# BIOCOMPLEXITY OF ECOSYSTEM HEALTH AND ITS MEASUREMENT AT THE LANDSCAPE SCALE

## A Research and Outreach Prospectus of Advanced Mathematical, Statistical, and Computational Approaches Using Remote Sensing Data and GIS

DEVELOPMENT AND IMPLEMENTATION OF A PROTOTYPE MARMAP
Remote Sensing Application, Technology and Education for
Multiscale Advanced Raster Map Analysis Program for
**Biocomplexity of Ecosystem Health and Its Measurement at the Landscape Scale**

G. P. Patil
Center for Statistical Ecology and Environmental Statistics
Department of Statistics
The Pennsylvania State University
University Park, PA 16802
http://www.stat.psu.edu/~gpp

***Research and Outreach Team***

**G. P. Patil**, Mathematical and Environmental Geospatial Statistics and Analysis
Department of Statistics, The Pennsylvania State University
**John Balbus,** Center for Risk Science and Public Health, George Washington University
**Gregory Biging,** Resource Information Technology and Landscape Ecometrics
College of Natural Resources, University of California, Berkeley
**Robert Brooks,** Cooperative Wetlands Center, The Pennsylvania State University
**Peng Gong,** Landscape Environmetrics, Integrated Assessment, and Visualization
College of Natural Resources, University of California, Berkeley
**Joseph JaJa**, Institute of Advanced Computer Studies, University of Maryland
**W. L. Myers**, Remote Sensing, Natural Resources, and Software
School of Forest Resources, The Pennsylvania State University
**David Rapport,** School of Environmental Design and Rural Development, University of Guelph; Editor-in-Chief, Ecosystem Health
**Orazio Rossi,** School of Environmental Sciences, University of Parma, Italy
**Ben Schneiderman,** Institute of Advanced Computer Studies, University of Maryland
**C. Taillie,** Computational Statistics and Stochastics, Department of Statistics, The Pennsylvania State University

-December 1, 2001-

## Contents

## 1. Introduction and Summary

This prospectus draws upon three innovative and integrative concepts and tools which together will provide the next generation of ecosystem health assessments at regional scales. The first lies in the concept of ecosystem health, which integrates across the social, natural, physical and health sciences and provides the basis for comprehensive assessments of regional environments. The second lies in the innovative stochastic technique for representing human disturbance and ecosystem response on the landscape. The third lies in representation of the spatial biocomplexity of landscapes through the application of echelon analysis to environmental assessment. This proposal shows how the integration of these three recent advances will provide powerful means of assessing environmental conditions at the watershed scales.

Human induced stressors affect biological and environmental processes along pathways with complex feedbacks whereby cumulative effects progressively impair the capacity of ecosystems to provide life support services essential to humanity. This complex of impairment constitutes ecosystem distress syndrome (EDS). Biocomplexity of EDS manifests itself through a wide variety of characteristics such as primary productivity, biodiversity, habitat suitability, ecological integrity, resilience, fragility, vulnerability, resistance, etc. The question of interest then is to study multidimensional biocomplexity dynamics of EDS through spatial organization and temporal behavior of measures of these characteristics.In order to make ecosystem health assessments effective, expressions of EDS must be captured rapidly, comprehensively, and economically which requires utilization of advanced remote sensing capabilities in conjunction with other available geospatial databases. These enormous data streams must be addressed by advanced stochastic modeling and innovative statistical methodologies. Further, the informational products of the analysis must be interpreted in light of the major attributes of ecosystem health: vigor (productivity), organization, and resilience, and in terms of changes in the availability of ecosystem ``goods" and ``services." This will be achieved with proposed research on multiscale stochastic models and statistical determinants of complexity in spatial structure of environmental and ecological factors that portend or signal onset of distress syndrome at landscape and regional scales. The focus will be on biophysical signs of the loss of biocomplexity with emphasis on the relationship of ecosystem distress on ecosystem goods and services.

Multiscale landscape fragmentation is an important manifestation of biocomplexity at a regional scale. The proposed research will provide a model-based inferential context for multiscale landscape fragmentation analysis, using hierarchical Markov transition matrices. The research will provide a framework for formal testing of important ecological hypotheses of distinct scaling domains, self-similarity, optimal landcover/landuse categories, and heterogeneity in fragmentation pattern for critical-area detection.

Recent developments in change detection using compressed multiband image data provide increased flexibility and practicality for systematic change detection on a regional basis. Combining such capability with spatial pattern analysis through 'echelons' will provide methodology for systematically monitoring spatial structure of multiband change across landscapes in order to profile characteristic broad scale regimes of change and to indicate trends in these regimes. The emergent spatial organization in terms of transition models and echelons will be coupled to landscape ecological assessment of biotope fragility, habitat suitability, guild-based ecological integrity, and watershed degradation. The regional scope of research will encompass Pennsylvania in conjunction with additional case studies in U.S. and abroad.

Major project goals for Pennsylvania include: (a) to represent the spatial complexity of key measures of ecosystem health at landscape and watershed scales; (b) to test the hypothesis that the most populated and industrialized Pennsylvania watersheds have a markedly different pattern of biocomplexity compared with the more pristine watersheds, (c) to show that the pattern of landscape fragmentation as revealed by the fragmentation and echelon analysis provides an ``early warning'' signal of regions at risk of ecological breakdown, (d) to show that the new and novel fragmentation and echelon analysis is a powerful tool to represent multiscale multi-attribute biocomplexity and that its application will allow synoptic measures of ecosystem health at a variety of scales ranging from physiographic provinces to watersheds of Pennsylvania, and (e) to offer guidance to classify watersheds of Pennsylvania for conservation, restoration, intervention, etc. by ecosystem health condition and vulnerability, and to prioritize them within each class.

The project goals and results will be achieved in a well-integrated disciplinary and cross-disciplinary effort coupled with matching educational activities and management plan.

## 2. Background and Motivation

In this project, we focus on novel methods to quantify ecosystem health at regional scales. This requires coming to grips with spatial biocomplexity, and representing the patterns of key parameters of ecosystem health, distress, and degradation at watershed scales. The central data sets we will work with are the Pennsylvania synoptic data banks based on National Gap Analysis, Breeding Bird Census and human census, soil erosion and pollution studies, etc., together with the remote sensing data from Landsat. Pressure-State-Response related hypotheses will be tested using the innovative transition and echelon approach leading to spatial complexity maps of biocomplexity representing pressure and response. The modeling approach introduced earlier in the NSF water and watersheds project will accomplish the uncertainty assessment as a result of the proposed development, refinement, and validation based on the proposed case studies in the country and abroad.

The fact that the earth's ecosystems have become overburdened is no longer in doubt (Arrow *etal*, 1995; Vitosek *etal*, 1997). Degradation is now pervasive at local, regional, and biospheric scales. The ready availability of remotely sensed data of the earth's surface from satellite imagery offers enormous potential to assess changes in the health of the earth's ecosystems, identify risks of further degradation, and opportunities for restoration. Thus far, however, little of this potential has been realized, owing to the lack of an appropriate conceptual framework which captures the biocomplexity of the system, including importantly the socio-economic, biophysical and human health dimensions and the lack of a new generation of statistical methodology that is adequate to represent the underlying biocomplexity and lead to achieving a predictive level of its understanding.

This prospectus draws upon three innovative and integrative concepts and tools which together will provide the next generation of ecosystem health assessments at regional scales. The first lies in the concept of ecosystem health, which integrates across the social (Costanza *etal*, 1998), natural (Rapport and Whitford 1999) and health sciences (Huq and Colwell, 1996; Epstein and Rapport, 1996), and provides the basis for comprehensive assessments of regional environments. The second lies in the innovative stochastic technique for representing human disturbance and ecosystem response on the landscape (Patil and Taillie, 1999a). The third lies in representation of the spatial biocomplexity of landscapes through the application of echelon analysis to environmental assessment (Myers, Patil, and Taillie, 1999).

This prospectus shows how integration of these three recent advances will provide powerful means of assessing environmental conditions at the watershed scales. The first provides the rationale for synoptic monitoring determining the viability of the regional landscape (Rapport *etal*, 2000). The second and third provide the advanced statistical ``tool-box'' that will enable highly reproducible quantitative assessments of ecosystem health at regional scales. In so doing, it will provide novel methods for representation of spatial biocomplexity in terms of key indicators of the health and resilience of regional ecosystems. The three in combination will enable representation and quantification of the inherent biocomplexity of regional ecosystems for environmental assessment and management (Rapport *etal*, 1999; Patil *etal*, 2000).

Attempts to assess the health of regions have suffered from several major limitations: lack of synoptic data, assessments based on field studies generally constrained to small areas employing classical statistical tests (e.g. Wichert and Rapport, 1998); lack of integration among the socio-economic dimension--often the major driver of ecological change (Vitosek et al., 1997), the biophysical dimension (Rapport and Whitford, 1999) and the human health dimension (Epstein and Rapport, 1996; Huq and Colwell, 1996); and the lack of appropriate new generation statistical methods (Patil and Myers, 1999) capable of capturing the high degree of complexity inherent in these regional systems.

These barriers can be breached by marrying the concept of ecosystem health (itself an integrative concept embodying socio-economic, biophysical, human health, and management dimensions) with advances in statistical methodology for representing the spatial complexity of key indicators of ecosystem health on a watershed basis. Considerable progress has been registered in identifying indicators of ecosystem health, distress, and degradation. See, for example: DeSoyza *et al.* (1997), Frohn (1997), Hansen and di Castri (1992), Hargis *et al.* (1997), Johnson *et al.* (1999), McKenzie *et al.* (1992), Milne (1992), Noss *et al.* (1999), O'Connell *et al.* (1998), O'Neill *et al.* (1996), Pearson and Gardner (1997), Radermacher (1999), Schumaker (1996), Sexton *et al.* (1999), Scott *et al.* (1990), Szaro *et al.* (1999), and White *et al.* (1997).

To demonstrate the feasibility and practicality of such assessments, we have chosen Pennsylvania as our key study area. Pennsylvania has been well mapped in terms of watersheds at different scales, ranging from 102 units for the State Water Plan to 9,855 units for individual named streams. These watershed units have been studied from different perspectives by different investigators, including non-point pollution, groundwater pollution potential, land cover, and animal habitats. It is immediately apparent that the Pennsylvania watersheds differ amongst each other in terms of ecology, geology, hydrology, degree of human influence, etc. Representing this complexity, synoptically, in a format that enables one to address questions of ecosystem health, integrity and resilience will be our key challenge and achievement. Using the Pennsylvania data we plan to address the following types of questions: What is the health status of a particular watershed and how does this compare with a similar but less stressed system? How has landscape health changed over time for particular watersheds or regions within them? To what degree is ecosystem degradation associated with cumulative effects from population growth and economic development within the watershed? Do changes in spatial biocomplexity of key indicators of ecosystem distress serve as an early warning sign of loss of resilience at regional scales? Which watersheds show the greatest degree of fragmentation? Do these watersheds also indicate a loss of ecosystem services such as water quality? Is the degree of fragmentation within watersheds correlated with the loss of ecosystems goods and services as measured by synoptic data on water quality, soil erosion, biodiversity, etc.?

Similar questions may be posed for data sets being developed within existing collaborative networks. For example, our approach would be applicable to data generated on the Mid Atlantic Region within the EPA Mid-Atlantic integrated assessment initiative; to data generated within the Map of Italian Nature initiative; and to data generated within projects to elucidate the impact of stress on the desert grasslands of USA (Whitford, 1998), and to data on transformation in a Finnish river and its estuary ( Hilden and Rapport, 1993; Hilden, 1998; and Rapport et al., 2000).

Human induced stressors affect biological and environmental processes along pathways with complex feedbacks whereby cumulative effects progressively impair the capacity of ecosystems to provide life support services essential to humanity (Daily, 1997). This complex of impairment constitutes ecosystem distress syndrome (EDS). Biocomplexity of EDS manifests itself through a wide variety of characteristics such as primary productivity, biodiversity, habitat suitability, ecological integrity, fragility, vulnerability, resistance, etc. The issue of interest then is to study multidimensional biocomplexity dynamics of EDS through spatial organization and temporal behavior of measures of these characteristics. In order to make ecosystem health assessments effective, the expressions of EDS must be captured rapidly, comprehensively, and economically which requires utilization of advanced remote sensing capabilities in conjunction with available geospatial databases. These enormous data streams must be addressed by advanced stochastic modeling and innovative statistical methodologies. The informational products of the analysis must be interpreted in light of current knowledge of the major attributes of ecosystem health: vigor, organization, and resilience (Mageau et al., 1995; Costanza et al., 1998a,b).

