To the Graduate Council:

I am submitting herewith a thesis written by Douglas Hamilton Cain entitled "A Multivariate Analysis of Metrics Describing Landscape Pattern and Structure." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Geography.

We have read this thesis
and recommend its acceptance:

Bruce A. Ralston

Sally P. Horn

Accepted for the Council:

C W Minkel

Associate Vice Chancellor and
Dean of The Graduate School

A Multivariate Analysis of Metrics Describing Landscape Pattern and Structure

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Douglas Hamilton Cain

August 1995

ACKNOWLEDGMENTS

I would like to acknowledge the great assistance provided by several people which allowed me to complete my research. I have benefited greatly from knowing the faculty and graduate students in the Department of Geography, and wish to express thanks to all of those people. I am grateful to my Thesis Committee, Ken Orvis, Bruce Ralston, and Sally Horn. Ken gave me invaluable guidance in the completion of this thesis; Bruce is responsible for the vast increase in technical knowledge which I now possess; and I would not have finished so soon if not for the discipline with which Sally's seminar provided me.

There are a number of other people whose assistance I must recognize. I could not have done this research without the friendship and guidance of Dr. Kurt Riitters. He and the Tennessee Valley Authority and the Environmental Protection Agency provided generous financial support.

And finally, I would like to thank the friends I have made in the Department of Geography. In particular, I could not have made it through without the friendship and support of Tracy Engel, Matt Casey, and Dennis Yankee. Thanks for being there.

ABSTRACT

Advances in both the availability of land cover maps from satellite sensors and computing methodology to process such maps means that it is now feasible to monitor some aspects of landscape condition nationwide from remotely sensed imagery. It is therefore desirable to identify a subset of landscape pattern metrics which meet the criteria for monitoring landscape condition and which can be estimated from satellite-derived land cover maps. Previous research revealed many redundancies among landscape metrics and reduced the apparent number of unique dimensions to a half-dozen, but that study used USGS Land Use Data Analysis (LUDA) maps which are qualitatively different from satellite-derived maps. The present study used factor analysis to test the stability of landscape metrics across sets of Landsat Thematic Mapper land cover maps for the Tennessee River and Chesapeake Bay watersheds that differ in resolutions, numbers of attributes, and methods of defining subset boundaries. Diversity and texture and fractal indices were more stable than shape and compaction indices, and they accounted for a significant proportion of the total variance among the land cover maps. Results indicate that individual shape and compaction indices may be too unstable to consistently describe landscape condition across huge data sets, but further research on this suite of metrics might reveal a small set that consistently captures the shape and compaction dimension(s).

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER I

INTRODUCTION

Growing concern over the destruction of natural resources and the increased fragmentation of landscapes has focused attention on the need to monitor landscapes. Remotely-sensed images provide one of the more feasible methods for addressing this issue. Such images contain extremely large amounts of information, which can be reduced to a manageable level of complexity by measuring discrete aspects of images using scalar indices, or metrics. Metrics quantify the patterns and structural aspects of an image, thus facilitating interpretation and comparison of the complex information it contains. The ability to quantify spatial patterns provides researchers with a tool to begin to analyze the links between the spatial patterns in a landscape and ecological processes.

Previous ecological studies have described the processes that create non-spatial patterns observed in biota, while landscape ecology broadens this field by examining relationships among spatial patterns, temporal patterns, and ecological processes (Turner, 1989). Landscape ecology includes system ecology, human geography, and spatial statistics. Systems ecology and human geography provide the mechanisms that influence landscape patterns, and spatial statistics (*i.e.* landscape metrics) comprise the tools for quantifying these patterns and measuring relationships between pattern and process (Turner, 1989).

Landscape metrics are scalar indices which quantify different aspects of landscape pattern such as texture (the coarseness of land cover patterns), or fragmentation (of land cover types) (O'Neill, 1994). They provide a method to transform the raw data contained in an image into a useful set of comprehensible numbers. Over fifty metrics have been defined that measure and describe landscape pattern and structure within remotely-sensed data. The relationships and correlations among these metrics are not yet well understood; many are redundant, or they may measure similar traits using different approaches.

For landscape ecology to be effective, its tools, including landscape metrics, must be tested to prove their utility and accuracy. Geographers are in a unique position to further this research because of its inherently spatial characteristics. One of the goals of geographic training and research is to simplify information and detect spatial patterns, which is precisely what landscape ecologists are attempting to do.

The Environmental Protection Agency (EPA) is testing the use of metrics as tools to assess the status, changes, and trends in landscapes (EPA, 1994a). Under the Environmental Monitoring and Assessment Program—Landscapes (EMAP-L), pilot studies have begun to evaluate, test, and demonstrate possible landscape monitoring approaches. As one part of the EMAP-L project, the EPA seeks to identify a small set of representative landscape metrics for use in monitoring and assessing landscape ecological condition (EPA, 1994a).

Riitters *et al.* (1995) identified a subset of metrics through factor analysis which were suitable for measuring structure and pattern in the USGS Land Use Data Analysis (LUDA) map set. That approach, however, does not guarantee that identical subsets will emerge when using maps derived from different sources. Two other common sources of landscape data are Thematic Mapper (TM) and Multi-Spectral Scanner (MSS) satellite images. These differ from LUDA maps, both in method of production and in spatial resolution. Satellite-derived images are more readily updated, using new images and efficient, automated classification processes, whereas the production of LUDA maps is a slow and highly subjective manual procedure. Because of this, the EPA, for the most part, plans to use TM or MSS images in its EMAP-L project. The EPA now desires either to verify that the factor analysis results derived from the LUDA maps are applicable to TM/MSS-derived land cover data, or else to identify different subsets of landscape metrics suitable for these different classes of data. The present study investigates how landscape metrics react to satellite-derived maps, and whether or not the subset of metrics which Riitters *et al.* (1995) identified, or a different subset, can reliably provide the same information provided by a larger number of landscape metrics.

STUDY OBJECTIVES

In this study I examined the stability of landscape metrics over different spatial resolutions, numbers of attributes, and methods of boundary determination. The goal was to first find out which, if any, metrics were stable across the range of conditions, and then to select a small subset from the stable metrics that, together, adequately accounted for most of the variance captured by all the metrics used in this study. The final phase was to test the effectiveness of the subset of metrics in distinguishing similar and dissimilar land cover maps.

To test the stability of the metrics, I used factor analysis on several data sets of varying spatial resolution, attributes, and sub-boundaries, created from two TM-derived land cover maps of the Tennessee River Watershed and the Chesapeake Bay Watershed. I compared the factor analysis results from the various data sets to determine which metrics behaved similarly most or all of the time. After determining which metrics were stable, I selected a small set and tested their appropriateness by seeing how well they discriminated groups defined by similar characteristics as measured by all the metrics used in the present study.

Because of the different methods of map production, I expected significant differences in how metrics react in LUDA maps and TM maps. Metrics that measure fragmentation of a landscape or the complexity of landscape-element shapes should yield lower values as spatial resolution becomes coarser, and metrics that measure dominance of a land cover or similarity to simple shapes should yield higher values. But will metrics that appear related to each other in one data set maintain the same relationship in other data sets? If they do, can a small set of metrics adequately represent these relationships? These are the questions I examined in this study.

BACKGROUND

Landscape Monitoring

O'Neill (1994) warns that the monitoring and assessing of landscape ecological condition should not be a policy driver nor primarily a management tool, but instead, a forecasting methodology analogous to the way the National Weather Bureau collects data, assesses weather conditions, and warns of possible impending danger. Metrics can detect changes in landscapes over space and time, and therefore they are useful in alerting scientists and managers to areas or regions that are changing rapidly, and possibly adversely. Once such a region is identified, program managers can allocate their resources, which are often limited, to the examination of the affected area in more detail, and ask, "does a major problem exist here?" For example, an index of diversity could detect overall changes in land cover proportion and evenness. If diversity diminishes over time, and another index which measures proportions of types of land cover indicates that the proportion of forest cover has decreased also, it could be a signal of impending significant habitat loss for wildlife in the region.

O'Neill *et al.* (1992a) discussed ecological monitoring programs in terms of the difficulties involved in their design and implementation. Such programs are particularly difficult to implement on the ecosystem level. Ecosystems are extremely complex, influenced by climate, topography, geology, organisms, and more, and modeling causal relationships resulting from particular inputs is dubious, at best. For example, the decline of red spruce forests in the northeastern United States can not be attributed to a single possible cause, but to many contributing causes including climatic change, insect pests, pathogens and diseases, stand dynamics, and atmospheric deposition (pollution) (Nicholas *et al.*, 1992).

O'Neill *et al.* identified three "challenges" to developing broad monitoring programs: the multiplicity of stated objectives, the complexity of ecosystems, and the ambiguity of defining ecosystem "health." Multiple objectives include detecting changes in extent and geographic distribution of ecological resources, assessing ecosystem condition, suggesting causality, and predicting future ecological vulnerability. The last two objectives pose great difficulties, and may be unachievable, in that "monitoring can suggest, but seldom demonstrate causality" (O'Neill *et al.*, 1992a), let alone predict accurately.

No simple set of indicators can capture the complexity of ecosystems if for no other reason than because outside influences are impossible to control. Educated guesses can be made by devising measurements that reveal trends destructive to ecosystems, but there will always be unexpected crises such as a catastrophic fire or insect epidemic. It is important to avoid raising unreasonable expectations of monitoring programs.

Assessments of ecological "health" are also extremely challenging. Providing decision-makers and the general public with information on the environmental condition of an area is the implicit goal of a monitoring program. However, opinion on the best method of assessment varies between different groups. Preservationists view as "unhealthy" any change in an ecosystem away from the natural, unmanaged state. On the other hand, those whose livelihoods depend upon the utilization of natural resources emphasize high productivity and regeneration of renewable resources (O'Neill *et al.*, 1992a).

Despite the challenges involved in ecological monitoring, especially on a broad scale, O'Neill *et al.* (1992a) state, "we have little choice but to begin." We must seek new ways to address the problems that monitoring poses. They suggest that the areas of integrative measures, remote telemetry, and data interpretation hold promise for innovation. Holistic measures, such as the measurement of the ratio of primary production to respiration (P/R) in aquatic ecosystems, cannot be the sole indicators of ecosystem health, but may aid assessment in conjunction with monitoring of populations of "indicator" (highly-sensitive and rare) species. Remote sensing through the use of telemetry, *i.e.* low maintenance instruments placed in the field to collect data, shows promise as a monitoring technique. The major drawback to this method of data collection is the extremely high initial cost of purchasing the large number of instruments required.

The scientific community has begun to investigate the use of landscape structure and pattern data in the assessment of ecosystem health. Reduced to numerical landscape indicators, such data can relate spatial patterns in land cover data, whether remotely-sensed or surveyed on the ground, to ecological processes within the landscape (O'Neill *et al.*, 1992a). Many indicators have been developed in the effort to numerically capture the spatial structure of a landscape and relate it to ecological processes. The goal of this study is to examine several of these indicators, or *metrics*, to determine which are most stable across different conditions such as map scale, the number of attributes in a land cover map, and different boundary determination procedures (or zoning; see below).

Turner (1990) states that both classification methodology and map scale require careful consideration in analyses of landscape structure. One possible problem that must be recognized in the present study is the modifiable areal unit problem (MAUP). Fotheringham and Wong (1991) found that results of multivariate statistical analysis of spatial variables can vary unpredictably as a function of aggregation technique. They aggregated census data at different levels and under different zoning techniques. Zoning refers to how the boundaries of sub-regions of an area are determined (*i.e.* census tracts, zip code areas, ecoregions, watersheds). Different aggregation levels are equivalent to different map scales, *i.e.* spatial resolutions. The authors demonstrated that increases in the aggregation level of data or variations in the zoning system resulted in dramatic changes in the values of some variables, making it possible to produce almost any desired result by aggregating the data in the appropriate way. The MAUP is not cause for researchers to cease conducting statistical analysis, but it is important to be aware of the problem, since it affects the outcome of the analysis.

A way of assessing the influence of MAUP is to document statistical results at different levels of aggregation, and with different zoning systems at the same level of aggregation, to distinguish between stable and unstable variables (Fotheringham and Wong, 1991). Defining an optimal zoning system may be difficult because a degree of subjectivity is involved. A zoning system that optimizes the interzonal variation of one variable may not be optimal for another.

Eliminating the MAUP may be impossible, but its influence can be reduced by avoiding the use of aggregated data (Fotheringham and Wong, 1991). In the case of remotely-sensed land cover data, it is thus important to use as fine a spatial resolution as possible. However, good reasons exist for using data aggregated at some level before employing landscape metrics. Computer data storage capacity can impose practical limits on resolution of raster-based images. Important information about ecological processes may exist at different scales. O'Neill *et al.* (1991) found a hierarchy of ecological patterns and processes to be detectable at different scales, requiring the aggregation of data for complete analysis of a landscape.

By studying patterns of vegetation along transects, O'Neill *et al.* (1991) examined the effects of scale on patterns of flora in different regions using univariate statistical methods based on principles derived from hierarchy theory. Hierarchy theory proposes "that spatial and temporal scales are the natural consequence of nonlinear biotic and abiotic interactions in complex ecological systems" (O'Neill, 1988). For example, at a regional scale, patterns may result from geomorphological processes, while at a finer scale, biotic interactions such as competition and grazing can overlay a different set of spatial patterns. The first step in applying hierarchy theory, according to O'Neill *et al.* (1991), is to determine whether multiple scales of vegetation patterns exist. Examining vegetation data from transects in three locations, O'Neill *et al.* did detect patterns superimposed at different scales, supporting the notion that multiple scales can be detected in a variety of ecological systems (O'Neill *et al.*,1991).

Landscape Metrics

O'Neill *et al.*(1992a) suggested the use of landscape metrics to monitor ecosystem health. There are several types of metrics, and many variants of each type. Some general types include contagion, dominance, shape/compaction, and fractal complexity (Turner and Gardner, 1991). Contagion metrics aim to quantify the extent to which land covers coalesce. High contagion equates with larger, more contiguous patches, and low contagion values represent a more fragmented landscape.

Dominance and diversity metrics index the prevalence of one or a few land covers. Common indices of this type are Simpson's and Shannon's diversity measures, which in different ways incorporate the number of land covers and the proportion of each into a single metric. Higher values of diversity indicate a more even representation of land cover types. Higher values of dominance indicate that one or a few land covers comprise a majority of the landscape.

Shape metrics quantify the departure of the shape of patches of a particular land cover type from simple geometric shapes such as circles or squares. If a patch is similar to a simple shape, this implies a significant human influence, as humans frequently tend to simplify the landscape (O'Neill *et al.*, 1988). Compaction metrics gauge the relationship between perimeter and area, such that smaller ratios indicate a simpler landscape structure, a correlate of significant human influence.

Fractal complexity relates to the fractal geometry of the landscape. Metrics measuring fractal dimensions estimate the shape complexity of patches as represented by logarithmic scaling. An example of a fractal dimension is Mandelbrot's "coastline length" algorithm (Mandelbrot, 1983), which measures the logarithmic relationship between area and perimeter. Fractal measures differ from shape metrics in that they do not measure complexity by comparing patches to simple shapes, but rather by estimating complexity from a regression of patch perimeter and patch area. The slope of the regression line determines the estimate of complexity of shapes. Larger values indicate more complex shapes, which has been shown to indicate less human development in an area (see below) (Krummel *et al.*, 1987).

Previous Use of Landscape Metrics

Once researchers identified metrics that capture the structure and pattern of landscapes, they searched for relationships between such structure and pattern and ecological processes. Predominantly natural landscapes are likely to be in better overall condition than highly-developed areas, and areas where landscape patterns are changing dramatically are likely to contain ecosystems which are changing the fastest (EPA, 1994a). Previous use of landscape metrics has primarily focused on modeling the effects of disturbance on the ability of animals to range and to migrate, modeling disturbance as a process, and measuring water quality (Turner, 1989). Recent research has modeled habitat suitability as a means of predicting potential biodiversity (EPA, 1994a).

Gardner *et al.* (1987) used percolation theory and neutral models to quantify the relationship between the landscape and animal movement, and between the landscape and disturbance. Percolation theory, which is the study of spatially random processes, provides a way of analyzing the effects of randomly varying the connections between spatially distributed patches of the same land cover (Milne, 1991; Gardner and O'Neill, 1991). Neutral models are hypothetical models of expected landscape patterns in the absence of a specific ecological process, such as fire spreading across a landscape; they test the relationship between pattern and process. Gardner *et al.* (1987) generated raster maps, *i.e.* neutral models, of sample landscapes, with varying proportions of a sample land cover. Using a simple metric, the proportion of a type of land cover in a landscape, they demonstrated the effect of structure on ecological processes. Random maps of differing probabilities for a particular cell type were used to study the ability of animals to move about a landscape, assuming movement was possible only within and between cells of a given type. The authors calculated and analyzed the number, size distribution, and fractal dimension of cell clusters of the same type. According to percolation theory, the largest cluster of one type of cell on a randomly generated map tends to "percolate" or connect the map from one side to the other as the proportion of cells of the one type approaches 60%. This would allow a hypothetical animal restricted to a single habitat type for foraging, shelter, and movement to disperse successfully across the landscape (Turner *et al.*, 1988).

Expanding on this study, O'Neill *et al.* (1988a) examined what they termed "resource utilization scales" for different animals. A resource utilization scale refers to the range of grid cells that a hypothetical organism can move through in its search for a resource, with the needed unit of resource located at the center of the grid cell. Some animals will have larger resource utilization scales, meaning they can traverse larger distances and can use a more sparsely distributed resource. Using a raster map with distance measured in cells, or pixels, an animal with a large range might cover four cells as opposed to one cell for a short-ranging animal, or the animal may be able to use or cross more than one land cover type to find its necessary resources. For example, an animal with a cell range of one may require a land cover with a probability of greater than 59% to move about successfully and forage for resources. Another animal may have a range of four heterogeneous cells, reducing the critical probability for connectivity to around 20%. This pattern can be thought of in terms of food chains in that organisms high on the food chain operate at a larger resource utilization scale.

In addition to animal movements, Turner *et al.* (1989) applied percolation theory in demonstrating that the impact of disturbance is influenced by spatial heterogeneity. Turner *et al.* define disturbance as "any relatively discrete event in time that disrupts ecosystem, community, or population structure and changes resources, substrate availability, or the physical environment." The authors found that frequency and intensity of disturbance affect differently structured landscapes in different ways. Randomly generated maps of different proportions of particular land covers (such as forest) provided insight to the management of disturbance-prone landscapes. In the context of a hypothetical disturbance such as fire or disease, rare, fragmented habitats covering lesser proportions of the landscape were resistant to high-intensity disturbances, but vulnerable to high frequencies of disturbance. Common habitats covering large proportions of the landscape were vulnerable to both low frequency and low intensity disturbances due to a higher connectivity of the landscape.

Krummel *et al.* (1987) studied disturbance using fractal analysis. They calculated a fractal dimension capturing the complexity of the relationship between perimeter and area of patches of land cover for forest patches in the Natchez Quadrangle LUDA map of central Louisiana and Southern Mississippi. The authors found that anthropogenic disturbance related to lower fractal dimensions which indicated simpler geometric shapes. Landscapes influenced by natural disturbance had more complex shapes.

