MULTISCALE ADVANCED RASTER MAP INFORMATION SCIENCE AND TECHNOLOGY
A Research and Outreach Prospectus of Advanced Mathematical, Statistical, and Computational Approaches Using Remote Sensing Data
DEVELOPMENT AND IMPLEMENTATION OF A USER FRIENDLY MARMAP SYSTEM

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

Joseph JaJa, Institute of Advanced Computer Studies (UMIACS)
University of Maryland
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
G. S. Biging, Resource Information Technology and Landscape Ecometrics
College of Natural Resources, University of California, Berkeley
Peng Gong, Landscape Environmetrics, Integrated Assessment, and Visualization
College of Natural Resources, University of California, Berkeley
Jaeyong Lee, Department of Statistics, The Pennsylvania State University
W. L. Myers, Remote Sensing, Natural Resources, and Software
School of Forest Resources, The Pennsylvania State University
Ben Shneiderman, Department of Computer Science and UMIACS
University of Maryland
Stephen Rathbun, Department of Statistics, The Pennsylvania State University
C. Taillie, Computational Statistics and Stochastics
Department of Statistics, The Pennsylvania State University
John Townshend, Geography, Remote Sensing, and Landscape Analysis
Department of Geography and UMIACS, University of Maryland

Partners
Featured in Sections 18-20 on Pages 23-24

November 29, 2001

Contents

Prospectus Summary

This project brings together a multidisciplinary team to develop advanced statistical and computational techniques for analyzing, assessing, and extracting information from raster maps. This information will provide a rigorous foundation to address a wide range of applications including disease mapping, emerging infectious diseases, landscape ecological assessment, land cover trends and change detection, watershed assessment, and map accuracy assessment. We will develop an advanced map analysis system that integrates these techniques with an advanced visualization toolbox, and use the system to conduct large case studies using rich sets of raster data, primarily from remotely sensed imagery, which are already available at the various sites participating in this project. As a result, we will be able to study and evaluate raster maps of societal, ecological, and environmental variables to facilitate quantitative characterization and comparative analysis of geospatial trends, patterns, and phenomena. In addition to environmental and ecological studies, our techniques and tools can be used for policy decisions at national, state, and local levels, crisis management, and protection of infrastructure.

Geospatial data form the foundation of an information-based society. The President's Committee on Environmental and Natural Resources (CENR, 1997) has indicated that remote sensing has been a vastly under-utilized resource involving a multi million dollar investment at the national level. Even when utilized, the credibility has been at stake, largely because of lack of tools that can assess, visualize, and communicate accuracy and reliability in timely manner and at desired confidence levels. This project aims at developing a methodological and computational foundation to make effective use of geospatial data in addressing regional and global environmental, ecological, and public health issues.

Our team involves researchers with a solid track record in a number of complementary areas that are at the core of this project. Our approach will build on the considerable work done by the individual members of our team to address in an integrated fashion the issues stated above, further develop and combine different methodologies paying particular attention to the related computational aspects, and integrate the resulting advances into a sophisticated system to be used on large scale studies. In particular, our research program includes the development of techniques for map modeling and analysis using Markov random fields, geospatial statistics, accuracy assessment and change detection, upper echelons of surfaces, advanced computational techniques for geospatial data mining, and advanced visualization techniques. A testbed that includes a powerful remotely accessible cluster with 10 TB of disks will be set up at the University of Maryland to test and validate the system, and to conduct the large theoretical and empirical investigations in collaboration with George Washington University, Penn State University, and the University of California at Berkeley, leading to a novel innovative breakthrough in the waiting.

Prospectus Description

Introduction and Motivation

Geospatial data form the foundation of an information-based society. The President's Committee on Environmental and Natural Resources (CENR, 1997) has indicated that remote sensing has been a vastly under-utilized resource involving a multi million dollar investment at the national level. Even when utilized, the credibility has been at stake, largely because of lack of tools that can assess, visualize, and communicate accuracy and reliability in a timely manner and at desired confidence levels.