Building on these collaborations, the Biocomplexity Integrated Research (BIR) prospectus considers research on statistical determinants and stochastic models of complexity in spatial structure of environmental and ecological factors that portend or signal onset of distress syndrome at landscape and regional scales. This will entail major augmentation, extension, and application of concepts and computational capabilities acquired so far. The background research has established a quantitative framework for elucidating and eliciting complexity in phenomena that constitute fields of spatially variable intensity, and also for transitions among states of qualitative conditions. The foundation methodologies have been explored in a preliminary manner for operability with remotely sensed multispectral data.

Multiscale landscape fragmentation in landcover/landuse is an important manifestation of biocomplexity at a regional scale. For ecosystem health assessment, it becomes important to characterize, compare, and classify the biocomplexity associated with landscape fragmentation at a landscape and watershed level. The proposed research will provide a model-based inferential context for multiscale landscape fragmentation analysis, using a series of stationary and reversible Markov transition matrices to generate a hierarchy of categorical raster maps at different resolutions.

Recent developments in change detection using compressed multiband image data provide flexibility and practicality for systematic change detection on a regional basis. Combining this capability with conceptual extensions of spatial pattern analysis through `echelons' provides a methodology for systematically monitoring spatial structure of spectral change across landscapes in order to profile characteristic broad scale regimes of change and to indicate trends in these regimes. Echelons are unique in providing direct hierarchical tree-based representations of spatial complexity across areas of varying intensity for biological and environmental variables.

The emergent spatial organization in terms of transition models and echelons will be coupled to landscape ecological assessment of biotope fragility, habitat suitability, guild-based ecological integrity, and watershed degradation. The regional scope of primary research will encompass Pennsylvania as a primary case study. Major project goals for Pennsylvania include: (a) to represent the spatial complexity of key measures of ecosystem health at landscape and watershed scales; (b) to test the hypothesis that the most populated and industrialized Pennsylvania watersheds have a markedly different pattern of biocomplexity compared with the more pristine watersheds, (c) to show that the pattern of landscape fragmentation as revealed by the fragmentation and echelon analysis provides an ``early warning'' signal of regions at risk of ecological breakdown, (d) to show that the new and novel fragmentation and echelon analysis is a powerful tool to represent multiscale multi-attribute biocomplexity and that its application will allow synoptic measures of ecosystem health at scales ranging from physiographic provinces to watersheds of Pennsylvania, and (e) to offer guidance to classify watersheds of Pennsylvania for conservation, restoration, intervention, etc., by ecosystem health and vulnerability and to prioritize them within each class.

The proposed research will contribute innovative model-based reproducible automated assessment and management of the biocomplexity of ecosystem health, distress, and degradation with a novel working quantitative toolbox of biocomplexity knowledge discovery techniques, developed and fine-tuned with a variety of case studies in the country and abroad. An urgent need for today is to achieve mathematical multiscale spatial modeling and analysis of categorical, ordinal, and numerical maps for environmental and ecological variables in a manner that facilitates quantitative comparative analysis for subregions of concern to resource managers and to environmental and ecological scientists in a timely manner.

## 3. Indicators of Biocomplexity of Ecosystem Health

**What Constitutes Ecosystem Health?** A healthy ecosystem has been defined as one that is free from ecosystem distress syndrome, maintains its organization and autonomy over time, and is resilient to stress (Costanza, 1992). Ecosystem Health can be assessed by indicators of vigor (productivity), organization and resilience (Mageau et al., 1995, Costanza et al., 1998). Ecosystem health assessments have been carried out for a number of ecosystems, generally based on extrapolation from limited field data. These include the Chesapeake Bay (Mageau et al., 1995) and other marine ecosystems (Rapport, 1989b; Hilden and Rapport, 1993), freshwater ecosystems (Wichert and Rapport, 1998), forested ecosystems (Yazvenko and Rapport, 1997), arctic ecosystems (Rapport \etal 1997) and desert grasslands (Whitford, 1998; Rapport and Whitford, 1999; Whitford et al., 1999). These studies confirm the sensitivity of indicators of vigor, organization, and resilience as measures of ecosystem health in ecosystems that have undergone degradation as a result of pressure from human activity. The association of a long history of intensified human activity in a watershed with increasing signs of degradation suggests that indicators are appropriate for monitoring health (and conversely degradation) in field situations (Hilden and Rapport, 1993).

**Assessing Ecosystem Health at Regional Scales**: The existence of multiple dynamic stable states for both natural and human-dominated ecosystems complicates the task of determining the extent to which ecosystem structure and function have been altered by human activity. Nonetheless, careful studies leave little doubt that ecosystem degradation has occurred in many systems, including forests (Yazvenko and Rapport, 1997), marine (Hilden and Rapport, 1993), fresh water (Wichert and Rapport, 1998), desert grasslands (Whitford, 1998) and many others. The documentation of health, or more often, its converse, pathology, is undertaken by looking at a group of indicators and comparing their values with norms established for healthy ecosystems. These norms are determined by comparisons between stressed and unstressed systems of similar type

(Rapport et al., 1985; Boswell et al., 1994) or for a system under intensifying pressure from human activity over time (e.g., Hilden and Rapport, 1993; Boswell et al., 1994).

**The Ecosystem Distress Syndrome** : Margalef suggested (1975, p.239) that ``All or most of the ways in which man interferes with the rest of nature produce coincident or parallel effects. [For example] diversity is reduced, horizontal transportation [of nutrients] is increased and the ratio of production/biomass is increased. The parallelism of change and its logical coherence represents a welcome simplification of the whole set of problems." Earlier, Leopold (1941) had proposed the concept of ``land sickness" to refer to signs of dysfunction exhibited in his native Wisconsin landscape in response to a variety of pressures from human activities. Leopold suggested that common signs of land degradation included soil erosion, loss of fertility, hydrological abnormalities, occasional irruption of certain species and mysterious local extinction of others, as well as qualitative deterioration in farm and forest products, the outbreak of pests and disease epidemics, and boom and bust wildlife population cycles. Over the past half century, and particularly in recent decades, many of the signs identified by both Leopold and Margalef on theoretical grounds have been confirmed in various empirical studies (Rapport et al., 1985; Rapport, 1989b).

Building on these insights, Rapport et al. (1985) proposed the ecosystem distress syndrome (EDS) analogous to Selye's biological distress syndrome. EDS identifies common structural and functional properties of ecosystems under stress. Based on a comparative study of different ecosystems, these authors identified common features of stressed systems, including altered productivity, nutrient cycling, reduced resilience, altered community dominance favoring “r” selected species (shorter reproductive cycles, smaller size), increase in non-native species (exotics), increased disease prevalence, increased instability in component populations, reduced biodiversity, etc. These properties have subsequently been validated in additional case studies (Hilden and Rapport, 1993; Rapport et al. 2000; Rapport and Whitford, 1999).

Using proxies for various signs of ecosystem distress (e.g., biodiversity, community dominance, sediment loads, nutrient status of receiving waters) and relating these to the available synoptic geospatial and remote sensing data will provide a quantitative portrait of each watershed relevant to assessing ecosystem health. To our knowledge, this will be the first time assessments of this nature have been attempted. Heretofore ecosystem health assessments have been largely based on field observations, generalized to larger systems. Such methods have the drawback of being limited in scope, expensive, and lacking in quantitative significance when extrapolated to larger regions.

**Relationship of Ecosystem Distress to Nature's Services**: This project will also allow exploration of a key element of biocomplexity, namely the relation between socio-economic activity and ecosystem status. The bridge that allows this integration is through the concept of “nature's services” (Cairns and Pratt, 1995; Costanza, 1997; Daily, 1997). Ecosystem degradation, as is well documented, is invariably accompanied by a decline in nature's services, such as potable water, biodiversity, productivity of crops, fisheries and wildlife, soil fertility and the like. Quantitative assessments of ecosystem health on a watershed basis should reflect the supply of nature's services on the same basis. Thus in highly compromised watersheds (as revealed by both multiscale fragmentation analysis and echelon analysis), we should also find the largest loss of nature's services. We will test this hypothesis by comparing the evaluations of health status with measures of water quality for those watersheds for which data are available on nutrient status and contaminants in water at the outflows.

**Hypotheses to be Tested**: 1) Human influence alters spatial complexity of landscapes as expressed in environmental indicator variables. In the initial stages, human influences tend to increase spatial complexity, but in the more advanced stages of ecosystem distress syndrome there is progressive spatial simplification (reduction of diversity). As openings are created in a forest matrix there is an associated increase in fragmentation and, initially, an increase in biodiversity. New biota are associated with the new habitat, but those species eventually may displace endemic species. As fragmentation increases, area sensitive species thus eventually decline. In such a case, the temporal increase in spatial complexity is a warning sign of impending simplification of biocomplexity.

Interpretation of spatial pattern in the context of this and ensuing hypotheses cannot be mechanistic, even though mathematical pattern extraction may be. For instance, overall biodiversity could increase in a fragmenting environment relative to a more pristine environment (Appalachian Plateau in southwestern Pennsylvania versus Applachian Plateau in northcentral Pennsylvania); further, urban areas are much simplified in terms of biodiversity and also categorical landcover. Watersheds that are in the process of developing spatial complexity, interpreted ecologically, can portend a transition to future simplification.

2) Human influence is hypothesized to alter self-similarity of spatial pattern with changing scale. Self-similarity has been used mainly to indicate a lack of human influence. Human influence is reflected in alteration of recurrence relations of patterns at different scales (i.e., fractal dimension). The recurrence pattern at different scales is one component of biocomplexity. A method for testing will be developed in terms of Markov transition models.

3) We hypothesize that assessment of ecosystem resilience or fragility can be done, at least in part, in terms of spatial complexity as revealed by echelon analysis of indicators of ecosystem distress.

4) We hypothesize that advanced landscape fragmentation with attendant loss of biocomplexity and alteration of spatial complexity will compromise an array of ecosystem goods and services. Testing this with respect to water quality can be achieved in a first-order manner by comparing composition in the outflow of heavily impacted watersheds with that of lightly impacted watersheds. Watersheds would be matched for similarity in other attributes such as underlying geology, vegetation types, physiography, etc.

# 4. Modeling and Simulation of Thematic Raster Maps

## 4.1 Disjunctive Indicator Geostatistical (DIG) Model.

This model is intended to facilitate the use of geostatistical methods in the analysis of categorical raster maps---maps in which the response at each raster cell (or grid point) is thematic instead of numerical. (Patil, 2001a; Patil and Taillie, 2001a). The DIG model has three main ingredients:

- A regular grid with lattice points $t$.
- A standard normal (Gaussian) process $Z(t)$ on the grid with correlation function $r(h)$. In practice, we adopt standard parametric forms for the correlation function, e.g., $r(h) = \exp(-l\,h)$ with parameter $l$.
- A partition $A_1, A_2, \cdots, A_k$ of the $Z$-axis with one partition set $A_i$ for each of the $k$ different categorical responses. This partitioning is referred to as the *transitionogram*.

The surface values $Z(t)$ are latent (or hidden) and are not observable. The model evaluates the *disjunctive indicators* of $A_1, A_2, \cdots, A_k$ on $Z(t)$ thereby determining a unique categorical response at grid point $t$ (see Figure 1). It is these categorical responses that are observed. Categorical responses at neighboring grid points are correlated due to spatial autocorrelation of the latent surface $Z(t)$.

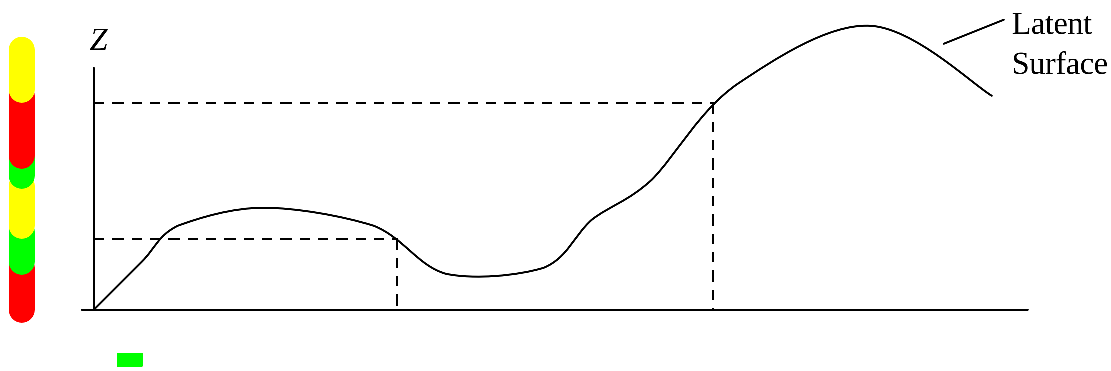


***Figure1.*** *Elevation of the latent surface is categorized according to the transitionogram on the left of the Z-axis.*

Using a standard Gaussian process for $Z(t)$ is not a severe limitation because the probability integral transform could be applied at each grid point with corresponding transformation of the partitioning sets $A_1, A_2, \cdots, A_k$ thereby ensuring marginal, if not joint, normality. Critical to the robustness of the model is the fact that the partitioning sets are not required to be intervals. Otherwise, the potential spatial transitions from one category to another category at adjacent cells would be too limited. Instead, each partitioning set can be a disjoint union of intervals so that distinct partitioning sets, $A_i$ and $A_j$, can interlace one another.

