MULTISCALE ASSESSMENT OF LANDSCAPES AND WATERSHEDS WITH SYNOPTIC MULTIVARIATE SPATIAL DATA IN ENVIRONMENTAL AND ECOLOGICAL STATISTICS

by G. P. Patil[1], W. L. Myers[2], Zhen Luo[1], G. D. Johnson[1], and C. Taillie[1]

[1]Center for Statistical Ecology and Environmental Statistics
Department of Statistics
[2]School of Forest Resources and
Environmental Resources Research Institute

The Pennsylvania State University
University Park, PA 16802

NSF Project Officer: Penelope Firth
EPA Project Officer: Barbara Levinson

Prepared with partial support from the NSF/EPA Water and Watersheds Program, National Science Foundation Cooperative Agreement Number DEB-9524722. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agency and no official endorsement should be inferred.

Invited Paper in the Special Issue of
Mathematical and Computer Modeling on Stochastic Models in Mathematical Biology

Technical Report Number 97-1202
TECHNICAL REPORTS AND REPRINTS SERIES
January 1998

MULTISCALE ASSESSMENT OF LANDSCAPES AND WATERSHEDS WITH SYNOPTIC MULTIVARIATE SPATIAL DATA IN ENVIRONMENTAL AND ECOLOGICAL STATISTICS

G. P. Patil[1], W. L. Myers[2], Zhen Luo[1], Glen D. Johnson[1], and Charles Taillie[1]

[1]Center for Statistical Ecology and Environmental Statistics
Department of Statistics
[2]School of Forest Resources and
Environmental Resources Research Institute

The Pennsylvania State University, University Park, PA 16802 USA

Abstract

The paper attempts to provide a multiscale assessment of landscapes and watersheds using synoptic multivariate spatial data. Multiscale assessment is a frontier problem in environmental and ecological statistics today. The paper briefly deals with univariate surface data, multivariate signal data, and multicover categorical data, and applies stochastic conceptualization involving dendrogram trees and conditional entropies with special reference to the landscapes and watersheds of Pennsylvania.

Prepared with partial support from the NSF/EPA Water and Watersheds Program, National Science Foundation Cooperative Agreement Number DEB-9524722. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agency and no official endorsement should be inferred.

Keywords: Dendrogram tree; Transition matrix; Conditional entropy; Landscape mapping; Echelon analysis

1 Introduction and Summary

The general nature of the approach to multi-scale characterization of landscapes in this paper is through analytically induced pattern deterioration. The basic idea is akin to survival of the fittest in natural selection whereby strong elements persist and weaker ones are progressively vanquished. Although spatial pattern is a complex of properties, one can think in terms of large coherent patches that stand in contrast to their neighbors as being the stronger elements, whereas small internally variable patches having relatively similar neighbors are the weaker ones. As with a large field of contestants in a marathon, the weaker will drop out early while the stronger will withstand considerable stress before being subordinated. Multiscale character is thus reflected in the rate and extent of deterioration for pattern elements under the influence of a progressive degenerative process.

The paper considers three kinds of spatial data layers. Surface data such as elevations or concentrations constitute one kind. Signal data such as remotely sensed reflected energy constitute the second kind. And categorical data such as land cover, soil, or geology constitute the third kind. Analytical methods appropriate to each kind of data have been conceptualized and software capability for their application is being developed. Echelon hierarchies of hillforms organized as dendrogram trees apply to surface data, with pruning of the tree being the degenerative process. PHASE formulation by clustering is appropriate for signal data, with information capacity and patch size constraints serving to induce pattern deterioration. Random and modal filters bring about progressive deterioration of spatial pattern in categorical data.

Progressive filtering of categorical data has been studied intensively, and conditional entropy profiles are utilized to capture the trajectory of pattern deterioration across scales. Transition matrix models have been formulated that exhibit pattern deterioration similar to land cover maps of selected watersheds in Pennsylvania. Model parameters are currently being estimated for 104 major watersheds of Pennsylvania. The fitted transition matrices will be studied for their ability to characterize and differentiate among watersheds.

PHASE formulation not only provides for multi-scale exploration of signal data, but also offers an alternative to conventional methods of handling and analyzing digital image data in remote sensing. It compresses multi-band image data into a single layer of patch type identifiers with accompanying table(s) of characteristics for patch types. PHASE compressed image data is a value-added product that is beyond the scope of usual copyrights on image data, thus permitting unlimited distribution. The compression is sufficient to accommodate 10 scenes of Landsat Thematic Mapper formulations covering Pennsylvania on a single CD-ROM along with a variety of supplemental vector GIS data.

2 Constrained Topological Stochastics of Landscapes and Watersheds: Conceptualizing a Comparative Approach for Quantitative Spatial Data

2.1 Landscape Pattern Comparatives in Environmental Management

There is growing recognition of importance for both diversity and spatial structure in landscapes relative to ecosystem processes [1]. Oversimplifying considerably, basic concerns revolve around complexity and proximity regarding alternate pathways for environmental processes. Food chains, for example, involve energy fixation and consumption/respiration by sequences of different organisms. Plant species serve as primary producers, and animal species serve as consumers. Not all animal species can consume all plant species, thus restricting the possible transitions of energy between species. For an animal species to persist as one of the links in one of the energy pathways, it must have an energy source. The more isolated and remote the different energy sources, the less likely that alternates can be exploited if a current source becomes unavailable. There is thus stability in complexity of interspersed biotic elements because a multiplicity of alternate process pathways is maintained. Stochastic transitions among biotic elements thus become less constrained as diversity and proximity increase.

Superficially, diversity would seem to be a simple matter of variety. Under scrutiny, however, diversity is anything but a simple matter because it involves variety of things in a context. Both the nature of things and nature of context quickly become problematic to capture in a manner that is relevant to environmental processes. Even for plants as the simplest context, clear individuality is often lacking and elementary presence/absence changes with areal scope of determination. The mobility of fauna gives them spatial/temporal transience, making when and where to count them somewhat arbitrary.

Major concern here is with landscape diversity, for which allotment of a locality to a type depends on the scale or areal scope for which the determination is made. In a landscape sense, the virtues of diversity are also not without limit. Fauna often require varying degrees of "habitat integrity," as for several avain species that occupy "forest interior" or deep woods landscape settings. The perception of diversity and spatial structure in landscapes is therefore partly a consequence of both perspective and the manner in which things are perceived. Landscape absolutes tend to be elusive and perhaps even illusive. However, a comparative sequence of perceptions at varying scales can reveal differences between landscapes that have major environmental implications.

