COMMUNITY ECOLOGY 4(2): 205-217, 2003
1585-8553/$20.00 © Akadémiai Kiadó, Budapest

Synoptic environmental indicators as image analogs for landscape analysis

W. L. Myers[1], G. P. Patil[2,4] C. Taillie[2], and D. C. Walrath[3]

[1] *School of Forest Resources and Penn State Institutes of Environment, The Pennsylvania State University, University Park, PA 16802, USA. Email: wlm@psu.edu*
[2] *Center for Statistical Ecology and Environmental Statistics, Department of Statistics, The Pennsylvania State University, University Park, PA 16802, USA.*
[3] *Penn State Institutes of Environment, The Pennsylvania State University, University Park, PA 16802, USA. Email: dwalrath@psu.edu*
[4] *Corresponding author. Email: gpp@stat.psu.edu*

Keywords: Change detection, Clustering algorithms, Environmental indicators, Geographic information systems, Image analysis, Landscape ecology, Remote sensing, Species habitat, Species richness.

Abstract: Spatially synoptic multivariate image data implicitly embody information on landscape pattern, for which analytical techniques of explicit pattern extraction are evolving. In parallel, a multiplicity of 'environmental indicators' is being generated in the arena of geographic information systems. Landscape ecological analysis offers substantial opportunity for configuring these indicators synoptically as cells over spatial extents and for stacking them into complementary sets of *image-structured multiple environmental indicators* whereby the values of the indicators become intensity analogs of brightness for spectral bands. As environmental signal analogs of multiband images, these data become available to image portrayal in both graytone and quasi-color renditions to reveal joint properties of pattern for visual interpretation. Likewise, many of the conventional image analysis operations can be conceived more broadly to allow their application in the indicator context. This includes combinatorial approaches such as calculation of an NDVI equivalent from indicator intensities. Similarly, supervised and unsupervised analyses can have meaningful application in the context of multiple environmental indicators. Furthermore, newer techniques of pattern-based image segmentation can also be applied. Application to habitat modeling for vertebrates from Gap Analysis shows the effectiveness of the approach.

Abbreviations: DE–Digital Elevation Model, GIS–Geographic Information System, NDVI–Normalized Difference Vegetation Index, PHASE–Palette Homogeneity Among Segmentation Elements, PSI–Progressively Segmenting Images, RHII–Regional Habitat Importance Index.

Introduction

Landscape ecological research using geographic information systems (GIS) has become a prolific source of quantitatively synoptic environmental indicators showing localized ecological variation over substantial regional extents. Biodiversity studies in Pennsylvania are illustrative in addressing species richness and habitat importance by guilds and other taxonomic or life history groupings. By *synoptic* we mean information that provides spatially comprehensive coverage of a landscape at some uniform degree of resolution. Synoptic information is typically represented using grid cells in which each cell carries a value for the indicator. Thus, data are sampled, aggregated, or interpolated at the cell level.

Environmental indicators frequently come in complementary sets, whereby the interpretation of a value for one indicator is conditioned by the value(s) of another indicator or indicators. A simple example is species richness for wetland species and upland species. Since wetland environments tend to occur more frequently in lowland settings, a low value for wetland species is often counteracted by a higher value for upland species. However, a landscape setting that is low for both can be considered as biotically impoverished. Therefore, using a single such indicator in isolation is not fully informative.

Although GIS is typically used for investigation of synoptic environmental indicator data, there is less than full compatibility of data characteristics with analytical capabilities. GIS are primarily geared to handle two major types of geospatial information. One type consists of a categorical map, possibly having companion tables of quantitative attributes for the categories. The other is level/gradient information on variables that can be considered as actual or virtual topographies in digital elevation models (DEM). Cellular values of quantitative environmental indicators do not fit the categorical mold directly. They also often fail to fit the gradient mold by virtue of having frequent abrupt changes at the edges of patches and corridors. This gives some new relevance to the classic issues of community versus continuum in ecology. If all environmental variations were in the nature of gradients, then patterns of landscape variation would be expressed as gradual transitions (ecotones) having an absence of definite edges. In this case, the methods of geostatistics (Isaaks and Srivastava 1989) would provide an obvious avenue for handling quantitative environmental indicators. However, mosaics of patch patterns are typical of many landscapes (Forman 1995, Forman and Godron 1986, McGarigal and Marks 1995). Both disturbance and abrupt changes in geological substrate, soil, slope, aspect, drainage and other environmental factors contribute to the patchiness and edginess of the landscape pattern. Regardless of cause, such discontinuities impose strong limitations on interpolative methods.

It has been observed (Walrath 2000) that patterns of variation in quantitatively synoptic environmental indicators have much in common with the spatial variability of spectral signals obtained in multiband remote sensing, exhibiting interplay between different scales of variation at both localized and regionalized levels. There are further similarities in the conditionality of interpretation. Like upland and wetland biota mentioned earlier, visible and infrared signals must be interpreted jointly to deduce the nature of land cover. Vegetation exhibits the particular combination of low intensity (absorbed) red light and high intensity (reflected) near-infrared radiation. Water and moist surfaces absorb both types of radiant energy. Dry bare mineral soil tends to be strongly reflective relative to both types of radiant energy (Frohn 1998, Gibson and Power 2000, Jensen 2000, Wilkie and Finn 1996). This further suggests the use of systematic coloration for display and interpretation of quantitatively synoptic information on multiple environmental indicators.

Recent research on synergistic use of remote sensing and GIS technologies for purposes of landscape change detection (Myers 2003, Myers, Patil and Taillie 2003) offers even greater potential for incorporating modified image analysis approaches into work with quantitatively synoptic multiple environmental indicators. Pattern-based compressive image segmentation is particularly interesting with regard to extraction of joint variation among multiple indicators without imposing the linearity constraints associated with commonly used multivariate methods such as principal components. Additionally, it induces full compatibility with categorical/tabular data structures of GIS.

Pennsylvania physiographic context

Since all of our examples are concerned with habitat relations of biota and landscapes, a brief overview of the Pennsylvania physiographic context can aid comprehension of the patterns that emerge. Figure 1 is a shaded relief map derived from a digital elevation model (DEM) showing the general topography and physiography of Pennsylvania. A DEM is a rectangular grid of cells giving elevation at the center of each cell. A shaded relief map is obtained as a mathematical model of shading from a virtual sun positioned at a specified elevation and direction. Hill-shading from a DEM is a commonly available function in a GIS. Other important terrain information such as slope and aspect can also be calculated readily from a DEM via GIS. North is at the top of the figure, but there is the appearance of a slight skew that is an artifact of the

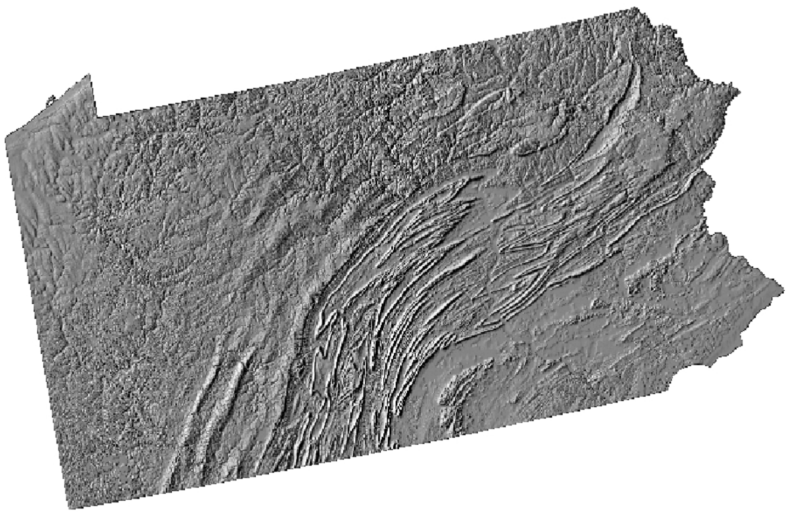

Figure 1. Shaded relief map showing the general physiography of Pennsylvania.