**Model Simulation**: Once the parameters of the DIG model are specified, unconditional simulation of maps is straightforward and reasonably fast. One generates a realization of the Gaussian surface $Z(t)$, via the usual Cholesky or spectral decomposition of the variance covariance matrix, and then evaluates the disjunctive indicators of $A_1, A_2, \cdots, A_k$ on $Z(t)$. The only obstacle here is the size of the map and corresponding size of the variance-covariance matrix of $Z(t)$. But, this is a well-studied issue in the geostatistical literature with one solution being the generation of $Z(t)$ in blocks according to the range of spatial dependence (Deutsch and Journel, 1998; Goovaerts, 1997). More difficult is *conditional* simulation in which categorical responses are specified at a fixed subset of locations $t$ and each simulated map must exactly reproduce these known responses while "filling-in" the unknown responses at other locations. Conditional simulation is important, for example, in thematic accuracy assessment. We propose to develop and implement a conditional simulation algorithm for the DIG model. Note that conditional simulation of Gaussian processes $Z(t)$ is quite standard in the geostatistical literature; the difficulty here is that we do not get to observe the conditioned portion of $Z(t)$, only its induced categorical values. This problem can be addressed by the method developed by Kozintsev and Kedem (2000) whereby, given the categories, an isotropic Gaussian field is simulated.

**Model Fitting:** Here, we suppose an actual categorical raster map is available as the data from which we must estimate the parameters of the DIG model. Since the likelihood function is intractable, we propose to fit the model by minimizing the discrepancy between appropriate empirical (calculated) map characteristics and their corresponding model predictions (which are functions of the model parameters). Two sets of characteristics appear promising:

- Marginal histogram of mapping-category frequencies
- Joint occurrence probabilities of pairs of categories at varying distances and directions (auto-association matrices). In fact, the auto-association matrices (see below) for all distances determine the indicator variograms and cross-variograms, and conversely.

It is somewhat unusual to have sets as unknown parameters to be estimated so the question arises as to how we can represent and vary $A_1, A_2, \cdots, A_k$ during optimization. Allowing these partitioning sets to be completely arbitrary does not appear to be computationally feasible. Therefore, we propose to use the probability integral transform to map the $Z$-axis to the unit interval. Next, we subdivide the unit interval into, say, 1000 equal subintervals (equivalent to 1000 equal-probability subintervals of the $Z$-axis) and assign categories to each of the subintervals. Each such assignment determines a partition $A_1, A_2, \cdots, A_k$ and we have to optimize over all the possible assignments and simultaneously over any unknown parameters of the correlation function $r(h)$. If we let $N_i$ be the number of subintervals to which category $i$ is assigned, then $N_i/1000$ is the model predicted marginal relative frequency of category $i$ so we can match this to the empirical relative frequency (to 3 decimal places) by fixing $N_i$ during optimization.

Thus, it remains to minimize the discrepancy between observed and model-predicted auto-association matrices $R_{ij}, \quad i, j = 1, \cdots, k$. We propose to use the Kullback-Liebler distance to measure the discrepancy. The model-predicted $R_{ij}$ are given by $\Pr[Z(t) \in A_i, Z(t') \in A_j]$ where the grid points $t$ and $t'$ are a distance $h$ apart. Since $A_i$ and $A_j$ are each finite unions of disjoint intervals, the above expression becomes a finite sum of bivariate normal probabilities of rectangles which can be computed using the tetrachoric expansion (Pearson, 1901). For the actual process of optimization we propose to examine two methods: (i) *genetic algorithms* as suggested by the chromosome-like structure of the transitionogram (Goldberg, 1989), and (ii) *simulated annealing* (Azencott, 1988, 1992; Gidas, 1995).

## 4.2 Hierarchical Markov Transition Matrix (HMTM) Model.

The proposed approach employs a series of Markov transition matrices to generate a hierarchy of categorical raster maps at successively finer resolutions. Each transition in the hierarchy may involve a different matrix, thereby modeling distinct, as well as smoothly ranging scaling domains. Even when data is available at only the finest resolution, the model is nonetheless identifiable and parameters can be estimated by exploiting a duality between hierarchical transitions in the model and spatial transitions at varying distance scales in the data map. See Johnson (1999), Johnson and Patil (1998), Johnson *et al* (1998, 1999ab, 2000), Patil *et al* (1999, 2000ab), and Patil and Taillie (1999, 2000abc).

**Auto-Association Matrices:** Consider a raster map of some attribute A and suppose this attribute has $k$ categorical levels denoted by $a_1, a_2, ....a_k$. For empirical description of the spatial dependence at varying distances in the map, we employ a series $\hat{R}_0, \hat{R}_1, \hat{R}_2, ...$ of $k \times k$ matrices. The matrix $\hat{R}_n$ is obtained by scanning the map and examining pairs of pixels which are $2^n$ pixels apart, either horizontally or vertically. The $i, j$ entry of $\hat{R}_n$ is the relative frequency of occurrence of response $(a_i, a_j)$ in such pairs of pixels. Thus, $\hat{R}_n$ is a symmetric probability table expressing empirically the auto-association of attribute A at distance $2^n$ across the map. The series, $\hat{R}_0, \hat{R}_1, \hat{R}_2, ...$, of auto-association tables is a categorical counterpart of the empirical variogram for numerical response data.

The HMTM model is a parametrized probability model for classified maps with the property that the parameters of the model can be estimated directly from the empirical auto-association matrices. The model generates a sequence $M_0, M_1, ...., M_L$ of categorical raster maps. Each map covers the same spatial extent, but successive maps are of increasingly finer resolution. The first map $M_0$ consists of a single pixel and, recursively, the pixels of $M_n$ are bisected horizontally and vertically to produce the pixels of $M_{n+1}$, giving rise to a "quadtree" of pixels (Samet, 1990). See Figure 2. Mapping categories are assigned to pixels of $M_n$ using Markov transition matrices. Suppose there are $k$ mapping categories (values), labeled as 1, 2,… $k$. At the coarsest scale, the assignment of a value to the single pixel of $M_0$ is generated from an initial stochastic probability vector $p^{[0]}$. Given the assignment of values to pixels of $M_n$, the assignment to $M_{n+1}$ is generated by a row stochastic transition matrix, $G^{[n,n+1]} = \left[G_{ij}^{[n,n+1]}\right]$, $i, j = 1, ....k$. Fix attention on a particular pixel of $M_n$ and let its value be $i$. The values $j$ for its four subpixels are generated by four independent draws from the distribution specified by the $i$ th row of $G^{[n,n+1]}$.

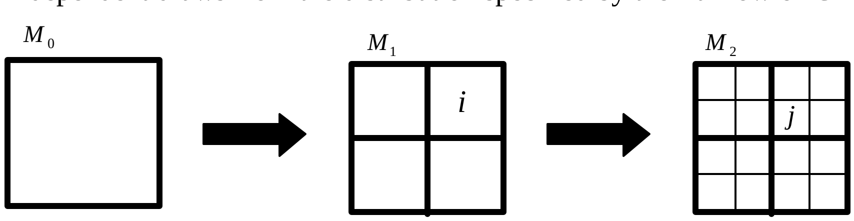

***Figure 2. Nested hierarchy of pixels. Each pixel of $M_n$ subdivides into four subpixels in $M_{n+1}$.***

Only a single floor resolution map $M_L$ may be available for analysis. From this single resolution map, we estimate model parameters by relating spatial scaling levels across $M_L$ to hierarchical levels in the model. With suitable restrictions on the model parameters, an identifiability theorem asserts that distinct sets of model parameters correspond to distinct probability distributions on $M_L$. The correspondence is accomplished analytically by relating the eigen-decomposition of the hierarchical transition matrices to the eigen-decomposition of the spatial auto-association matrices. See Patil and Taillie (1999, 2000abc).

Unconditional simulation of floor resolution maps can be done directly using the hierarchy of transition matrices and is very fast. Conditional simulation is more difficult and is accomplished by applying MCMC methods on the entire quadtree of pixels with nodal neighborhoods consisting of parent and sibling pixels. Thus, HMTM is a Markov random field on the quadtree.

### 4.3 Markov Random Fields.

The DIG and HMTM models are defined in terms of specific procedures for generating realizations—which make simulation fast and conceptually straightforward. *Markov random field* (MRF) models, on the other hand, specify a parametric family of probability distributions on the set $\Omega$ of all thematic raster maps of given size and with given set of categorical responses. This probability distribution has the Gibbs form

$$p(x) = \exp[-H(x)]/Z, \qquad x \in \Omega,$$

where $Z$ is the normalizer and $x$ ranges over all possible maps in $\Omega$. Parametric forms are specified for the "energy" function $H(x)$ that expresses the strength of association among the categorical responses in neighboring pixels. See Barone *et al.* (1990), Bremaud (1999), Cressie (1991), Geman (1990), Geman and Geman (1984), Gimel'Farb (1999), and Winkler (1995) for detailed discussion.

Gibbs sampling and other MCMC variants are employed for simulation of Markov random fields; see Geman and Gemen (1984), Metropolis *et al.* (1953), and Newman and Barkema (1999). In contrast with the DIG and HMTM models, conditional simulation for MRF models is no more difficult than unconditional simulation. Model fitting has been discussed by, for example, Besag (1974), Guyon (1995), and Younes (1988, 1991). Both simulation and model fitting are computationally demanding for MRFs. However our previous work has shown that parallel computing can be used to substantially speed up these computations. See Bader, JaJa, and Chellappa (1995).

## 5. Applications of Raster Map Models

The research will examine the following issues:

- **Map characterization and discrimination:** The eigen-decomposition of the auto-association matrices will be studied for map characterization and discrimination. Using Principal Components methodology as in Slud et al. (2000), we can derive from the

HMTM model low-dimensional numerical features of a landscape, which can be examined over space and time, and with respect to cross-classification by gross geographical and environmental features.

- **Fragmentation profiles:** The fragmentation profile is a graphic display of the persistence of spatial pattern across spatial scales (Figure 3). See Johnson (1999), Johnson and Patil (1998), Johnson *et al* (1998, 1999ab, 2000), Patil *et al* (1999, 2000), and Patil and Taillie (1999, 2000abc). We will study profile responsiveness to variation of parameter values in the DIG/HMTM/MRF map models.

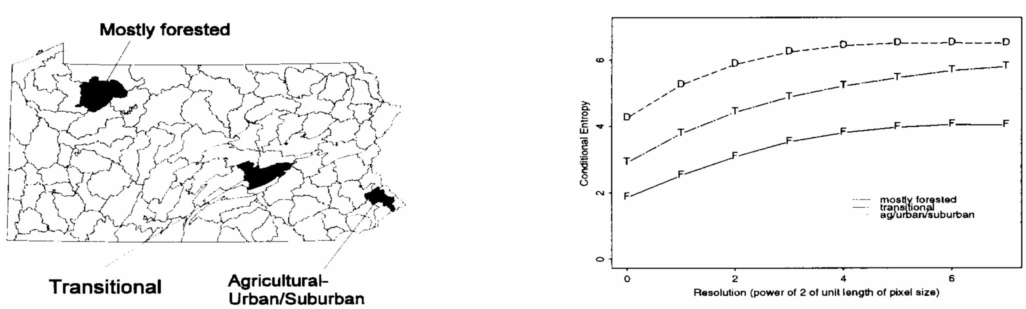


***Figure 3.*** *Fragmentation profiles for three Pennsylvania watersheds with distinct landcover patterns: mostly forested, transitional and mostly deforested (ag/urban/suburban).*

- **Simulation modeling:** Maps can be simulated using the DIG/HMTMMRF models, thereby providing an excellent vehicle for model-based inference in thematic map analysis including goodness of fit tests and nested tests for parameter reduction, as well as tests of scientific hypotheses such as self-similarity and distinct scaling domains.