A stochastic view of landscapes is entirely consonant with the aforegoing indeterminacy. Since pattern is a multifarious concept, however, stochastic models encompass widely varying fidelity relative to actual landscape patterns. Stochastic constraints serve to particularize models for greater fidelity to specific sorts of landscape characteristics. Given a set of land cover categories mapped on a regular tesselation, for instance, one can compile patch size distribution without regard to patch type as a basis for modeling. Such a model would capture one characteristic of land cover pattern, and models of this nature could serve to distinguish grossly different landscapes relative to mean or variance for patch size. It could have equal fidelity, however, for landscapes having small patches of herbaceous cover with large patches of forest cover versus large patches of herbaceous with small patches of forest. Further constraining the model to recognize different patch size distributions by cover type would resolve this particular sort of ambivalence. However, the kinds of process interactions that occur and propagate across a landscape depend heavily on which kinds of elements are adjacent to which other types of elements. Much greater fidelity of model to landscape is therefore obtained by constraining to an observed edge apportionment between types.

These sorts of landscape properties are usually specified in terms of a rather large suite of so-called landscape metrics [2] which are substantially geometric as with patch attributes like size and edge/area ratio. Most of these metrics either directly or indirectly express topological properties of landscape pattern. With a raster (regular rectangular lattice) representation, for example, a patch is a sequence of cell-to-cell transitions of the same type and patch size is associated with sequence (run) length. Likewise, edge/area ratio is associated with ratio of unlike to like transitions – whether on a global basis or for sets of cells that are self-run members. Correspondingly, juxtaposition and edge apportionment are similarly expressive relative to landscape pattern. Topological analysis and constrained topological stochastics can thus serve for modeling and comparative analysis of landscapes.

One of our precepts is that practical utility for such models will require that they be data-referenced. Constraining a model to match aggregate (global and/or marginal) characteristics of a regular landscape tessellation provides a vehicle for data referencing that permits either relaxation or tightening according to the purpose. This is reminiscent of constraining marginal totals in analysis of cross-tabulated data using contingency tables. One of the ultimate purposes in this is to determine whether two or more landscapes are within some latitude of stochastic pattern variants. A second purpose is to determine whether the landscapes in a study region apparently organize into distinctive patterns, which can then be investigated relative to causality and possible need for intervention. Our deliberations are directed toward these ends, although still being short of closure on the ultimate purposes.

2.2 Eliciting Explicit Topology from Implicit Structure of Spatial Data

Need for landscape models to be data-referenced does not necessarily imply that primary datasets pertaining to landscapes should be modeled directly. The topological characteristics of a tessellated environmental response surface variable are largely implicit and not easily discerned from simple cartographic renditions of the data as maps. The level of difficulty is greater still for understanding of spatial structure in multivariate synoptic environmental signals such as those obtained from remote sensing systems. Therefore, one focus of our research has been on systematic strategies for eliciting explicit topological characteristics from the information regarding topological structure that resides implicitly in regularly tessellated spatial datasets. These strategies involve determining topological component parts or fractions of a spatial lattice that permit decomposition and reconstitution. These incorporate and extend usual ideas of spatial segmentation for tesselations.

Echelon decomposition of surface variables as topological hillforms:

Our strategy for topological decomposition of surface variables is referred to as "echelons" by virtue of being reminiscent of military troop movements in echelon formation. It is essentially a hierarchical hillform organization that is intrinsic to the spatial structure of the variable, and therefore completely objective [3,4]. Echelons divide a (real or virtual) terrain into structural components consisting of peaks, foundations of peaks, foundations of foundations, and so on in an organizational recursion. Saddles determine the divisions between objects. Each object is numbered for individual reference. The peaks constitute one series of structural components, being numbered in decreasing order of summit elevation. The foundations constitute a second series of components that are likewise numbered in order of decreasing top level, starting with the next number after that assigned to the lowest peak. When ties occur within a series, numbering of tied components is decided arbitrarily.

The echelon concept has proven to be most easily understood in terms of a receding floodwaters analogy, whereby the point of beginning is a completely submerged (virtual) topography. As floodwater recedes, elevated structures of the (virtual) terrain will rise above the water level as isolated islands. As recession of waters progress, occasions of abrupt island merger will occur. At each such occasion, one imagines "slicing off" each elevated structure just above the level of merger, assigning the same number to both surfaces of the cut zone (using the next available number), and then setting the cut portion back in place. Proceeding in this manner will divide the virtual terrain exactly into echelon components.

The echelons comprise a structural hierarchy of organizational orders. The orders of the hierarchy are assigned and numbered in the same manner as for a network of streams and tributaries. Thus peaks are akin to unbranched tributaries, and have order 1. A foundation for two or more order 1 objects is of order two. Likewise, a foundation for two or more order 2 objects is of order 3. A low order object sharing a foundation with a higher order object does not increase the order of the foundation. This is like the case of an unbranched tributary entering a higher order stream. Alternatively, this can be considered as an extended family of terrain components having a genealogy similar to that of an extended human family. In this case, however, a component that (a)rises from another is more aptly termed an "ascendant" than an "descendant." Likewise, the "parent" component is better termed "founder."

Note that structures akin to volcanic craters will emerge as atolls (rings) that gradually close with further recession. Such closure is not an occasion for cutting, so echelon components can have pits that would "hold water" if the terrain material were not porous. Since pits of this nature are likely to be of structural interest, a full structural analysis should include a complementary set of "bathymetry" echelons obtained by the analytical equivalent of making a cast with the terrain as a mold. Pits in the terrain become first-order echelons of the bathymetric mold.

Echelons are determined solely by organizational complexity and not by either absolute "elevation" or steepness. A rather large suite of form attributes can be determined for any given echelon component. Such attributes include areal extent of the basal slice, vertical relief, etc. Determination of several such form attributes may depend on an interval scale of measure for the vertical dimension, but the structure as echelons depends only on an ordinal (ordered) scale of measure.