Lambert conformal conic map projection for the conterminous United States.

The eastern boundary of Pennsylvania is irregular as it follows the Delaware River. The only other side having a natural irregularity is the northwestern corner that juts up to the shore of Lake Erie. The balance of the northern edge borders New York. The western edge borders Ohio. The Piedmont Plateau comprises the southeastern section of Pennsylvania. This is a generally low-lying area consisting of geologic materials that have weathered to soils that are quite fertile. The relatively gentle topography coupled with fertile soils and positioning along Chesapeake Bay have resulted in extensive human development for urban and agriculture with consequent deforestation (Miller 1995, Myers 2000).

The Piedmont is bordered to the northwest by a very long and broad valley called the Great Valley. South Mountain is a major topographic feature that is interposed between the Piedmont and the Great Valley along the southern edge of the state. The Great Valley marks the beginning of the strongly folded Ridge and Valley Region that wraps around South Mountain and the Piedmont from the south-central area of the state to the northeast, terminating in the long finger of the heavily urbanized Wyoming Valley. The ridges of this region comprise the Appalachian Mountains, which is a somewhat confusing name since the Appalachian Plateau lies farther to the west and north encompassing the Allegheny Mountains. The ridge portions of the Ridge and Valley are heavily forested, but the valleys are mostly agricultural with some urban development.

The largest physiographic region is the Appalachian Plateau occupying the western and northern parts of the state, and rather sharply separated from the Ridge and Valley by the steep Allegheny Front. This is a very thick and old geologic area comprised predominantly of horizontally oriented sandstone that weathers to rather thin and infertile soils. The topography in this region is due primarily to continued erosion over geologic ages, with some areas having more resistant and less fractured rock. The Allegheny Mountains and Laurel Highlands extend northward as broad ridge-like structures from the southern edge of the state. The portion of the Appalachian Plateau occupying the western part of the state is considerably lower than the remainder that occupies the more central and northern part of the state. A band on the Plateau across the northern portion of the state was subject to glaciation and has its topography modified accordingly. The central and northern portions of the state in the high plateau regions have resisted development and reverted largely to forest after repeated clearings. The western low plateau area has substantial forests, but they are often quite fragmented.

Major urban areas occur more or less in the corners of the state and along the eastern edge. These include Philadelphia in the southeast, Pittsburgh in the southwest, Erie in the northwest, Scranton/Wilkes-Barre in the Wyoming Valley of the northeast, and Allentown-Bethlehem in the east. The Susquehanna is the major river system of eastern Pennsylvania flowing into the Chesapeake Bay. The Allegheny River system is extensive in western Pennsylvania, joining the Monongahela to form the Ohio River flowing out of the southwest. The northwest corner of the state drains to Lake Erie.

Visualization via combinatorial color

Humans have a tricolor visual perception apparatus, with all other colors being induced by mixed response of red sense, green sense, and blue sense receptors. For example, yellow is perceived when red sense and green sense receptors are stimulated equally with no contribution from blue sense receptors. White or shades of gray are perceived when all three kinds of receptors are stimulated equally. This is the basis for generating many different colors on computer screens using only red, green, and blue (RGB) display elements. Even though it involves the human senses, this perceptual color mixing is mathematical. The mathematics of combinatorial color is extensively exploited for visualization in image analysis, and is subject to similar exploitation in working with multiple environmental indicators.

Red, green and blue are the *primary* colors, and a multi-color composite image can be produced in two ways. One way is direct or multiband mode, in which a different signal is used to generate each of the primary colors. One spectral signal or environmental indicator is selected for representation as red, another to be represented as green and a third as blue. The signals to be used in this manner must be scaled in a manner that is appropriate for the color display device, which is usually from a minimum of 0 to a maximum of 255. A second way is indirect or pseudo-color map mode whereby a categorical value is used to select a row from an attribute table and quantitative column attributes are selected to generate the color components. One column is used to determine red, a second column for green, and a third column for blue. In this case, the column entries must be appropriately scaled for the color display device.

Walrath (2000) used composite color with GIS for exploratory multiscale analysis of avian distributions in Pennsylvania based on a breeding bird atlas comprising

190 species compiled over 4928 blocks with each block encompassing 3.75 minutes of longitude and 2.5 minutes of latitude (Brauning 1992). Bird community composition was studied in terms of seven habitat guilds: edge (EDG), deciduous mixed forest (DMF), coniferous forest (COF), herbaceous-cultivated (HEC), herbaceous-grassland (HEG), urban-developed (UDL), and water/submerged-land (WSL). Red color was determined by number of species in the two herbaceous guilds, green color was determined by number of species in two forest guilds, and blue color was determined by number of species in the wetlands guild. Greenish tones thus indicated dominance of forest dwelling bird communities, reddish tones indicated grassland-agricultural communities, bluish tones indicated wetland communities, and yellowish tones indicated mixed upland communities. The coloration was accomplished by a sophisticated scheme of GIS shade-set palettes. Polygons were delineated around areas of uniform coloration, and the polygons were then overlaid on land cover maps for verification of consistency between habitat structure and community composition. The change in coloration was studied as blocks were successively aggregated to represent broader landscape scales.

Gap analysis with 'biobands'

The success of Walrath's exploratory work encouraged further development of the approach to analyze comparative patterns of habitat richness and rarity for vertebrates in the state of Pennsylvania through the process of Gap Analysis. Gap Analysis is a nationwide program of biodiversity assessment in the United States entailing a coarse scale geographic approach to conservation that relies heavily on computer-based geographic information systems and related information technologies (Davis et al. 1990, Olivieri and Backus 1992, Scott et al. 1993). Gap Analysis uses land-cover maps derived from remote sensing in combination with various layers of map information and spatially specific databases along with knowledge-based models of biological characteristics for each species to arrive at mappings of potential habitat. Conservation stewardship status of habitat is then considered according to land ownership and mode of management. The overall objective of Gap Analysis is to identify 'gaps' in the conservation 'safety net' for sustaining biological diversity. Attention is focused on vertebrates under the assumption that their habitat needs will also serve to a large extent as surrogates for those of other organisms.