Anthropogenic disturbance is a major area of focus for landscape ecologists, and Hunsaker *et al.* (1992) showed that agricultural development has a negative effect on water quality. They used landscape metrics to predict water quality in southern Illinois, measuring proportions of land uses, edge relationships, dominance, contagion and other metrics from LUDA maps. The authors regressed the measurements against electrical conductivity data, a surrogate of water quality, for streams and rivers in the Southern Illinois area. High R-squared values resulted, indicating predictive relationships between these metrics and conductivity. In addition, the authors also found they could predict water quality better when using data from the entire watershed as opposed to using only near-stream corridor data, which only takes into account land uses within a certain distance of a stream. This indicates that land uses located further from a stream may affect water quality to a greater degree than previously thought (Hunsaker *et al.*, 1992).

Disturbance affects the connectivity and health of ecosystems, and animals depend upon certain types of land cover for necessary resources and for freedom of movement. The types and amounts of natural land cover available for wildlife utilization affect the biodiversity of ecosystems. Large patches may act as reservoirs of resources, maintaining a population or populations on a landscape. Species diversity relates to landscape patch size (Forman and Godron, 1986), suggesting that patch area is an important determinant of species diversity. Increased fragmentation of a landscape into many isolated patches has been shown to reduce biodiversity (Gardner *et al.*, 1993), as has the loss of edge between two natural land covers (Wilson, 1988). Edges provide unique habitats which are often associated with rare and endangered species. Rare species are often effective alarms indicating ecosystem damage because they show stress more quickly than other species (O'Neill *et al.*, 1992a).

Land cover maps created from remotely-sensed images provide valuable insight into habitat diversity. Riitters and O'Neill (1994) modeled habitat diversity using a sliding-window routine on a land cover map of the Chesapeake Bay Watershed. Based on a square window defined to be a hypothetical organism's home range, the authors calculated habitat requirements such as a certain percentage of a particular natural cover, and they wrote the result to a new pixel in a new map. Repeating this process for each pixel created a new map of habitat suitability. Inputting the estimated requirements for different organisms into the model through adjusting the window size and by using different landscape indicators, scientists can inventory habitat availability for species with different edge and interior habitat requirements (EPA, 1994a). While the model used hypothetical organisms, the goal is to eventually focus on actual organisms.

The modeling of habitat suitability, disturbance, and animal movement, as well as the measurement of water quality, demonstrate that landscape metrics show promise as tools for monitoring and assessment. As progress continues in linking ecological processes to the pattern and structure of landscapes, many landscape metrics are being used without knowledge of their behavior under different conditions. Different spatial resolutions, numbers of attributes, and methods of boundary determination may significantly affect how metrics behave, and, to avoid making incorrect evaluations of ecological processes, it is important to know which metrics are stable and which are not under these circumstances. The present study analyzes the stability of several landscape metrics using land cover maps of two regions, the Tennessee River Watershed and the Chesapeake Bay Watershed.

CHAPTER II

STUDY AREAS

PHYSIOGRAPHY AND LAND COVER

Tennessee River Watershed

The Tennessee River Watershed (Figure 2-1) drains 10,606,000 hectares in six states, with headwaters in southwest Virginia, western North Carolina, and northern Georgia. Tributaries converge to form the Tennessee River in east Tennessee near Knoxville, and the river then flows southwest to Chattanooga before it turns west in the northern part of Alabama, and then north in Mississippi. The river cuts across western Tennessee and into western Kentucky, where it flows into the Ohio river.

Several physiographic regions make up the Tennessee River Watershed (Figure 2-2). The Appalachian Mountains and a ridge and valley zone of structurally-controlled hills and drainage characterize the eastern half of the watershed. The ridge and valley zone gives way to the Cumberland Plateau as one moves west, which is replaced by the Mississippi Embayment in the extreme west.

The predominant land cover in the watershed is forest (60%), with agricultural uses (crops and pasture) ranking second with 34% cover. Water covers just over 4% of the region, and urban areas account for 1.4%. Other categories such as mines and barren areas occupy the remaining fraction of a percent (Figure 2-3).

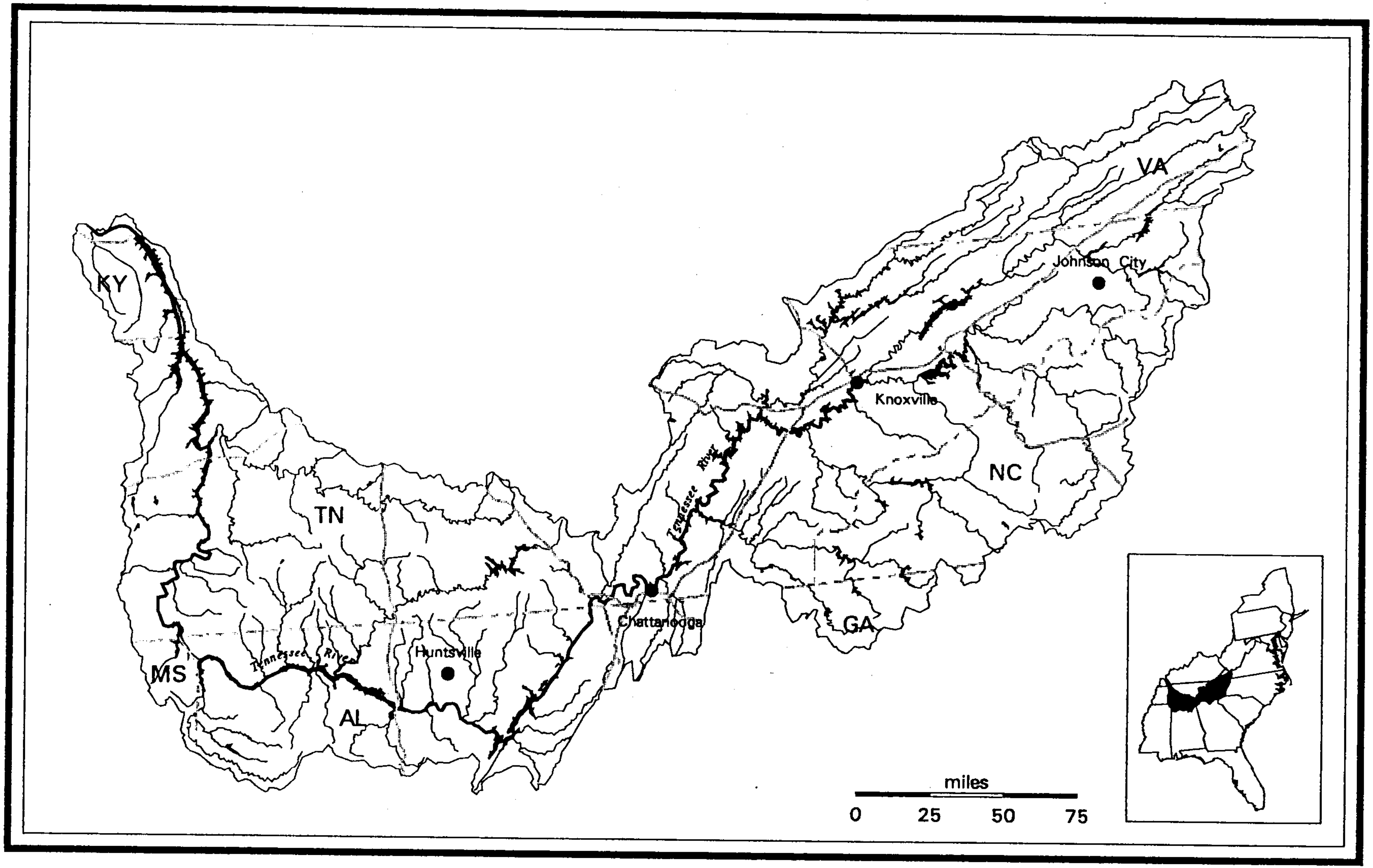


Figure 2-1: Tennesse River Watershed Study Area.

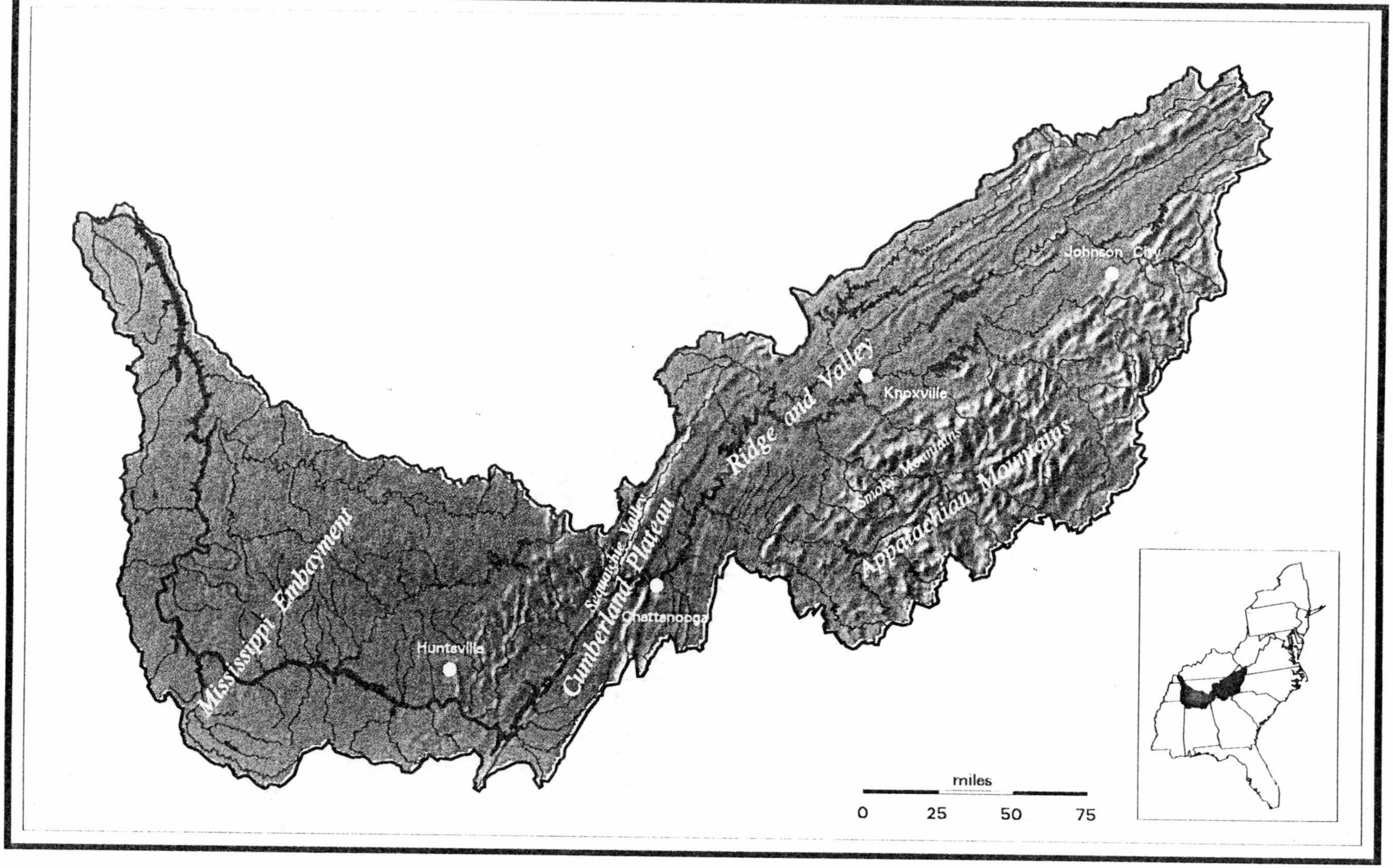


Figure 2-2: Shaded Relief Map of the Tennessee River Watershed with Physiographic Regions.

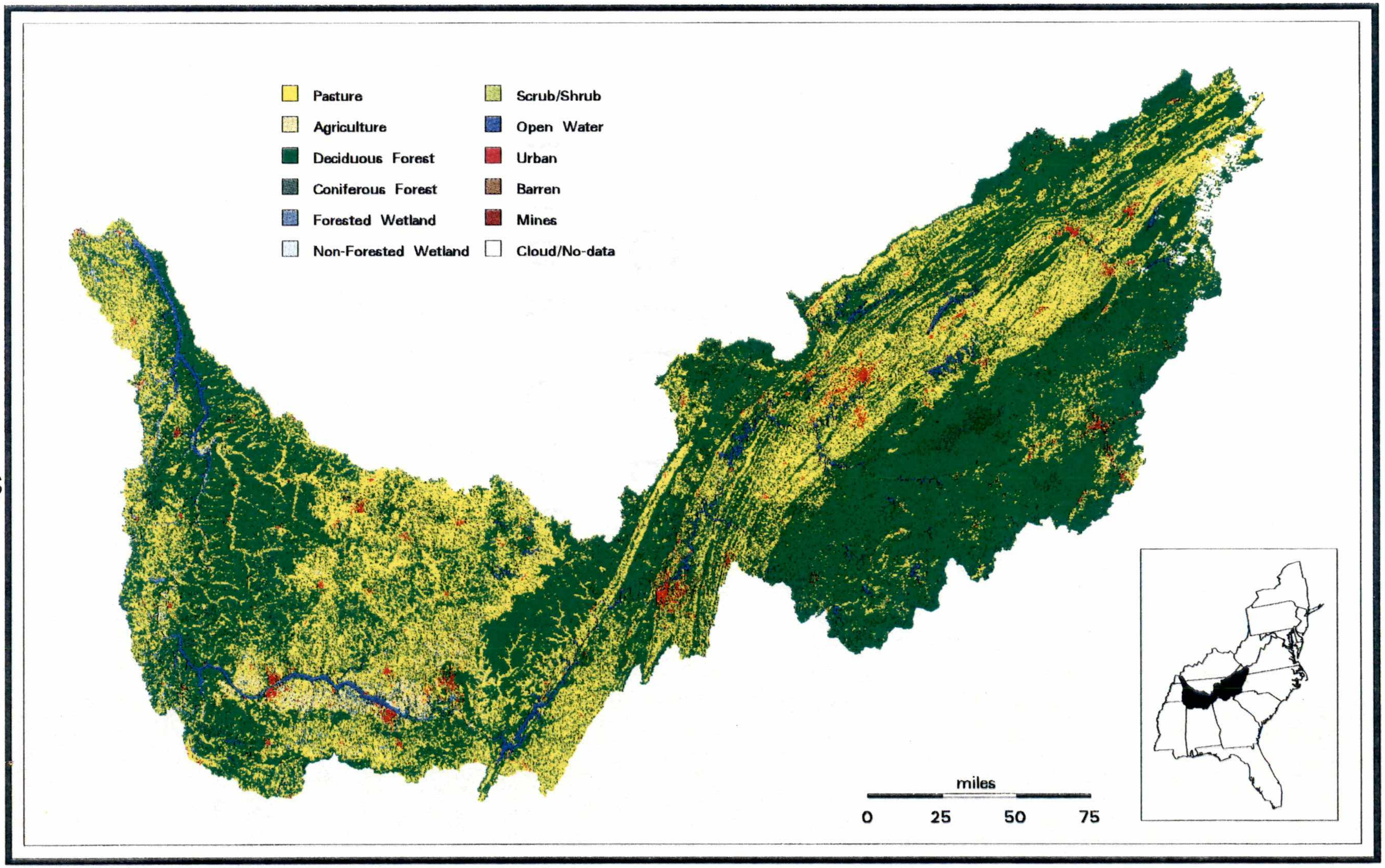


Figure 2-3: 12-Class Land Cover Map of the Tennessee River Watershed.

Chesapeake Bay Watershed

The Chesapeake Bay Watershed (Figure 2-4) drains 17,759,000 hectares in six states, largely via the Potomac and Susquehanna Rivers. The watershed drains an area from south central New York through Pennsylvania and Maryland to central Virginia. It also includes the eastern section of West Virginia and the western part of Delaware.

Several physiographic regions also comprise the Chesapeake Bay Watershed (Figure 2-5). These include the Appalachian Plateau in the north, the ridge and valley province along the western half of the area, and piedmont and coastal plain in the southeastern third of the region.

Forest accounts for 55% of the land cover in the watershed, and agricultural/pasture cover comprises 33%. Water covers about 7.5% of the watershed, due largely to the Chesapeake Bay itself, which is a mix of freshwater, tidewater, and seawater. Urban land uses occupy a much higher proportion of the landscape (4.5%) than they do in the Tennessee River Watershed. Exposed or barren land covers the remainder of the area, at 0.3% (see Figure 2-6).

LAND COVER CLASSIFICATION

Tennessee River Watershed

ERDAS, Inc. classified a composite Landsat TM image of the Tennessee River Watershed into a 12-class land cover map (Figure 2-3) using an iterative, multi-stage

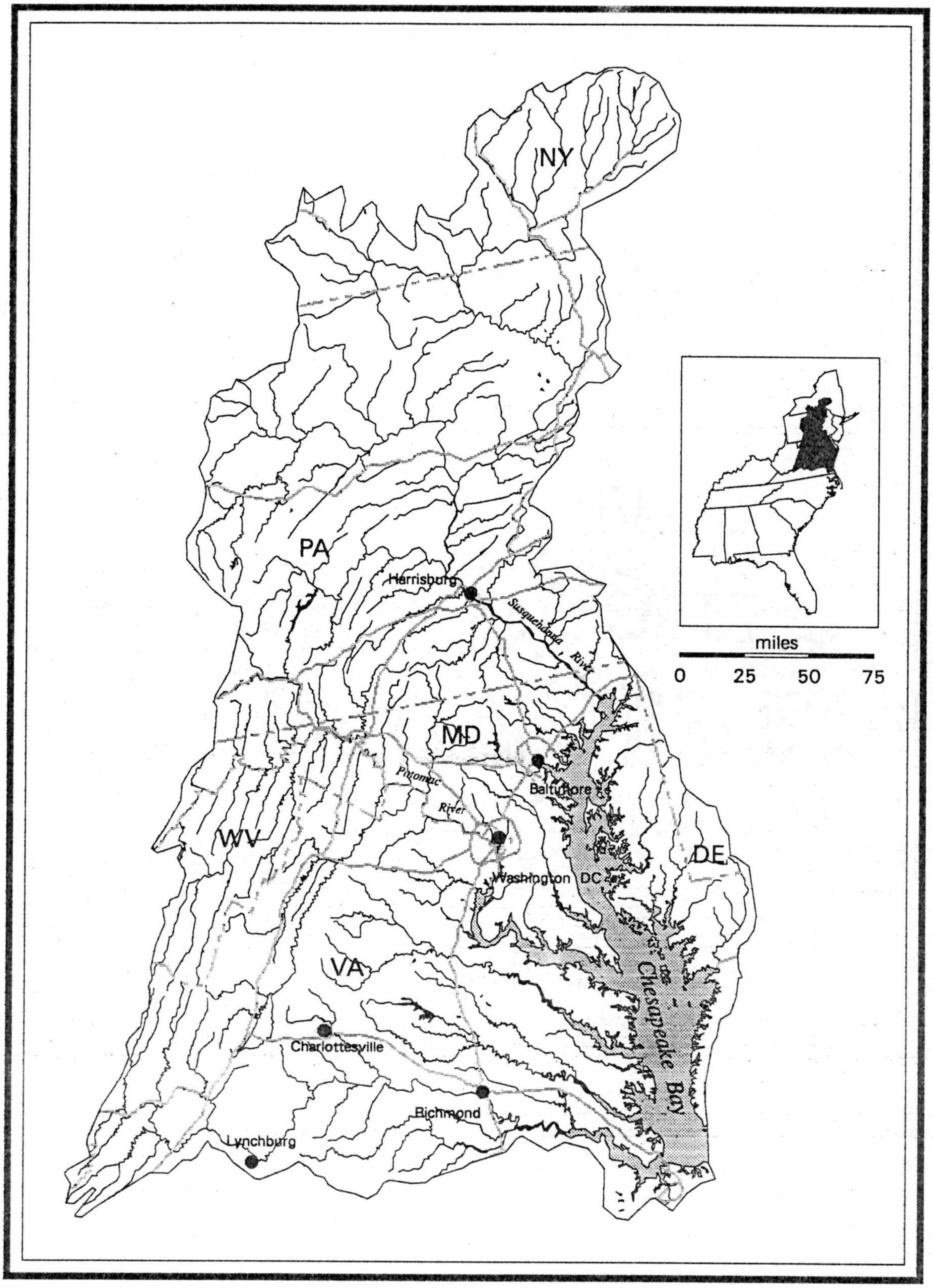


Figure 2-4: Chesapeake Bay Watershed Study Area.