The genealogical analogy extends to dendrogram analysis of the family tree. Much of the terminology for the ECHELON file derives from such context. A CROTCH of the tree is a node at which all links have equal order, and thus induces an increase of order. A TIP is a first-order echelon. A TWIG is a sector of the tree involving only first- and second-order echelons. A STEM is a sector of the tree over which order remains constant aside from first-order ascendants. A BRANCH is a sector of the tree that does not have the simplified form of TIP, TWIG, or STEM.

Evidential avain species richness for Pennsylvania in hexagonal cells encompassing 635 square kilometers each provides a context illustrating echelons as reported by Myers, et al. [5]. The hexagonal tesselation was developed by the U.S. Environmental Protection Agency [6]. The Nature Conservancy was commissioned by EPA to compile evidence of breeding occurrence for avian species by hexagonal cell, with the data being made available for biodiversity research at Penn State University. Figure 1 has dual entries for hexagons, with the upper being number of breeding species based on all evidence and the lower being hexagon identification number. The dual entries for hexagons in Figure 2 show number of breeding species as the lower and numbered echelon components as the upper. Extremities of the echelon organizational tree are also indicated by proportional shape symbols inscribed in Figure 2 hexagons. A triangle symbolizes the summit level of a first-order (peak) echelon. Diamonds symbolize the lowering (subsident) flanks of first-order echelons. Circles symbolize second-order foundations (twigs) for first-order echelons. The extensive integral forest habitat zone in the plateau region of northcentral Pennsylvania is readily apparent, as are also several regions of major mountain/ridge habitat.

Figure 1: Pennsylvania hexagons with breeding bird species richness as upper entry and hexagon ID# as lower entry.

Figure 2: Pennsylvania hexagons with breeding bird species richness as lower entry and echelon ID# as upper entry. Inscribed proportional symbols for selected echelon components: triangle as summit level of order 1 (peak) echelons; diamond for subsident flanks of order 1 echelons; circle for order 2 foundations of order 1 echelons.

In addition to facilitating automated determination and mapping of major structural features in surfaces, echelons are intended to provide a framework for studying realizations of stochastic models in a comparative manner. The effect of adding or removing noise and/or factor influence terms on surface complexity can be seen in echelon organization, particularly relative to branching characteristics of "family tree" dendrograms.

PHASE decomposition of synoptic multivariate environmental signals:

The topology of synoptic multivariate environmental signals such as those from remote sensing is considerably more obscure than for a univariate response surface. Patchiness in the raw data is poorly expressed because of the low likehood that a complete data vector will be repeated exactly. There is, however, usually considerable degree of local damping in multidimensional variation. The intercellular variation within a local area of forest, for example, is substantially less than the variation between forest cells and herbaceous cover cells. This is essentially a latent patchiness. The latent patchiness can be made more explicitly evident by restricting the scope for expression of variation, and effectively setting up a competition for variational carrying capacity. This is achieved via a cluster-based PHASE compression strategy, where PHASE is an acronym for Pixel Hyperclusters Approximating Spatial Ensembles [7,8]. Pixel is the image analyst's term for the cells of a rectangular grid (raster). In the current context, spatial ensembles are the latent patch types of a landscape mosaic.

Although it is more heuristic than analytic, the basic idea underlying the PHASE approach should be fairly intuitive. If a landscape has a mosaic pattern of recurring patch types, then environmental data collected in patches of a given sort should be considerably more alike than those collected in different sorts of patches. This should hold true regardless of how many variables are measured, as long as the variables are directly or indirectly related to the nature of the patch. If the patches are well sorted, the average values measured over all instances of a patch type should represent any particular instance of the patch type reasonably well. If most instances of a patch type are suitably well represented by averages for the patch type, then substituting averages for individual values should retain the general character of spatial pattern in the landscape. The PHASE strategy is thus to sort observational elements in a manner consistent with patch types, then tag the members of each sort and replace their individual values by averages. The entire set of observational data is thereby distilled to a set of tags that mark each element as being of a particular sort, along with a table of average values by sort.

The sorting (clustering) tactics become less crucial as the pattern of (latent) patchiness becomes more strongly expressed. Assuming the observational elements to be cells (pixels) in a rectangular grid, the major influence of spatial pattern is through edge/area ratio relative to cell size. The more edgy the pattern and the larger the cell, the greater will be the occurrence of intracellular mixtures (mixed pixels) that tend to confound whatever sorting tactic may be used. The degree to which segregation in sorting corresponds to patch types will influence the degree of fidelity between original spatial pattern among individual elements and that expressed among elements in terms of averages. Failure to distinguish between the more similar types of patches will tend to generalize/smooth/blur the pattern. Segregating members of a patch type (more sorts than patch types) will involve redundancy but not appreciably affect pattern retention. Distortion of pattern occurs primarily when sorting associates some members of a patch type with exotics of one kind and other members with exotics of another kind.

In order to maintain salient visual characteristics of landscape pattern it is also important to give some preference in sorting to the unusual, at least to the degree that distinctive linears constituting minor overall components of the landscape are not unduly subdued. Roads, riparian elements, right-of-way clearings, small water bodies, and localized clearings are often defining components of landscape pattern both visually and ecologically.

Clustering has a long history of use in the image analysis literature under the umbrella of "unsupervised classification." The current usage is different, however, since focus is on approximation rather than classification. The "hyper" in "hypercluster" refers to extraction of considerably more clusters (groups) than has typically been done for purposes of unsupervised classification. PHASE embryonics lie in the hypercluster work of Kelly and White [9] with the Khoros Group at the Los Alamos National Laboratory and their Spectrum software for interactive classification of image data, but their intent is fairly limited relative to the conceptual scope of the PHASE approach.

The protocol for PHASE clustering is an extension of the ISODATA method [10] which starts with a specified number of cluster "seeds" and associates pixels with seeds by a migrational gravity-like technique. One of the PHASE extensions is a search algorithm for seeds that are well dispersed, in order to obtain better representation of extremes among the ultimate clusters. A second PHASE extension is capability for weighting of cell vectors, which provides for clustering of clusters using mean vectors and cluster frequencies. A third PHASE extension is a cluster splitting facility to ensure that the grouping is not dominated by small sets of extreme values.