Consider an imminent 21*st* Century scenario: What message does a multi-categorical map have about the large landscape it represents? And at what scale, and at what level of detail? Does the spatial pattern of the map reveal any societal, ecological, environmental condition of the landscape? And therefore can it be an indicator of change? How do you automate the assessment of the spatial structure and behavior of change to discover critical areas, hot spots, and their corridors? Is the map accurate? How accurate is it? How do you assess the accuracy of the map? How do we evaluate a temporal change map for change detection? What are the implications of the kind and amount of change and accuracy on what matters, whether climate change, carbon emission, water resources, urban sprawl, biodiversity, indicator species, human health, or early warning? And with what confidence? The proposed research is expected to find answers to these questions and a few more that involve multi-categorical raster maps based on remote sensing and other geospatial data. The specific issues are deep, but concrete and finite.

This project builds on the substantial research performed by the participating investigators in a wide range of complementary areas to achieve three major goals. The first is to advance the state of the art in statistical and computational techniques for analyzing and mining raster data. In particular, we will conduct research on (i) raster map modeling and simulation using a variety of Markov random field (MRF) models; (ii) surface topology and echelon analysis; (iii) multicriteria decision support; and (iv) advanced computational techniques in segmentation, classification, and geospatial data mining. The second major goal is to integrate these techniques into a Multiscale Advanced Raster MAP (MARMAP) system coupled with advanced visualization tools. User need assessment and the introduction of novel information visualization techniques will play a major role in the development of this system. The third major goal is to conduct large-scale studies in a number of ecological, environmental, and public health domains. While these studies are currently underway and are funded separately, we believe that the new methodologies and computational techniques will enable major advances and new insights at a much faster rate. We intend to develop several monographs that describe the state-of-the-art of the research covered by the case studies, emphasizing the role of the methodologies and computational tools developed under this project.

Proposed Research
Development of Fundamental Methodologies and Computational Techniques

1. Modeling and Simulation of Thematic Raster Maps

1.1 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 $Z(t)$.

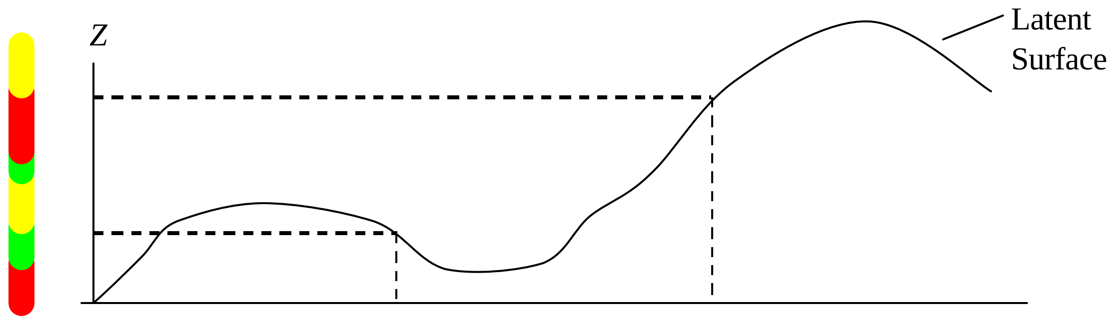

Figure1. *Elevation of the latent surface is categorized according to the transitionogram on the left of the Z-axis.*

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}, $i, j = 1, \cdots, k$. 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).

1.2 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 $a_1, a_2, \ldots. a_k$. For empirical description of the spatial dependence at varying distances in the map, we employ a series $\hat{R}_0, \hat{R}_1, \hat{R}_2, \ldots$ 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, $\hat{R}_0, \hat{R}_1, \hat{R}_2, \ldots$, 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 $M_0, M_1, \ldots, M_L$ 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, $G^{[n,n+1]} = \lfloor G_{ij}^{[n,n+1]} \rfloor$, $i, j = 1,\ldots,k$. 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 ith row of $G^{[n,n+1]}$.

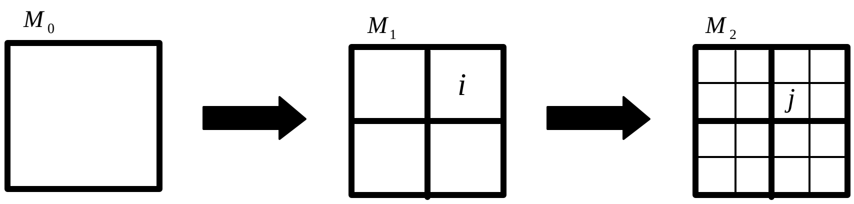

Figure 2. Nested hierarchy of pixels. Each pixel of M_n subdivides into four subpixels in M_{n+1}.

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 accomplishedby 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.