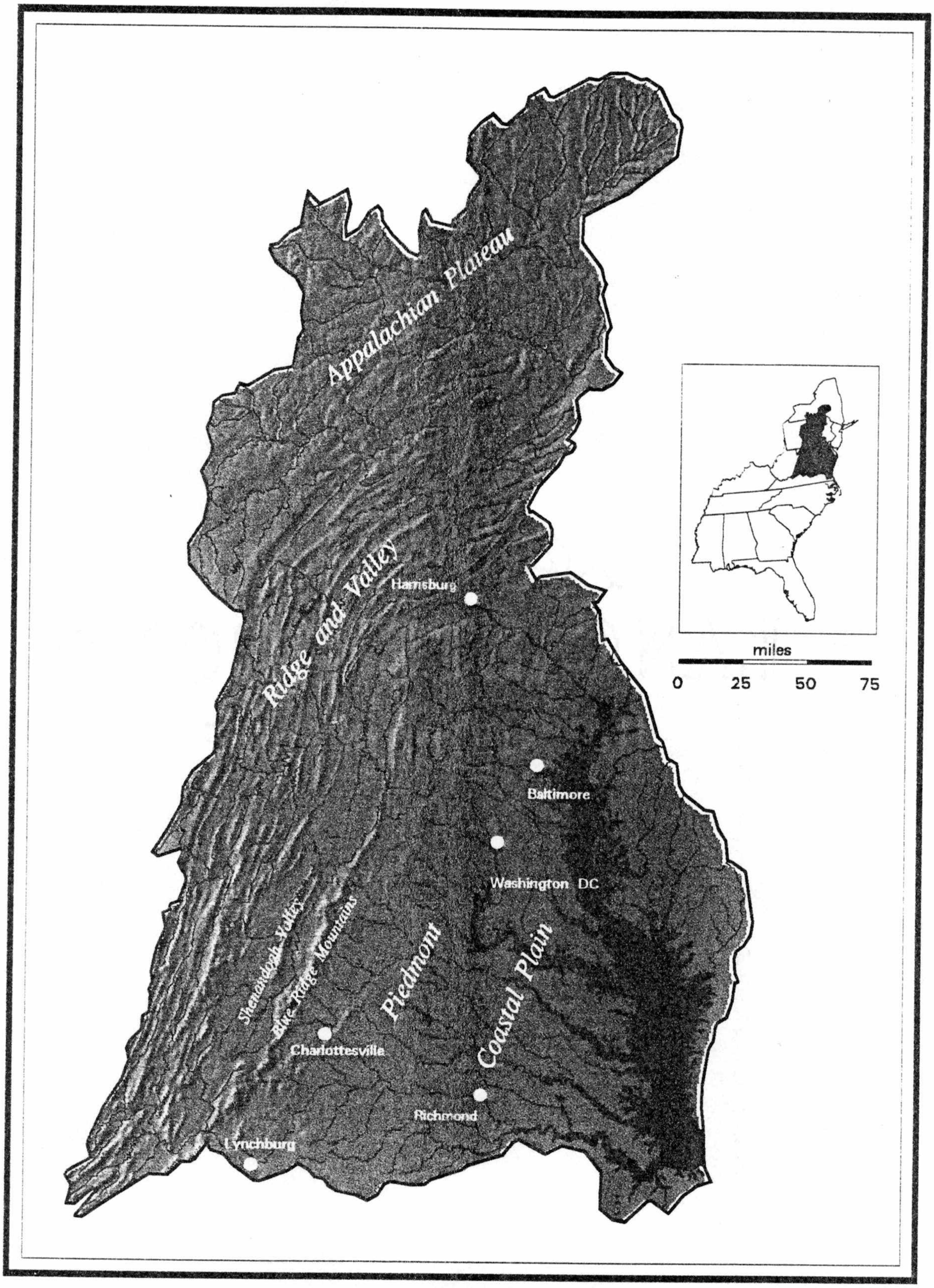


Figure 2-5: Shaded Relief Map of the Chesapeake Bay Watershed with Physiographic Regions.

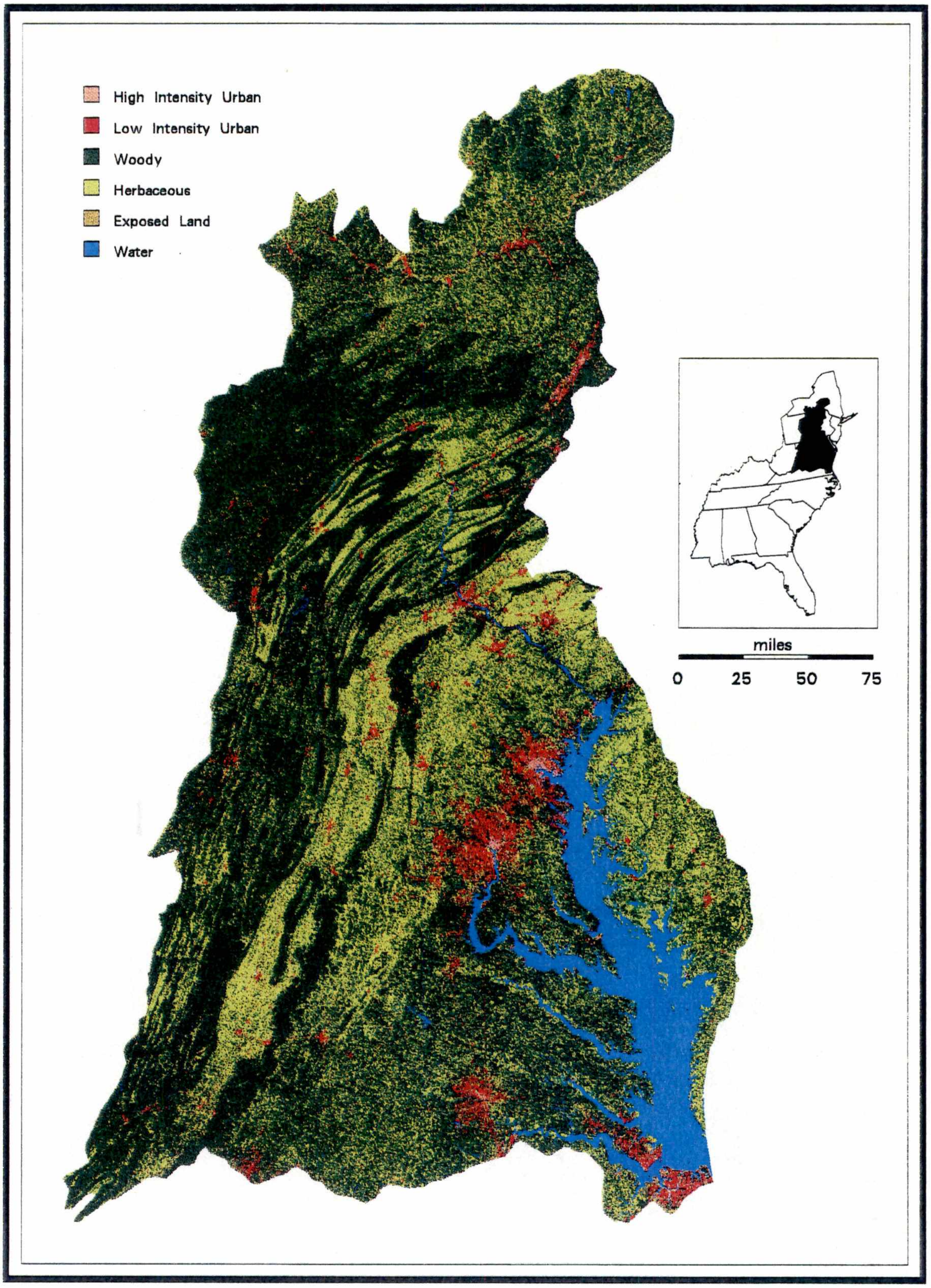


Figure 2-6: 6-Class Land Cover Map of the Chesapeake Bay Watershed.

process (TVA, 1992). Ground-truthing included use of aerial photography, wetland maps and other supporting information such as location data for mines, *etc.* Supervised analysis of the image identified initial spectral signatures for each class or category of land cover. To identify spectral signatures that this supervised method may have overlooked, ERDAS personnel also ran an unsupervised spectral clustering of the data. Next, using 1:24,000 scale maps as guides, they drew polygons around urban areas and masked them from the original image, then reclassified them using a spectral clustering method. Keeping only urban-classified areas from this clustering, they then overlaid the urban area polygons on the previously classified image.

To assess accuracy, ERDAS and TVA employees sampled ground-truthed areas and compared them to the computer-classified results. The overall classification achieved a minimum accuracy of 85% (TVA, 1992).

Chesapeake Bay Watershed

A team at the Lockheed Corporation classified the Landsat TM image of the Chesapeake Bay Watershed into a 6-class land cover map (Figure 2-6) using an unsupervised/supervised approach. They defined spectral classes using an unsupervised spectral clustering method, and then used these classes as a starting point for a supervised classification (Wickham, 1995).

The Lockheed team assessed the accuracy of the classification by comparing sampled areas that were ground-truthed (using high-altitude, infra-red aerial photography) with computer-classified results. They achieved an overall classification accuracy of 80%, which is somewhat lower than the accuracy results of the Tennessee River Watershed Land Cover map (EPA, 1993).

ASSESSMENT OF LAND COVER SUITABILITY

To determine the detail and accuracy that the classified images capture, and to perform an independent ground-truth assessment of ecosystem stress, I visited several of the areas covered by the images. While this was not a comprehensive assessment of every area analyzed, it helped me judge the real-world effectiveness and applicability of using classified images to monitor ecosystem health. Areas that I visited included the Sequatchie Valley and Cumberland Plateau in the center of the Tennessee River Watershed, the area between Chattanooga and Johnson City, and the Smoky Mountains (Figure 2-2). In the Chesapeake Bay Watershed I examined the land cover on the eastern side of the Blue Ridge Mountains between Lynchburg, Virginia, and Washington, D.C., and the southern portion of the Shenandoah Valley (Figure 2-5).

Tennessee River Watershed Land Cover Map Assessment

The Tennessee River Watershed land cover map depicts a definite pattern of corridor patches of deciduous forest (Figure 2-3) from the northeast to the south central section. The ridge and valley topography restricts human development, and the land cover map represents the actual landscape well. Landscape metrics measuring characteristics such as contiguity (contagion) and shape complexity should reflect the long stretches of forest along the ridges and the agricultural and pasture development in the valleys by indicating a lower fragmentation and more simple shapes in this area than in other landscapes. Measuring this structure is important because the forest cover along the ridges could be beneficial to wildlife, allowing animal movement over great distances. However, while not extremely common in the areas of the Tennessee River Watershed which I observed, significant residential development that would not show up on the images could occur along the ridges under the canopy of the forest, and could hinder animal movement.

Possible impacts on the forest ecosystems that do not show up in the land cover map are the conditions and types of forests. For example, natural forests may provide more suitable habitat for some wildlife species than cultivated forests (*i.e.* tree plantations), but this differentiation is not made on the land cover maps.

Traveling along the eastern edge of the Cumberland Plateau (Figure 2-2), one can see that small towns have developed at the base of the plateau, creating the scattered urban class that is visible in the Tennessee River Watershed land cover map (Figure 2-3), and the degree of human impact should be measurable using landscape metrics that analyze shape complexity and diversity of land cover type. The Sequatchie Valley, which cuts the watershed in half (Figure 2-2), is a wide, flat, unroofed anticline that stretches from just west of Chattanooga to the northeast paralleling the Cumberland Plateau escarpment. Pasture and agricultural uses are the main types of land cover in this valley, and the land cover map (Figure 2-3) represents this well.

The coniferous forest cover in the central portion of the eastern section of the land cover map (Figure 2-3) shows the spruce and spruce-fir zones in the Smoky mountains

(Figure 2-2). As in the land cover map, the mountainous terrain scatters the actual coniferous forests, with the coniferous forest growing at the higher elevations. A particularly useful method of measuring stress on this coniferous forest might be to monitor over time the proportion of an area occupied by this land cover type and how fragmented or contiguous the patches are, allowing wildlife managers to detect any rapid and dramatic change in forest composition.

Chesapeake Bay Watershed Land Cover Map Assessment

Traveling from Lynchburg, Virginia north to the Washington, D.C. area, I found a different pattern in the Chesapeake Bay Watershed than that just described for the ridge and valley area of the Tennessee River Watershed. The Blue Ridge Mountains form the western edge of the piedmont area (see Figure 2-5), where agricultural development is more scattered, because it is not restricted by numerous, abrupt ridgelines (see Figure 2-5). Forest is much more fragmented and patchy in this area, but seems to be more connected by riparian and hedgerow corridors, which do not appear on the land cover map unless they are very wide (greater than 30 meters).

Closer to the Washington, D.C. area, urban development increases, as expected. The number of residences, or farmsteads, seems to remain constant, but commercial properties are interspersed between them with increasing frequency as the distance to Washington, D.C. decreases. One can see this in the land cover map southwest of the capitol, where the urban class scatters out towards Charlottesville (Figure 2-6). Several metrics measuring proportion of landscape occupied by a land cover, contagion, and diversity should reflect this increase in development. Diversity of land cover should increase in the transition zones between heavily-developed and less-developed areas because classes common to both should be present.

Heading south on the western side of the Blue Ridge Mountains in the Shenandoah Valley, I noticed much more open land under either pasture or cultivation. This corresponds with the land cover map, however once again, there is a lot of riparian growth and hedgerow development which did not get captured by the image, but is probably very significant in allowing wildlife movement.

While general patterns on the Chesapeake Bay Watershed land cover map seem accurate, they are indeed general. Much information is lost because so few classes exist. In particular, vital information is lost by only having one forested class and one herbaceous class. Distinguishing between deciduous and coniferous forests can be useful because they affect wildlife diversity differently. The northern section of the Blue Ridge Mountains is almost exclusively deciduous forest, yet in the southern section there is a mix of deciduous and coniferous patches. These sections are very different in composition, but this is not detectable on the land cover map.

The herbaceous class encompasses the agricultural development (cultivated land and pasture). While the two land cover types may have similar erosional qualities, cultivated land is much more likely to generate higher water pollution rates due to higher fertilizer and pesticide use, and to tillage. Including both land covers in one class may take away some of the benefit of using land cover images and metrics describing landscape structure and pattern by incorporating more uncertainties into the results.

Summary of Suitability Assessment

Overall, large scale stresses on the environment seem to be detectable from the land cover maps. Human development often has a negative impact on ecosystems, and it is desirable to measure the amount of human influence on a landscape. Where human development of the land is most pronounced, it is evident that patterns are simpler, that pasture and agricultural fields have been cut out of the forest in the most efficient manner possible, *i.e.* that edges are straighter than natural edges. Patches of forest disappear when human development is greatest, and this leaves fewer areas for wildlife to occupy. Very small patches may not be able to sustain a population of some species, so the resolution of the image may have some use in showing only patches of significant size. On the other hand, riparian and hedgerow corridors of some thickness may be significant in connecting the landscape, and may be visible only at finer resolutions. A balance must be found, because at one scale the riparian growth and smaller patches may be visible, and this could represent information allowing assessment of habitat for edge species. At a coarser scale which removes the majority of riparian corridors and small patches, assessment of habitat suitability for interior species may be possible.

Where greater areas of agricultural production exist, larger amounts of fertilizer and pesticides are likely to end up in the water draining a landscape. Unless natural filtering by forest cover is provided, water quality may suffer. Since Hunsaker *et al.* (1992) demonstrated that water quality correlates to the area under agricultural use and not so much to the presence of riparian growth, the fact that the land cover maps do not capture riparian growth may not be detrimental to the ability to detect agricultural stress on an area.

However, these land cover maps miss some significant stresses on ecosystems. Roads and electrical transmission lanes (linear objects) cut through forests, cause significant erosion at stream crossings, and limit the ability of wildlife to move about the landscape (O'Neill, 1994). With the exception of some sections of interstate highways or other roads with significant development around them, roads do not show up on the land cover maps.

Other important stresses such as illegal dumping areas and point source pollution of streams do not get captured by the land cover maps. The land cover maps alone cannot detect certain microscale influences on ecosystem health, but incorporating them with a vector-based Geographic Information System (GIS) which captures linear and point features may improve assessments. The land cover maps could be useful for detecting regional patterns and general characteristics, and the vector-based GIS could add the power to analyze finer-scale stresses such as the impact of roads and point sources of pollution.

CHAPTER III

METHODS

To test how metrics measure data of different spatial resolutions, numbers of attributes, and different ways of being divided into observations (subset maps) for statistical analysis, I created several data sets from digital land cover maps of the Tennessee River Watershed and the Chesapeake Bay Watershed. Calculating metric scores for each subset map in each data set, I measured the independent contributions of the metrics to total variance in the data set using factor analysis methods similar to those used by Riitters *et al.* (1995). After determining which metrics were stable, I selected several which I thought consistently explained or accounted for a large proportion of the variance among the different subset maps. Grouping similar subset maps of a particular data set into classes based on the variance calculated from all of the metrics, I tested the ability of the smaller set of metrics to differentiate between subset maps and correctly classify them into appropriate classes using discriminant analysis. Figure 3-1 outlines the steps of the research design. I conducted all statistical procedures using the statistical package SAS (SAS, 1982).