PHASE compression extracts 255 clusters, which is the maximum that can be coded with one byte allocated to each cell when zero is reserved as a missing data flag. Contiguous sets of cells belonging to the same cluster become topological units called "clustiles" to avoid the implication that cluster patches are necessarily conformant to landscape patches. The entire set of cells belonging to a particular cluster and encompassing all of its clustiles is termed a clustope. The 255 clusters thus decompose the dataset both qualitatively and spatially, with the clusters being in the nature of anonymous qualitative categories and the clustopes being topological subspaces. This decomposition is objective in the sense of being informationally determined in a systematic and replicable manner, which provides an equally objective basis for comparison of compressed datasets with regard to topological properties. The PHASE compressed datasets are likewise candidates for comparative stochastic modeling with relevance for the several kinds of constraints discussed earlier.

The PHASE residuals due to within cluster variation also have the important consequence of placing the compression beyond any copyright on the raw data since exact restoration is precluded. PHASE compressions of remotely sensed image data can thus be used in ways that may be prohibited for the parent data. PHASE compressions can be treated as pseudocolor raster maps for purposes of display via geographic information systems (GIS) software. Figure 3 is a portion of a PHASE compressed Landsat thematic mapper (TM) scene encompassing Harrisburg, PA rendered as grayscale image in terms of near infrared reflectance.

3 Topological and Distributional Effects of Constriction

Most of the usual analytical methods for pattern recognition [10] can be modified for application to a PHASE compression of multivariate spatial data, including supervised and unsupervised classification of digital image data into thematic mapping categories [11]. Categorical landscape map data referenced in an ensuing section on multiresolution stochastic transition models was derived from PHASE compressed satellite-based remotely sensed data by a hybrid of unsupervised and supervised classification methods which assign clusters to environmental mapping categories.

Since there are fewer environmental mapping categories than hyperclusters, the environmental mapping necessarily has a simplified topology relative to that for the parent PHASE formulation with fewer and larger patches. Thus patch size distribution, edge apportionment, etc. all undergo alteration in greater or lesser degree. It is possible that some mapping categories correspond to single clusters, but this implies that other categories are comprised of many clusters. Such a reduction in separately recognized types of mapped entities is herein called "quantal constriction." Constriction of mapping types can arise algorithmically by reclustering (clustering of clusters) as well as by thematic classification. The nature of the topological change induced by constrictive reduction from T to t types will depend on the topological properties of the

Figure 3: Section of PHASE compressed Landsat TM satellite image over Harrisburg, PA as seen in near infrared reflectance.

pattern prior to reduction. If the (T-t) original types suppressed are both rare and also occur in relatively few small patches, then overall character of the constricted pattern will be little different from the original. On the other hand, constrictive combination of several major types having low contrast and preferentially shared edges (like forest variations) can result in a greatly simplified pattern.

This suggests a second-order transitional approach to characterizing and comparing spatial patterns of landscapes as expressed in PHASE formulations [12]. A regular sequence of quantal constrictions by reclustering will induce progressive pattern simplification. In much the same way that the performance of a student on a battery of standardized examinations will provide a capability profile, so also will the rate and nature of pattern simplification under constriction provide a multifold characterization of pattern. An informationally based constrictive strategy consists of a geometric sequence of one-bit reductions in number of quantal types. A PHASE is thus an original 8-bit constriction. A 7-bit constriction by reclustering reduces the number of (non-missing) types to 127. A 6-bit constriction further reduces that number to 63, and so on until a 1-bit (binary) mapping is ultimately reached.

Whereas quantal constriction operates via a reduction schedule for number of types, it is also possible to constrict by mitigating against small patches without reducing the number of types. We call this spatial constriction as opposed to quantal constriction. Of course, there is the further possibility of constricting dually in both quantal and spatial modes. A deterministic approach to spatial constriction can be used to enforce a minimum patch size, which is often called a “minimum mapping unit” or mmu by cartographers. This consists of recursive topological annexation of small patches by larger neighbors with which they share the greatest edge. Although seemingly straightforward, this can be computationally problematic for large lattices. A stochastic mode of spatial constriction is considered in the next section.

A rigorous challenge for stochastic modeling of landscapes is to contrive models so that they not only meet scale-specific topological constraints, but also have particular trajectories of pattern simplification under quantal and spatial constriction regimes. The role of null models [13] in prospective landscape hypothesis testing must also be kept in view.

4 Multiscale Stochastic Spatial Simplification of Landscape Mappings

The concepts of scale, resolution, and pattern simplification are intertwined. One way to mimic change of scale by "zooming out" is to "smooth out" minor details in an image or map. Since the definitions of major features and minor features depend on the scale chosen, it is desirable to have a sequence of simplified mappings corresponding to different scales. This problem has been studied recently by computer scientists and mathematicians under the name "scale-space" (Weickert [15]). Multiscale approaches have also been proposed in the literature of image segmentation (Koepfler *et al.* [16]).

The goal is to smooth out minor features (patches) in an image, perhaps of a grayscale nature with a patch being a collection of side-or corner-connected pixels. Minor features are either of small size or low contrast or both. Feature elimination should be by merger with neighboring patches. The smaller that features are, the earlier they should disappear in the process of multiscale simplification. It is also desired to avoid the tendency of many common filters to blur boundaries of features that persist. Some kind of adaptive filter is thus needed. A useful analogy is with viewing a landscape from progressively higher altitudes, whereby small patches lose identity faster than larger ones and attrition is more rapid for low contrast edges than ones of high contrast. There should be no splitting of patches.

Commonly-used filters with a fixed window size, such as the median filter, are not suitable for this purpose, since they will blur the boundaries of features corresponding to coarse resolutions when smoothing out features of finer resolutions. In the first column of Figure 4, five simplified images of a 64 by 64 landscape image using the median filter with a 3 by 3 window are given. The effects of blurring and rounding are obvious.