The Pennsylvania investigation used the habitat mappings for species to conduct comparative analysis among six taxonomic groups on a grid of one-kilometer cells (Myers et al. 2000). The taxonomic groups were chosen on the basis of similarity in life history as (1) mammals, (2) birds, (3) amphibians, (4) snakes and lizards, (5) turtles, and (6) fishes. Two measures were used in doing the comparisons: (a) potential species richness according to habitat suitability and (b) a regional habitat importance index (RHII). Species richness was determined as the number of species for which suitable habitat was present in the cell. The RHII takes into account overall habitat scarcity along with scarcity of habitat in conservation areas (Myers et al. 2001) to focus on species that are vulnerable to habitat loss and/or catastrophic events.

The quantitative RHII perspective lends particular emphasis to species that couple overall habitat scarcity with low representation in conservation areas and difficulty of finding habitat outside existing conservation areas by which to enhance the level of protection. The formula for computing RHII for a given species is:

RHII = 100 x proportion of nonhabitat in entire state
x proportion of nonhabitat in conservation areas
x proportion of nonhabitat outside conservation areas.

The first proportion in the RHII formulation expresses overall scarcity of habitat. The second proportion further captures the 'gaps' in the sense of Gap Analysis where habitat for a species is sparse in existing conservation areas. The third proportion expresses lack of further conservation opportunity for a species. An importance index was computed for each cell by adding together the RHII values of those species for which suitable habitat was present in the cell. This index proved quite effective for flagging cells that contain several species at risk according to official lists for Pennsylvania.

Image-structured layers that might be called 'biobands' were then compiled from the species (habitat) richness and RHII mappings by interpolative adjustment of the ranges so that the maximum value became 250 for encoding into a byte. These biobands were then 'stacked up' separately for the two types of measures to obtain a pair of six-band image-structured, multiple indicator datasets.

Figure 2 shows the bioband of mammalian species richness for Pennsylvania. Darker tones depict areas of higher species richness. Many of Pennsylvania's mammals are very widely distributed, giving some degree of gray tone to most parts of the state. Several of the species, however, are largely restricted to the more heavily forested regions of the mountains and high plateau. Thus, these areas appear in darker tones.

Figure 3 shows the corresponding layer for fish species. More species of fishes occur in the larger rivers. Larger rivers naturally flow through the major valley ar-

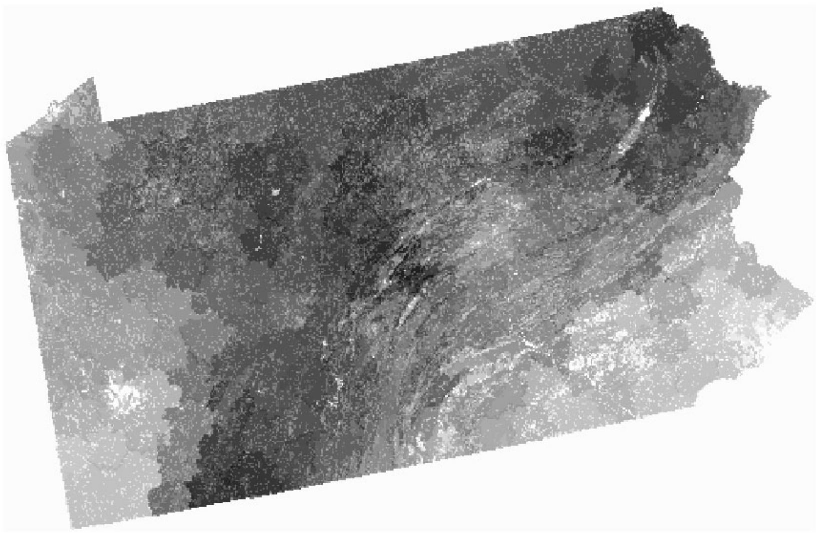

Figure 2. Image-structured layer showing distribution of mammalian species richness potential in Pennsylvania according to suitable habitat, with darker tones indicating habitat for greater numbers of species.

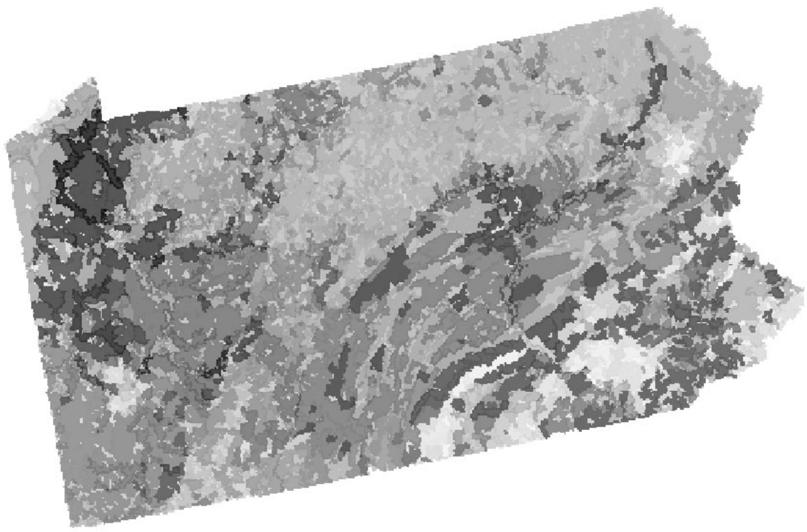

Figure 3. Image-structured layer showing distribution of fish species richness potential in Pennsylvania according to suitable habitat, with darker tones indicating habitat for greater numbers of species.

eas. The fish species that inhabit the smaller streams in the higher elevations are a relatively small and characteristic group (Argent et al. 2003). Even this small group may be absent from streams that are degraded by factors such as disturbance and acidification. The fish species richness bioband makes it evident that northwestern Pennsylvania is a region of particular habitat diversity. By comparison, the larger valleys in the southeast are relatively impoverished due to the effects of intensive agricultural and urban development over long periods.

Figure 4 shows the distribution of mammalian habitat in terms of the RHII regional habitat importance index, and Figure 5 shows the RHII view for fish. Since many of the mammalian species are habitat generalists with low RHII, the pattern is responsive mostly to a few species having limited ranges and/or re-introduction programs. The large hexagonal cells used in delimiting ranges also influence the pattern. Approximately one-third of fish species in Pennsylvania are considered threatened or endangered, so a considerable number of species find expression in the RHII pattern that is controlled more by watersheds than by range hexagons. In both cases, RHII exhibits a stronger inter-regional contrast than species richness.