CREATION OF DATA SETS

I created five separate map sets from the original, 12-class, 25-meter-resolution Tennessee River Watershed (TRW) Land Cover map (TVA, 1992). First, I split the original map into four equal rectangles, or quadrants. This process was repeated for each rectangle until 256 quadrants were created. Figure 3-2 shows a diagram of the splitting process and naming conventions for the data set creation. An overlap between quadrants of seventy-five cells was included to ensure that features near quadrant boundaries would

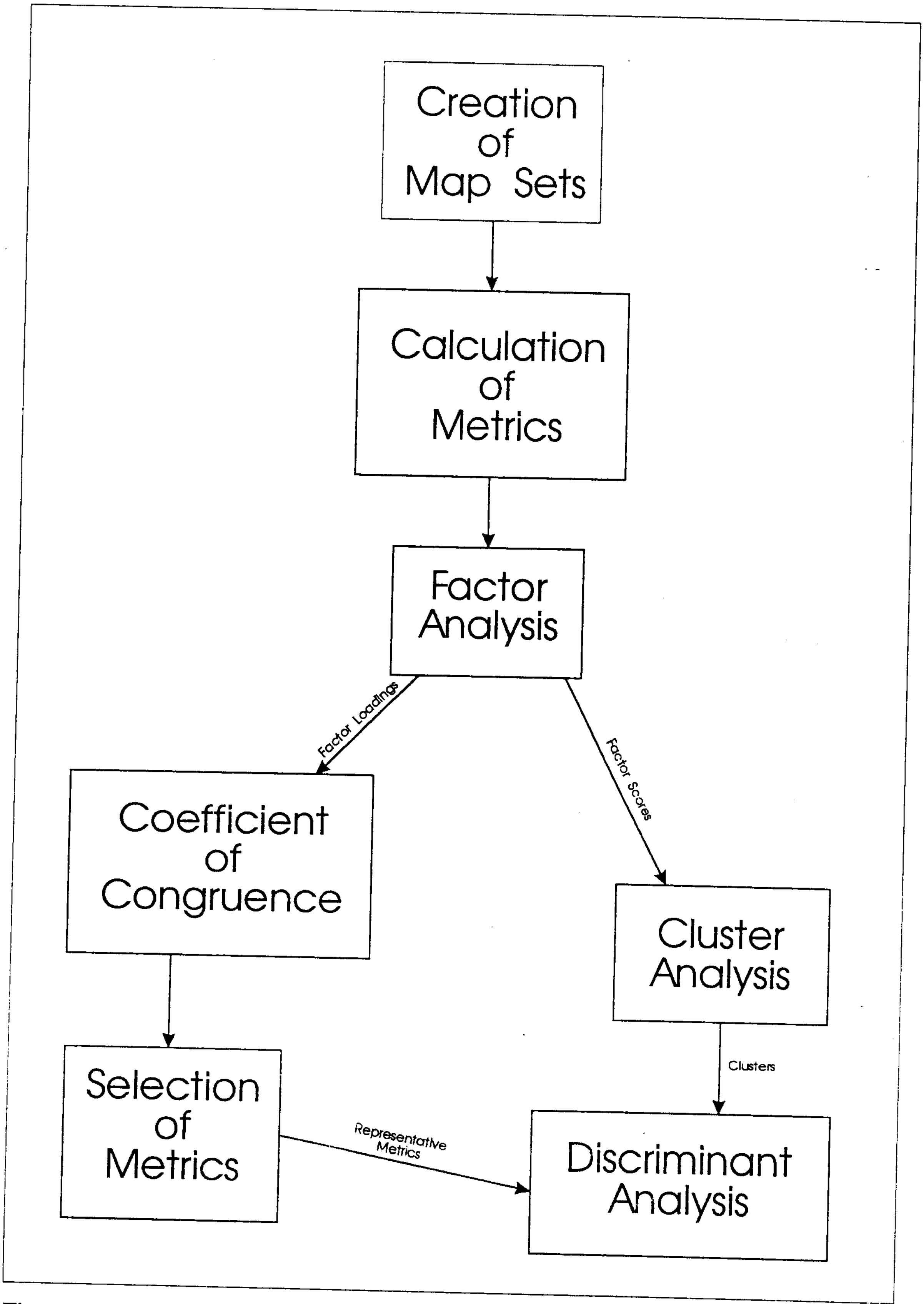


Figure 3-1: Flowchart of procedures and analytical steps taken in this study.

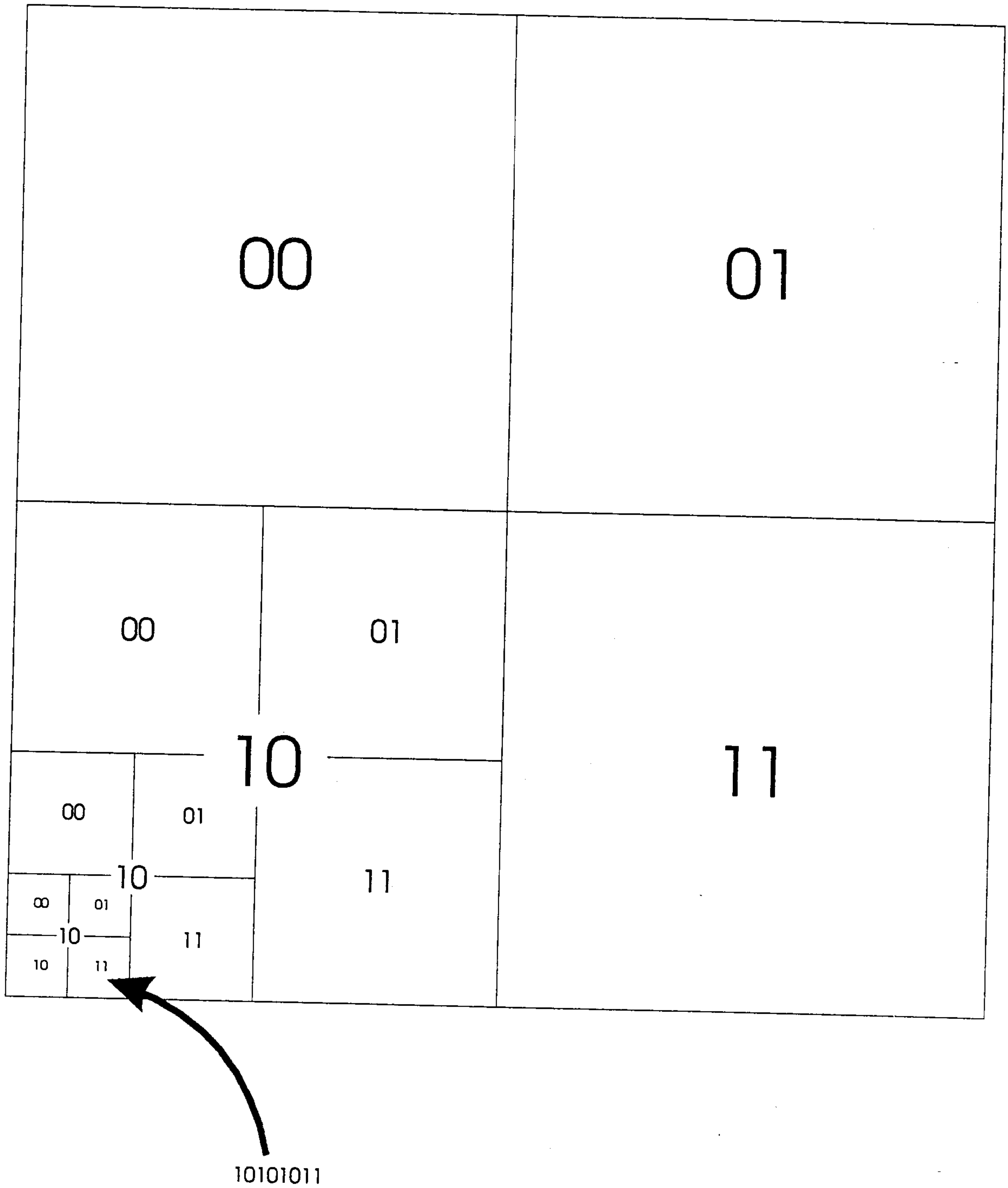


Figure 3-2: Diagram of quadrant-splitting procedure. The image is split into four quadrants, which are further split into four more, etc. Each observation is referenced by the numbers of each quad in the hierarchical tree.

be included. Each of the 256 quadrants had an extent of 1800 by 1100 cells, which at a cell resolution of 25 meters was almost 1200 km^2. Quite a few subset maps had large areas of *NO DATA* (outside the watershed) values, and some had cells that represented cloud cover, so I rejected subset maps with 15% or more *NO DATA* or cloud cover values. A total of 85 subset maps remained (Figure 3-3), and composed the TRW original set (Table B-1).

To examine the effect of spatial scale on the stability of landscape metrics, I created map sets with different resolutions. Because different methods of aggregation produce different results, I created three different resampled map sets from the TRW original set using two methods. For two of the resampled data sets, I used a majority-rule method with a moving 5 by 5 cell window wherein the window is coded with the land cover type with the largest proportion of cells in the window. I created one set with a cell size of 125 meters by compressing the 5 by 5 window to a single cell, such that each subset map had an extent of 360 by 250 cells. The second majority-rule-resampled set was left uncompressed at 25-meter resolution, such that all cells in a 5 by 5 window had the same land cover type. The third set was resampled to a resolution of 125 meters using a nearest-neighbor algorithm as provided by the Geographic Information System software Arc/Info (ESRI, 1992). This method coded the new, compressed cell with the value of the center cell of the 5 by 5 window. Each subset map in this TRW low-resolution nearest-neighbor set had an extent similar to the subset maps in the TRW low-resolution majority-rule set (Table B-1).

To analyze whether metrics behave consistently with data sets of different numbers of attributes, I created an additional map set from the TRW original set in which the twelve land cover classes were recoded into five classes (Table B-1). These five new classes approximately matched the land cover classes of the Chesapeake Bay Watershed map (see below). Pasture, agriculture, and scrub/shrub were reclassified as non-forested;

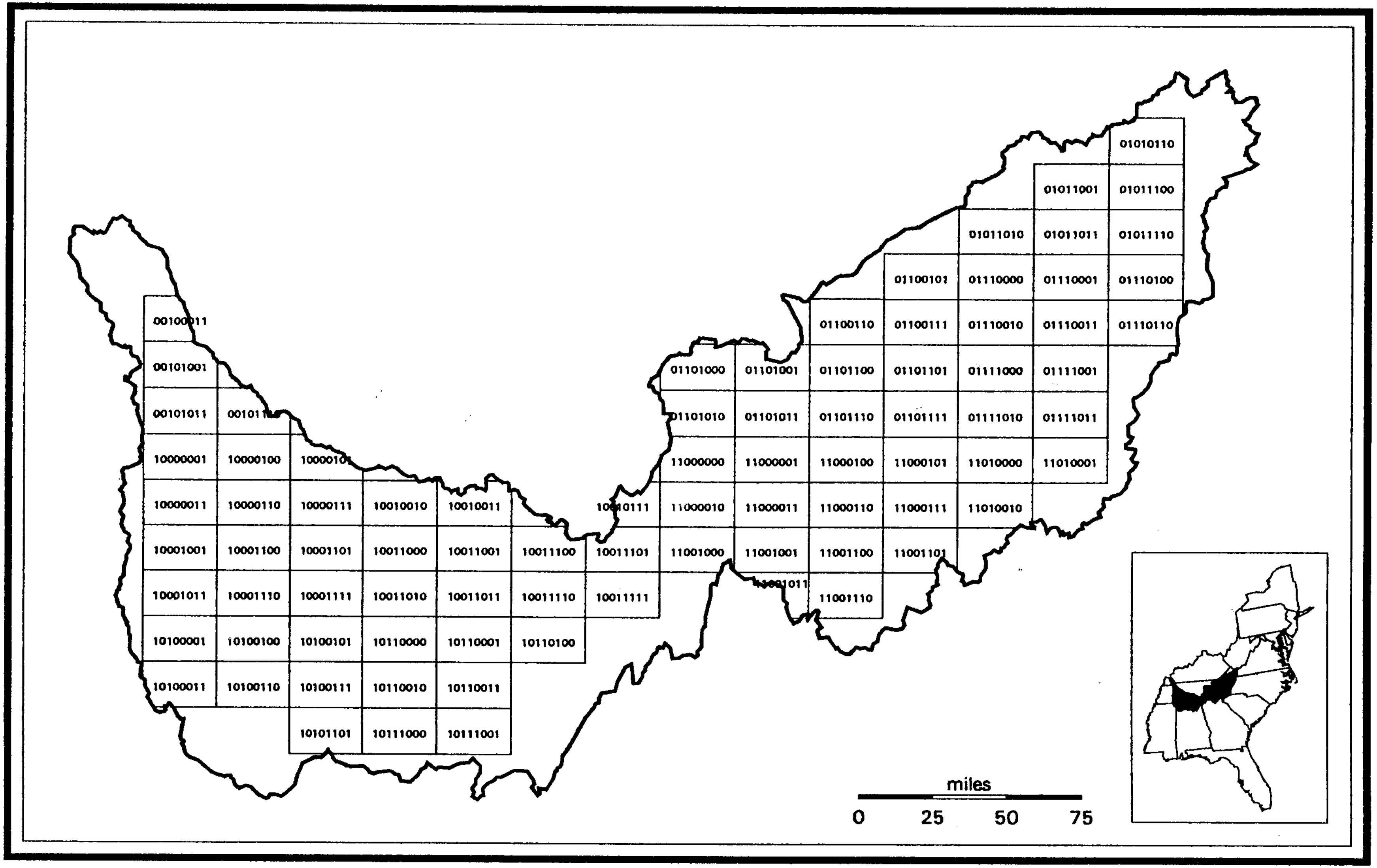


Figure 3-3: Tennessee River Watershed Divided into 85 Quadrants.

deciduous and coniferous forests as forested; barren land and mines as exposed land; and open water, forested wetland, and non-forested wetland as water cover. Some differences may remain between these classes and the Chesapeake Bay Watershed classes because it is possible that the team that classified the Chesapeake Bay Watershed used different methods of classifying water, "herbaceous," and other classes.

To compare the behavior of metrics between two regions, I created a comparable data set using the Chesapeake Bay Watershed (CBW) data. In order to match the number of classes in the TRW 5-class set, I first condensed the two urban classes in the full-coverage, 6-class, 25-meter resolution CBW Land Cover map into a single class. For the comparison set I split the resulting 5-class map into 256 quadrants. Once again, I rejected subset maps having more than 15% *NO DATA* and cloud cover values; a subset of 97 subset maps remained (Figure 3-4). Each subset map had an extent of 1500 by 1900 cells (just under 1800 km^2), 25-meter resolution, and five land cover classes (Table B-1).

To test metric stability under different methods of boundary determination, I created one final data set covering the same region but using a different splitting approach. I cut out subset maps from the condensed (5-class), full-coverage CBW Land Cover map using sub-watershed boundaries as defined by U.S.G.S. 8-digit Hydrologic Units (Sember *et al.*, 1987) (Figure 3-5). Sub-watersheds enclosing areas greater than 1800 km^2 were selected so that the subset maps would be comparable in size to the maps split into quadrants. This final 25-meter, 5-class data set included 46 subset maps (Table B-1).

CALCULATION OF METRICS

Metrics were selected to include a wide variety of the types currently being used in landscape ecology. For a discussion of the metrics used in this study, see Appendix A. Several computer programs currently exist which calculate landscape metrics for images. Worthy of mention are *FRAGSTATS* (McGarigal and Marks, 1994), *SPAN* (Turner,

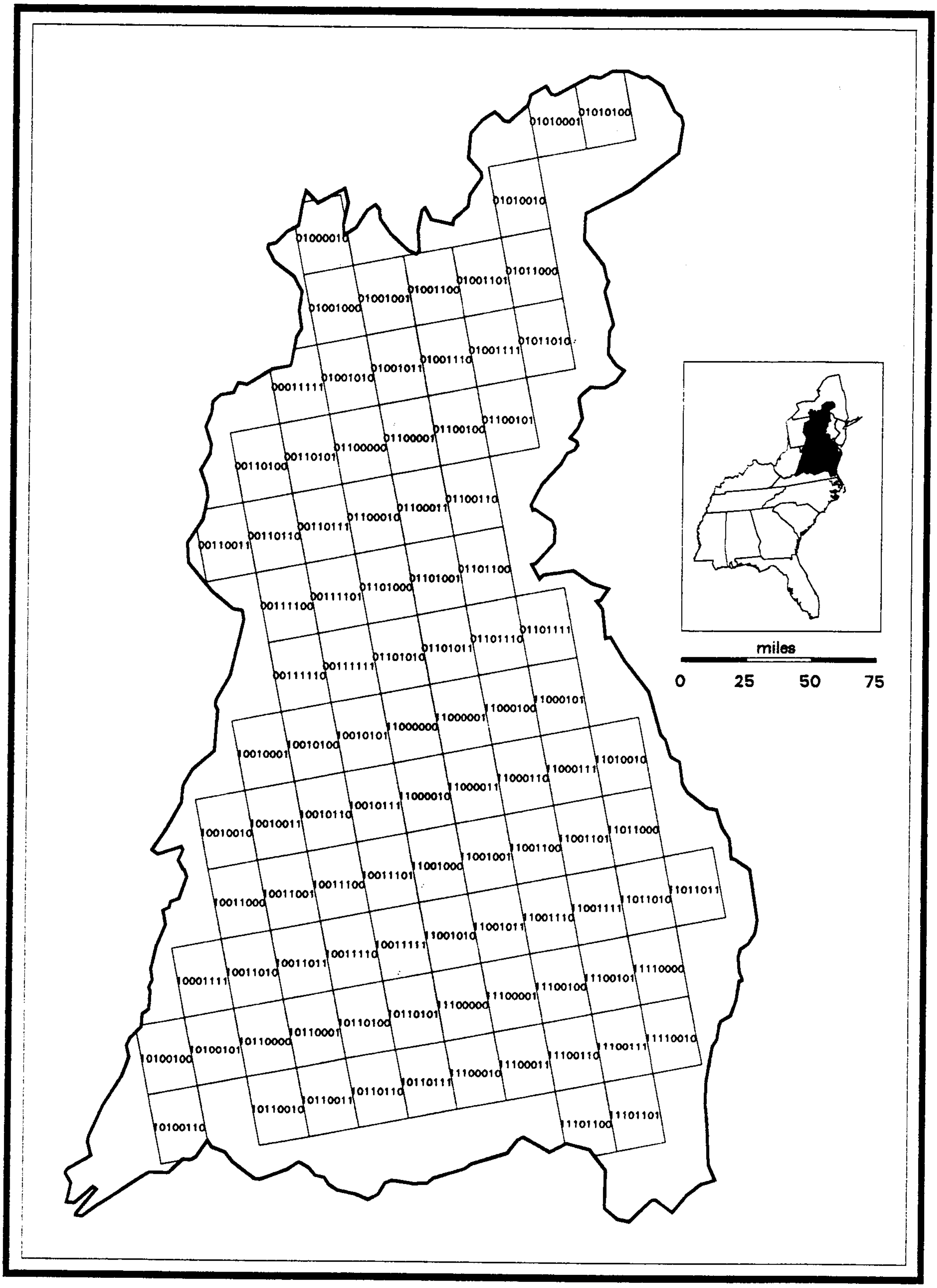


Figure 3-4: Chesapeake Bay Watershed Divided into 97 Quadrants.