We propose a penalized smoothing method to simplify an image. It uses the same statistical model as used in image restoration and segmentation. It may be interpreted as a Bayesian method with a Markov random field prior. But it may well be interpreted as a variational method. Let $\{x_{ij}\}$ denote the original image where x_{ij} is the label or level at pixel (i, j). A simplified or smoothed version of the original image is defined as the minimizer $\{z_{ij}\}$ of the following global objective function:

$$\sum_{i,j} d(x_{ij}, z_{ij}) + \lambda \sum_{(i,j)\sim(k,l)} \phi(z_{ij} - z_{kl}) \tag{1}$$

where $(i, j) \sim (k, l)$ means that two pixels indexed by (i, j) and (k, l) are neighbors. $d(x_{ij}, z_{ij})$ is a function measuring some kind of distance between two labels or levels

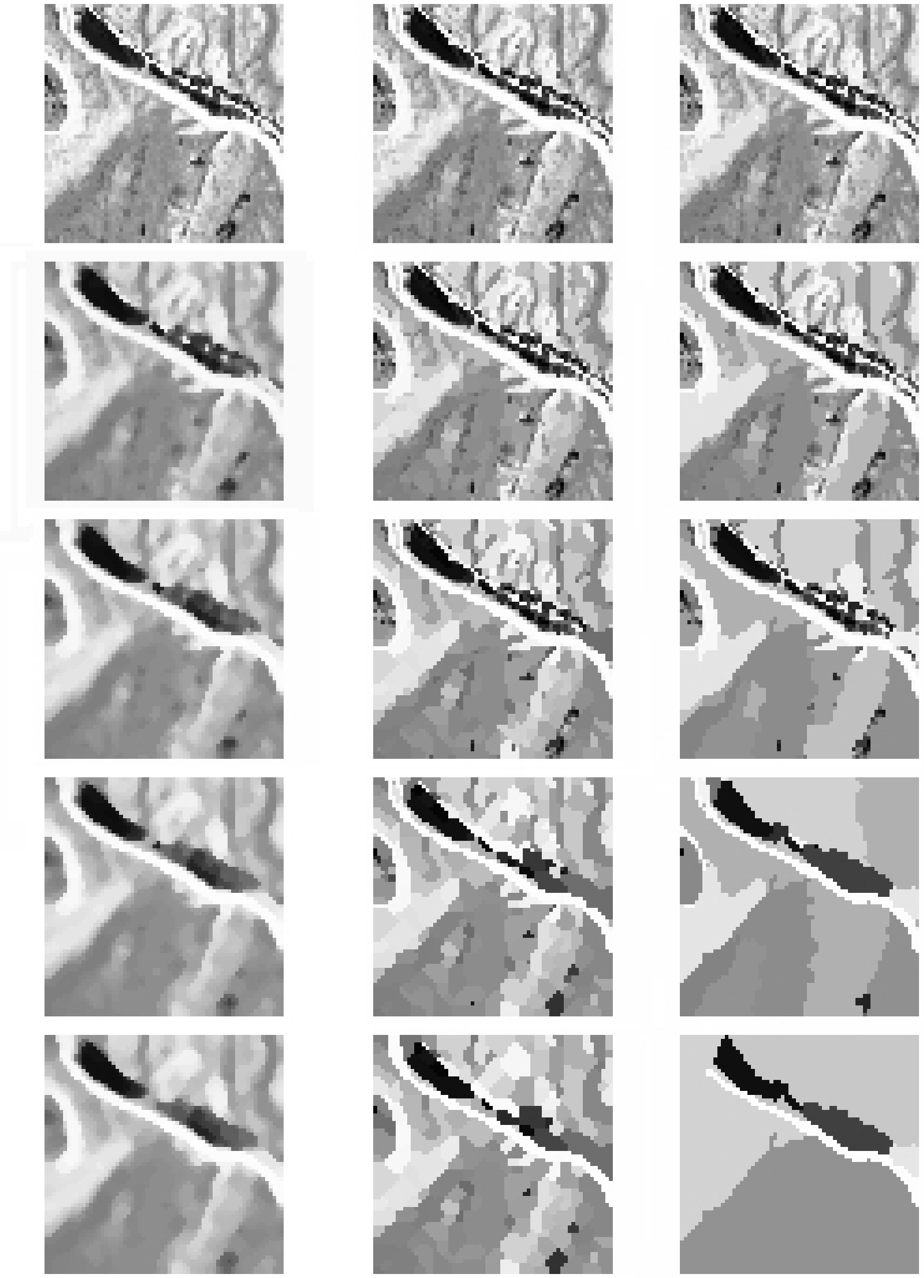

Figure 4: First column: simplified images by a median filter; Second column: simplified images by ICM with five choices of λ, 2,4,8,16,32; Third column: simplified images by patchwise updating algorithm with five choices of λ, 2,4,8,16,32.

x_{ij} and z_{ij}, and ϕ is function measuring some kind of contrast between neighboring pixel labels. λ controls the amount of smoothing or simplification. λ is an index of scale and a large λ corresponds to a coarser scale. There are many possible choices of d and ϕ For illustration purpose, we consider only $d(x, z) = (x - z)^2$ and $\phi(t) = I_{t \neq 0}$ in this section.

In principle, now we have a sequence of simplified images at different scales, i.e., corresponding to different λ's. But finding a minimizer of (1), especially a global one, is not a trivial job. Simulated annealing has been a popular method to solve this combinatorial optimization problem. In theory it converges to a global minimizer with a suitable annealing schedule. In practice, however, because of its slow convergence, approximations or local minimizers are computed instead. ICM (Besag, 1986) is one popular choice. It goes through all the pixels in the image one by one and compares a pixel only with its neighbors to find the best value to replace the current one for the purpose of reducing the objective function (1). It is a coordinate descent method for minimizing a multivariate function. In general it cannot be guaranteed to converge to a global minimizer, nevertheless it gains its popularity because of its simplicity, its quick convergence to a local minimizer, and its similarity to other window-based image processing techniques.

An important property of ICM is its relative insensitivity to the choice of λ. This is viewed as a robust property of ICM since choosing an appropriate λ in image segmentation or restoration is not very easy. ICM moves very slowly in the configuration space of the image and settles easily at a local minimum close to the initial image. The discrete nature of pixel values in digital images makes it even more so. Therefore

ICM tends to give a locally modified version of the initial image. For our purpose of multiscale simplification, however, ICM is not suitable precisely because of this robust property. As an illustration, five smoothed images corresponding to five different values of λ are plotted in the middle column of Figure 4. We can see that the one with the largest λ (the bottom one) is still not too far away from the initial one. If we want to simplify the image further, then ICM cannot help, even with a larger and larger λ. As a matter of fact, for this particular image, any λ larger than 256 results in the same image corresponding to $\lambda = 256$ and this most simplified image by ICM is still not very different from the bottom one in the middle column. Another problem that can be seen from Figure 4 is that ICM with larger and larger λ's tends to round patches and hence change their original shapes.