- **Patch structure:** Patch structure is a powerful indicator of spatial pattern and many of the FRAGSTATS (McGarigal and Marks, 1995) measures of spatial pattern are patch-based. The proposed research will examine the responsiveness of selected patch-based metrics to spatial dependence versus dominance as well as to abrupt changes in model parameters at different hierarchical levels (scaling domains).

- **Statistical detection of heterogeneity in spatial pattern:** A local determination of pattern will be made using appropriate scalar and vector measures. Sampling distributions of these measures and corresponding local $p$-values will be obtained by simulation from the globally fitted DIG/HMTM/MRF models.

- **Thematic accuracy assessment:** The effect of spatial pattern on estimation of the error matrix and associated parameters will be studied by conditional simulation using the raster map models to generate classified maps with varying spatial patterns of error.

- **Bivariate raster map analysis for thematic change detection:** The proposed MARMAP system will provide bivariate modeling and simulation capability to help with thematic change detection. The bivariate DIG model employs a single latent surface with the two overlaid transitionograms. In the HMTM and MRF approaches, the parametric modeling needs to reflect the cartesian product structure of the responses.

## 6. Surface Topology, Upper Level Sets, and Echelons of Surfaces

Quantitative spatial data are important inputs of many environmental process models for determining future implications of current resource use, policies, and interventions. It is therefore desirable to have a systematic means of determining spatial organization in mappings of quantitative variables. Echelons present means for objectively determining quantitative spatial structure for direct mapping either with or without computer-assisted visualization (Myers *et al*, 1995, 1997, 1999; Johnson *et al*, 1998; Kurihara *et al*, 1999; Patil and Taillie, 1999; Smits and Myers, 2000). Thus, they can facilitate analysis of implications of errors associated with environmental models that take quantitative layers as input, or produce quantitative output layers, or both.

**Echelons of Spatial Variation:** The spatial variables for echelon analysis can be considered as topographies, whether real or virtual. Echelons divide the (virtual) terrain into structural entities consisting of peaks, foundations of peaks, foundations of foundations, and so on in an organizational recursion. Saddles determine the divisions between entities. Each entity is assigned an echelon number for identification purposes. See Myers, Patil, and Taillie (1999).

Consider, for example, the terrain depicted in profile in Figure 4a. The numbered entities are called echelons. Echelons are determined directly by organizational complexity in the spatial variable and determine a family tree as illustrated in Figure 4b. The number of "ancestors" for an echelon is a local measure of regional complexity. The echelons also comprise a structural hierarchy of organizational orders in the same manner as for a network of streams and tributaries (Rodriguez-Iturbe and Rinaldo, 1997). Since most echelon trees are much too complicated for visual study as dendrograms, characterization and comparison of echelon trees is done through analytical processes such as pruning. See Myers, Patil and Taillie (1999).

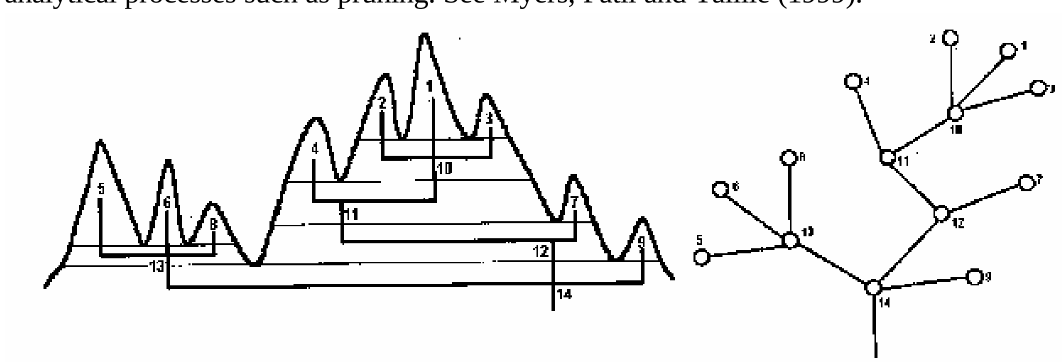

***Figure 4. Echelon decomposition of a surface and associated echelon tree.***

*Proposed Research***:** The proposed research will advance the analytical utility of echelons. A major question concerning quantitative spatial variables with respect to many applications is whether there are substantial sectors of the surface having particularly high or particularly low values relative to the mean level. An echelon family would be seen as a candidate for focus if the probability of its extent receiving observed amounts is less than the criterion under a random distribution of quantity over area. Since echelon determination is computationally intensive, there would be further advantage in capability to extract principal families from partially determined echelons. Echelons may also be determined after filtering the surface variable to smooth local variability. The degree of change in the echelon structure as a result of filtering is indicative of the sensitivity or insensitivity to errors in the data. Filtering strategies will be explored for the purpose of assessing robustness of spatial structure to errors in the surface variable. A further line of research for a variety of applications involves methodology for comparative study of spatial complexity as expressed by a suite of echelon indicators. Each indicator can be treated as a synthetic sensor band. These pseudo-sensor bands can be assembled as synthetic multi-band complexity image datasets for the region in question. Segmentation of the synthetic multi-band data will extract prevailing patterns of complexity among the several indicators of ecosystem health.

## 7. Multiple Indicators, Partial Ordering, and Multicriteria Decision Support: Comparisons and Rankings without Integration---Some Statistical and Visual Tools

We address the question of ranking a collection $S$ of elements when a suite of indicator values is available for each member of the collection (Patil, 2001b; Patil and Taillie, 2001b). The elements can be represented as a cloud of points in a multidimensional space, but the different indicators typically convey different comparative messages and there is no unique way to rank the elements. The traditional approach of combining the indicators in some fashion has well-known severe limitations. We take the view that the relative positions in indicator space determine only a *partial ordering* (Fishburn, 1985; Neggers and Kim, 1998; Trotter, 1992) and work with Hasse diagrams (Neggers and Kim, 1998; Di Battista, 1999) of the partial order to study the collection $\Omega$ of all rankings that are consistent with the partial order. Such rankings are said to be *admissible* and are called *linear extensions* of the partial order. One can then pose such questions as the following:

1) What is the smallest (i.e., best) possible rank that can be assigned to a given element $a \in S$? What is the largest (worst) rank?
2) How many rankings from $\Omega$ assign rank 1 (best) to element $a \in S$? Rank 2? etc.
3) If rankings are chosen at random (with equal probability) from $\Omega$, what is the probability that element $a \in S$ receives a rank of $i$ or better?

The answer to the first question lets us associate an interval of possible ranks to each element in $S$. The intervals can be very wide, however. Noting that ranks near the endpoints of each interval are infrequent under admissible rankings, the answer to the second question provides a frequency or probability distribution over the interval of possible ranks. These distributions, called *rank-frequency distributions*, turn out to be unimodal (in fact, log-concave).

The third question leads to a canonical and objective procedure for ranking the members of $S$. The answer to the question is given by the cumulative distribution function (CDF) of the corresponding rank-frequency distribution. However, these CDFs can be ordered using the so-called "stochastic ordering" of cumulative distribution functions. This provides a new partial order on $S$, which *extends* (is consistent with) the original partial order. We call this process for extending the partial order the *cumulative rank frequency* (*CRF*) *operator*. The CRF operator can be iterated. In all cases studied to date, repeated application eventually results in a *linear* ordering of $S$ (see Figure 5) but it is not known if this is true in full generality. The research would examine this issue.

In most cases of practical interest, the number of linear extensions in $\Omega$ is too large for complete enumeration. For example, the UNEP HEI (Human Environment Index) data set involves three environmental indicators for 141 countries of the world. Elementary combinatorics shows that the number of linear extensions satisfies $8.6\times10^{105} \leq \#(\Omega) \leq 1.9\times10^{243}$ which is beyond foreseeable computational capabilities for direct enumeration. However, Markov Chain Monte Carlo (MCMC) methods, applied to the uniform distribution on $\Omega$, would allow us to *estimate* the normalized rank-frequency distributions needed to apply the CRF operator. See Aldous (1987), Brightwell and Winkler (1991) and Karzamov and Khachiyam (1991). The research would develop and implement the computational tools needed for application of MCMC.

Finally, the elements under comparison may be spatial regions; for example: countries across a continent or across the entire globe, watersheds within a state, or census tracts in a metropolitan area. In such cases, an echelon analysis of the partial order can be carried out by letting the successive levels in the Hasse diagram determine the newly exposed cells in the falling-water-level echelon model. This will provide a visualization tool for displaying and studying spatial connectivity and corridors among the highs and lows in the partial order.

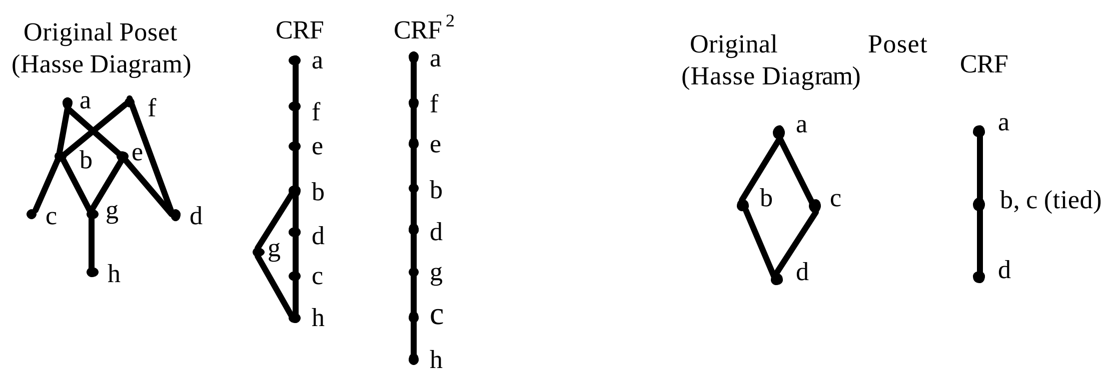


***Figure 5****. The three diagrams on the left show the linearizing effect of the CRF operator. The two diagrams on the right show how ties can emerge during linearization. A* ***poset*** *is a partially ordered set.*

## 8. Spatial Scan Statistic based on Upper Level Sets and Echelons of Surfaces

The spatial scan statistic was developed for detecting geographic clusters of disease that are statistically significant with respect to some larger geographic area within which the cluster is embedded (Kulldorf, 1997; Kulldorf and Nagarwalla, 1994). All potential zones are evaluated from a list that is created by starting with each original mapping unit and expanding a circle to incorporate increasingly larger areas that include other mapping units. After doing this for each mapping unit, an extraordinarily large list of candidate zones have been analyzed, whereby many zones are overlapping. While it is possible for the spatial scan statistic to pinpoint the general location of a cluster, its exact boundaries remain uncertain.

Echelon analysis will be used in conjunction with the spatial scan statistic in order to more clearly delineate cluster boundaries, since echelon families identify the spatial connectivity of a response surface. For example, two isolated first order echelons may be connected by a common second order echelon, as identified by “saddle point” mapping units. Echelons at any hierarchical level may be tested for statistical significance by the spatial scan statistic approach. Therefore, the combination of these two different methods will result in the determination of spatially disjoint areas of significantly elevated disease rates. Essentially, echelon analysis mechanizes and objectifies the way a person may look at a thematic or PRISM map and quickly determine a reasonable set of candidate zones, while eliminating many other zones as obviously uninteresting.

## 9. Geospatial Data Compression, Segmentation, and Classification

From both theoretical and practical perspectives, landscapes have a mosaic nature with particular pattern elements emerging at different scales. This compound mosaic nature is fundamental as a basis for landscape ecology. Since spectral reflectance mirrors the compositional character of land cover, digital image data also have latent informational structure as spatial mosaics. Each multi-band digital image dataset has an intrinsic *integral scale* due to the resolution element (pixel) over which spectral reflectance is sampled or intermixed as a composite by the sensor. Practical extraction of mosaic pattern can be conducted at three information levels of scale above the *integral* scaling level.

At the broadest level, mosaic pattern can be extracted for predominantly *perceptual* purposes. Most portrayals of images via computer displays are geared toward a one-byte informational level entailing something on the order of 256 tonal elements. For practical purposes, this can be considered as perceptual *macroscale*. More detailed mosaics that can serve a variety of practical analytical purposes span a *mesoscale* range encompassing perhaps two orders of magnitude increase in number of compositional elements. Beyond this is *microscale* level of spatial variability that can be considered as informational noise for most practical purposes that image data might serve. Variation at this level of detail can be captured in a statistical manner without retaining further spatial specificity of compositional elements.

