λ) thus represents the total graph complexity that can be explained by these factorial relationships. If the spectral radius of the SAG is greater than Λ, there exists unidentified or unexplained complexity. If $\lambda_1 < \Lambda$, the SAG is over determined and graph complexity can be fully explained by identified SFS. Over determination occurs because SFS may contain redundant information (Phillips, 2013a). For instance, a topographic sequence may represent variations in elevation, slope, moisture, or insolation.

2. Theory

If the spatial pattern of soil can be wholly explained by soil forming factors, then spatial adjacency should be entirely determined by SFS. Note the difference between a spatially explicit approach and the spatial adjacency analysis. For example, in a spatially explicit assessment of, e.g., soil variation in relation to slope curvature, an unmeasured depression or concavity within an otherwise convex slope segiment could result in an apparently anomalous soil type. However, as long as the relationship between curvative and soil type is incorporated in an SFS for the study area, then failure to identify or measure a local variation in curvature will not affect the graph representation.

As noted above, a SFS can be represented as a linear sequential subgraph, a type of graph where $m = N - 1$ and the nodes are arranged in a chain. Classical linear vegetation succession models, for example, have this form, along with most catenary relationships and factor sequences in pedology. The maximum largest eigenvalue for a graph where $m = N - 1$ is

$$\lambda_{1,}(max) = \left[2\ (N-1)^2/N\right]^{0.5} \tag{3}$$

Thus as the size of the graph (length of the sequence *N*) increases, the maximum spectral radius increases as the square root of N. Several different connected graph types have $N - 1$ edges, including radiation-type structures (i.e., Phillips, 2011a). However, λ_1 for a chain or linear sequential graph structure is highly constrained such that $\lambda_1 = 1$ for $N = 2$, and for $N > 3$:

$$\lim_{N\to\infty} \lambda_1 = 2 \tag{4}$$

Therefore, for any SFS subgraph of a SAG

$$1 \le \lambda_1 (G') < 2 \tag{5}$$

Using Φ to denote the minimum number of SFS needed to fully account for the complexity of a SAG as indicated by its spectral radius,

$$\mathrm{INT}(\lambda_1/2 + 1) \le \Phi \le \mathrm{INT}(\lambda_1) + 1 \tag{6}$$