Considering what we expect from such simplified images, a natural updating scheme is based on patches instead of pixels. That is, each time we consider a pixel for updating, we will consider it together with all the pixels connected to it with the same value. Each patch of the simplified image will be updated to another common value if the new patch value reduces the value of (1). In other words, the current value of a patch R in $\{z_{ij}\}$ will be replaced by the minimizer of

$$c(z) := \sum_{(i,j)\in R} d(x_{ij}, z) + \lambda \sum_{(i,j)\in R:(k,l)\sim(i,j)} \phi(z_{kl} - z). \tag{2}$$

Note that this is not just an ICM with pixels replaced by patches because the patches are not fixed ahead of time, they will change in the process of updating. This clearly has some similarity with the region-based approaches in image segmentation. See, for example, Zhu and Yullie (1995) and references therein. A very similar approach based also on combining the variational method and region-merging method is proposed by Koepfler et al. (1994) for image segmentation.

Applying this patch-wise updating algorithm to the same image used before, we have five smoothed images corresponding to the same five λ's used by ICM. They are given in the right column of Figure 4. The original image is used as the initial image for the finest scale, and then each simplified image is used as the initial image for the next coaser scale. This sequence of smoothed images clearly represent simplifications at different scales. And more importantly, no blurring or rounding effects exist even with large λ's.

5 Multiresolution Stochastic Transition Models of Geographic Landscapes

5.1 Landscape Modeling

Contemporary ecological science is largely motivated by the need to preserve biodiversity. This requires habitat assessment at multiple spatial scales ranging from small field sites to landscape and regional level analysis. The larger scale analysis typically involves characterizing various aspects of spatial pattern in spatially synoptic land use grids (raster maps) which are derived from remote sensing. Understanding the stochastic nature of these spatial patterns does, however, require simulation so that independent realizations of some neutral form of a landscpe can be obtained.

An image analyst might be inclined to invoke random field modeling (i.e. [19,20]).

For this approach, consider a lattice of $s = 1, \cdots, N$ pixels where the value of each pixel is a random variable, X_s. An actual value is realized for each pixel, $X_s = x_s$, according to a probability model. The usual approach is to use a Markov model whereby

$$P(X_s = x \mid X_{s'} = x_{s'}, \text{ for all } s' \neq s) = P(X_s = x \mid \mathbf{X}_{\delta s} = \mathbf{x}_{\delta s})$$

for some neighborhood δs around pixel s. Starting with an initial lattice that has randomly assigned states, such a model needs to be run a sufficient number of times to achieve convergence, after which subsequent simulations can be treated as independent realizations of the same image type. The simulations are thus fully defined by the probability model whose parameters require definition or estimation.

Besides the difficulties of parameter selection, random field modeling does not readily allow the simulation of patterns at different scales in a manner that finer scaled patterns are constrained by coarser scaled patterns. Natural landscapes are known to reveal such hierarchically constrained patterns (for example, see [21,22]) and this is predicted by ecological hierarchy theory [23].

Simple binary random maps were initially used by Gardner, et al. [24,25] as neutral models of landscape structure. For a two-dimensional raster framework of m by m cells, the “success” of each cell is obtained as an independent Bernoulli trial with probability P of success. Therefore, the number, size and shape of a single patch type is controlled simply by adjusting the value of P. Percolation theory [26] predicts that as the number of raster cells goes to infinity, the critical “success” probability goes to $Pc = .5982$, meaning that a continuous patch of “successful” cells will extend from one side of the map to the opposite side when $P > Pc$. Actually, Gardner, et al. [25] showed that percolation occurs when $P = 0.6$ for a map that is only 20 x 20 raster cells.

An hierarchical scaling component was introduced to binary map simulation by Lavorel, Gardner and O'Neill [27,28], who identify this approach as equivalent to the fractal curdling procedure [29]. This is accomplished by specifying the number of scales, L, the number of units, m_i for $i = 0, \cdots L$ within each scaling level, and the proportion of successes, P_i for $i = 0, \cdots L$ within each scaling level. The initial matrix was m_1 x m_1 cells, with a probability of success within each cell set equal to P_1. For each successful cell at the scale 1, the cell was subdivided into m_2 x m_2 cells and successful cells at scale 2 were chosen at random with probability P_2. This was repeated a third time by subdividing successful level-2 successes into m_3 x m_3 cells and choosing successful cells at scale 3 with probability P_3. For such hierarchical simulations, although the fraction of successes in the final map may be less than the percolation threshold ($P < .5982$), the "successful" sites may still connect all the way across the map, a phenomenon that is also observed with actual landscapes [24].

5.2 Hierarchical Simulation for Multiple Category Landscapes

A natural extension of the approach discussed above was introduced by Johnson, Myers and Patil [30] for simulating hierarchically structured landscapes consisting of K land cover categories. As before, let L be the number of scales and let m_i be the number of cells at scale i along the side of a cell from scale $i-1$. For the most coarse scale, level

0, define a multinomial probability vector as

$$\mathbf{G_0} = [P_1, \cdots, P_K]$$

and for subsequent finer scales ($i = 1, \cdots, L$), define a probability transition matrix as

$$\mathbf{G}_{i-1,i} = \begin{bmatrix} G_{i11} & \cdots & \cdots & \cdots & G_{i1K} \\ \vdots & \ddots & & & \vdots \\ \vdots & & G_{ijh} & & \vdots \\ \vdots & & & \ddots & \vdots \\ G_{iK1} & \cdots & \cdots & \cdots & G_{iKK} \end{bmatrix}$$

where G_{ijh} equals the probability of going from category j at the $i-1$ scale to category h at the i^{th} scale. The notation $\mathbf{G}$ is used to indicate that this is a model used to "generate" a hypothetical landscape. Once the transition probabilities are established, a landscape is simulated by the following protocol.

1. Start with one pixel and assign a category to it according to the probability rule defined by $\mathbf{G_0}$. Choice of $\mathbf{G_0}$ is rather arbitrary.
2. Split the "parent" pixel into m_1 by m_1 "children" pixels.
3. Assign a category to each child pixel, conditional on the parent pixel category, according to the probability rule defined by $\mathbf{G}_{0,1}$.
4. For each subsequent finer scale, $i = 2, \cdots, L$, repeat steps 2 and 3 by first splitting each $i-1$ "parent" pixel into m_i by m_i "children" pixels, then assigning categories to the "children" pixels, conditional on the parent pixel categories, according to the probability rule defined by the stochastic transition matrix $\mathbf{G}_{i-1,i}$.