Pattern extraction by PSI/PHASE segmentation using proxy processes

The primary purpose of casting synoptic data on multiple environmental indicators in the manner of multiband image information is to facilitate extracting patterns of

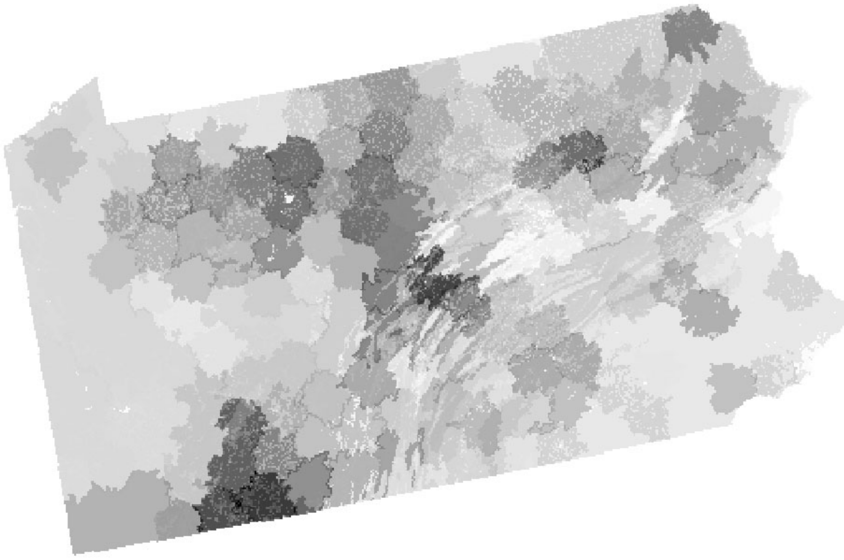

Figure 4. Image-structured layer showing distribution of mammalian species habitat importance in Pennsylvania according to RHII, with darker tones indicating more important habitats.

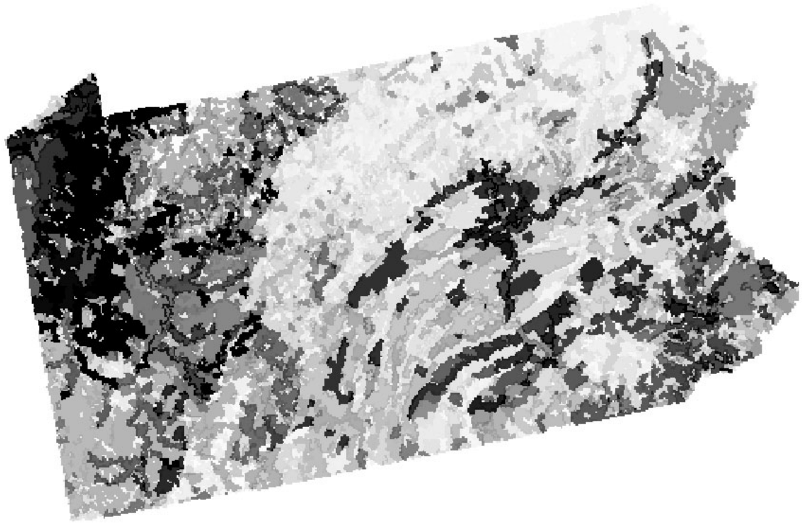

Figure 5. Image-structured layer showing distribution of fish species habitat importance in Pennsylvania according to RHII, with darker tones indicating more important habitats.

joint variation among the indicators. Each of the indicators is analogous to a band of spectral information, and the cells or partitions are analogous to pixels of image data. An effective means of extracting patterns of joint variation is to locate segments of the image over which the set of band (indicator) values is similar.

This context can be conceived as a multidimensional response space with each indicator being an axis in which the set of indicator values for a cell determines the endpoint of a vector from the origin. If cells (pixels) have similar response patterns, then the endpoints of their vectors will form a cloud or constellation in this response space. The patterns of occurrence in this space can be generalized by replacing an entire cloud or constellation by a single vector that is centrally located in the cloud. This central vector then becomes a proxy for all other vectors that comprise the cloud. Generalization of the response patterns over the vector space in this manner is referred to here as a *proxy process*. Statistical clustering becomes a proxy process if all members of each cluster are assigned a common representative set of response values, such as the cluster centroid.

Research on landscape change detection and habitat classification has produced a proxy process method for doubly segmenting image-structured data that also serves needs of pattern extraction in working with multiple indicators (Myers 2003, Myers, Patil and Taillie 2003; URL at www.environment.psu.edu/publictions/2003_6/2003_6_web.pdf).

This approach extracts two layers of joint variation from the multiple layers of indicators. The coarse scale layer is compatible with categorical/tabular GIS data structures. The other layer provides finer scale detail regarding joint variation of the original indicators.

A strategic goal of the process is to capture the major differences within the GIS-compatible layer using 250 proxy response vectors. The number 250 is chosen both to suit GIS raster map viewers and to fit within a byte of computer storage. The first step in achieving this is to find 250 dispersed response vectors that serve as initial proxies. This search begins with the first 250 different response vectors, and determination of the closest (Euclidean distance) pair among them. The remaining response vectors (not currently among the 250) are then examined to see if they could replace a member of the closest pair so as to increase the closest distance. After a replacement, the new closest pair is determined and the process continued. Each cell (pixel) is then assigned to one of these initial proxies, thus producing a provisional segmentation.

The 250 provisional segments are then refined in a cyclic process of progressive splitting. The goal of refinement is twofold: to avoid very large segments, and to divide very diverse segments into more uniform parts. The splitting is accomplished by polar partitioning. The shortest response vector in each (sub)segment is the low pole, and the longest response vector is the high pole. Partitioning consists of associating each cell (pixel) with the pole to which it is closest (Euclidean distance). The proxy for a partition is the average of its poles. A criterion for priority of splitting is computed as the product of segment size (number of cells) times the squared (Euclidean) distance between the poles. Thus, the criterion prefers to split segments that are either exceptionally large in size or that exhibit large within-segment contrast. A segment can be split if its criterion value exceeds a threshold equal to 1/16 of the current mean segment size. The user sets the number of splitting cycles. The default setting of nine cycles typically results in 2000–2500 refined segments. These refined segments are primary PSI segments.

The primary PSI segments are then regrouped into 250 final secondary (coarser) segments for the GIS layer. GIS displays use the same color for all primary segments within a secondary segment. Therefore, the secondary segments are called *PHASE segments* as an acronym for Palette Homogeneity Among Segmentation Elements. The regrouping uses a modified single linkage strategy. The modifications address need to preserve contrast, impose a minimum segment size, and limit the number of primaries in a secondary to 255. The latter constraint allows the primaries to be handled as subdivisions of secondaries. The proxy for a secondary is the mean of the proxies for its primaries. The PHASE segments are ordered and numbered according to the lengths of the proxy vectors. A map of the PHASE segments keyed to segment number is prepared in GIS compatible format, and the proxy values are entered into auxiliary attribute tables.

The simplest PHASE display treats the (ordered) segment numbers as being intensity or brightness values in a graytone image. A multitude of color displays can also be prepared in which the hues are determined by the component values of the proxy vectors. A graytone display of the (proxy) image for species richness is shown in Figure 6. A companion display for the RHII habitat importance indices is shown in Figure 7. These figures make it quite evident that the RHII regional habitat importance indexing approach is much more focused on particular portions of the state.