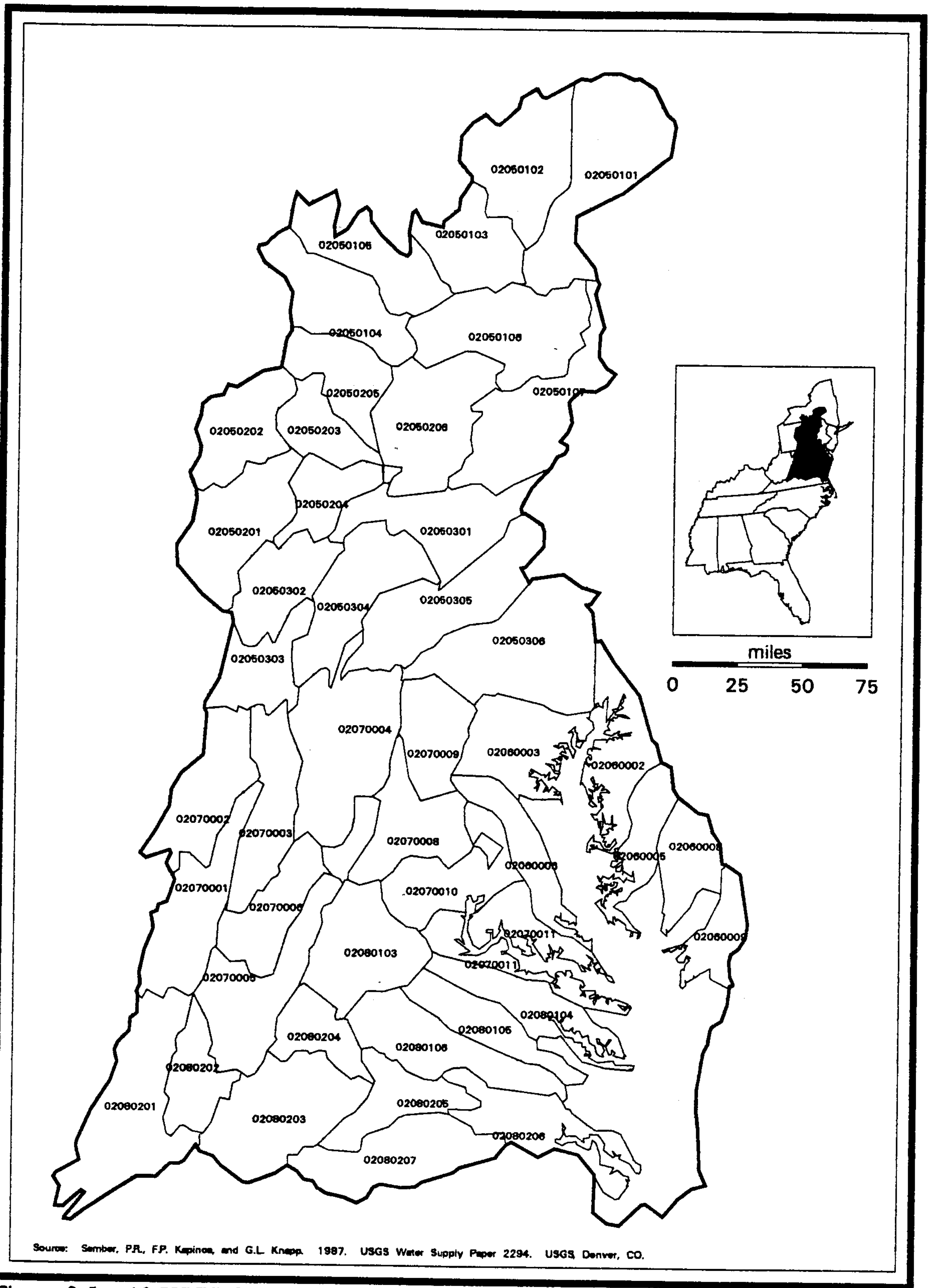


Figure 3-5: 46 USGS 8-Digit Hydrologic Units (Sub-Watersheds) of the Chesapeake Bay Watershed.

1990), and *r.le* (Baker and Cai, 1992). All of these programs analyze image pixels, edges between pixels, and patch attributes, and use this information to calculate several different metrics quantifying different attributes of the image. I used a program developed by Riitters *et al.* (1995) to calculate metrics for each subset map.

FACTOR ANALYSIS

Factor analysis is a method that reduces a large set of variables to a smaller set of 'factors', for which new scores can be created for the data analyzed. Johnston (1980) suggests that replacing one set of variables with another is desirable in order 1) to identify groups of inter-correlated variables; 2) to reduce the number of variables being studied; and/or 3) to rewrite the data set in an alternative form. Factor analysis finds the common variance shared by different variables and then uses these patterns to correlate them to a series of vectors, or factors. If many variables correlate highly with each other, then a single vector in geometric space can embody them. A rotation of the factors can define groups better, and many rotation methods are available. Ultimately a table of variables' loadings upon factors is produced, and groups of similarly-behaving variables can be identified. Another method similar to factor analysis is principal components analysis, which makes no assumption about common and unique variance. I investigated both techniques for use in this study.

In addition to which method of factor analysis to use, other considerations were how many factors to extract and what type of rotation to use. An eigenvalue is the sum of the squared loadings of each variable on a given factor. When expressed as a proportion by dividing by the total number of variables, this reveals the variance explained by the factor. It is common practice to select factors with eigenvalues greater than one when determining how many factors to extract (Johnston, 1980). Another practice is to select enough factors to explain a certain portion of the total variance. I selected enough factors to explain at least 95% of the variance in the TRW original set, and at least 90% of the variance in the other map sets.

Two main types of rotation methods exist, oblique and orthogonal. In geometric terms, oblique rotations allow factors to be built at less than 90 degrees from each other. This means that the factors may be correlated to some extent. An orthogonal rotation forces the factors to be at right angles, or to be entirely uncorrelated. This may not make sense if two factors should be correlated (for example, education and income), so I examined different types of rotations.

COEFFICIENT OF CONGRUENCE

Coefficients of congruence offer a method of comparing different factors or components generated from different data sets. Two or more matrices of loadings which use common sets of variables can be compared to determine which factors are consistently similar, or stable. Coefficients of congruence are basically correlations of the vectors of the factor or component loadings, calculated as follows (Johnston, 1980):

$$G_{ij} = \frac{\sum_{k=1}^{n} L_{ki} L_{kj}}{\sqrt{\sum_{k=1}^{n} L_{ki}^2 \sum_{k=1}^{n} L_{kj}^2}}$$

where

L_{ki} and L_{kj} are the factor/component loadings for variable k in matrix i and matrix j, respectively;

n is the number of variables; and

G_{ij} is the coefficient of congruence between matrices i and j.

A matrix of the coefficients can be constructed to compare how similar the factors or components from different data sets are to each other. The measure approaches a value of one when the loadings are proportional, and is effective in revealing a resemblance that may not be apparent due to the arbitrariness of the rotations (McDonald, 1985). For example, if two factor analyses were compared using the coefficient of congruence and they were very similar, they would have high values across the diagonal of the congruence matrix. This would indicate that corresponding factors or components had similar loadings for each of the variables in the analysis. Coefficients of congruence were used in this study to compare the different factor analyses and to determine similarities, if any, between them.

SELECTION OF REPRESENTATIVE METRICS

Once the factor analyses were complete, I attempted to identify groups of metrics that loaded together on the factors, and selected from those groups the metrics which correlated most strongly with the factor. The selection was based on the consistency of metrics in loading on a particular factor, and in loading together on different factors, and also on the congruence between factors. Metrics were selected to adequately represent each significant factor as closely as possible.

CLUSTER ANALYSIS

Cluster analysis is a method of classification which groups observations together based on minimizing within-group variation and maximizing between-group variation. Clusters of subset maps were created based on the factor scores from the factor analyses of the land cover subset maps, so as to capture as much variance as possible. The cluster analyses were done so that the ability of certain metrics to separate maps into groups could be tested using discriminant analysis.

DISCRIMINANT ANALYSIS

Discriminant analysis allows a set of variables to be tested for their ability to classify or discriminate observations into previously created groups (Johnston, 1980). In other words, it tests how well the variables can distinguish groups of observations, which may have been arbitrarily created (such as political regions), or in this case created using cluster analysis. The main advantages of this technique are that it reveals observations that the variables misclassify, and that it produces a table of the proportion of observations classified correctly.

Two procedures were used, a stepwise analysis, and one that used all input variables as a complete set. The stepwise method tested for statistical significance of each input variable by incorporating one variable at a time in order of its level of discriminatory power as measured by Wilks' *lambda* (SAS, 1982). Results indicate whether a variable contributes significantly to discriminating between groups.

The goal of the procedure that uses all the input variables is to determine how well the set of variables differentiates between groups (which in this case were defined using cluster analysis). The procedure indicates the number of misclassified observations, and these can be examined in an attempt to find reasons for the failure of the set of variables to distinguish between groups. I used discriminant analysis to test the ability of the selected subset of metrics to predict membership in the clusters. This in effect determined how well the subset of metrics represented all of the metrics used in this study.

In summary, to test the stability of landscape metrics across different spatial resolutions, numbers of attributes, and methods of boundary determination, I created several data sets from two land cover maps of the Tennessee River Watershed and the Chesapeake Bay Watershed, and calculated metric scores for each subset map in each data set. I used several multivariate statistical methods, first, to identify which landscape metrics were similar, and then, to test the ability of a subset of those metrics to accurately represent all of the metrics. Chapter IV presents and discusses the results of the analyses.

CHAPTER IV

RESULTS AND DISCUSSION

Correlations among the landscape metrics indicated that relationships exist among several sets of metrics, particularly those measuring texture/fragmentation and those measuring fractal complexity. Results from the factor analyses and examination of the coefficients of congruence confirmed these relationships. I selected a subset of representative metrics, and, using discriminant analysis, determined that the subset was effective in re-creating groups formed by clustering the maps based on full set factor scores.

SUMMARY STATISTICS

Tables B-2 through B-8 give simple summary statistics for each data set. Many of the metrics did not have normal distributions, and as shown in histograms (Figure B-1 and B-2), some metrics had extreme outliers. The coefficient of variance is an expression of the standard deviation as a percentage of the mean, which presents a measure of relative variability (Silk, 1979), and I included it to provide a comparison between metrics within a particular map set. Because the Tennessee River Watershed (TRW) high-resolution majority-rule set was resampled using a window size of 5 by 5 cells, but the pixel size did not change, the metric measuring the proportion of patches encompassing more than five pixels (P005) had zero variance since *all* patches were greater than the 5-cell minimum (see Appendix A).

Tables B-9 through B-15 show the correlation matrices for the data sets. Several metrics consistently correlated strongly in all cases. This should be a good predictor of some of the stronger relationships that can be expected to emerge from the factor analyses. The Shannon and Simpson indices consistently correlated at 0.9 or higher, which suggested they would load together in factor analysis, and in fact they did. Sum of diagonals (SUMD), number of patches (NPAT), largest patch size (LPAT), and average patch size (PSIZ) tended to correlate highly with the previously mentioned indices, but not always above 0.9. In the Chesapeake Bay Watershed (CBW) quadrants and sub-watersheds data sets, the equivalent correlations of the same set of metrics dropped noticeably, indicating a possible regional difference, or perhaps a difference in the classification of the images. The fractal measures of perimeter-area scaling, by edges and by cells (OEFC, OCFC), correlated at 0.8 or higher with each other, and they also loaded together in the factor analysis.

FACTOR ANALYSIS

Johnson and Wichern (1982) recommend performing principal components and maximum likelihood factor analyses for comparison to test the stability of the results. I investigated different methods of factor analysis and several rotations using metrics calculated for the TRW original set. The initial analysis used a principle components method with a varimax rotation and no specification of the number of components. The first five components had eigenvalues above one, and six components explained 95% of the variance.

After that initial analysis, I ran principal components analysis and maximum-likelihood and principal-factor methods of factor analysis with varimax, promax, and quartermax rotations, extracting five and six factors. Loadings for the different methods and rotations were similar (see Table 4-1). The promax rotation showed that forcing orthogonality on the factor structure did not impose a false relationship on the variables. With this information, I chose the principal components method using varimax rotation for analysis of all data sets, extracting six components because this number of components explained 95% of the variance in the TRW original set. In subsequent data sets, six components contributed to cumulative proportions of variances explaining 92% to 97%. For the purposes of this paper, I use the terms *components* and *factors* interchangeably.

Tables C-1 through C-7 present the results of the factor analyses of the different map sets. A summary table (Table 4-2) illustrates high and moderate loadings of metrics on factors in different data sets. Using the factor matrices to detect metrics which tended to group together across different data sets, I found four general groups. The first group describes texture (contagion/diversity). As expected from their high correlations, the Simpson and Shannon indices (SIDI, SIEV, SHDI, SHEV, SIDA, SICO, SHDA and SHCO) and the metric measuring highest land cover proportion (PMAX) load strongly together, and they consistently load on the first factor. Additionally, in all the TRW data sets, indices measuring the sum of diagonals (SUMD), number of patches (NPAT), largest patch size (LPAT), average patch size (PSIZ), and proportion of patches encompassing

Table 4-1: Comparison of different methods of factor analysis, rotation, and number of factors for the Tennessee River Watershed original data set.

	FACTOR LOADING								
METRIC	pcn28v	pcn6v	pcn6q	pcn6p	pfn6v	pcn5v	pcn5q	pcn5p	mln5v
NTYP	G	F	F	F	e	E	E	E	nhl
PMAX	A	A	A	A	A	A	A	A	A
SIDI	A	A	A	A	A	A	A	A	A
SIEV	A	A	A	A	A	A	A	A	A
SHDI	A	A	A	A	A	A	A	A	A
SHEV	A	A	A	A	A	A	A	A	A
SUMD	A	A	A	A	A	A	A	A	A
SIDA	A	A	A	A	A	A	A	A	A
SICO	A	A	A	A	A	A	A	A	A
SHDA	A	A	A	A	A	A	A	A	A
SHCO	A	A	A	A	A	A	A	A	A
FDDA	A	A	A	A	A	A	A	A	A
NPAT	af	A	ab	A	A	A	af	A	aD
LPAT	A	A	A	A	A	A	A	A	A
PSIZ	a	A	ab	A	A	A	ab	A	ad
P005	af	ab	aB	A	ab	ab	aB	A	D
PA-1	B	B	B	B	B	B	B	B	B
DSTA	B	B	B	B	B	B	B	B	B
PA-2	d	d	ad	d	c	c	C	C	ac
NASQ	D	D	d	D	C	C	C	C	C
RGYR	B	B	B	B	B	B	B	B	b
PTRD	C	C	aC	C	D	D	D	D	E
ABRR	D	D	D	D	C	C	C	C	C
ABSR	bc	bc	C	bC	bd	bd	D	bD	be
ACCR	bc	bc	abc	bc	bd	bd	abd	bd	be
ALAR	C	C	C	C	D	D	D	D	E
OCFC	E	E	E	E	ce	C	C	C	c
OEFC	bE	bE	be	e	bc	C	C	C	c

Notes: See Appendix A for explanation of metric abbreviations

See Table B-1 for description of data set.

Letters A - G correspond to Factors 1 - 7

Uppercase letters signify a loading of 0.7 or higher

Lowercase letters signify a loading between 0.5 and 0.7

nhl — No High Loadings

pc — principle components analysis

pf — principle factor analysis

ml — maximum likelihood factor analysis

n# — number of factors

v — varimax rotation

q — quartimax rotation

p — promax rotation

Table 4-2: Summary of factor analysis results for seven data sets using principal components analysis with varimax rotation.

	DATA SET						
METRIC	T12_25	T12_125M	T12_25M	T12_125N	T5_25	C5_25Q	C5_25W
NTYP	F	E	E	f	F	F	nhl
PMAX	A	A	A	A	A	A	A
SIDI	A	A	A	A	A	A	A
SIEV	A	A	A	A	A	A	A
SHDI	A	A	A	A	A	A	A
SHEV	A	A	A	A	A	A	A
SUMD	A	A	A	A	A	d	A
SIDA	A	A	A	A	A	A	A
SICO	A	A	A	A	A	A	A
SHDA	A	A	A	A	A	A	A
SHCO	A	A	A	A	A	A	A
FDDA	A	F	A	F	A	aD	a
NPAT	A	A	A	A	a	B	b
LPAT	A	A	A	A	A	d	af
PSIZ	A	A	A	A	ae	B	B
P005	ab	A	nhl	A	ac	B	B
PA-1	B	c	B	C	C	B	Bc
DSTA	B	cd	D	C	C	B	Bc
PA-2	d	B	B	B	ad	C	E
NASQ	D	B	B	B	B	C	d
RGYR	B	B	B	B	E	B	B
PTRD	C	C	bD	D	B	B	C
ABRR	D	B	B	B	b	cd	D
ABSR	bc	C	B	cd	bc	B	C
ACCR	bc	B	B	Bd	B	C	b
ALAR	C	bC	B	D	B	b	C
OCFC	E	D	C	E	D	be	B
OEFC	bE	D	C	E	D	B	B

Notes: See Appendix A for explanation of metric abbreviations
See Table B-1 for explanation of data set abbreviations
Letters A - F correspond to Factors 1 - 6
Uppercase letters signify a loading of 0.7 or higher
Lowercase letters signify a loading between 0.5 and 0.7
nhl — no high loadings

more than five pixels (P005) also tend to load on the first factor, but NPAT, PSIZ, and P005 shift to the second factor and group with the perimeter-area scaling indices (OCFC, OEFC) and other metrics in the CBW data sets, and SUMD and LPAT shift to the fourth factor in the CBW quadrants set (Tables C-6 and C-7). As mentioned above, this could indicate a difference between the regions, or it could be a result of the classification differences between the two map sets.

The metric measuring density-area scaling (FDDA) lost importance and shifted to the sixth factor in the TRW low-resolution resampled data sets, which could reflect sensitivity to scale. It is important to keep in mind that this metric exhibited extreme outliers for these two map sets (Figure B-1).

Second in importance in explaining variance, usually loading in the second and third factors, is a group which describes patch shape and compaction. Metrics that fall into this group are perimeter-area ratio (PA-1), Gardner's D-statistic (DSTA), adjusted perimeter-area ratio (PA-2), normalized area (NASQ), radius of gyration (RGYR), patch topology (PTRD), bounding rectangle ratio (ABRR), boxside ratio (ABSR), circumscribing circle ratio (ACCR), and longest axis ratio (ALAR). This group is not very stable, however, in that metrics load on the second and third factors in different map sets without consistency. In each case, the two factors are probably measuring a similar dimension, but the results of the factor analyses may be skewed by large differences in some metrics, like largest patch size (LPAT), average patch size (PSIZ), and number of patches (NPAT). The scores for these metrics vary widely due to the differences in scale and number of classes in the data sets.

Certain metrics do tend to load on the same factor in this patch shape/compaction group. Perimeter-area ratio (PA-1) and Gardner's D-statistic (DSTA) match up well together, but seem sensitive to change in numbers of attributes and to scale differences. They move from the second factor in the TRW original set and in both CBW sets to the third factor in the TRW 5-class set and the TRW resampled sets. Patch topology (PTRD), boxside ratio (ABSR), and longest axis ratio (ALAR) load together in most of the data sets, but they shift from factor to factor, and seem sensitive to recoding, resampling, and different methods of boundary determination. Adjusted perimeter-area ratio (PA-2), normalized area (NASQ), and bounding rectangle ratio (ABRR) usually load together on the same factor, and they load strongly on factor 2 in all the resampled sets. This indicates that these metrics may be resistant to different methods of aggregation. Radius of gyration (RGYR) and circumscribing circle ratio (ACCR) tend to behave in parallel, but differ in the recoded map set and in the CBW quadrants set. They load consistently, however, across the resampled data sets, which could be a result of the diminishing detail and complexity causing similar reactions from each metric, since both compare patch shape and compaction to the shape of a circle (see Appendix A).

Third in importance is a fractal-dimension group. The perimeter-area scaling indices (OCFC, OEFC) group together consistently, and usually without other metrics loading heavily on the same factor. Across map sets they tend to move from factor to factor. The fact that they load on the second factor in both CBW data sets indicates that they may be stable across different methods of determining boundary. High correlations between the two metrics predicted this group.

Number of types (NTYP) consistently loaded strongly on the fifth or sixth factor by itself in all of the data sets except the CBW sub-watersheds data set. Its importance is therefore the least in explaining the variance of the map sets, but it could be significant especially on the two low-resolution data sets and the TRW 5-class and CBW sub-watersheds data sets because the communality for NTYP drops below 0.8 (Tables C-2, C-4, C-5 and C-7). The communality value shows the proportion of variance that a variable has in common with other variables, and the lower values for these data sets indicate that NTYP is not as redundant a measure at low resolution or with only a few types of land cover. The number of types could be an important measure of transition zones in a landscape, where natural and man-made land covers mix together, but this should also be reflected in diversity measures. Since this metric is used in calculation of diversity/contagion metrics, I had expected that it would load on the same factor with them.