The resulting simulated landscape is controlled by $\mathbf{G}_{i-1,i}$, which is allowed to vary from one scale to the next. On the one hand, varying $\mathbf{G}_{i-1,i}$ allows the study of known scale-dependent changes in a landscape fragmentation; while on the other hand, applying one matrix $\mathbf{G}$ at all scales maintains a more neutral model for simulating a landscape under the null hypothesis of a self-similar fragmentation pattern across all scales, which is characteristic of a fractal model.

As an example, consider the following transition matrix, which was created through some heuristic rules to simulate 8-category land use grids that yield similar properties to a mostly forested watershed in Pennsylvania (see [30]). One random simulation of a "neutral" landscape to the size of 128 by 128 pixels is seen in Figure 5.

5.3 Similarity Dimension of the Multiresolution Stochastic Process

When Mandelbrot [29, chapter 23] discusses *random curdling*, it is presented in the simple case of binary mapping, whereas a grid of base b consists of b^E subintervals, for an embedding euclidean space of dimension E $\in \{1,2,\cdots\}$. Each subinterval is assigned a "hit" (a success, a curd) with probability p, and misses (failures, or wheys) are assigned with probability $1-p$. For those subintervals that are successes, they are in turn divided into b^E subsubintervals and the process is repeated henceforth. For any surviving subinterval at any stage of the process, the expected number of surviving subsubintervals is $E[N] = pb^E$. The Mandelbrot similarity dimension is then redefined

Table 1: Probability transition matrix based on a mostly forested landscape; Rounded to two decimal places for display.

	W	C	M	B	VP	PH	AH	TU
W	0.45	0.05	0.10	0.34	0.02	0.02	0.02	0.01
C	0.01	0.50	0.10	0.34	0.01	0.01	0.01	0.01
M	0.01	0.05	0.55	0.34	0.01	0.01	0.01	0.01
B	0.01	0.05	0.10	0.79	0.01	0.01	0.01	0.01
VP	0.02	0.05	0.10	0.34	0.45	0.02	0.02	0.01
PH	0.02	0.05	0.10	0.34	0.02	0.45	0.02	0.01
AH	0.02	0.05	0.10	0.34	0.02	70.02	0.45	0.01
TU	0.02	0.05	0.10	0.34	0.02	0.02	0.02	0.45

W = water, C = conifer forest, M = mixed forest,

B = broadleaf forest, VP = vegplex,

PH = perennial herbaceous, AH = annual herbaceous,

TU = terrestrial unvegetated

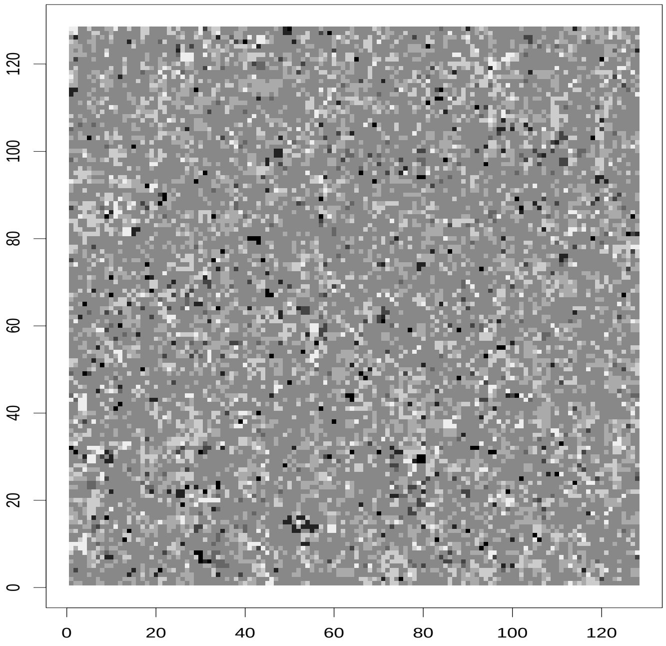

Figure 5: A random simulation of a landscape using the stochastic matrix in table 1; the land use categories are represented by decreasing grayscale intensity, following the order of land cover categories as listed in Table 1; For example, water (W) is black and terrestrial unvegetated (TU) is white.

as

$$D = \frac{\log(E[N])}{\log(b)} = E + \frac{\log(p)}{\log(b)} \tag{3}$$

By Equation 3, it is clear that the the similarity dimension has an uper bound of E. Although negative values can result from Equation 3, Mandelbrot points out that this implies that a random curdling process with $D \leq 0$ goes to the empty set as the process continues, and every empty set has a dimension $D = 0$.

Since our stochastic transition model is an extension of fractal-like curdling, we seek to characterize the similarity dimension associated with a given transition probability matrix.

For the n^{th} resolution, as we zoom in from scale 0 to L, let P_i equal the probability that a pixel selected at random is labeled as category i for $i = 1, \cdots, K$. Let P_{ij} equal the probability that a "child" pixel of the $n+1$ scale is labeled as category j, given that its "parent" pixel of the n^{th} scale is in category i for $(i,j) = 1, \cdots, K$. Following Johnson, Tempelman and Patil [31], we can then define the conditional entropy between these two maps as

$$H = -\sum_{i=1}^{K} P_i \sum_{j=1}^{K} P_{ij} \log_2 P_{ij} \ , \tag{4}$$

where $P_{ij} \log_2 P_{ij}$ equals 0 when $P_{ij} = 0$. This conditional entropy is bound by

$$0 \leq H \leq \log_2(K) \ ,$$

where the lower bound is achieved when for all $P_i > 0$, $P_{ij} = 1$ and $P_{ik} = 0$ for all $j \neq k$, and the upper bound is achieved when for all $P_i > 0$, each P_{ij} is equal and therefore equals $1/K$. In other words, the lower bound is achieved when each parent pixel can only be subdivided into children pixels of one category, and the upper bound is achieved when there is an even distribution of categories amongst children pixels that are nested within parent pixels of a common parent category.

If the "zooming in" process is in a stationary phase with respect to the marginal distribution of the K categories, and the same probability transition matrix is being applied at each transition to a finer resolution, then the process is said to be *stochastically self-similar*.