Compound environmental indices

The PHASE tabular attributes are easily combined into compound indexes akin to those commonly used in remote sensing. A commonly used index for remote sensing is the so-called *Normalized Difference Vegetation Index* (NDVI), based upon the fact that red light is strongly absorbed by plants for photosynthesis and near-infrared radiation is strongly reflected. Thus, NDVI = (Infrared – Red)/(Infrared + Red) should be large where healthy vegetation is present, and small in the absence of vegetation. The normalizing denominator helps balance the effects of shading on lighted versus dark sides of ridges. An analogous 'upland index' for the biobands can be formulated by treating mammals as the analog of infrared and fish as the analog of red. Since there are more upland mammal species in Pennsylvania than there are wetland species, higher values of the index (mammals – fish)/(mammals + fish) should indicate upland biota, whereas lower values should indicate wetland biota. The denominator should at least partially compensate for variation in overall richness. Similarly, the sum of mammals, birds, and snakes-lizards should be indicative of terrestrial habitats, whereas the sum of fish, amphibians and turtles should be indicative of wetland habitats. Figure 8 shows a display of the upland index, with lighter areas representing upland biota.

Generally, magnitude of an environmental indicator can be thought of as analogous to brightness of spectral bands, but 'intensity' is probably a more appropriate term than brightness for this context. Figure 9 is an intensity image for total of fish, amphibians and turtles. Likewise, Figure 10 is an intensity image for total of mammals, birds and snakes-lizards. Tricolor combinatorial displays of in-

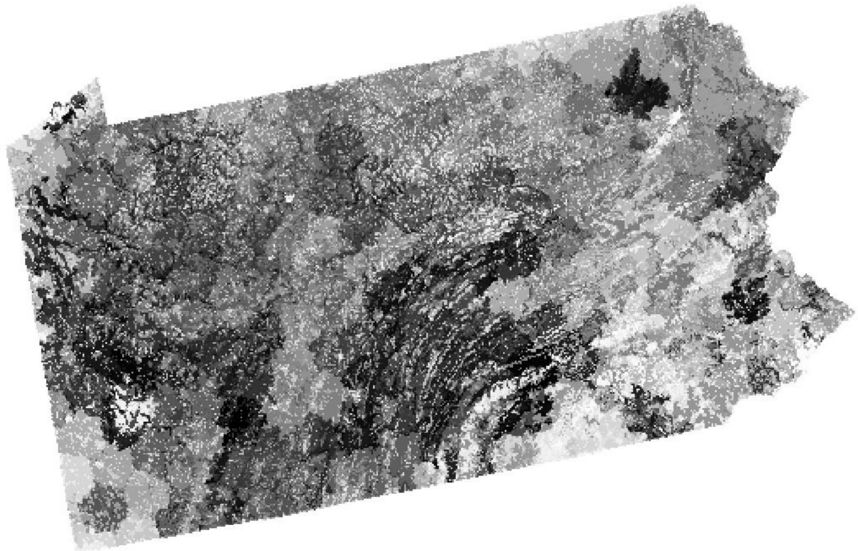

Figure 6. Graytone display of PHASE segments for potential species richness biobands with darker areas indicating habitat for greater numbers of species.

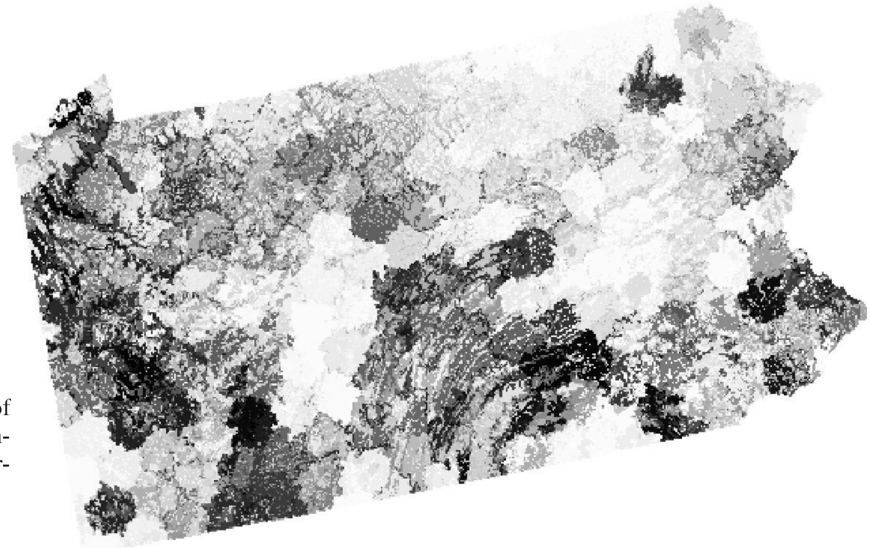

Figure 7. Graytone display of PHASE segments for habitat importance biobands with darker areas indicating higher importance.

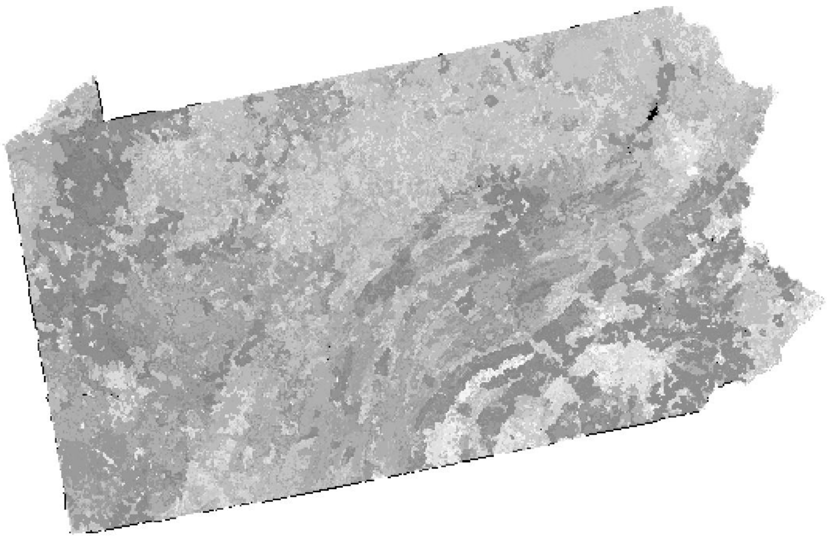

Figure 8. Upland index analog of NDVI with lighter tones indicating upland biota and darker tones indicating wetland biota.