The process of mosaic pattern extraction is one of image segmentation, where the operative partitioning takes place in the spectral domain. With inspiration from recent hyper-clustering approaches to image data, a learning strategy for **p**rogressively **s**egmenting **i**mages (**PSI**) has been conceived and implemented in a manner that generates dual-scale mosaics as approximating compressions of multi-band image datasets. A coarse PHASE (Palette Homogeneity Among Segmentation Elements) one-byte mosaic serves perceptual purposes for image rendering and also indexes 250 subsets of a finer mosaic contained in a separable second byte that serves analytical purposes (Myers, 2000).

The PSI mosaics have proven particularly advantageous for purposes of detecting changes in landscapes over time from periodic image acquisitions. The PSI approach supports a variety or both conventional and non-conventional change detection strategies. Mosaic analogs of all conventional image approaches are available. Combinatorial segmentation of multi-temporal image data sets can serve to isolate inconsistencies of landscape appearance over time. Indirect comparison of spatial segmentation patterns allows analysis of change using different sensing systems of over time that would be impossible under conventional approaches.

In addition to visual interpretation and change detection, thematic classification can be conducted on a segment basis as opposed to the conventional pixel basis. This entails hybridization of supervised and unsupervised techniques of classic image analysis. Segment-wise classification can be accomplished much more rapidly, however. Coupling change detection and segment-based classification offers prospects for highly automated updating of thematic maps from repetitive imagery.

Generation of PSI mosaics has been implemented for conventional computing platforms with heavy reliance on transfer of image data between disk and RAM memory. The process is computationally intensive, and typically entails an overnight run for a large image. We have done extensive work on the development of efficient and portable parallel algorithms involving the processing of images and raster maps. See Helman and JaJa (1995), Bader and JaJa (1996), Fallah-Adl et al (1996), Kalluri et al (1999, 2000, 2001). We plan to extend these techniques for the generation of PSI mosaics and their applications to change detection and thematic classification on large volumes of image data.

We will also develop new fuzzy classification algorithms in which transitional pixels can have multiple class membership. In particular we propose to extend the Amo-Montero-Biging fuzzy classification model (Amo et al 2000) to utilize surrounding contextual information as a second step in an adaptive fuzzy classification scheme. As a result, we will develop a hybrid adaptive classifier having the merits of both contextual classification and multiclass membership.

## 10. Data Structures and Algorithms for the Exploration of Raster Maps

This component of the project focuses on the development of efficient data structures and algorithms to explore associations between environmental phenomena and spatial patterns, building on the quantitative outcomes of the statistical models, and developing higher level models for detecting changes and finding interesting spatio-temporal patterns and trends. This requires the explicit discovery of spatio-temporal patterns based on parameter values that have been derived through the use of some of our statistical analysis techniques and models such as HMTM or echolons. In fact, a recent study by the NASA Earth Science Information Partnership (ESIP) that includes all the major data centers for earth sciences reveals that all major scenarios of data mining or knowledge discovery of spatio-temporal data involve a core component that requires the fast determination of patterns and regions over which a certain number of parameter values satisfy certain constraints, for example the values fall within certain ranges or that they remain within certain bounds over a certain time period.

In a recent work, we addressed the problem of quickly identifying regions for large scale multivariate raster maps. See JaJa and Shi (2001). We developed novel data structures and algorithms that are based on strong theoretical techniques and that have been validated by extensive experimentations over a wide range of data sets including the high-resolution Landsat TM. These techniques enabled the identification of various patterns and regions very quickly. Our techniques rely on an efficient representation of the raster maps using a combination of a specially designed R-tree built around the parameter values and spatial decomposition of the region into subregions described by their boundaries. We have shown that querying over arbitrary range values of any subset of the parameters can be done extremely quickly allowing real-time interactions even for the large data sets.

This project will extend these techniques in a number of directions which include: (1) the use of density-based sampling techniques to create a hierarchy of multi-resolution maps organized in a pyramidal structure such that only the coarsest possible resolution will be accessed as needed; (2) the development of spatio-temporal variant of R-trees that can be used in conjunction of the statistical models for quickly assessing accuracy and detecting changes; and (3) the generalization of these techniques to heterogeneous raster data, including multi-resolution maps.

## 11. Interface Design and Visualization Toolbox

A major goal of this effort is to develop a visualization interface integrated with software tools based on various statistical techniques and models developed by the investigators on this project. Information visualization and interface design are critical to making effective use of the various techniques and models. In fact, the proposed activities will produce complex surfaces and patterns that are key to understanding the structure of the landscape and make the right inferences. An effective set of information visualization tools will be essential to gain a deeper understanding of various outcomes and their relationships to spatial patterns and trends. Such outcomes include fragmentation profiles, simulation outcomes, patch structures, error distribution, change detection, spatial variation and regional indicators, thereby enabling users to examine their interrelationships and dependencies in a visual setting. Our goal will be to promote the discovery of inherent structures and patterns, build and test hypotheses, enable the detailed study of particular facets and dimensions of the data, and provide means to visually assess the utility and accuracy of the statistical and computational techniques developed.

The University of Maryland Human-Computer Interaction Lab (HCIL) is internationally recognized for their pioneering work in interface design and information visualization. During the past few years, the HCIL has developed highly interactive interfaces for EOSDIS and the Census Bureau using the principles of dynamic queries and query preview. See Ahlberg and Shneiderman (1994), Asahi et al (1995), Fredrikson et al (1999), Tang and Shneiderman (2001). Dynamic queries have been shown to an effective technique to browse complex information and encourage exploration, as well as to find patterns and exceptions. We will expand this work to develop an advanced interface for map analysis and exploration integrated with visualization tools such as map overlays and mosaicking and coupled with the GIS ESRI ARC-Info for which the University of Maryland has a site license. We will also combine our successful user-controlled strategies for information visualization with dynamic aggregation to enable rapid exploration of alternative hypotheses, detection of fundamental patterns, and identification of interesting outliers.

Our approach will be to work with domain specialists to identify their needs and frequent tasks. A phased implementation will allow us to implement simple algorithms at first and then embed more sophisticated algorithms. As our implementations mature we propose to conduct usability tests with the domain specialists to reface the interfaces and demonstrate efficacy.

## 12. Landscape Patterns, Change Detection, and Accuracy Assessment

**Atlantic Slope Watersheds and Land Cover Study:** The northeastern Atlantic Slope encompasses many ongoing investigative efforts dealing with watersheds and land cover, the most recent of which is the large Atlantic Slope Consortium project sponsored by EPA to study watershed and landscape linkages**.** Pennsylvania watersheds have been mapped at several scales through EPA and NSF sponsored research. The Multiresolution Land Characteristics (MRLC) land cover mapping work covers the entire northeast Atlantic Slope region. The Coastal Change Analysis Project (C-CAP) tracks land cover changes in the coastal zone. This wealth of geospatial information is augmented at global scale by the Global Land Cover Facility (GLCF) housed at the University of Maryland Institute for Advanced Computer Studies (UMIACS) and the Land cover Land use Change (LCLUC) thrust within NASA's Earth Science Enterprise (ESE). The capabilities of the MARMAP system will be applied to integrative studies of landscape change and ecosystem integrity over this region. This will include remapping land cover in Pennsylvania and developing regional coverage of image maps for general usage with GIS by natural resource managers.

**China Landscape Change Detection:** Investigators at Berkeley have been engaged in cooperative studies of land cover change in Beijing and Shenzhen, China using remote sensing – see Gong et al (1996). Investigators at Penn State University have likewise been cooperating with NASA scientists to develop advanced techniques of forest landscape change detection in northern China using remote sensing. Both programs of research have made available substantial amounts of field information for purposes of verification. The advanced facilities of MARMAP will be applied in these contexts to determine the levels of technological improvement that have been achieved in the present project.

## 13. Geographic Surveillance, Disease Mapping, and Evaluation

**Disease Mapping and Evaluation:** Disease data occur either as individual case events or as groups of case events (count data) within areal units, such as census tracts, zip codes, counties, etc. Any disease map must be considered with the appropriate background population which gives rise to the incidence. Maps answer the question: where? The maps in conjunction with the underlying data reveal spatial patterns not easily recognized from lists of statistical data. For example, use of remote sensing data and other relevant geospatial data can help evaluate surrounding landscape characteristics that may be precursors for vector-borne diseases leading to early warning, involving landscape health, ecosystem health, and human health. Investigators at Berkeley are searching for the habitats of snails that cause for the prevalence of schistosomiasis in western China using remotely sensed data, see Seto et al (2001). Algorithms developed in this study can be used to improve snail habitat characterization in 1-4 m resolution satellite imagery. This case study will involve collaboration with NASA and CDC on several infectious and non-infectious diseases of current interest. Also, the Penn State group is beginning to work with NCHS with regard to their national cancer data, and the GW group is investigating communities in the DC area with high incidence and mortality of breast cancer. These studies will benefit from the application of Hasse diagrams and corresponding rank frequency distributions; however, the large number of objects to be ranked based on multiple criteria will require estimation of normalized rank frequency distributions using MCMC methods. These studies will also involve applications of spatial scan statistics based on upper level sets and echelons of surfaces.

**Geographical Surveillance of Sudden Oak Death in California**: The Sudden Oak Death (SOD) Phytophthora sp. was first reported in 1995 and has been rapidly spreading in California in 6 coastal counties. Monitoring the changing pattern of oak death in the past 6-7 years plays an important role in studying the disease transmission. SOD has recently been isolated from Quercus agrifolia Nee (coast live oak) and Quercus kelloggii Newb. (black oak), both in the black oak group (subgenus Erythrobalanus); and from Lithocarpus densiflorus (Hook.& Arn.) Rehd. (tanoak). Change detection algorithms proposed in this study will be used by the Berkeley group to monitor the location and infection pattern of SOD.

## 14. Urban Heat Islands and Urban Sprawl

**Urban Heat Island Initiatives:** The urban heat island may be visualized as a temperature dome on urban area. It contributes to the formation of ozone, which is a major urban air pollutant that has serious human health consequences. Analysis of thermal energy characteristics helps us understand how we can modify the city landscape to lessen the impacts of the urban heat island and its subsequent effects on air quality. Current research by NASA and EPA is using remote sensing data to analyze the relationship between land use patterns and urban heat island development. A NASA initiative is in place that uses spacecraft and aircraft remote sensing data together with other relevant geospatial data on a local scale to help quantify and map urban sprawl, landuse change, air quality, and their impact on human health, such as pediatric asthma. This case study will involve collaboration with NASA, EPA, CDC, etc. A case study for Washington DC Urban Heat Island will be led by the GW group. There are three main objectives:

(1) Characterization of thermal landscape in the Washington metropolitan area. This aims at evaluating not only the strength of the urban heat island but also the spatial variance within the heat island.
(2) Evaluation of the relative roles of land cover characteristics and urban structures. This involves the quantification of land cover characteristics and urban structures such as percent impervious surfaces, biomass density, urban canyon geometry, and roadway density.
(3) Linking localized thermal characteristics to human health outcome. This attempts to directly and indirectly link illnesses, such as asthmatic attacks and heat strokes, to thermal stress.

These studies will involve applications of spatial scan statistic based on upper level sets and echelons of surfaces together with applications of posets, Hasse diagrams, and the resultant rank orderings and prioritizations (Patil and Taillie, 2001b)

## 15. Multiple Indicators, Comparisons, and Rankings

**UNEP State of the Environment Case Study:** The United Nations Environment Program (UNEP) has planned to initiate an Annual Report on the State of Environment, nationwide and worldwide. This case study will involve collaboration with UNEP, EPA, NCHS, etc., where interest is current in the ability to be able to accomplish rankings and rank intervals for a collection of elements with the multicriteria multiple indicators using project-based methods and tools involving partially ordered sets, Hasse diagrams, rank frequency distributions, and rank orderings consistent with the basic data matrix. The collection of elements may be watersheds, clusters, states, health service areas, ecoregions, etc. (Patil and Taillie, 2001b).

**Investigation of Schistosomiasis in China:** In addition to the related work described under disease mapping, the GW group will be studying how the temporal changes around the Three Gorges Dam (TGD) across China's Yangtze River will impact the basic ecological factors that drive the evolution of vertor-parasite genetics and different modes of schistosome transmission to man. These factors include mode of schistosome transmission, human infectivity rates, population rates, snail densities, etc. The techniques developed under this project will be used to prioritize and select sites for monitoring and to develop maps of endemic area.

### References


Ahlberg, C. and Shneiderman, B. Visual Information Seeking: Tight coupling of dynamic query filters with starfield displays , Proc. of ACM CHI94 Conference (April 1994), 313-317 + color plates.