1.3 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 Ω 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 Ω. 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).

2. 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.

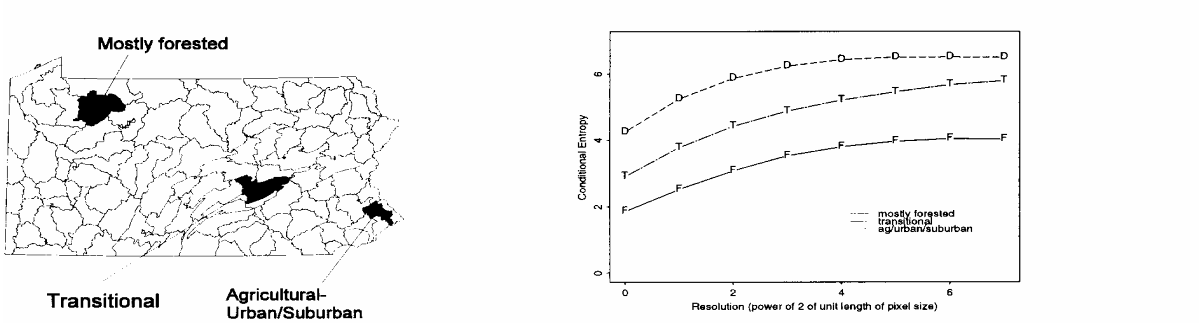

Figure 3. *Fragmentation profiles for three Pennsylvania watersheds with distinct landcover patterns: mostly forested, transitional and mostly deforested (ag/urban/suburban).*

- **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.

3. 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).

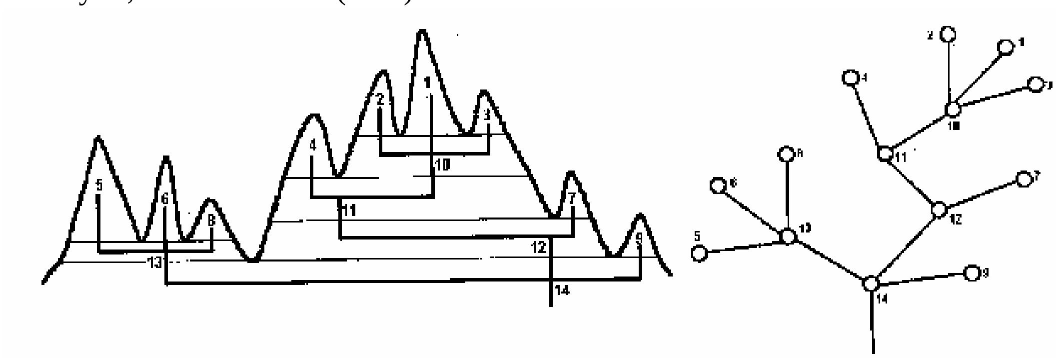

Figure 4. Echelon decomposition of a surface and associated echelon tree.

*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.

4. 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 Ω 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 Ω assign rank 1 (best) to element $a \in S$? Rank 2? etc.
3) If rankings are chosen at random (with equal probability) from Ω, 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 Ω 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 Ω, 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.

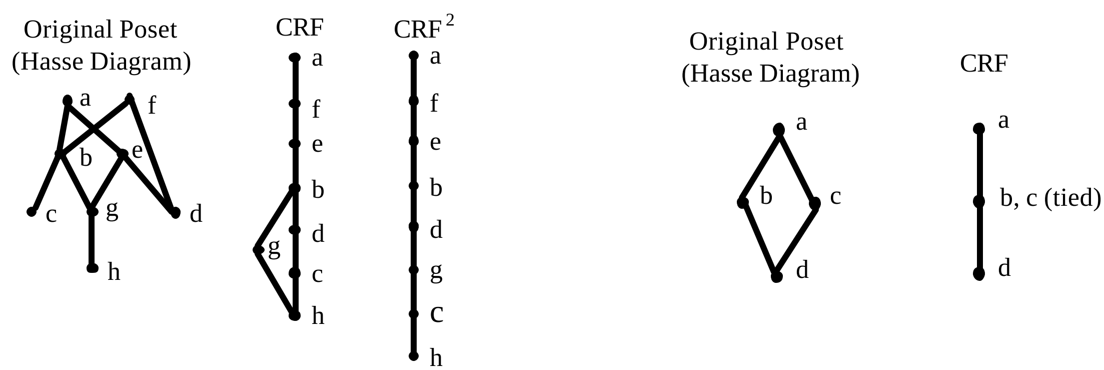

Figure 5*. The three diagrams on the left show the linearizing effect of the CRF operator. The two diagrams on the right show how ties can emerge during linearization. A* ***poset*** *is a partially ordered set.*

5. 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.