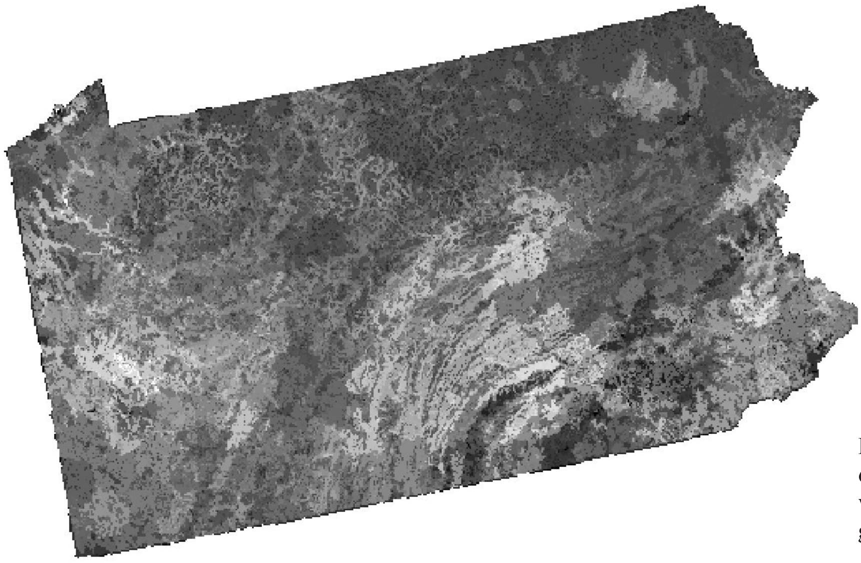

Figure 9. Intensity image for total of fish, amphibians and turtles with lighter areas indicating greater richness potential.

Figure 10. Intensity image for total of mammals, birds and snakes-lizards with lighter areas indicating greater richness potential.

dices are much more effective than simple graytone displays. In this case, a particularly informative rendering was obtained by making the upland index appear as red, with slightly subdued amphibian richness being green, and enhanced turtle richness being blue. A rendering that helps to differentiate various lowland settings is to represent amphibian richness as red, turtle richness as green, and fish richness as blue. An NDVI analog can be worth considering whenever two environmental indices vary alike for certain conditions while varying in opposition for other conditions.

Categorical classification

One of the more important roles for imaging is as a medium for mapping, whereby different portions of landscapes are designated as belonging to one or another of a mutually exclusive set of categories or classes. This is conventionally called thematic mapping in the context of remote sensing. Thematic mapping can be automated in varying degrees, but usually entails a partnership between human analyst and computer algorithms for pattern recognition (James 1985, Tso and Mather 2001). The more automated classification scenarios are conventionally designated as being either supervised or unsupervised.

The 'supervised' scenario for computer-assisted mapping from multiband data has a skilled image analyst acquire so-called 'ground truth' information from maps, documents, and/or field work to locate samples of each category in an image display to serve as 'training sets' from which to extract statistical characterizations of signal patterns that serve as 'signatures' for quantitative pattern matching to assign category designations for other cells in the image.

The 'unsupervised' scenario delays the appeal to ground truth until the later stages of the mapping process. In fact, it is not even necessary to have particular categories in mind at the beginning. Instead, a computational similarity (or dissimilarity) analysis is performed as a 'clustering' operation that segregates the pixels into several groups with the members of any given group having substantial similarity of signal patterns (vectors). A sample of each cluster group is then investigated via ancillary information to determine its composition, and appropriate labels are thereby attached. Thus, the supervised approach starts with a predetermined interpretive legend and some specific knowledge of samples, whereas the unsupervised approach acquires specific information as needed and may develop the interpretive legend in the course of the investigation.

The PSI/PHASE approach incorporates ideas of both the supervised and the unsupervised approaches. The pattern-based segmentation with dual levels of detail incorporates both divisive and agglomerative clustering concepts from statistics. However, the numbers of segments are far more numerous at both levels of detail than is typical of clustering for unsupervised analysis. The coarser level of segmentation with PHASE clusters is akin to what is sometimes called 'hyperclustering' in the context of image analysis, but finer level aggregations of PSI primary segments are much more numerous than hyperclustering would entail. The multiplicity of segments invites application of supervised ideas at these levels, with the proviso that distributional assumptions would need to be modified in order to become formally statistical. Thus, the PSI/PHASE approach can be considered in several respects as both hybridizing and extending more conventional approaches.

PHASE maps having the appearance of images enable a mode of interactive mapping that is not available with actual multiband image data files. This also requires fairly sophisticated GIS software facilities for map displays. This scenario entails overlaying things in a viewer. A PHASE map rendered as a pseudocolor image is brought into the viewer as a base layer. A second copy of the PHASE segment map is then placed on top of the first, and made entirely transparent so the user sees the base layer. A cursor query facility is used to determine what PHASE segment number resides at a location of interest. The legend for the temporarily 'transparent' PHASE map in the top layer is then modified to give this segment a distinctive color that appears superimposed on the image-like map below. The user can thus examine the spatial pattern for that segment and its relationship to other segments. This combination of information is often sufficient to assign the segment to one of the legend categories for which mapping is underway. The top layer can be turned on and off so that the developing thematic map can be seen in relation to the mimicked image in the underlying PHASE map.

Figure 11 illustrates a map that was developed in this manner using the bioband dataset of multiple species richness indicators. The intent for mapping was to determine the occurrence of chronic degradation in upland and lowland habitats of Pennsylvania. Lowlands are recognized as having strong representation of aquatically associated species. This gives rise to four categories considered as: (1) viable lowland habitats, (2) viable upland habitats, (3) degraded lowland habitats, and (4) biotically impoverished. The general paucity of biota does not provide a basis for distinguishing upland from lowland in the fourth category. The third category is distinguished by generalist species in the uplands with some remnants of distinctive lowland species.

Discordance detection as change detection analog

A variety of methods for detecting changes by comparing signal characteristics of companion images collected at different times become available when temporally matched sets of multiple environmental indicators are cast in the form of image-structured datasets (Bruzzone and Prieto 2000, Chen et al. 2003, Coppin and Bauer 1996, Gong 1993, Lunetta and Elvidge 1998, Mas 1999, Myers, Patil and Taillie 1999, Rogan et al. 2003, Singh 1989). The more straightforward of these entail some variation on what has come to be known as *change vector analysis*.

Change vector analysis is an extension of the simple idea of analyzing differences in the respective bands of multiband image data collected at an earlier time and a later time. The respective band differences are treated as components of a multidimensional change vector. The lengths and directions of the change vectors are then analyzed as indicators of change. The change vector approach is readily adapted for use with segmented image data (Myers 2003, Myers, Patil and Taillie 2003).

Segmentation also makes possible a special version of the change vector idea that allows comparison of different sets of multiple indicators to determine their spatial discordance and concordance. This is based on matching of the spatial structure of the segmentation patterns, and then expressing the inconsistencies in terms of one or the other set of indicators. One of the segment maps is chosen as the base of comparison. For purposes of explanation, let this be indicator set A. The other segment map then becomes B. The matching is done at the level of PHASE secondary segments. Each PHASE segment in A occupies a particular set of pixel positions. The same positions in B are scanned to determine which of the B segments occupies the greatest amount of the area covered by this A segment. This dominant B segment thus becomes a B-counterpart of A. Quantitative expression of discordance at a particular position is then obtained by calculating a difference vector between the attributes of the actual B segment and the B-counterpart of that A segment. Figure 12 is a comparison of this nature between the species richness indicators and the RHII habitat importance indicators.