I generated factor scores for each map set, and I mapped the scores for the first factor (Figures C-1 through C-7). Spatial patterns for the first factor are similar in data sets for both study areas. The only noticeable difference among the TRW data sets is that the TRW original set has similar patterns to the other map sets covering the same area, but as a photographic negative. The scores for this map are opposite in sign, which only means that the variables loaded on the opposite ends of the factor, a common, non-significant happening in factor analysis. The important thing is that the patterns are similar to patterns in the other map sets for the region. In the CBW quadrants set, a rough pattern emerges which is similar to the pattern seen in the CBW sub-watersheds set.

Factor scores for subsequent factors did not create reproducible spatial patterns, and this further indicates that some metrics are not loading consistently together or on the same factor.

COEFFICIENT OF CONGRUENCE

The coefficients of congruence support the results of the factor analyses. Table 4-3 shows all the diagonals of the congruence coefficients for each factor in each data set. Factor 1 shows high congruence across all the map sets, indicating that variables are loading similarly on all of the first factors. Factor 2's congruence matrix shows a split, where the second factors for the TRW original set and the two CBW data sets are similar, and the TRW 5-class and resampled data sets all have high congruence on factor 2. When comparing corresponding factors, congruence deteriorates after the second factor, with no discernible patterns showing across all seven data sets.

This may be partly because when some data sets are compared to each other, more than one element of difference is being introduced. For example, if one of the TRW resampled data sets was compared to one of the CBW data sets, comparisons are being made across different spatial scales, numbers of attributes, and different regions. It is difficult to separate out how each difference is affecting the outcome of the coefficient of congruence.

As first determined by the examination of the factor analyses, the coefficients of congruence confirm that the first factor is stable across all data sets, and that some similar metrics are loading on the second factor for some of the sets, but most of the metrics are

Table 4-3: Coefficients of Congruence for six factors among seven data sets. See Table B-1 for explanation of data set abbreviations.

DATA SET	T12_25	T12_125M	T12_25M	T12_125N	T5_25	C5_25Q	C5_25W
				FACTOR 1			
T12_25	1.000						
T12_125M	**-.911**	1.000					
T12_25M	**-.958**	**.872**	1.000				
T12_125N	**-.905**	**.999**	**.867**	1.000			
T5_25	**-.991**	**.913**	**.958**	**.908**	1.000		
C5_25Q	**-.885**	**.915**	**.862**	**.912**	**.913**	1.000	
C5_25W	**-.897**	**.926**	**.868**	**.924**	**.922**	**.993**	1.000
				FACTOR 2			
T12_25	1.000						
T12_125M	-.429	1.000					
T12_25M	-.376	**.818**	1.000				
T12_125N	-.334	**.991**	**.800**	1.000			
T5_25	.071	**.646**	**.799**	**.653**	1.000		
C5_25Q	**.804**	-.293	-.082	-.236	.393	1.000	
C5_25W	**.795**	-.377	-.200	-.318	.226	**.937**	1.000
				FACTOR 3			
T12_25	1.000						
T12_125M	**.655**	1.000					
T12_25M	-.015	-.242	1.000				
T12_125N	.003	**.632**	-.334	1.000			
T5_25	.127	**.608**	-.340	**.865**	1.000		
C5_25Q	**.607**	.361	-.136	-.083	-.110	1.000	
C5_25W	**.615**	**.939**	-.170	**.621**	**.641**	.358	1.000
				FACTOR 4			
T12_25	1.000						
T12_125M	-.455	1.000					
T12_25M	.325	-.477	1.000				
T12_125N	.220	-.298	**.627**	1.000			
T5_25	**-.724**	**.817**	-.243	-.206	1.000		
C5_25Q	-.086	-.016	-.148	-.229	-.248	1.000	
C5_25W	**.681**	-.311	.105	.269	-.412	**-.632**	1.000
				FACTOR 5			
T12_25	1.000						
T12_125M	-.069	1.000					
T12_25M	.213	**.686**	1.000				
T12_125N	**.920**	-.023	.189	1.000			
T5_25	-.319	.274	-.024	-.191	1.000		
C5_25Q	.423	-.033	.018	.427	-.280	1.000	
C5_25W	.292	.517	.354	.181	.302	.110	1.000
				FACTOR 6			
T12_25	1.000						
T12_125M	-.068	1.000					
T12_25M	.353	.172	1.000				
T12_125N	-.463	**.847**	.110	1.000			
T5_25	**.764**	-.093	-.214	-.453	1.000		
C5_25Q	**.808**	-.106	.009	-.478	**.694**	1.000	
C5_25W	.089	-.050	-.545	-.218	.344	.480	1.000

not stable in loading together or on the same factor (with the exception of the perimeter-area scaling metrics, which do load together consistently).

COMPARISON TO PREVIOUS STUDY

As mentioned in Chapter I, Riitters *et al.* (1995) conducted a factor analysis using metrics calculated for LUDA maps. Explaining about 87% of the variation in 26 landscape metrics, the analysis extracted six factors as dimensions of pattern and structure. The authors interpreted these factors as composite measures of average patch compaction, overall image texture, average patch shape, patch perimeter-area scaling, number of attribute classes, and large patch density-area scaling. They selected metrics which correlated highly with each of the factors and suggested them for use as representative metrics. These metrics were average perimeter-area ratio (PA-1), contagion (SHCO), standardized patch shape (NASQ), patch perimeter-area scaling (a metric derived from OCFC), number of attributes (NTYP), and large patch density-area scaling (a fractal estimate of patch mass, not used in this study) (Riitters *et al.*, 1995).

Comparison of the results of this study and the study by Riitters *et al.* show similarity in the texture and fractal perimeter-area scaling dimensions. The lack of consistency for the other dimensions demonstrates that many of the metrics are unstable in their reactions to different spatial scales, numbers of land cover classes, and methods of boundary determination. One reason for this instability could be the subjectivity involved in the production of LUDA maps. Interpretation of land cover is totally manual, and the resolution is coarser than in TM-derived land cover maps. Shape and compaction do emerge from the analysis of the seven different data sets, but metrics measuring these dimensions often load on the same factor, indicating that the same dimension is perhaps being measured in different ways. This does not show up in the LUDA map study because Riitters *et al.* conducted a factor analysis on only one data set.

SELECTION OF METRICS

To select a subset of representative metrics that would sufficiently account for variation in a landscape, I first rejected metrics which seemed particularly unstable or did not consistently have strong loadings across several data sets. Once again referring to Table 4-2, density-area scaling (FDDA) reacted strongly to the change in resolution in the two resampled, compressed data sets. Proportion of area with patches containing more than five pixels (P005), largest patch size (LPAT), adjusted perimeter-area ratio (PA-2), bounding rectangle ratio (ABRR), boxside ratio (ABSR), circumscribing circle ratio (ACCR), and number of patches (NPAT) all loaded only moderately in two or more data sets, indicating that their influence on the variance in the observations was inconsistent and unstable on at least one type of difference (resolution, number of classes, zones/regions). Number of types (NTYP) almost always loaded strongly on the same factor, but that factor explained only a small portion of the variance, and was probably reflecting itself.

Texture, compaction/shape, and fractal complexity are dimensions that seem to emerge from the factor analyses with some consistency, so I selected metrics which measured these aspects of landscape structure and pattern. Since the texture dimension accounted for such a large proportion of the variance in all the data sets, I selected two metrics from this group, highest land cover proportion (PMAX) and Shannon contagion (SHCO). I selected these two metrics from all the highly loading metrics to ensure that most of the variance in this factor was captured because these were the least correlated with each other, and were on opposite sides of the factor in geometric space.

The congruence analyses showed a splitting into two subgroups of high similarity for the second factor, which I previously determined was a composite of the separate compaction and shape dimensions found in the study by Riitters *et al.* (1995). One group included the TRW original set and both CBW sets, and the other included the TRW 5-class and resampled data sets. To ensure that characteristics of both groups were represented, I selected a strongly-loading metric for each subgroup. Gardner's D-statistic (DSTA) and normalized area (NASQ) loaded either highest or second-highest most frequently for the first and second subgroups, respectively, and were selected to represent the shape and compaction dimension.

The final discernible dimension that emerged from the factor analyses was the fractal dimension. Perimeter-area scaling by cells (OEFC) was selected over perimeter-area scaling by edges (OCFC) because it loaded higher more often. The importance of this metric may vary significantly, because fractal dimension varied in its importance in explaining variance across the different map sets.

CLUSTER ANALYSIS

To find out which subset maps in each set were similar to each other, I input factor scores into cluster analyses of the map sets. I used Ward's minimum variance method of clustering, which adds up, over all variables, the ANOVA sum of squares, or euclidean distance, between two clusters, and then follows a hierarchical method of merging observations one at a time to minimize within-cluster sum of squares (SAS, 1982). The process is subjectively stopped at a point where a large increase in the semi-partial r-squared value occurs. The semi-partial r-squared value is the decrease in the proportion of variance accounted for by joining two clusters (SAS, 1982). Tables C-8 through C-14 show the clustering histories used to determine the number of clusters kept for each data set.

Figures 4-1 through 4-7 show the clusters for the different data sets. While cluster analysis does not group observations together based on any spatial criteria, spatial grouping is apparent in the seven data sets. For example, in the TRW 5-class (Figure 4-5), one can see spatial patterns in the clusters which roughly correspond to the physiographic regions (Figure 2-2). This demonstrates that the metrics used to create the factor scores successfully distinguish regions and biomes using landscape pattern and structure. Examining the cluster means for this data set reveal that clusters have characteristics that correspond with the patterns in the land cover map (Table 4-4).

Cluster 1 has the highest average score on factor 2, which in this data set represents metrics which measure patch shape and compaction (see Table C-5). This cluster is represented by mostly forested or herbaceous land cover, and topography is flatter than in other parts of the Tennessee River Watershed.

Cluster 2 has a large negative average score for factor 2, indicating more complex shapes. The mean score for factor 1 (texture/contagion) is not very high, so this cluster

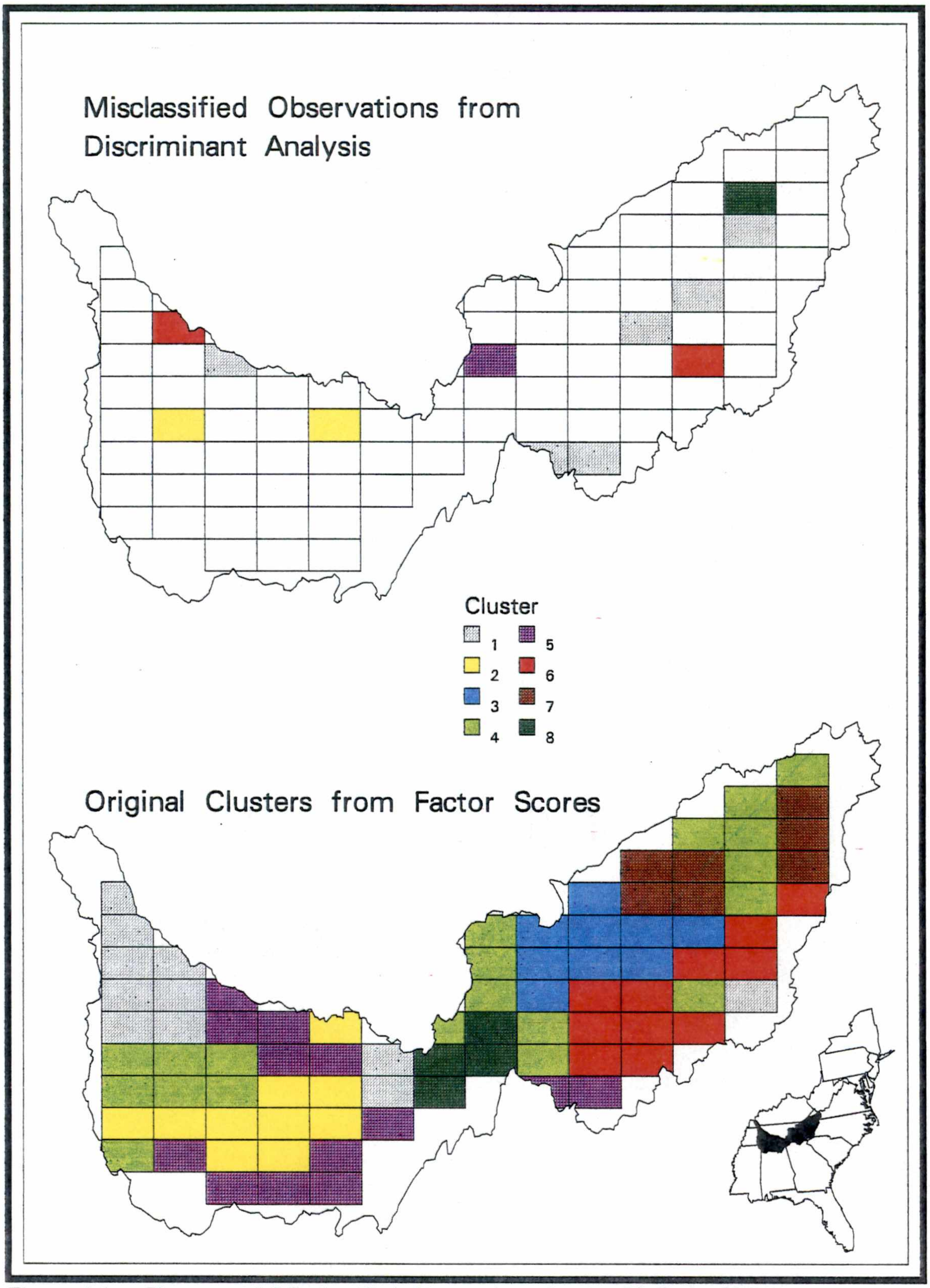


Figure 4-1: Clusters and Misclassified Observations for the 12-Class, 25-Meter-Resolution Tennessee River Watershed Data Set.

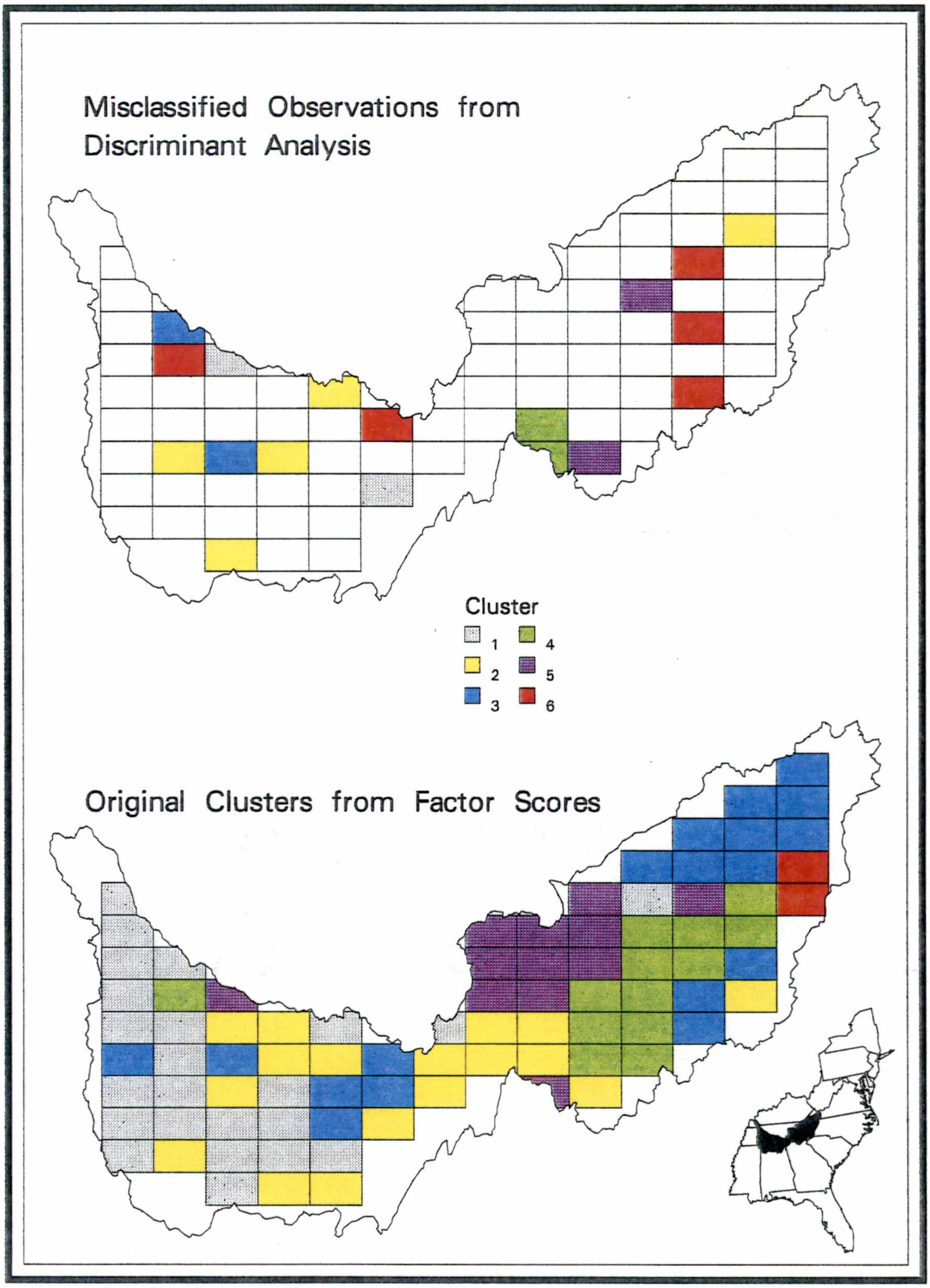


Figure 4-2: Clusters and Misclassified Observations for the 12-Class, 125-Meter-Resolution, Majority-Rule-Resampled Tennessee River Watershed Data Set.