6. 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.

7. 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.

8. 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.

Case Studies

9. 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.

10. 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.

11. 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)

12. 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.

13. Prospectus Outcomes and Deliverables

Year 1:
1. User need assessment and preliminary interface design for MARMAP coupled with some tools for DIG/HMTM/MRF modeling and simulation, echelon analysis, and multi-criteria decision support.
2. Set up a remotely accessible Linux cluster environment including 10 TB of disks, dedicated to supporting the MARMAP system and the large case studies.
3. Train research teams for case studies in concepts and methods of analysis at the core of the MARMAP system. Initiate work on various case studies.
4. Develop specific families and subfamilies of parametric models for the DIG/HMTM/MRF models and develop strategies for efficient extraction of upper level echelons.

Year 2:
1. Develop and implement model-fitting and unconditional simulation algorithms for DIG/HMTM/MRF models.
2. Development of visualization tools for the echelon analysis and the multi-criteria decision-making techniques.
3. Develop high performance algorithms and software for region analysis based on value range search, PSI mosaic generation, and hybrid adaptive classification.
4. Develop MCMC simulation algorithms for the uniform distribution on the set of linear extensions of a poset.
5. Conduct research on the various case studies using the MARMAP system and provide feedback on the utility of various modeling, simulation, analysis, and visualization tools. Initiate the writing of the four monographs.

Year 3:

1. Refine the various tools and include tools for simulation using the DIG/HMTM/MRF models, with emphasis on MCMC methods.
2. Develop conditional simulation algorithms for the DIG/HMTM/MRF models and MCMC tools for simulating weighted (non-uniform) distributions on the collection of linear extensions of a poset and corresponding weighted CRF operators.
3. Develop fuzzy analysis tools to handle soft data in the multi-criteria setting.
4. Extend the value range search techniques to handle non-integer data (gaussian/lognormal responses).
5. Develop algorithms to determine association rules and trends with emphasis on land cover land use study.
6. Include progress on various case studies into the four monographs.

Year 4:
1. Assess and compare different indicators of change for change detection analysis.
2. Develop multi-criterion echelon analysis of Hasse Diagrams and posets.
3. Integrate methods for conditional simulation for various models, determination of association rules and trends, and advanced visualization tools into the MARMAP system.
4. Complete the monograph that focuses on landscape patterns, change detection, and accuracy assessment, illustrating the techniques used to derive various results.

Year 5:
1. Refine MARMAP system based on feedback from case studies and integrate tools developed in Year 4.
2. Use the system to complete work on case studies.
3. Validational "Show and Tell" Workshops in Washington DC, including wide publicity of the MARMAP system to various related communities.
4. Complete monographs on the three remaining case studies: Geographic Surveillance and Disease Mapping; Urban Heat Islands and Urban Sprawl; Multiple Indicators and Decision Support.

14. Research Training and Education

An essential part of this prospectus is to introduce the concepts and the methods at the core of the MARMAP system to researchers in the ecology, environment, and public health communities. This effort will be greatly facilitated by the inclusion of prominent researchers from these communities on our team and the availability of the MARMAP system that will provide an easy way to apply these methods to various domains. In addition, there is significant interest from various potential partners we contacted for us to conduct training courses on a regular basis, covering the main techniques behind the MARMAP system and its use. We believe several interested parties will provide funding for these training courses, to be offered in addition to those given during the annual summer workshop.

In graduate education, the investigators will make serious efforts to integrate some of the techniques and methods into the wide range of related graduate courses offered through the four universities. In particular, we believe that the students in these interdisciplinary courses will be excited to use the MARMAP system and experiment with the innovative computing paradigms. Also, the graduate students supported will be required to contribute to the tutorials offered during each summer workshop, in addition to presenting their research progress.

As stated under prospectus outcomes, we intend to develop four monographs covering the techniques and tools and their particular applications in each of the four major case studies. The material for these monographs will be developed over the years, and tested incrementally through various graduate and training courses offered by the investigators. The graduate students will be heavily involved in this process.