Local anomalies in regional context

Segments constitute zones over which there is substantial consistency of multivariate response, as evidenced by substantial retention of landscape pattern in the foregoing figures despite the fact that all cells in a segment share the same proxy response vector. This is particularly noteworthy in view of the fact that information on location of responses is not considered explicitly in the proxy processes by which the surrogate response vectors are developed. Information on location is retained in the PSI and PHASE proxy processes only for purposes of mapping, and does not directly influence choice of surrogate response vectors. Thus, these proxy processes are of a regional nature.

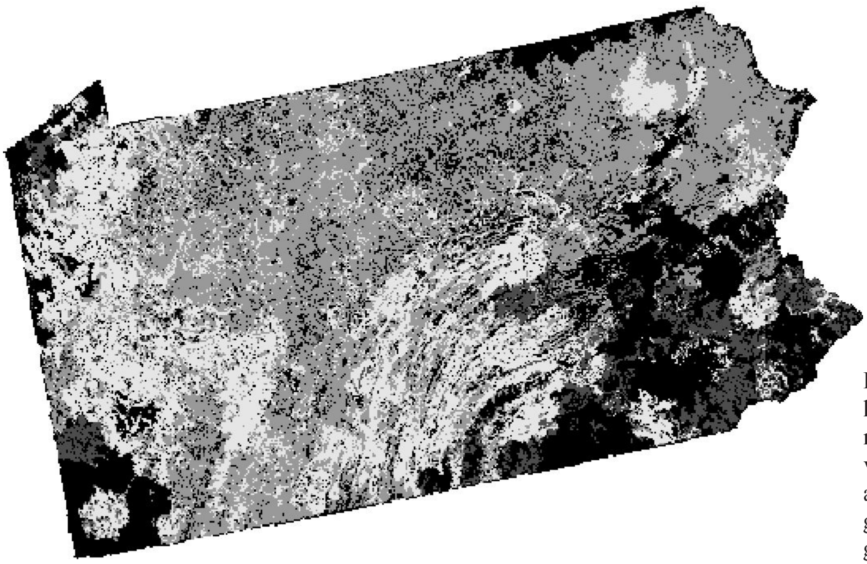

Figure 11. Generalized map of habitat degradation in Pennsylvania from richness of habitats for vertebrate species. Light gray = viable lowland habitats; medium gray = viable upland habitats; dark gray = degraded lowland habitats; black = biotically impoverished.

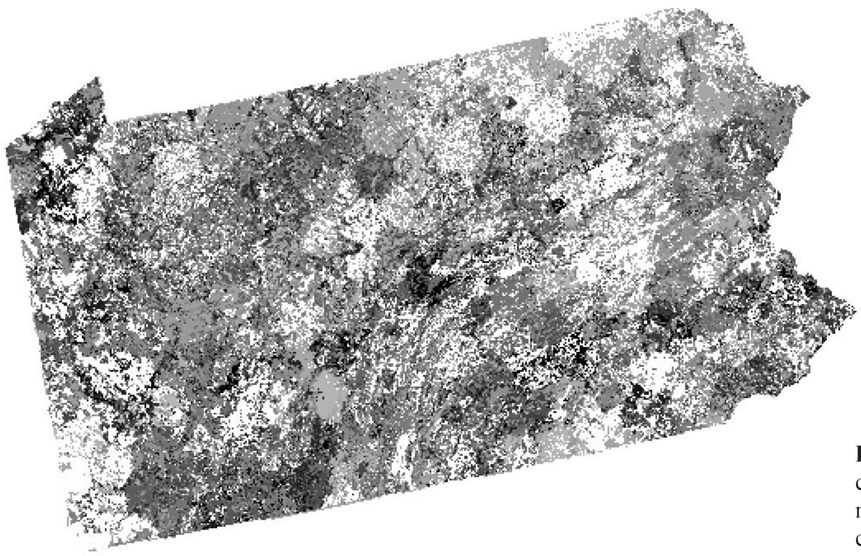

Figure 12. Spatial counterpart comparison of habitat diversity mappings based on potential species richness as opposed to regional habitat importance index (RHII). Darker areas have greater difference in local emphasis.

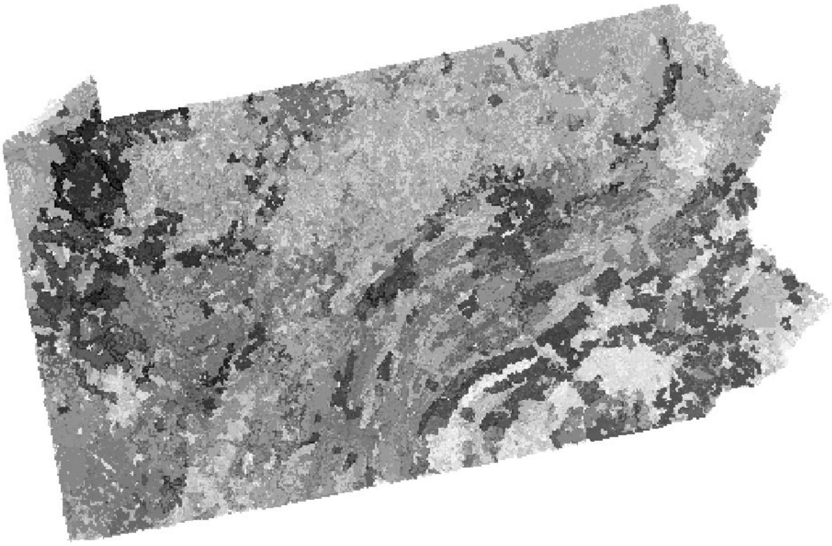

Figure 13. Image layer of primary segment values for fish species, with darker tones indicating greater richness potential.

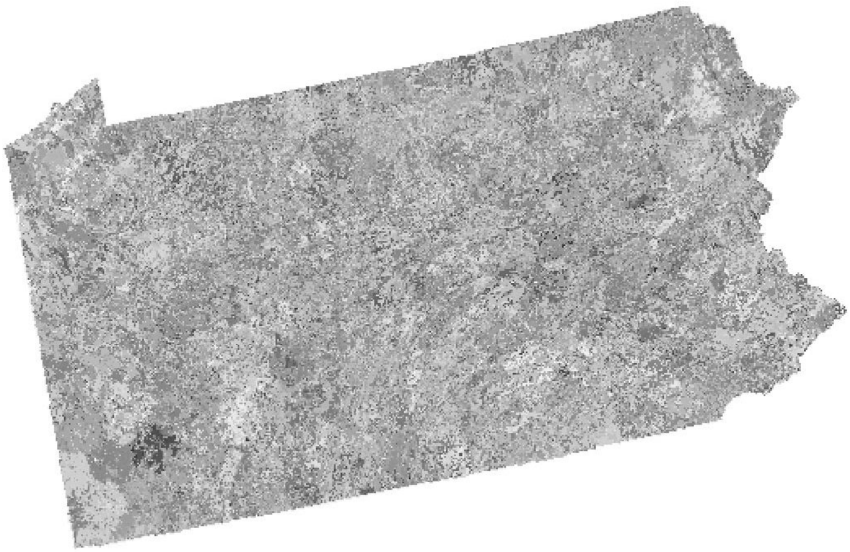

Figure 14. Intensity image showing magnitude of residual local variation for species richness indicators, with darker tones representing larger residuals.