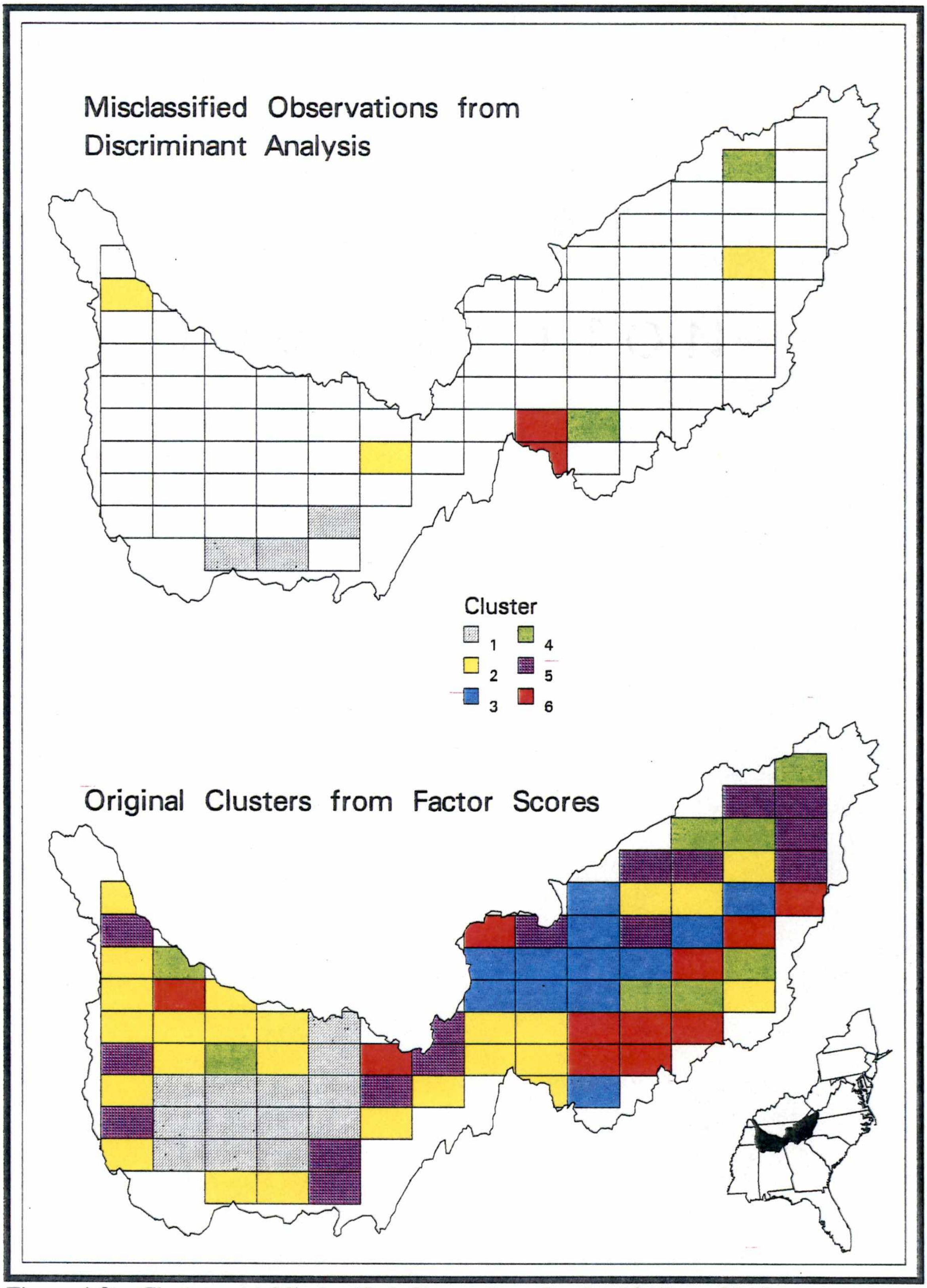


Figure 4-3: Clusters and Misclassified Observations for the 12-Class, 25-Meter-Resolution, Majority-Rule-Resampled Tennessee River Watershed Data Set.

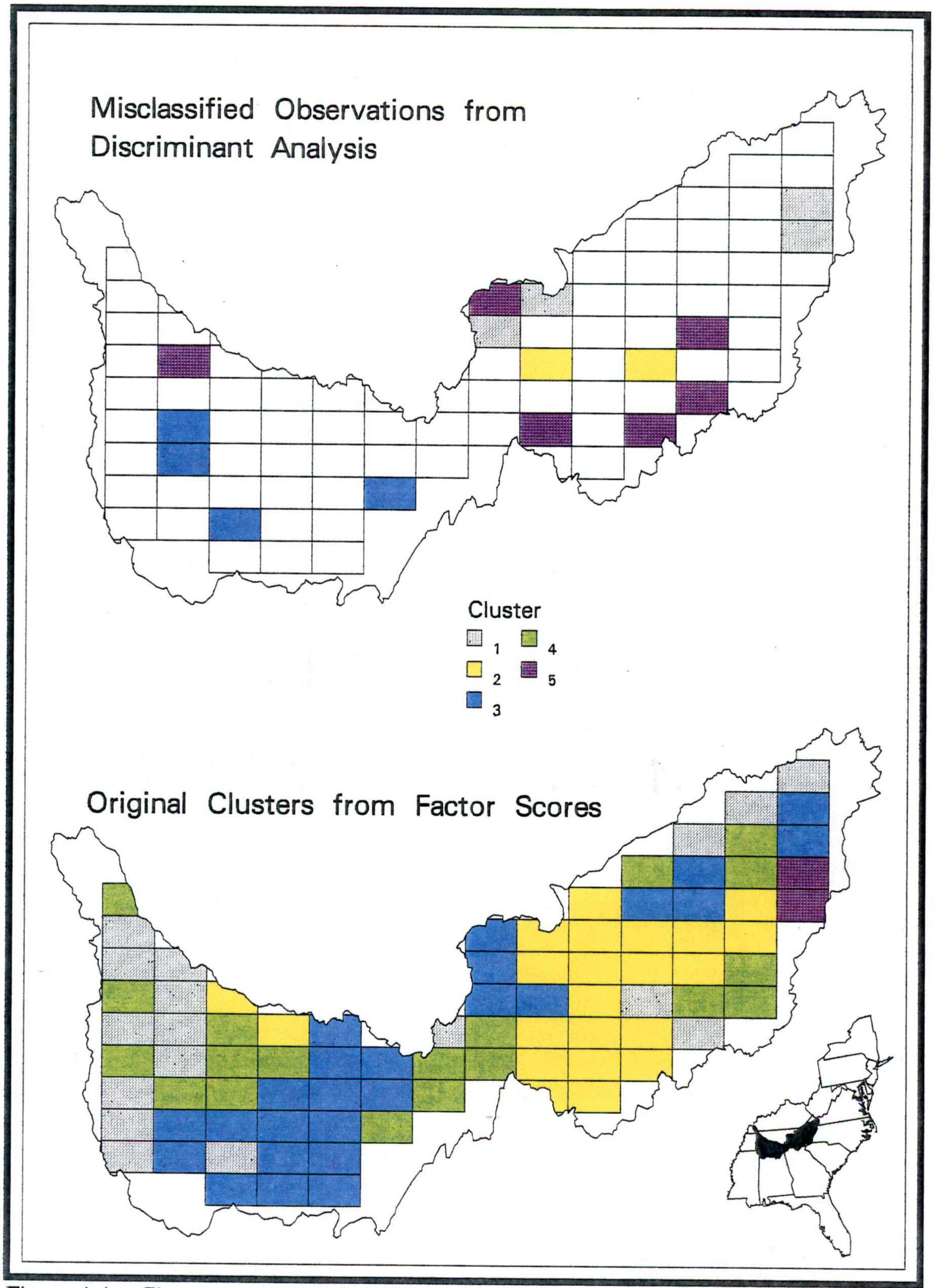


Figure 4-4: Clusters and Misclassified Observations for the 12-Class, 125-Meter-Resolution, Nearest-Neighbor-Resampled Tennessee River Watershed Data Set.

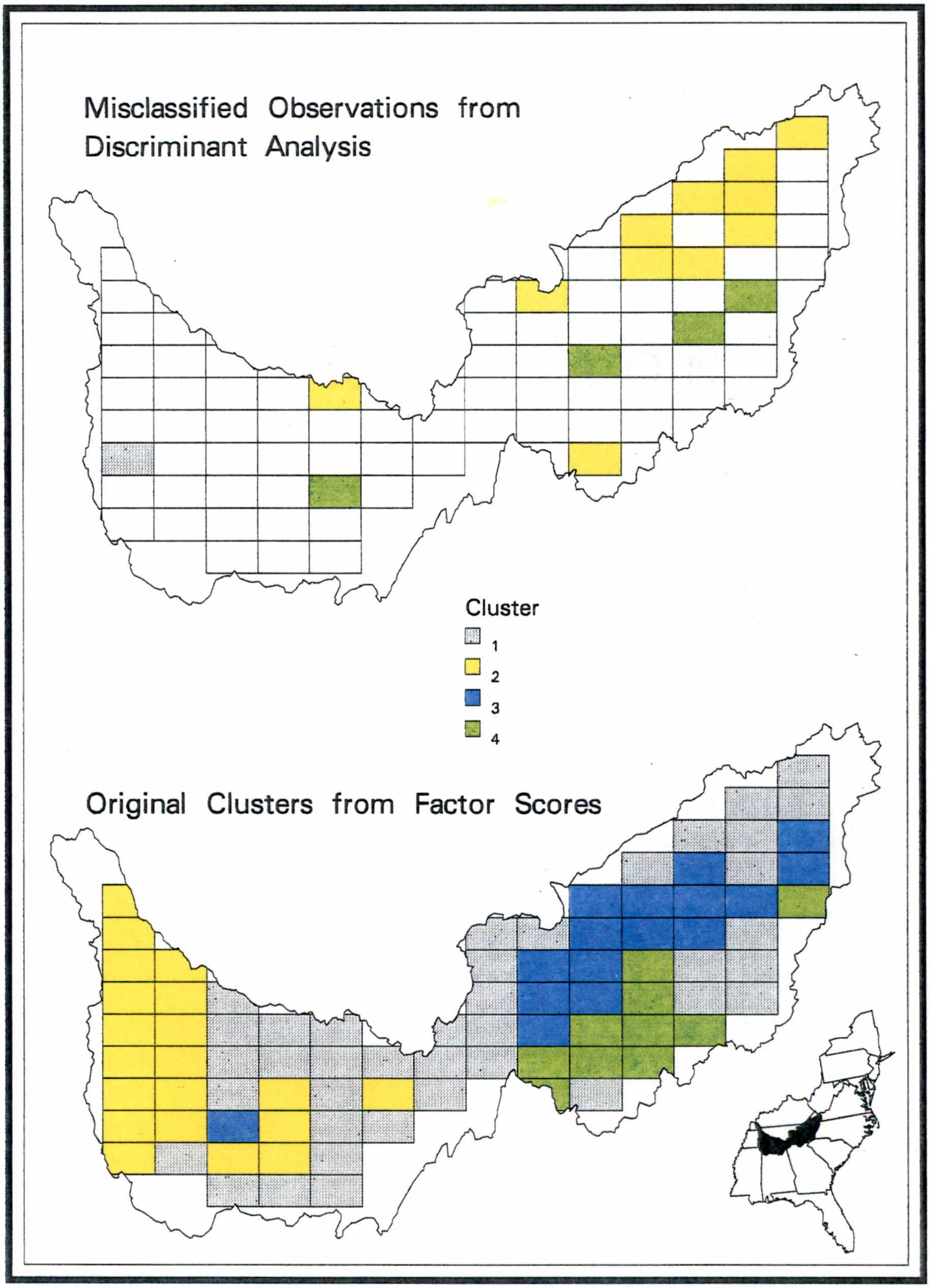


Figure 4-5: Clusters and Misclassified Observations for the 5-Class, 25-Meter-Resolution Tennessee River Watershed Data Set.

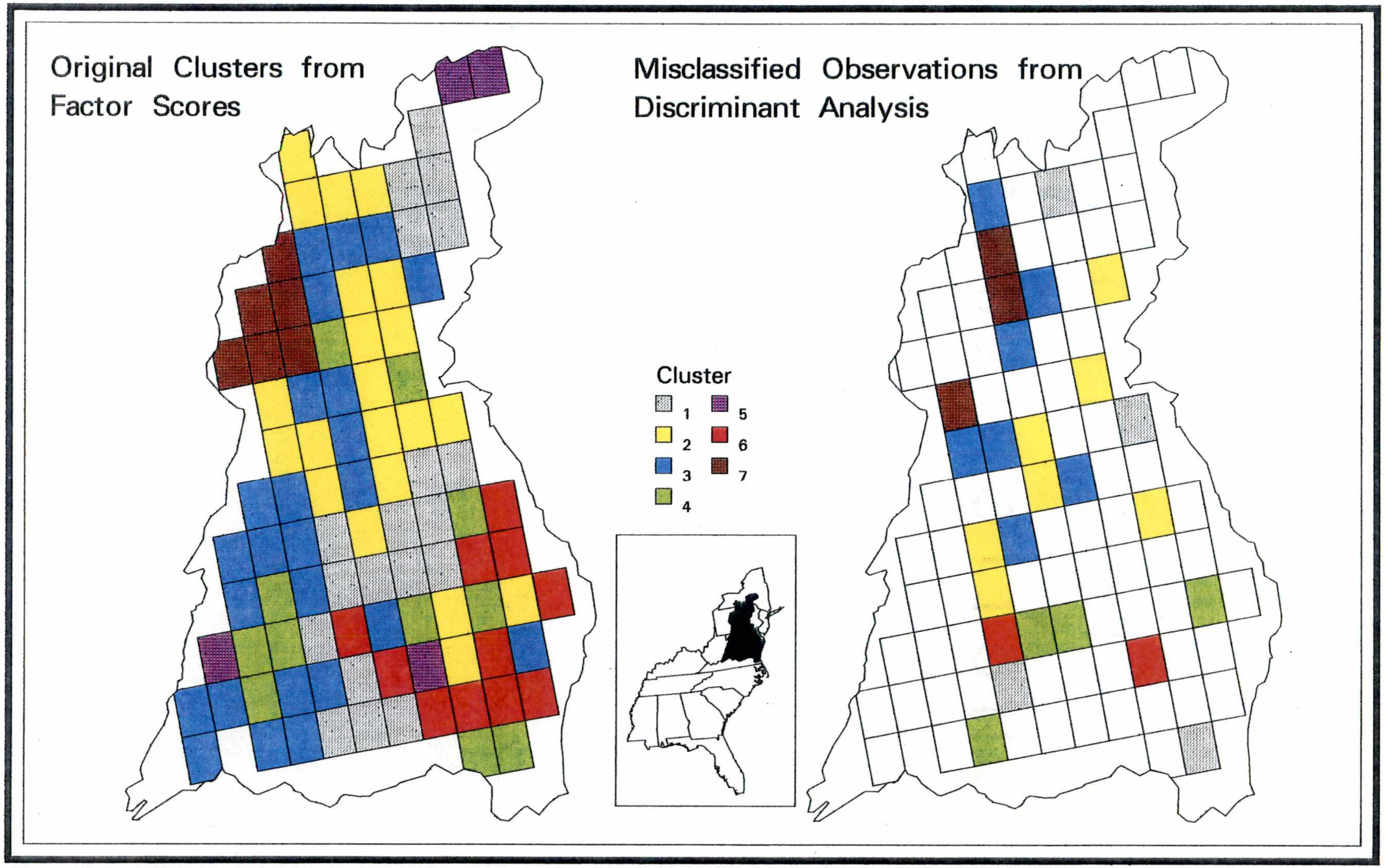


Figure 4-6: Clusters and Misclassified Observations for the 5-Class, 25-Meter-Resolution, Quadrant-Split Chesapeake Bay Watershed Data Set.

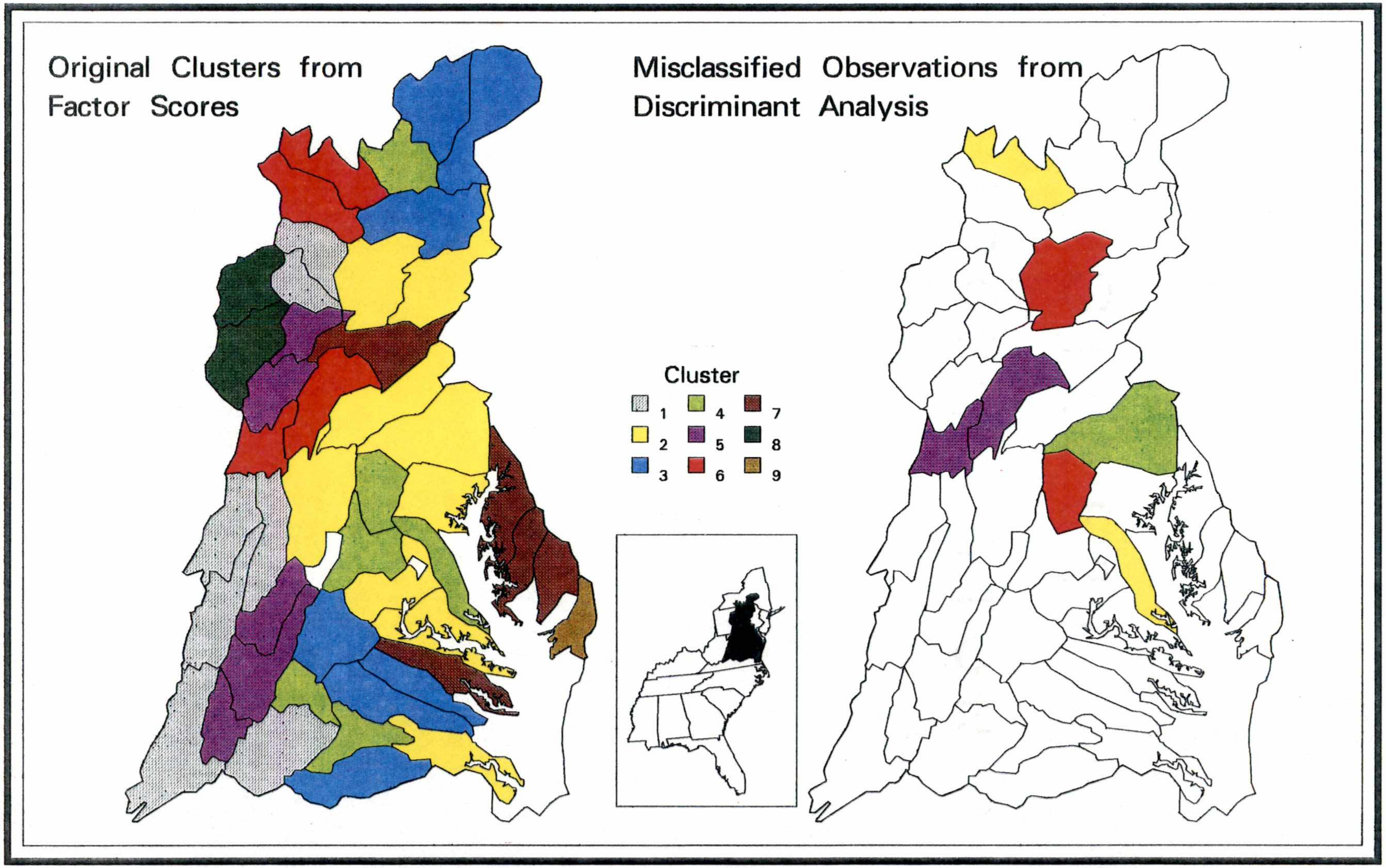


Figure 4-7: Clusters and Misclassified Observations for the 5-Class, 25-Meter-Resolution, Sub-Watershed-Split Chesapeake Bay Watershed Data Set.

Table 4-4: Cluster Means and Standard Deviations for the Tennessee River Watershed 5-class data set. See Table B-1 for description of data set.

Cluster	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Frequency
			Mean Factor Score				
1	.202	.676	.151	.341	.009	.269	39
2	.192	-1.156	.067	.025	.376	.699	20
3	.438	-.292	-.148	-.480	-.234	-1.561	16
4	-1.819	.195	-.460	-.566	-.524	.106	11
Minimum	-1.819	-1.156	-.460	-.566	-.524	-1.561	
Maximum	.438	.676	.151	.341	.376	.699	
			Standard Deviation				
1	.767	.703	.993	1.017	.916	.723	
2	.680	.533	.854	.956	.682	.268	
3	.812	.653	1.076	.716	1.029	.636	
4	.528	.873	1.079	.886	1.528	.737	

may be capturing the more complex nature of the western section of the Tennessee River Watershed, in which the Tennessee River and its tributaries are more meandering, and in which the level of human development is lower than in the east.

The ridge and valley area of the Tennessee River Watershed (Figure 2-2) corresponds with Cluster 3 (Figure 4-5). Cluster 3 has the highest positive texture/contagion score, which is probably due to the fragmented nature of the long corridors of forest and agricultural land covers, and to the high diversity of the mixture of natural and anthropogenic land covers, in the eastern half of the watershed. The fractal dimension in this data set is represented by Factor 4 (Table C-5), and Cluster 3 has a low value for this dimension, reflecting the simpler shapes of the area. This combined with the higher diversity reveal a larger amount of human development, but also reflect the topography.