15. Management Plan

The overall management of the project will be the responsibility of the PI, Dr. Joseph JaJa. He brings a substantial administrative experience in managing research centers and large projects, including his current position as the Director of UMIACS whose research programs are funded at the level of $15M-$20M a year. He has been involved with large partnerships, including his current activities in the Earth Science Information Partnership and the National Partnership for Advanced Computational Infrastructure. The management team of the project will consist of the PI and the co-PIs who will set short-term and long-term directions and goals, implementation plans, assess project progress, and establish collaborative mechanisms among the participating investigators. The goals stated in the section on Project Outcomes and Deliverables will serve as the initial general guidelines for the management team. Moreover, each member of the management team will lead the coordination of the research efforts in a thrust area as follows:

- Computational Tools and System Integration – Dr. JaJa.
- Statistical Modelin g and Techniques – Dr. Patil.
- User Need Assessment and Visualization Tools – Dr. Shneiderman.
- Ecological and Environmental Applications – Dr. Biging.
- Public Health Applications – Dr. Balbus.

The management team will be in constant communication through email and phone calls, and will meet twice a year in conjunction with the two annual workshops planned for the project. These meetings will focus on assessing progress, adjusting goals and directions as appropriate, and setting new goals.

The two planned workshops, one around the middle of the academic year and the other during the summer, are expected to be an integral part of the project, which will allow each investigator to describe his/her work and how it relates to the overall goals of the project, and to help the investigators to better integrate their expertise to help evolve the multiscale advanced raster map analysis system with home base development case studies, national and international case studies, and validation federal case studies. Beginin g the second year of the project, the summer workshop will include special tutorial sessions covering the MARMAP system and the methodologies behind it. Researchers from nearby local organizations will be invited to attend these tutorials. Typically, the summer workshop will be held around Washington DC containing many government and industrial labs with strong interest in the main theme of this project.

Our project will have a Theory and Practice Advisory Council (TAPAC), consisting of members who are well-known for project-related expertise and insight for administrative guidance. The TAPAC will ensure that science and user interest are represented throughout the project. It will also be a strong source of relevant domain expertise. Section I provides the enthusiastic acceptance letters from many of the anticipated members of TAPAC, separated into a Science Advisory Board and a User Advisory Board. The TAPAC is expected to meet once a year during the summer workshop and will be asked to provide a report to the management team evaluating the progress made on the project and suggesting ways to adjust directions as they see fit. The management team will take their evaluation very seriously and decide on future plans as appropriate.
Finally, the concept of mobility and interactive visits will be fully explored and implemented between participating faculty, graduate students, and postdocs across the four universities. Each university group

will have weekly miniseminar(s) on relevant themes of the project involving local faculty, graduate students, and post docs. These will be carefully strengthened from time to time with visiting collaborators from participating institutions to keep the individual and collective momentum and synergy in progressive development. Every effort will be made to iteratively accomplish the upward spiral of horizontal and vertical research and training integration.

16. International Collaboration

This project will have an impressive international collaboration with scientific leaders responsible for major raster map and geospatial analysis programs in their countries such as Canada, China, Costa Rica, Italy, Netherlands, Spain, and Sweden. The international synergies and benefits to be gained from these collaborations include the considerable enrichment of our case studies that will strengthen the empirical and validation aspects of our project. The Supplement to this proposal (Section I) contains a good number of support letters from potential international collaborators. Here we illustrate two current international collaborations.

Biging and Gong jointly have two international projects. In the first project, they are working with researchers in the Argentinean Space Agency (CONAE) and The Forest Division of the Argentinean Secretariat of Agriculture, Livestock, Fisheries and Food to use the EO-1 satellite to classify and monitor development of forest plantations in the Patagonia region. In the second study, they are working with researchers from the Complutense University in Madrid and from the European Commission-Joint Research Center to utilize fuzzy mathematics to develop improved classification algorithms for remotely sensed data. In addition, Gong has several international projects in China and Japan. In China, he directs the International Institute for Earth System Science (ESSI) at Nanjing University where a number of global change projects related to key environmental issues in China are ongoing. He collaborates with Sichuan Institute of Parasitic Diseases on schistosomiasis control with the use of GIS and remote sensing. He also collaborates with Beijing Normal University and Chinese Academy of Sciences on land cover and land use change and carbon cycle modeling efforts. He also works with the researchers at Chinese University of Hong Kong and National Chiao-Tong University at Taiwan on hyperspectral remote sensing. He collaborates with PASCO, Japan on the use of GIS in public health.