The PHASE GIS layer makes the spatial zones of common properties observable at the coarser secondary level of segmentation. However, the finer primary level of segmentation is not readily handled in this manner. The most convenient way of making the finer segments of common variation observable is to generate an image layer for each indicator comprised of the proxy values for the primary segments. This is illustrated for the fish species richness indicator in Figure 13. Multivariate spatial displays for purposes of visualization at this level of segmentation are most readily rendered using color composite image capabilities of remote sensing software systems rather than GIS. The ideal is to have a software system that combines GIS mapping and multiband color composite image display facilities.

Since the PSI/PHASE proxy processes of pattern extraction examine variability across the entire regional extent, it is possible that they may filter out some coherent local variation in particular landscape settings. Detection of such local anomalies relative to regional context may offer landscape insights and/or investigative opportunities. Detection of special local circumstances may be approached as an image analysis problem of change detection. In this case, however, the 'first' image is the original multilayer dataset of environmental indicators prior to segmentation, whereas the 'second' image is the multiband restoration by proxy values from the segmentation. The differences constitute a sort of residual variation not accounted for by segmentation. This is shown for the sets of species richness indicators in Figure 14.

Local variation of possible interest will take the form of connected groups of cells having darker tones or graded tonal variation, whereas lack thereof is indicated by random tonal variation in the residual image. Local occurrence of distinctive variation in the vicinity of Pittsburgh is apparent in southwestern Pennsylvania, but there appears to be little basis for further generalization.

Acknowledgements: Prepared with partial support from NASA Biospheric Sciences Branch, Goddard Space Flight Center, U.S. EPA Star Grant for Atlantic Slope Consortium Cooperative Agreement, and the NSF Digital Government Program, Division of Experimental and Integrative Activities, Directorate for Computer and Information Science and Engineering. The contents have not been subjected to Agency review and therefore do not necessarily reflect the views of the Agencies and no official endorsement should be inferred.

References

Argent, D., J. Bishop, J. Stauffer, R. Carline and W. Myers. 2003. Predicting freshwater fish distributions using landscape-level variables. *Fisheries Research* 60: 17–32.

Brauning, D. 1992. *Atlas of Breeding Birds in Pennsylvania*. University of Pittsburgh Press, Pittsburgh, PA.

Bruzzone, L. and D. Prieto. 2000. Automatic analysis of the difference image for unsupervised change detection. *IEEE Transactions on Geoscience and Remote Sensing* 38: 1171–1182.

Chen, J., P. Gong, C. He, R. Pu and P. Shi. 2003. Land-use/land-cover change detection using improved change vector analysis. *Photogrammetric Engineering and Remote Sensing* 69: 369–379.

Coppin, P. and M. Bauer. 1996. Digital change detection in forest ecosystems with remote sensing imagery. *Remote Sensing Reviews* 13: 207–234.

Davis, F. W., D. M. Stoms, J. E. Estes, J. Scepan and J. M. Scott. 1990. An information systems approach to biological diversity. *International Journal of Geographic Information Systems* 4: 55–78.

Forman, R. T. T. 1995. *Land Mosaics: The Ecology of Landscapes and Regions*. Cambridge University Press, Cambridge, U.K. 632 p.

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

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

Gibson, P. and C. Power. 2000. *Introductory Remote Sensing: Principles and Practices*. Taylor and Francis, New York.

Gong, P. 1993. Change detection using principal component analysis and fuzzy set theory. *Canadian Journal of Remote Sensing* 19: 22–29.

Isaaks, E. and R. M. Shrivastava. 1989. *An Introduction to Applied Geostatistics*. Oxford Univ. Press, New York.

James, M. 1985. *Classification Algorithms*. John Wiley & Sons, London, UK.

Jensen, J. 2000. *Remote Sensing of the Environment: An Earth Resource Perspective*. Prentice-Hall, Upper Saddle River, NJ.

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

Mas, J. 1999. Monitoring land-cover changes: a comparison of change detection techniques. *International Journal of Remote Sensing* 20: 139–152.

McGarigal, K. and B. Marks. 1995. FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. General Technical Report PNW 351, U.S. Forest Service, Pacific Northwest Research Station.

Miller, E. W., Ed. 1995. *A Geography of Pennsylvania*. The Pennsylvania State University Press, University Park, PA.

Myers, W. 2000. Landscape scale ecological mapping of Pennsylvania forests. Research Rept. ER2002, Environmental Resources Research Institute, The Pennsylvania State University, Univ. Park, PA 16802 USA.

Myers, W. 2003. Doubly segmented images for pattern-based approach to change detection. Final report on NASA Research Project NAG5-1054. Research Report PSIE 2003-6, Penn State Institutes of Environment, The Pennsylvania State University, Univ. Park, PA 16802 USA. 90 pp. + CD-ROM.

Myers, W., J. Bishop, R. Brooks, T. O'Connell, D. Argent, G. Storm, J. Stauffer, Jr. and R. Carline. 2000. The Pennsylvania GAP Analysis Final Report. The Pennsylvania State University, Univ. Park, PA 16802.

Myers, W., J. Bishop, R. Brooks and G. P. Patil. 2001. Composite spatial indexing of regional habitat importance. *Community Ecology* 2: 213–220.

Myers, W., G. P. Patil and C. Taillie. 1999. Conceptualizing pattern analysis of spectral change relative to ecosystem status. *Ecosystem Health* 5: 285–293.

Myers, W., G. P. Patil and C. Taillie. 2003. Doubly segmented proxy images for multi-scale landscape ecology and ecosystem health. *Community Ecology* 4:163-183.

Olivieri, S. T. and E. H. Backus. 1992. Geographic information systems (GIS) Applications in biological conservation. *Biology International* 25: 10–16.

Rogan, J., J. Miller, D. Stow, J. Franklin, L. Levien and C. Fisher. 2003. Land-cover change monitoring with classification trees using Landsat TM and ancillary data. *Photogrammetric Engineering and Remote Sensing* 69: 793–804.

Scott, J. M., F. Davis, B. Custi, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'Erchia, T. C. Edwards, Jr., J. Ulliman and R. G. Wright. 1993. Gap analysis: A geographic approach to protection of biological diversity. *Wildlife Monographs* No. 123.

Singh, A. 1989. Digital change detection techniques using remotely sensed data. *International Journal of Remote Sensing* 10: 989–1003.

Tso, B. and P. Mather. 2001. *Classification Methods for Remotely Sensed Data*. Taylor and Francis, New York.

Walrath, D. 2000. Multiscale analysis of avian distributions in Pennsylvania. Master of Science Thesis in Ecology, The Pennsylvania State University, Univ. Park, PA 16802 USA.

Wilkie, D. and J. Finn. 1996. *Remote Sensing Imagery for Natural Resources Monitoring: A Guide for First-Time Users*. Columbia University Press, New York.