Aldous, D. On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing. *Probability in the Engineering and Informational Sciences*, **1**, 33–46, 1987.

Arrow, K., Bolin, B., Costanza, R., Dasgupta, P., Folke, C., Holling, C.S., Jansson, B-O., Levin, S., Maler, K, Perrings, C., Pimentel, D. Economic growth, carrying capacity and the environment. *Science,* **268**, 520--521, 1995.

Asahi, T., Turo, D., and Shneiderman, B. Using treemaps to visualize the analytic hierarchy process. *Information Systems & Research* 6, 4 (December 1995), 357-375.

Azencott, R. Simulated annealing. Seminaire Bourbaki, No. 697. *Asterique*, **161–162**, 223–237, 1988.

Azencott, R., ed. *Simulated Annealing*. Wiley, New York, 1992.

Bader, D. and JaJa, J. Parallel Algorithms for Image Histogramming and Connected Components with an Experimental Study. *Journal of Parallel and Distributed Computing*, 35(2):173-190, 1996.

Bader, D., JaJa, J., and Chellappa, R. Scalable Data Parallel Algorithms for Texture Synthesis and Compression using Gibbs Random Fields. *IEEE Transactions on Image Processing*, 4(10):1456-1460, 1995.

Bader, D., JaJa, J., Harwood, D., and Davis, L. Parallel Algorithms for Image Enhancement and Segmentation by Region Growing with an Experimental Study. *Journal of Supercomputing*, 10(2):141-168, 1996.

Barone, P., Frigessi, A., and Piccioni, M. *Stochastic Models, Statistical Methods, and Algorithms in Image Analysis*. Lecture Notes in Statistics, No. 74. Springer-Verlag, New York, 1990.

Besag, J. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, **36**, 192-236, 1974.

Basharin, G. P. On a statistical estimate for the entropy of a sequence of independent random variables. *Theory of Probability and its Applications*, **4**, 333--336, 1959.

Berman, A., and Plemmons, R. J. *Nonnegative Matrices in the Mathematical Sciences*. Siam, Philadelphia, 1994.

Bissonnette, J. A. (ed). *Wildlife and Landscape Ecology: Effects of Patterns and Scale*. Springer, New York, 1997.

Boswell, M. T., O'Connor, J., and Patil, G. P. A crystal cube for coastal and estuarine degradation: Selection of endpoints and development of indices for use in decision making. In *Handbook of Statistics Volume 12: Environmental Statistics*, G. P. Patil, and C. R. Rao (eds). North Holland/Elsevier Science Publishers, New York and Amsterdam. pp. 771-790, 1994.

Bremaud, P. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, New York, 1999.

Brightwell, G. and Winkler, P. Counting linear extensions. *Order*, **8**, 225–242, 1991.

Cairns Jr., J and Pratt, J. R. The relationship between ecosystem health and delivery of ecosystem services, In: D.J. Rapport, C. Gaudet and P. Calow (eds), Evaluating and Monitoring the Health of Large-Scale Ecosystems, pp. 273-294. Springer-Verlag, Heidelberg, 1995.

CENR Integrating the Nation's Environmental Monitoring and Research Networks and Programs: A Proposed Framework. National Science and Technology Council, Committee on Environment and Natural Resources. Washington, DC, 1997.

Chen, S. J., and Hwang, C. L. *Fuzzy Multiple Attribute Decision Making: Methods and Applications*. Springer-Verlag, Berlin, 1992.

Chiaromonte F. Structures and exhaustive reductions: a general framework for the simplification of multivariate data. *Journal of the Royal Statistical Society-B*, 1999. (Submitted).

Chiaromonte F., Cook R.D. Sufficient dimension-reduction and graphics in regression. *Statistical Science*, 1999. (Submitted).

Christakos, G., and Hristopulos, D. T. *Spatiotemporal Environmental Health Modelling*. Kluwer, Boston, MA, 1998.

Cook, D., Buja, A., Cabrera, J., and Hurley, C. Grand tour and projection pursuit. *Computational and Graphical Statistics*, **4(3),** 155--172, 1995.

Cook, D. *Regression Graphics*, Wiley, New York, 1998.

Costanza, R. (ed). *Ecological economics: the science and management of sustainability*. Columbia University Press, New York, 1991.

Costanza, R. Toward an operational definition of ecosystem health. In *Ecosystem Health*: *New Goals for Environmental Management*, R. Costanza, B. G. Norton, and B. D. Haskell, eds. Island Press, Washington. pp 3-19, 1992.

Costanza, R., Mageau, M., Norton, B., and Patten, B. C. What is sustainability? In *Ecosystem Health*, D. Rapport, *et al*, eds. Blackwell Science, Malden, MA, pp. 31--239, 1998a

Costanza, R., Mageau, M., Norton, B., and Patten, B. C. Predictors of ecosystem health. In *Ecosystem Health*, D. Rapport, *et al*, eds. Blackwell Science, Malden, MA, pp. 240--250, 1998b.

Costanza, R., and Maxwell, T. Resolution and predictability: An approach to the scaling problem. *Landscape Ecology*, **9**, 47--57, 1994.

Cressie, N. A. C. *Statistics for Spatial Data*. John Wiley & Sons, New York, 1991.

Daily, G. (ed) *Nature's Services: Societal Dependence on Natural Ecosystems*. Island Press, Washington, 1997.

Del Amo, A., Montero, J., and Biging, G.S., Classifying Pixels by means of Fuzzy Relations, *International Journal on General Systems*, 29(4):605-621, 2000.

DeSoyza, A. G., Whitford, W. G., and Herrick, J. E. Sensitivity testing of indicators of ecosystem health. *Ecosystem Health*, **3(1)**, 44--53, 1997.

Deutsch, C. V. and Journel, A. G. *GSLIB: Geostatistical Software Library and User's Guide*, second edition. Oxford University Press, Oxford, 1998.

Di Battista, G., Eades, P., Tamassia, R., and Tollis, I. G. *Graph Drawing: Algorithms for the Visualization of Graphs*. Prentice Hall, Upper Saddle River, New Jersey, 1999.

Dieudonne, J. *Foundations of Modern Analysis*. Academic Press, New York, NY, 1969.

EPA. The index of watershed indicators. EPA-841-R-97-010, United States Environmental Protection Agency, Office of Water, Wash., DC, 1997.

Fallah-Adl, H., JaJa, J., Liang, S., Kaufman, Y., and Townshend, J. Fast Algorithms for Removing Atmospheric Effects from Remotely Sensed Imagery. *IEEE Computational Science & Engineering*, 66-77, Summer 1996.

Filar, J. A., Ross, N. P., and Wu, M. L. Environmental assessment based on multiple indicators. CEIS, U.S. EPA, Washington, DC, pp. 1--30, 1999.

Fishburn, P. C. *Interval Orders and Interval Graphs: A Study of Partially Ordered Sets*. Wiley, New York, 1985.

Forman, R. T. T., and Godron, M. *Landscape Ecology*. John Wiley & Sons, New York, 1986.

Fredrikson, A., North, C., Plaisant, C., Shneiderman, B. Temporal, geographical and categorical aggregations viewed through coordinated displays: A case study with highway incident data. Proc. 1999 Workshop on New Paradigms in Information Visualization and Manipulation, ACM New York (November 1999), 26-34.

Frohn, R. C. *Remote Sensing for Landscape Ecology: New Metric Indicators for Monitoring, Modeling and Assessment of Ecosystems*. Lewis, Boca Raton, 1998.

Gardner, R. H., O'Neill, R. V., and Turner, M. G. Ecological implications of landscape fragmentation. In *Humans as Components of Ecosystems, The Ecology of Subtle Human Effects and Populated Areas*. M. J. McDonnell and S. T. A. Pickett, eds. Springer-Verlag, New York, 1993.

Geman, D. Random fields and inverse problems in imaging. *Lecture Notes in Mathematics*, No. 1427, pp. 117-193. Springer-Verlag, New York, 1990.

Geman, S. and Geman, D. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721--741, 1984.

Gidas, B. Metropolis-type Monte Carlo simulation algorithms and simulated annealing. In *Topics in Contemporary Probability and its Applications*, J. Laurie Snell, ed., pp. 159–233, 1995.

Gimel'Farb, G. L. *Image Textures and Gibbs Random Fields*. Kluwer, Boston, 1999.

Gong, P., Shi, P., Pu, R., and Guo, H., *Earth Observation Systems and Earth system Science*, Science Press, Beijing, China, 208p.

Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, Massachusetts, 1989.

Goovaerts, P. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, Oxford, 1997.

Graham, A. *Nonnegative Matrices and Applicable Topics in Linear Algebra*. Ellis Horwood Limited, Chichester, 1987.

Grossi, L., Zurlini, G., and Rossi, O. Statistical detection of multiscale landscape patterns. *Ecological and Environmental Statistics*\/, 2000. (To appear).

Guerra, C. Vision and image processing algorithms. Chapter 22 of *Algorithms and Theory of Computation Handbook*\/, M. J. Atallah, ed. CRC Press, Boca Raton, 1999.

Gustafson, E. J. and Parker, G. R. Relationships between landcover proportion and indices of spatial pattern. *Landscape Ecology*, **11**, 65--77, 1992.

Guttorp, P. *Stochastic Modeling of Scientific Data*. Chapman \& Hall, London, 1995.

Guyon, X. *Random Fields on a Network: Modeling, Statistics, and Applications*}. Springer-Verlag, New York, 1995.

Hansen, A. J., and di Castri, F. (eds). *Landscape Boundaries.: Consequences for Biotic Diversity and Ecological Flows*. Springer-Verlag, New York, 452 pp, 1992.

Hargis, C. D., Bissonette, J. A., and David, J. L. Understanding measures of landscape pattern. In *Wildlife and Landscape Ecology: Effects of Pattern and Scale*. J. A. Bissonette, ed. Springer, New York, pp. 231--261, 1997.

Hargis, C. D., Bissonette, J. A. and David, J. L. The behavior of landscape metrics commonly used in the study of habitat fragmentation. *Landscape Ecology*, **13**, 167--186, 1998.

Hartfiel, D. J. *Markov Set Chains\/*. Springer, New York, 1998.

Haslett, J., Bradley R., Craig, P., Unwin, A., and Wills, G. Dynamic graphics for exploring spatial data with application to locating global and local anomalies. *The American Statistician*, **45(3),** 234--242, 1991.

Hastings, H. M. and Sugihara, G. *Fractals---A User's Guide for the Natural Sciences\/*. Oxford University Press, New York, 1993.

Helman, D., and JaJa, J. Efficient Image Processing Algorithms on the Scan Line Array Processor. *IEEE Transactions on PAMI*, 17(1):47-56, 1995.

Hilden, M. Who framed the Kyronjoki? In *Ecosystem Health*, D. J. Rapport, R. Costanza, P. Epstein, C. Gaudet, R. Levins, eds. Blackwell Science, Malden, MA, pp. 348--355, 1998.

Hilden, M. and Rapport, D. J. Four centuries of cumulative impacts on a Finnish rier and its estuary; an ecosystem health approach. *J. Aquatic Ecosystem Health*, **2**, 261--275, 1993.

Hwang, C. L., and Yoon, K. *Multiple Attribute Decision Making: Methods and Applications: A State-of-the-Art Survey*, Springer-Verlag, Berlin.

JaJa, J. and Shi, Q. Efficient Techniques for Exploring Geospatial Data, submitted for publication, 2001.

Johnson, G. D. *Landscape Pattern Analysis for Assessing Ecosystem Condition: Development of a Multi-Resolution Method and Application to Watershed Delineated Landscapes in Pennsylvania.* Ph.D. Thesis, The Pennsylvania State University, University Park, PA, 1999.

Johnson, G. D., Myers, W. L., and Patil, G. P. Stochastic generating models for simulating hierarchically structured multi-cover landscapes. *Landscape Ecology*, **14**, 413-421, 1999.

Johnson, G. D., Myers, W. L., and Patil, G. P. Predictability of surface water pollution loading in Pennsylvania using watershed-based landscape measurements. *Journal of the American Water Resources Association*, 2000. (Submitted)

Johnson, G. D., Myers, W. L., Patil, G. P., and Taillie, C. Multiresolution fragmentation profiles for assessing hierarchically structured landscape patterns. *Ecological Modeling*, **116**, 293--301, 1999a.

Johnson, G. D., Myers, W. L., Patil, G. P., and Taillie, C. Characterizing watershed-delineated landscapes in Pennsylvania using conditional entropy profiles. *Landscape Ecology*, 1999b.