Under the condition of stochastic self-similarity, it has been introduced by Johnson, Tempelman and Patil [31], and proven rigorously by Tempelman [32], that a similarity dimension can be defined for the stochastic transition model as

$$D = E + \frac{EH_p}{\log b} \tag{5}$$

The last term on the right is the stochastic component of the dimension of the system. For our particular application, $E = 2$ and $b = 2$.

As the conditional entropy increases, the corresponding similarity dimension decreases. This is related to the increasing dispersion of the land use categories. As the connectedness of any land use category increases, so does its fractal dimension, as measured by a variety of ways [31]. Increased connectedness, or intra-category contagion, also decreases the conditional entropy.

References

1. R.T.T. Forman, *Land Mosaics: The Ecology of Landscapes and Regions*, Cambridge University Press (1995).

2. K. McGarigal and B. Marks, FRAGSTATS: Spatial pattern analysis program for quantifying landscape structure, USDA, Forest Service, Pacific Northwest Research Station, General Technical Report PNW-GTR-351 (1995).
3. W. Myers and G.P. Patil, Spatial clustering of polygonal maps having ordinal attributes for landscape analysis, *Computing Science and Statistics* **27** 147-154 (1996).
4. W. Myers, G.P. Patil and K. Joly, Echelon approach to areas of concern in synoptic regional monitoring, *Environmental and Ecological Statistics* **4**(2) 131-152 (1997).
5. W. Myers, R. Brooks, G. Storm and J. Bishop, Pennsylvania Gap analysis report – 1995, Report ER9602, Environmental Resources Research Institute, Penn State University, University Park, PA 16802 (1996).
6. D. White, A. Kimerling and W.S. Overton, Cartographic and geometric components of a global sampling design for environmental monitoring, *Cartography and Geographic Information Systems* **19**(1) 5-22 (1992).
7. W. Myers, PHASE approach to remote sensing and quantitative spatial data, Report ER9710, Environmental Resources Research Institute, Penn State University, University Park, PA 16802 (1997).
8. W. Myers, G.P. Patil and C. Taillie, PHASE formulation of synoptic multivariate landscape data, Technical Report No. 97-1102, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA 16802 (1997).

9. P. Kelly and J. White, Preprocessing remotely-sensed data for efficient analysis and classification, Applications of Artificial Intelligence 1993: Knowlege-Based Systems in Aerospace and Industry, Proceedings SPIE 1993, pp. 24-30 (1993).

10. J.T. Tou and R.C. Gonzalez *Pattern Recognition Principles*, Addison-Wesley, Reading, MA (1974).

11. T. Lillesand and R.W. Kiefer, *Remote Sensing and Image Interpretation*, 3rd edition, Wiley, New York (1994).

12. W. Myers, G.P. Patil and C. Taillie, Exploring landscape scaling properties through constrictive analysis, Technical Report No. 97-1101, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA 16802 (1997).

13. N.J. Gotelli and G.R. Graves, *Null Models in Ecology*, Smithsonian Institution Press, Washington, D.C. (1996).

14. W.L. Myers, G.P. Patil and C. Taillie, Adapting quantitative multivariate geographic information system data for purposes of sample design: the PHASE approach, Technical Report No. 97-1201, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA (1997).

15. J. Weickert, A review of nonlinear diffusion filtering. In *Scale-Space Theory in Computer Vision*, (Edited by B. H. Romeny, L. Florack, J. Koenderink, and M. Viergever), pp. 3-28, Springer, Berlin (1997).

16. G. Koepfler, C. Lopez and J.M. Morel, A Multiscale Algorithm for Image Segmen-

tation by Variational Method. *SIAM Journal of Numerical Analysis* **31** 282-299 (1994).

17. J. Besag, On statistical analysis of dirty pictures (with discussions). *Journal of the Royal Statistical Society, Series B* **48** 259-302 (1986).
18. S.C. Zhu and A.L. Yuille, Region Competition and its Analysis: A Unified Theory for Image Segmentation, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **18**(9) (1996).
19. B.D. Ripley, *Statistical Inference for Spatial Processes*, Cambridge University Press, Cambridge (1988).
20. R.C. Dubes and A.K. Jain, Random field models in image analysis. *Journal of Applied Statistics* **16**(2) 131-165 (1989).
21. N.B. Kotliar and Wiens, Multiple scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity, *Oikos* **59** 253-260 (1990).
22. R.V. O'Neill, R.H. Gardner and M.G. Turner, A hierarchical neutral model for landscape analysis. *Landscape Ecology* **7**(1) 55-61 (1992).
23. R.V. O'Neill, A.R. Johnson and A.W. King, A hierarchical framework for the analysis of scale, *Landscape Ecology* **3**(3/4) 193-205 (1989).
24. R.H. Gardner, B.T. Milne, M.G. Turner and R.V. O'Neill, Neutral models for the analysis of broad-scale landscape pattern, *Landscape Ecology* **1**(1) 19-28 (1987).
25. R.H. Gardner, M.G. Turner, V.H. Dale and R.V. O'Neill, A percolation model of ecological flows. In *Landscape Boundaries: Consequences for Biotic Diversity and*

Ecological Flows: Ecological Studies 92 (Edited by A. J. Hansen and F. DiCastri), pp. 259-269, Springer-Verlag, New York (1992).

26. D. Stauffer, *Introduction to Percolation Theory*, Taylor and Francis, London (1985).

27. S. Lavorel, R.H. Gardner and R.V. O'Neill, Analysis of patterns in hierarchically structured landscapes. *Oikos* **67** 521-528 (1993).

28. S. Lavorel, R.H. Gardner and R.V. O'Neill, Dispersal of annual plants in hierarchically structured landscapes, *Landscape Ecology* **10**(5) 277-289 (1995).

29. B.B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman and Co., New York (1983).

30. G.D. Johnson, W.L. Myers and G.P. Patil, Stochastic generating models for simulating hierarchically structured multi-cover landscapes, Technical Report No. 97-0702, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA 16802 (1997).

31. G.D. Johnson, A.K. Tempelman and G.P. Patil, Fractal based methods in ecology: a review for analysis at multiple spatial scales, *Coenoses* **10**(2-3) 123-131 (1995).

32. A. Tempelman, Multiscale analysis of pattern, part 1: Stochastic self-similar multiscale processes and their dimensions, Draft Technical Report 97-0801, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA. (1997).