17. References Cited

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.

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.

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.

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.

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

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.

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.

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

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.

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

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.

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.

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

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

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., 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.

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. *International 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.

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

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.

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.

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.

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.

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., 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 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. 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.

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.

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

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

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.

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)

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.

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

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

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.

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.

18. Project Investigators and Their Affiliations

Principal Investigator

Joseph JaJa, University of Maryland Institute for Advanced Computer Studies (UMIACS)

Co-Principal Investigators

Ben Shneiderman, Department of Computer Science and UMIACS, University of Maryland
G. P. Patil, Center for Statistical Ecology and Environmental Statistics, PSU
John Balbus, Center for Risk Science and Public Health, GW
Gregory Biging, Department of Biometrics, Berkeley

Co-Investigators

University of Maryland

Abram Kagan, Department of Mathematics
Ben Kedem, Department of Mathematics
Eric Slud, Department of Mathematics
John Townshend, Department of Geography and UMIACS

Penn State University

Naomi Altman, Department of Statistics
Jaeyong Lee, Department of Statistics
Jia Li, Department of Statistics
Wayne Myers, Forest Biometrics and Co-Director of the Office for Remote Sensing of Earth Resources
Steve Rathbun, Department of Statistics
Charles Taillie, Department of Statistics

George Washington University

Sudip Bose, Department of Statistics
Ivan Cheung, Department of Geography
George Davis, Department of Microbiology and Tropical Medicine
Douglas Fuller, Department of Geography
Paul Levine, Department of Epidemiology/Biostatistics
Hosam Mahmoud, Department of Statistics
Reza Modarres, Department of Statistics
William Roper, Department of Civil and Environmental Engineering
Prem Sarin, Environmental and Occupational Health
Curtis Tatsuoka, Department of Statistics
Thomas Wilkie, Department of Microbiology and Tropical Medicine

University of California, Berkeley

Peng Gong, Department of Biometrics

19. Theory and Practice Advisory Council

The Theory and Practice Advisory Council (TAPAC) will consist of the Science Advisory Board and User Advisory Board. Almost all the members have agreed to participate on these boards – see letters of support under item 4.

Science Advisory Board:

1. Rich Birdsey, USFS
2. Larry Cox, NCHS
3. Michael Goodchild, UCSB
4. Peter Guttorp, U Washington
5. Glen D. Johnson, NYS
6. Menas Kafatos, GMU
7. Thomas Mace, NASA
8. Kanti Mardia, UK
9. Reagan Moore, SDSC
10. Bernard C. Patten, U Georgia
11. Phil Ross, EPA OEI
12. Paul Switzer, Stanford
13. Darrel Williams, NASA

> The proposed work relies on spatial data analysis and change detection methodologies that can be applied at many different spatial scales. We need such a raster map information system. Research highlighted in the proposal should significantly advance our capability to conduct the national and regional assessments. We are excited about the potential to work with you on resolving these key technical issues using the information technology at its frontier.

> We are very excited about your proposal. Your activities hold great promise. The initiative is extremely timely and will have a great impact on the future of the subject. Raster maps will be increasingly important in many areas in the coming years.

User Advisory Board:

1. Carolyn Adams, NRCS
2. George Hanuschak, NASS
3. Bruce Jones, EPA ORD
4. John Kelmelis, USGS
5. Gregory T. Koeln, EARTHSAT
6. Thomas Loveland, USGS
7. Kamlesh Lulla, NASA
8. Nancy Maynard, NASA
9. Dale Quattrocchi, NASA
10. Ashbindu Singh, UNEP
11. Paul G. Szerszen, SAIC
12. K. Thirumalai, DOT

> We look forward to working with this distinguished group on this well-conceived proposal that will build decision-making analytical tools using information technology adapted to larger geographic study areas and make use of digital data from multiple sources. We do need a multiscale advanced raster map information system in place as soon as possible.

20. International Collaborators

1. Enrico Feoli, Italy
2. Gerard Heuvelink, Netherlands
3. Koji Kurihara, Japan

4. Ajith Perera, Canada
5. Bo Ranneby, Sweden
6. Sonia Rodriguez, Costa Rica
7. Orazio Rossi, Italy
8. Alfred Stein, Netherlands