Johnson, G. D., Myers, W. L., Patil, G. P., and Taillie, C. Quantitative characterization of hierarchically scaled landscape patterns. *Environmental and Ecological Statistics*, 2000.

Johnson, G. D., Myers, W. L., Patil, G. P., and Walrath, D. Multiscale analysis of the spatial distribution of breeding bird species richness using the echelon approach. In *Assessment of Biodiversity for Improved Forest Planning,* P. Bachmann, M. Kohl, and R. Paivinen, eds. Kluwer Academic Publishers, pp. 135--150, 1998.

Johnson, G. D., and Patil, G. P. Quantitative multiresolution characterizations of landscape patterns for assessing the status of ecosystem health in watershed management areas, *Ecosystem Health*, 4(3), 177--187, 1998.

Johnson, N. C., Malk, A. J., Szaro, R. C., and Sexton, W. T. (eds). *Ecological Stewardship: A Common Reference for Ecosystem Management, Volume I.* Elsevier Science, Oxford, UK, 1999.

Jones, K. B., K. H. Riitters, J. D. Wickham, R. D. Tankersley, Jr., R. V. O'Neill, D. J. Chaloud, E. R. Smith, and A. C. Neale. An ecological assessment of the United States Mid-Atlantic Region: A landscape atlas. EPA/600/R-97/130, United States Environmental Protection Agency, Office of Research and Development, Research Triangle Park, North Carolina, 1997.

Kalluri, S., JaJa, J., Bader, D., Zhang, Z., Townshend, J., and Fallah-Adl, H. High Performance Computing Algorithms for Land Cover Dynamics Using Remote Sensing Data. *International Journal of Remote Sensing*, 2000.

Kalluri, S., Zhang, Z., JaJa, J., Liang, S., and Townshend, J. Characterizing Land Surface Anisotropy from AVHRR Data at a Global Scale Using High Performance Computing. I*nternational Journal of Remote Sensing*, 22(11), 2171-2191, 2001.

Karzanov, A. and Khachiyan, L. On the conductance of order Markov chains. *Order*, **8**, 7–15, 1991.

Kijima, M. *Markov Processes for Stochastic Modeling\/*. Chapman \& Hall, London, 1997.

Knuth, D. E. *The Art of Computer Programming. Vol* I: *Fundamental Algorithms \/* (second edition). Addison-Wesley, Reading, MA, 1973.

Kozintsev, B. and Kedem, B., Generation of “Similar” Images from a Given Discrete Image, Journal of Computational and Graphical Statistics, 9, 286-302, 2000.

Krummel, J. R., Gardner, R. H., Sugihara, G., O'Neill, R. V., and Coleman, P. R. Landscape patterns in a disturbed environment. *Oikos*, **48**, 321--324, 1987.

Kulldorf, M. 1997. A spatial scan statistic. *Communications in Statistics – Theory and Methods,* **26(6)**: 1481-1496.

Kulldorf, M. and Nagarwalla, N. 1994. Spatial disease clusters: detection and inference. Statistics in Medicine, 13.

Kurihara, K., Myers, W. L., and Patil, G. P. The relationship of the population and land cover patterns in Tokyo area based on remote sensing data. *Technical Report 99-1103*, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA., 1999.

Lange, K. *Numerical Analysis for Statisticians*. Springer, New York, 1999.

Leopold A. Wilderness as a land laboratory. *Living Wilderness,* **3**, 1941.

Levin, S. The problem of pattern and scale in ecology. *Ecology*, **73**, 1943--1967, 1992.

Levin, S. A., Grenfell, B., Hastings, A., Perrelson, A. S. Mathematical and computational challenges in population biology and ecosystem science. *Science*, **275**, 334--343, 1997.

Lord, J. M., and Norton, D. A. Scale and the spatial concept of fragmentation. *Conservation Biology*, **2**, 197--202, 1990.

Lunetta, R. S. and Elvidge, C. D. (eds). *Remote Sensing Change Detection: Environmental Monitoring Methods and Applications*. Ann Arbor Press, Ann Arbor, MI, 1998.

Mageau, M.T., Costanza, R. and Ulanowicz, R.E. The development and initial testing of a quantitative assessment of ecosystem health. *Ecosystem Health*, **1**, 201--213, 1995.

Margalef, R. Human impact on transportation and diversity in ecosystems. How far is extrapolation valid? In *Proceedings of the First International Congress of Ecology. Structure, Functioning and Management of Ecosystems*, The Hague. Sept 9-14, 1974. Centre for Agricultural Publishing and Documentation, Wageningen, Netherlands, pp. 237--241, 1975.

May, R. M. The effects of spatial scale of ecological questions and answers. In *Large Scale Ecology and Conservation Biology*, P. J. Edwards, R. M. May, and N. R. Webb, eds. Blackwell Scientific Publicaitons, London, UK, 1994.

McGarigal, K. and Marks, B. FRAGSTATS: Spatial pattern analysis program for quantifying landscape structure. *General Technical Report PNW-GTR-351. Portland, OR, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station,* 1995.

McKenzie, D. H., Hyatt, D. E., and McDonald, V. J. *Ecological Indicators, Volume 1*. Elsevier Applied Science, London and New York, 148 pp, 1992.

McKenzie, D. H., Hyatt, D. E., and McDonald, V. J. *Ecological Indicators, Volume 2*. Elsevier Applied Science, London and New York, 148 pp, 1992.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087-1091, 1953.

Metzger, J. P., and Muller, E. Characterizing the complexity of landscape boundaries by remote sensing. *Landscape Ecology*, **11(2)**, 65--77, 1996.

Meyn, S. P., and Tweedie, R. L. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993.

Milne, B. T. Lessons from applying fractals models to landscape patterns. In Qu*antitative Methods in Landscape Ecology\/,* M. G. Turner and R. H. Gardner, eds. Springer-Verlag, Berlin, 1991.

Milne, B. T. Indications of landscape condition at many scales. In *Ecological Indicators, Volume 2*. D. H. McKenzie, D. E. Hyatt, V. J. McDonald, eds. Elsevier Applied Science, London and New York, pp 883--895, 1992.

Mumford, D.
What makes images special from a statistical viewpoint. Computer Vision Day, Department of Statistics, Penn State University, University Park, PA, 1997.

Myers, W. L. PHASE approach to remote sensing and quantitative spatial data. Technical Report ER 9710, Environmental Resources Research Institute, Pennsylvania State University, University Park, PA, 1997.

Myers, W. L. Remote sensing and quantitative georide in PHASES (Pixel Hyperclusters as Segmented Environmental Signals), release 3.4. Technical Report ER9901, Environmental Resources Research Institute, Penn State University, University Park, PA 16802, 1999.

Myers, W. L. PHASE-Based Broad-Area Landscape Change Analysis. Final report on NASA Reserch Project NAGS5-6713. Environmental Resources Research Institute Research Report ER2005. Penn State University, University Park, PA 16802, 2000.

Myers, W. L., Patil, G. P., and Joly, K. Echelon approach to areas of concern in synoptic regional monitoring. *Environmental and Ecological Statistics*, **4(2),** 131--152, 1997.

Myers, W. L., Patil, G. P., and Taillie, C. Comparative paradigms for biodiversity assessment. Invited paper at the IUFRO Symposium in Chiang-mai, Thailand. In *Measuring and Monitoring Biodiversity in Tropical and Temperate Forests*, T. J. Boyle and B. Boontawee, eds. CIFOR, Bogor, Indonesia, pp. 67--85, 1995.

Myers, W. L., Patil, G. P., and Taillie, C. Conceptualizing pattern analysis of spectral change relative to ecosystem health. *Ecosystem Health*, **5(4),** 285--293, 1999.

Neggers, J. and Kim, H. S. *Basic Posets*. World Scientific, Singapore, 1998.

Newman, M. E. J. and Barkema, G. T. *Monte Carlo Methods in Statistical Physics*. Oxford University Press, Oxford, 1999.

Noss, R. F., Slosser, N. C., Strittholt, J. R., and Carroll, C. Some thoghts on metrics of ecological integrity for terrestrial ecosystems and entire landscapes. Advisory Report to EPA, pp.~1--124, 1999.

O'Connell, T. J., Jackson, L. E., and Brooks, R. P. The bird community index: A tool for assessing biotic integrity in the mid-Atlantic highlands. Final report to U.S. EPA, No. 98-4, Penn State Cooperative Wetlands Center, Forest Resources Lab, Penn State University, University Park, PA, 1998.

O'Neill, R. V. Perspectives in hierarchy and scale. In *Perspectives in Ecological Theory*, J. Roughgarden, R. M. May, and S. A. Levin, eds. Princeton University Press, Princeton, NJ, 1989.

O'Neill, R. V., DeAngelis, D. L., Waide, J. B., and Allen, T. F. H. *A Hierarchical Concept of Ecosystems*. Princeton University Press, Princeton, 1986.

O'Neill, R. V., Hunsaker, C. T., Jones, K. B., Riitters, K. H., Wickham, J. D., Schwartz, P. M., Goodman, I. A., Jackson, B. L., and Baillargeon, W. S. Monitoring environmental quality at the landscape scale; using landscape indicators to assess biotic diversity, watershed integrity and landscape stablity. *Bioscience*, **47**, 513--519, 1997.

O'Neill, R. V., Hunsaker, C. T., Timmins, S. P., Jackson, B. L., Jones, K. B., Riitters, K. H., and Wickham, J. D. Scale problems in reporting landscape pattern at the regional scale. *Landscape Ecology*, **11**, 169--180, 1996.

O'Neill, R. V., Johnson, A. R., King, A. W. A hierarchical framework for the analysis of scale. *Landscape Ecology*, **3**, 193--205, 1989.

Parkan, C. Measuring the performance of hotel operations. *Socio-Economic Planning Science*, **30(4)**, 257--292, 1996.

Parkan, C., and Wu, M. L. Process selection with multiple objective and subjective attributes. *Production Planning and Control*, (To appear), 1998a.

Parkan, C., and Wu, M. L. Measuring the performance of operations of Hong Kong manufacturing industries. *European Journal of Operational Research* (To appear), 1998b.

Patil, G. P. Statistical ecology and environmental statistics for cost-effective ecological synthesis and environmental analysis. In *Modern Trends in Ecology and Environment*, R. S. Ambasht, ed. Backhuys Publ., The Netherlands, pp. 5--36, 1998.

Patil, G. P. Invited Plenary Address at the Portuguese Statistical Congress, Ponte Delgada, Portugal. November 2001a.

Patil, G. P. Invited Plenary Address at the Brazilian Ecological Congress, Porto Allegre, Brazil, November, 2001b.

Patil, G. P., Johnson, G. D., Myers, W. L., and Taillie, C. Multiscale statistical approach to critical-area analysis and modeling of watersheds and landscapes. In *Statistics for the 21st Century: Methodologies for Applications of the Future*, C. R. Rao and G. J. Szekely, eds. Marcel Dekker, Inc., New York, pp. 293–310, 2000.

Patil, G. P., and Myers, W. L. Guest Editorial: Environmental and ecological health assessment of landscapes and watersheds with remote sensing data. *Ecosystem Health*, **5(4)**, 221--224, 1999.

Patil, G. P., Myers, W. L., Luo, Z., Johnson, G. D., and Taillie, C. Multiscale assessment of landscapes and watersheds with synoptic multivariate spatial data in environmental and ecological statistics. *Mathematical and Computer Modeling*, 1999. (In press).

Patil, G. P., and Rodriguez, S. Environmental and ecological regional policy research with remote imagery and geospatial information: Issues, approaches, and examples. Technical Report 99-1102, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA, 1999.

Patil, G. P., and Taillie, C. A Markov model for hierarchically scaled landscape patterns. In *Bull. of the International Statistical Institute*, Volume 58, Book 1. pp. 89--92, 1999.

Patil, G. P., and Taillie, C. Topological concepts and definitions for echelons and echelon trees. Technical Report 99-0602, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA, 1999b.

Patil, G. P., and Taillie, C. Modeling and interpreting the accuracy assessment error matrix for a doubly classified map. Technical Report 99-0502, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA., 2000a.

Patil, G. P., and Taillie, C. A multiscale hierarchical Markov transition matrix model for generating and analyzing thematic raster maps. Technical Report 2000-0603, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA., 2000b.

Patil, G. P., and Taillie, C. Analytic solution of the regularized latent truth model for binary maps. *Technical Report 2000-0601, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA.*, 2000c.

Patil, G. P., and Taillie, C. A disjunctive indicator geostatistical model for categorical raster maps. Technical Report 2001-0901, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA., 2000a.

Patil, G. P., and Taillie, C. Multiple indicators, partially ordered sets, and linear extensions: Multi-criterion ranking methods. Technical Report 2001-1001, Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Penn State University, University Park, PA., 2000b.

Pavlidis, T. *Structural Pattern Recognition\/*. Springer-Verlag, Berlin, 1977.

Pearson, K. Mathematical contributions to the theory of evolution--VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London, Series A*, **200**, 1-66, 1901.

Pearson, S. M., and Gardner, R. H. Understanding neutral models: Useful tools for standing landscape patterns. In *Wildlife and Landscape Ecology: Effects of Pattern and Scale*, J. A. Bissonette, ed. Springer, New York, pp. 215--230, 1997.

Picket, S. T. A. and White, P. S. (eds). *The Ecology of Natural Disturbance and Patch Dynamics*. Academic Press, Orlando, FL, 1985.

Porter, D., and Stirling, D. S. *Integral Equations: A Practial Treatment, from Spectral Theory to Applications*. Cambridge University Press, Cambridge, UK, 1990.

Prakasa Rao, B. L. S. *Nonparametric Functional Estimation*. Academic Press, New York, 1983.

Quattrochi, D. A., and Goodchild, M. F. (eds). *Scale in Remote Sensing and GIS\/*. Lewis Publishers, Boca Raton, 1997.

Radermacher, W. Indicators, green accounting and environment statistics--Information requirements for sustainable development. International Statistical Review, **67(3)**, 339--354, 1999.

Ramakomud, A. Change detection using hyperclustered data: the spatial averaging approach. Master of Science Thesis, Penn State Univ., Univ. Park, PA, 1998.

Rapport, D. J. Evolution of indicators of ecosystem health. *Applied Science*, **1**, 121--134, 1992.

Rapport, D. J. Ecosystem health: an emerging integrative science, In *Evaluating and Monitoring the Health of Large-Scale Ecosystems*, : D. J. Rapport, C. Gaudet and P. Calow (eds), Springer-Verlag, Heidelberg, pp 5--31, 1995.

Rapport, D. J. What constitutes ecosystem health? *Perspectives in Biology and Medicine*, **33**, 120--132, 1989a.

Rapport, D. J. Symptoms of pathology in the G'ulf of Bothnia (Baltic Sea): ecosystem response to stress from human activity. *Biol. J. Linn. Soc.*, **37**, 33--40, 1989b.

Rapport, D. J. Editorial: Ecosystem health, ecological integrity, and sustainable development: Toward consilience. *Ecosystem Health*, **4(3)**, 145, 1998.

Rapport, D. J., Christensen, N., Karr, J. R., and Patil, G. P. The centrality of ecosystem health in achieving sustainability in the 21st century: Concepts and New Approaches to Environmental Management. *Human Survivability in the 21st Century: Transactions of the Royal Society of Canada*, University of Toronto Press, pp. 3--40, 1999.

Rapport, D. J., Hilden, M., and Weppling, K. Restoring the health of the earth's ecosystems: A new challenge for the earth sciences. *Episodes*, 2000. (In press)

Rapport, D. J. and Regier, H. A. Disturbance and stress effects on ecological systems. In *Complex Ecology*, B. C. Patten and S. E. Jorgensen, eds. (Memorial volume in honour of G. VanDyne) Prentice-Hall, Englewood Cliffs, NJ, pp. 397--414, 1995.

Rapport, D. J. , Hilden, M. and Roots, E. F. Transformation in arctic ecosystems under stress. In *Disturbance and Recovery in Arctic Lands: An Ecological Perspective*, R. M. M Crawford, ed., Kluwer, pp. 73--90, 1997.

Rapport, D. J., Regier, H. A., and Hutchinson, T. C. Ecosystem behavior under stress. *The American Naturalist*, **125**, 617--640, 1995.

Rapport, D. J. and Whitford, W. G. How ecosystems respond to stress: Common properties of arid and aquatic systems. *BioScience*, **49(3)**, 193--203, 1999.

Retherford, J. R. *Hilbert Space: Compact Operators and the Trace Theorem*. Cambridge University Press, 1993.

Riitters, K. H., O'Neill, R. V., Hunsaker, C. T., Wickham, J. D., Yankee, D. H., Timmins, S. P., Jones, K. B. and Jackson, B. L. A factor analysis of landscape pattern and structure metrics. *Landscape Ecology*, **10**, 23--29, 1995.

Riitters, K. H., O'Neill, R. V., and Jones, K. B. Assessing habitat suitability at multiple scales: A landscape-level approach. *Biological Conservation*, **81**, 191--202, 1997.

Robert, C. P., and Casella, G. *Monte Carlo Statistical Methods*. Springer, New York, 1999.
Ronse, C. and Devijver, P. A. *Connected Components in Binary Images: The Detection Problem\/*. Research Studies Press, Letchworth, England, 1984.

Rodriguez-Iturbe, I., and Rinaldo, A. *Fractal River Basins: Chance and Self-Organization*. Cambridge University Press, Cambridge, UK, 547 pp., 1997.

Ronse, C. and Devijver, P. A. *Connected Components in Binary Images: The Detection Problem\/*. Research Studies Press, Letchworth, England, 1984.

Rudin, W. *Functional Analysis*. McGraw-Hill, New York, NY, 1973.

Saaty, T. L. *Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World* (1999/2000 edition), 3rd rev. ed., vol. 2, RWS Publications, Pittsburgh, 1999.

Samet, H. *Applications of Spatial Data Structures: Computer Graphics*, Image *Processing, and GIS*. Addison-Wesley, Reading, MA., 1990.

Schinazi, R. B. *Classical and Spatial Stochastic Processes*. Birkhauser, Boston, 1999.

Schumaker, N. H. Using landscape indices to predict habitat connectivity. *Ecology*, **77(4)**, 1210--1225, 1996.

Schweitzer, P. J. A survey of aggregation-disaggregation in large Markov chains. pp. 63--88 in Stewart, 1991.

Scott, D. W., and Thompson, J. R. Probability density estimation in higher dimensions. In *Computer Science and Statistics: Proceedings of the Fifteenth Symposium on the Interface*, J. E. Gentle, ed. North-Holland, Amsterdam, pp. 173--179. 1983.

Scott, J. M., Csuti, B., Smith, K., Estes, J. E., and Caicco, S. Gap analysis of species richness and vegetation cover: an integrated conservation strategy for the preservation of biological ddiversity. In *Balancing on the Brink: A Retrospective on the Endangered Species Act*. Island Press, Washington, DC, 1990.

Scott, J. M., Davis, F., Csuti, B., Noss, R., Butterfield, B., Groves C., anderson, H., Caicco, S., D'Erchia, F., Edwards, T. C., Ulliman, J., and Wright, G. Gap analysis: a geographic approach to protection of biological diversity. *Wildlife Monographs*, 123, 1993.

Serfozo, R. *Introduction to Stochastic Networks\/*. Springer, New York, 1999.

Seto, E., Xu, B., Liang, S., Spear, R., Gong, P., Wu, W., Davis, G., Qiu, D., and Gu, X., The Use of Remote Sensing for Predictive Modelling of Schistosomiasis in China, *PR&RS*, 76(11), 2001.

Sexton, W. T., Malk, A. J., Szaro, R. C., and Johnson, N. C. (eds). *Ecological Stewardship: A Common Reference for Ecosystem Management, Volume III.* Elsevier Science, Oxford, UK, 1999.

Shapiro, L. G.
Connected component labeling and adjacency graph construction. Pages 1--30 in *Topological Algorithms for Digital Image Processing\/*, T. Y. Kong and A. Rosenfeld, eds. North-Holland Publishing Company, Amsterdam, 1996.

Silverman, B. W. *Density Estimation for Statistics and Data Analysis*}. Chapman and Hall, London, UK. 1994.

Sinclair, A. *Algorithms for Random Generation and Counting: A Markov Chain Approach\/*. Birkhauser, Boston, 1993.

Slud, E., Stone, M., Smith, P. and Goldstein, M., Jr. Principal components representation of the two-dimensional tongue surface. To appear in *Phonetica*, 2000.

Smits, P. C., and Myers, W. L. Echelon approach to characterize and understand spatial structures of change in multi-temporal remote-sensing imagery. *IEEE Trans. Geoscience and Remote Sensing,* 2000. (Under revision)

Stewart, W. J. (ed.) *Numerical Solution of Markov Chains\/*. Dekker, New York, 1991.

Stewart, W. J. (ed.) *Computations with Markov Chains\/*. Kluwer, Boston, 1995.

Stewart, W. J. Numerical methods for computing stationary distributions of finite irreducible Markov chains. In *Computational Probability\/*, W. K. Grassmann, ed. Kluwer, Boston, pp. 81--110, 2000.

Szaro, R. C., Johnson, N. C., Sexton, W. T., and Malk, A. J. (eds). *Ecological Stewardship: A Common Reference for Ecosystem Management, Volume II*. Elsevier Science, Oxford, UK, 1999.

Tang, L. and Shneiderman, B. Dynamic aggregation to support pattern discovery: A case study with web logs. Department of Computer Science Technical Report, University of Maryland, College Park, MD (March 2001). Short version appears in Proc. Discovery Science 2001, Springer.

Tran, L., and Duckstein, L. Comparison of fuzzy numbers using a fuzzy distance measure. 1999.

Trotter, W. T. *Combinatorics and Partially Ordered Sets*. Johns Hopkins University Press, Baltimore, 1992.

Tufte, E. R. *Envisioning information*. Graphics Press, Cheshire, CT, 1990.

Turner, M. G., and Gardner, R. H. (eds). *Quantitative Methods in Landscape Ecology*. Springer-Verlag, New York, 1991.

Turner, M. G., O'Neill, R. V., Gardner, R. H., and Milner, B. T. Effects of changing spatial scale on the analysis of landscape pattern. *Landscape Ecology*, **3**, 153--162, 1989.

Urban, D. L., O'Neill, R. V., and Shugart, H. H., Jr. Landscape ecology; a hierarchical perspective can help scientists understand spatial patterns. Bioscience, **37**, 119--127, 1987.

Vitousek, P. M., Mooney, H. A. Lubchenco, J. and Milillo, J. M. Human domination of earth's ecosystems. *Science*, 277, 494--499, 1997.

White, D., Minotti, P. G., Barczak, M. J., Sifneos, J. C., et al. Assessing risks to biodiversity from future landscape change. *Conservation Biology*, **11(2)**, 349--360, 1997.

Whitford, W. G. The Desert Grasslands. In *Ecosystem Health*, D. J. Rapport, R. Costanza, P. R. Epstein, F. C. Gaudet, and R. Levins, eds. Blackwell Science, pp 313--323, 1998.

Wichert, G. and Rapport, D. J. Fish community structure as a measure of degradation and rehabilitation of riparian systems in an agricultural drainage basin. *Environmental Management*, **22**, 425--443, 1998.

Widom, H. *Lectures on Integral Equations*. Van Nostrand Reinhold Company, New York, 1969.

Wiens, J. A. Spatial scaling in ecology. *Functional Ecology*, **3**, 385--397, 1989.

Wilson, E. O. *Consilience: The Unity of Knowledge*. Knopf, NY, 332 pp., 1998.

Winkler, G. *Image Analysis, Random Fields and Dynamic Monte Carlo Methods: A Mathematical Introduction*, Springer, New York, 1995.

Wise, S., Haining, R., and Signoretta, P. The role of visualization for exploratory spatial data analysis of are-based data. *Proc. Fourth International Confrenece on Geocomputation (GeoComputation '98)*, Bristol, UK, 1998.

Yang, X. D. An improved algorithm for labeling connected components in a binary image. In *Computer Vision and Image Processing*, Academic Press. pp.\ 555--569, 1992

Yazvenko, S. B. and Rapport, D. J. The history of Ponderosa pine pathology: implications for management. *J. Forestry*, **95**, 16--20, 1997.

Younes, L. Estimation and annealing for Gibbsian fields. *Annales de l'Institut Henri Poincare--Probabilities et Statistiques*, **24**, 269-294, 1988.

Younes, L. Maximum likelihood estimation for Gibbsian fields. In *Spatial Statistics and Imaging*, A. Possolo, ed., pp. 403-426, 1991.

Zeleny, M. *Multiple Criteria Decision Making*, McGraw-Hill, New York, 1982.

Zhang, Z., Kalluri, S., JaJa, J., Liang, S., and Townshend, J. High Performance Algorithms for Global BRDF Retrieval. *IEEE Computational Science & Engineering*, 5(4):16-29, 1998.