The final cluster in the TRW 5-class set represents the Appalachian Mountains (Figures 2-2 and 4-5). The texture/contagion (Factor 1) score is lowest in this cluster (Table 4-4), as is the score for the fractal dimension (Factor 4), meaning low diversity of land cover and large, simple patches are characteristics of this cluster. A reason for the lower texture/contagion score could be the lower level of human development due to the rougher topography and to the existence of forest reserves and national parks in this area. The low fractal dimension score could be due to larger patches of forest causing area to have greater influence than perimeter in the calculation of the metrics measuring this aspect (see Appendix A).

While the other cluster maps are not as immediately interpretable, observations in the same clusters do tend to group together spatially. The TRW cluster maps show quite different clustering patterns when compared to each other (Figure 4-1 through 4-5), which may, with the exception of those loading on the first factor, be due to the metrics loading on different factors in different sets. Comparison of the CBW data sets to each other shows similar results, that clustering patterns are different (Figure 4-6 and 4-7). This indicates that factor scores used to create the clusters are not similar, which again supports the notion that different metrics are loading on different factors.

One possible reason for the differences in the cluster maps is that each variable (factor score) in the cluster analyses receives an equal weighting (Johnston, 1980). This implies that each factor has equal weight in the classification procedure, but as mentioned previously, the first two factors accounted for most of the variance. So a cluster analysis which weights the factor scores according to proportion of variance that the corresponding factor explains might yield more appropriate groupings.

It is also possible that the clusters are simply not optimal (assuming that an optimal cluster is defined as one which minimizes within-group variation and maximizes between-group variation). The hierarchical method of determining clusters groups observations to make the minimum change in the overall within-group variation, but once an observation is grouped, it cannot be reassigned to a more appropriate group created later in the process. An observation may actually be closer to the group centroid, or mean location in multi-dimensional space, of one cluster, but has already been classified into another cluster. This may have an impact on discriminant analysis.

Another possible reason that clusters might not be optimal is that the factors used to create the input variables (factor scores) may not be orthogonal despite the rotation forced upon them (Johnston, 1980). For cluster analysis, distances between observations and group centroids are based upon the Pythagorean theorem, which requires dimension axes to intersect at right angles (*i.e.* to be orthogonal) to accurately calculate lengths. If factors are not orthogonal, lengths may be distorted. The results from the initial factor analyses in the TRW original set demonstrated that an oblique rotation did not change the results drastically, so the possibility of inaccurate clustering due to the lack of orthogonality is unlikely.

DISCRIMINANT ANALYSIS

Discriminant analysis on all the data sets allowed me to test how well the chosen subset of metrics could correctly predict clusters created from full-set factor scores. At the 0.15 significance level the stepwise discriminant analyses of the TRW low-resolution nearest-neighbor set and the CBW sub-watersheds set did not show highest land cover proportion (PMAX) as significant, while in the CBW quadrants set, Shannon contagion (SHCO) was not entered into the analysis by the procedure. I still kept these metrics for the sake of consistency in analyses of all data sets using a discriminant procedure that forced inclusion. Tables C-15 through C-21 show how well each data set's clusters were predicted using only the five subset metrics. The CBW sub-watersheds set had the best results, with 85% of the maps correctly classified. The CBW quadrants had the worst results with only 66% correctly classified.

To see how much improvement would occur in classification, I added one more metric, Shannon edge diversity (SHDA). This metric was selected because although as contagion increases, diversity should generally decrease, it is not a completely linear relationship. Contagion will go up as the classes in an image become more connected, but as long as the number of classes and relative proportions of pixels in each class stay the same, diversity will remain unchanged. So the fact that the contagion and diversity metrics loaded on the same factor does not mean that they will always correlate with each other. The addition of the sixth metric, a texture metric, may also contribute more emphasis to the first factor, which accounted for the largest proportion of variance in the factor analyses.

In addition to the metrics not accepted by the previous stepwise discriminant analyses, Shannon edge diversity (SHDA) was not accepted as sufficiently discriminating for the CBW quadrants set. Despite this, it was included in the forced-entry discriminant analysis for consistency. Tables C-15 through C-21 show the success rates of the different discriminant analyses in predicting cluster membership. Figures 4-1 through 4-7 show the misclassified maps as compared to the original cluster maps. Table 4-5 summarizes the differences in the classification rates of five and six-metric discriminant analyses. Because Shannon edge diversity (SHDA) was not accepted in the stepwise discriminant procedure for the CBW quadrants set, I did not expect much improvement in its classification rate. But it created an increase in proportion of correct classifications (+6%). The inclusion of Shannon edge diversity (SHDA) greatly improved correct classification in the TRW

Table 4-5: Successful Classification Rates from discriminant analyses for five and six representative metrics. See Table B-1 for description of data sets.

Data Set	Five Representative Metrics	Six Representative Metrics	Overall Improvement
TRW original data set	82%	86%	+4%
TRW low-resolution majority-rule data set	72%	79%	+7%
TRW high-resolution majority-rule data set	78%	88%	+10%
TRW low-resolution nearest-neighbor data set	80%	81%	+1%
TRW 5-class data set	69%	81%	+12%
CBW quadrants data set	66%	72%	+6%
CBW sub-watersheds data set	85%	85%	+0%

5-class set (+12%) and in the TRW low-resolution majority-rule (+10%). However, despite the fact that the stepwise procedure did not reject the new metric for the CBW sub-watersheds set, its classification rate did not improve when the metric was incorporated into the discriminant analyses.

As mentioned above, a possible reason for many of the misclassified observations could be that clusters were not optimal. In many cases in all of the data sets, the misclassed observations are adjacent to observations correctly grouped into a particular cluster. Differences in the semi-partial r-squared value used to determine clusters were small, and the groups could be very similar to each other. Misclassified observations might be very close to centroids from more than one group. Observations falling into early-forming clusters are locked into those clusters even if they are closer to centroids of clusters that form later. This could be the case for the subset maps misclassified into cluster 2 in the TRW 5-class set (Figure 4-5). The original cluster for many of these maps was cluster 1, but many were "misclassified" into cluster 2, which formed after cluster 1 in the original, "correct" analysis.

Of course, misclassifications could also be due to the inability of the subset metrics to discriminate some observations sufficiently. This might explain misclassified observations which were not spatially close to correctly classified observations of the same cluster, as are many of the maps in the TRW low-resolution majority-rule set (Figure 4-2). Additionally, some presently undetectable dimension or characteristic of landscape structure and pattern may be reflected in whole or in part by the full-set factor scores, but the subset metrics do not capture the variance of this aspect of the subset maps. On the other hand, some undetectable dimension of the landscape may not be captured by currently used metrics at all, and discriminatory power of any subset of metrics may be weakened because of this.

Overall, successful classifications rates were high, indicating that the subset metrics were effective in differentiating between observations. The subset of metrics was selected based upon results from factor analyses and examination of coefficients of congruence, which allowed identification of metrics that measured similar traits of land cover maps. The quantitative evaluation of the subset metrics' ability to predict clusters created from full-set factor scores was designed to provide a qualitative assessment of how well the subset metrics represented the full set of metrics. The subset metrics sufficiently captured a large portion of the total variance of all the metrics, and thus provide a reasonable description of currently measurable landscape characteristics; however, many metrics were unstable in measuring landscape characteristics. Chapter V addresses this in more detail, and presents the conclusions of this study.

CHAPTER V

CONCLUSIONS

In this study I started with a large set of 28 metrics that describe landscape pattern and structure, and two land cover maps from which I created seven data sets of varying spatial resolution, numbers of attributes, and methods of subset-boundary determination. Because many of these metrics measure similar aspects of landscapes, my goal was to find a subset of metrics that adequately captured the information contained in all 28 metrics. I conducted factor analyses to discover which metrics consistently grouped together on the same factors, and checked these initial results with coefficients of congruence, which confirmed that the first factor (texture) was stable across all the data sets. The second and third factors shared certain metrics which represented patch shape and compaction and that loaded sometimes on one factor, sometimes on the other. A third dimension emerged, fractal complexity, which was usually represented by the fourth or fifth factor. The metric which measured the number of land cover types in a map consistently emerged on the sixth factor, but this factor accounted for little variance.

These results compare favorably with those of the study by Riitters *et al.* (1995), which used LUDA maps in a factor analysis of landscape metrics. Certain dimensions are identifiable in both studies. Diversity/contagion metrics are very stable in measuring the texture of a map. They account for a large part of the variation among different subset maps, and while the metrics I selected were effective in discriminating the clusters created from factor scores, this does not mean that other texture metrics are any less suitable for use in describing a landscape. The texture dimension always seems to be important in a landscape, regardless of spatial resolution, number of attributes, or method of determining internal boundaries.

Fractal complexity, while not always loading on the same factor, does stand out as a stable dimension. The fractal metrics of patches, as used in this study, consistently group together, and even though they moved from factor to factor, the factor always explained a sizeable amount of variance.

Patch shape and compaction are definitely important aspects of a landscape; however, the metrics used in this study were not stable in measuring these dimensions in different map sets. Some metrics measuring compaction and shape grouped together in some data sets and did not in others. This is a suite of metrics that needs more research. Perhaps factor analyses on these metrics in isolation would yield more information on their behavior.

One of the objectives of this study was to find a subset of metrics that sufficiently measures pattern and structure in TM-derived maps. I selected six metrics which seemed to best represent the texture, fractal complexity, and patch shape and compaction dimensions. If these metrics proved successful in capturing a significant proportion of the variance of the map sets in this study, then at a minimum the subset of metrics could be substituted for the larger set of 28 metrics for analysis of these and further map sets. However, I do not draw the conclusion that this is an optimal subset. It is possible that other metrics would be just as suitable for measuring pattern and structure. In addition, other aspects of a landscape could be discovered if additional indicators were used. For example, none of the metrics used here measured inter-patch distances, which are known to be important in metapopulation dynamics models (Merriam *et* al., 1991; Opdam, 1991).

Metapopulation dynamics is an area for which the use of landscape metrics seem extremely well suited. Metapopulations are spatially structured populations consisting of distinct sub-populations of a species, separated by space or barriers, and connected genetically via dispersal movements. Metapopulation theory has been used to analyze the effects of habitat fragmentation on a wide variety of organisms, including insects, small mammals, and birds (Merriam *et al.*, 1991; Opdam, 1991). Characteristics include a turnover of local populations going extinct and becoming reestablished, resulting in a shifting distribution pattern. Opdam (1991) states that organisms are negatively influenced by disappearance or fragmentation of natural areas. Such trends not only diminish remaining total habitat but also increase the distances between remaining patches of suitable habitat. In addition, erection of artificial barriers such as roads and urban areas acts to increase the effective distances for some organisms to near-infinity. These processes have resulted in local extinctions of individual species becoming more frequent, a phenomenon that can be expected to occur on a global scale. Landscape metrics, which can capture the connectivity of a landscape, have direct application to the study of metapopulation dynamics (Gardner and O'Neill, 1991).

Acknowledging the possibility that other useful metrics may exist, an additional objective of this study was to test how well the subset of metrics captured the overall variance of a data set as determined by factor analysis of the full set of metrics. I created groups using cluster analysis with the factor scores as the input variables, and then used the subset metrics in discriminant analyses of the clusters. First using five metrics, successful classification rates ranging from 66% to 85% resulted. Incorporation of an additional texture metric improved correct classification by 10-12% in two sets, by 1-7% in four data sets, and not at all in the remaining set. The subset of metrics captured from a minimum of 72% to a maximum of 88% of the variance. Also worthy of mention is the fact that stepwise discriminant analyses indicated that two metrics would not be significant discriminators in some of the data sets, yet classification rates still improved when they were incorporated.

These are fairly good rates, and this is assuming that clusters are optimal. The possibility does exist that the clusters were not optimal, because the cluster analysis gives equal weight to each variable in the cluster analyses despite the different proportions of variance accounted for by different factors (Johnston, 1980). The cluster procedure assumes that the scales upon which each variable is measured are the same. Also, the hierarchical clustering method used did not allow an observation to be switched to a newly-defined group once it had been classified, even if it was more similar to the newer group. In other words, an observation may have been closer in geometric space to the mean location of a cluster formed late in the process, but since its location was already part of the calculation of the older cluster's mean, the observation could not be switched to the newer cluster. This may have caused some maps to be misclassified, but I believe the number of misclassifications caused by either of the above two possibilities is minimal.

One further possibility is that the classifications made by the discriminant analyses using the subset of metrics could actually be more appropriate.

The subset of metrics was effective in discriminating between groups in all the data sets used in this study, but many landscape metrics, especially those measuring shape and compaction, did not load consistently on the same factors. Despite this instability, the use of metrics to measure landscape pattern and structure with the goal of measuring ecological processes holds some promise. We can extract important information from landscapes with metrics when the results of measurements are examined in the context of a particular region and the spatial scale at which they were taken. Monitoring or measuring ecological processes that occur at broad scales can best be done using information that has been generalized (as in land cover maps). Some broad-scale processes that are hard to measure any other way can be measured using land cover maps, such as the relation of river water quality to the amount of agricultural use throughout a landscape, or the relation of anthropogenic disturbance to the complexity of the observed patch shapes in a landscape.

However, ecosystem health cannot be fully ascertained by examining broad-scale patterns only. Processes at finer scales are equally important aspects of the health of an ecosystem. Also, linear and point objects of compatible scale, such as roads and streams, or smoke stacks and illegal dumps, are not going to show up on a land cover map, yet these are obviously important influences on ecosystems. The incorporation of linear and point data with the broader-scale land cover maps would improve the effectiveness of a monitoring system. Combined with a vector-based Geographic Information System, landscape structure metrics which analyze land cover maps derived from satellite data or from other methods provide an invaluable tool for scientists to research, understand, and constructively manage ecological processes.

REFERENCES

REFERENCES

Baker, W.L., and Y. Cai. 1992. *The r.le programs for multiscale analysis of landscape structure using the GRASS geographical information system.* Landscape Ecology. 7:291-302.

ESRI. 1991. Arc/Info User's Guide. Environmental Systems Research Institute, Inc., Redlands, CA.

Fotheringham, A.S., and D.W.S. Wong. 1991. *The modifiable areal unit problem in multivariate statistical analysis.* Environment and Planning A. 23:1025-1044.

Formann, R.T.T, and M. Godron. 1986. Landscape Ecology. John Wiley & Sons Inc., New York, NY.

Gardner, R.H., B.T. Milne, M.G. Turner, and R.V. O'Neill. 1987. *Neutral models for the analysis of broad-scale landscape pattern.* Landscape Ecology. 1:19-28.

Gardner, R.H., R.V. O'Neill. 1991. *Pattern, Process, and Predictability: The Use of Neutral Models for Landscape Analysis.* in Turner, M.G., and R.H. Gardner (editors), Quantitative Methods in Landscape Ecology. Springer-Verlag, New York, NY.

Gardner, R.H., R.V. O'Neill, and M.G. Turner. 1993. *Ecological implications of landscape fragmentation.* in McDonnell, M.J. and S.T.A. Pickett (editors), Humans as Components of Ecosystems, Springer-Verlag, New York, NY.

Garrett, W.E. (editor). 1985. Atlas of North America. National Geographic Society, Washington D.C.

Hunsaker, C.T., D.A. Levine, S.P. Timmins, B.L. Jackson, R.V. O'Neill. 1992. *Landscape Characterization for assessing regional water quality.* pp. 997-1006 in McKenzie, D.H., D.E. Hyatt, and V.J. McDonald (editors), Ecological Indicators, Elsevier, NY.

Johnson, R.A., and D.W. Wichern. 1982. Applied Multivariate Statistical Analysis. Prentice Hall, Englewood Cliffs, NJ.

Johnston, R.J. 1980. Multivariate Statistical Analysis in Geography, John Wiley & Sons Inc., New York, NY.

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

Li, H., and J.F. Reynolds. 1993. *A new contagion index to quantify spatial patterns of landscapes.* Landscape Ecology. 8:155-162.

Lovejoy, S. 1982. *Area-perimeter relation for rain and cloud areas.* Science. 216:185-187.

Magurran, A.E. 1988. Ecological diversity and its measurement. Princeton University Press, Princeton, NJ.

Mandelbrot, B.B. 1983. The fractal geometry of nature. W.H. Freeman, New York, NY.

McDonald, R.P. 1985. Factor Analysis and Related Methods. Lawrence Erlbaum Associates, Hillsdale, NJ.

McGarigal, K., and B.J. Marks. 1994. Fragstats, Spatial Patterns Analysis Program for Quantifying Landscape Structure, Oregon State University, Corvallis, OR.

Merriam, G., K. Henein, and K. Stuart-Smith. 1991. *Landscape Dynamics Models.* in Turner, M.G., and R.H. Gardner (editors), Quantitative Methods in Landscape Ecology. Springer-Verlag, New York, NY.

Milne, B.T. 1991. *Lessons from applying fractal models to landscape patterns.* in Turner, M.G., and R.H. Gardner (editors), Quantitative Methods in Landscape Ecology. Springer-Verlag, New York, NY.

Nicholas, N.S., S.M. Zedaker, and C. Eagar. 1992. *A comparison of overstory community structure in three southern Appalachian spruce-fir forests.* Bulletin of the Torrey Botanical Club. 119(3):316-332.

O'Neill, R.V. 1988. *Hierarchy theory and global change.* pp. 29-45 in Rosswall, T., R.G. Woodmansee, and P.G. Risser (editors), Scales and Global Change. Wiley, NY.

O'Neill, R.V., J.R. Krummel, R.H. Gardner, G. Sugihara, B. Jackson, D.L. DeAngelis, B.T. Milne, M.G. Turner, B. Zygmunt, S.W. Christensen, V.H. Dale, and R.L. Graham. 1988. *Indices of landscape pattern.* Landscape Ecology. 1:153-162.

O'Neill, R.V., B.T. Milne, M.G. Turner, and R.H. Gardner. 1988a. *Resource utilization scales and landscape pattern.* Landscape Ecology. 2:63-69.

O'Neill, R.V., S.J. Turner, V.I. Cullinan, D.P. Coffin, T. Cook, W. Conley, J. Brunt, J.M. Thomas, M.R. Conley, and J. Gosz. 1991. *Multiple landscape scales: An intersite comparison.* Landscape Ecology. 5:137-144.

O'Neill, R.V., R.H. Gardner, M.G. Turner, and W.H. Romme. 1992. *Epidemiology theory and disturbance spread on landscapes.* Landscape Ecology. 7:19-26.

O'Neill, R.V., C.T. Hunsaker, and D.A. Levine. 1992a. *Monitoring challenges and innovative innovative ideas.* pp. 1443-1460 in McKenzie, D.H., D.E. Hyatt, and V.J. McDonald (editors), Ecological Indicators. Elsevier, NY.

O'Neill, R.V. 1994. Personal Interview.

O'Neill, R.V., C.T. Hunsaker, S.P. Timmins, B.L. Jackson, K.H. Riitters, and J.D. Whickham. 1994. *Quantifying landscape status and trends at the regional scale.* Unpublished manuscript.

Opdam, P. 1991. *Metapopulation theory and habitat fragmentation: a review of holarctic breeding bird studies.* Landscape Ecology. 5:93-106.

Pickover, C.A. 1990. Computers, Pattern, Chaos and Beauty, Graphics from an Unseen World. St. Martin's Press, New York, NY.

Riitters, K.H. 1994. Personal Interview.

Riitters, K.H., and R.V. O'Neill. 1994. *Spatial changes in habitat and their impacts on plant and animal communities in the Chesapeake Bay Watershed.* Unpublished manuscript.

Riitters, K.H., R.V. O'Neill, C.T. Hunsaker, J.D. Wickham, D.H. Yankee, S.P. Timmins, K.B. Jones, and B.L. Jackson. 1995. *A Factor Analysis of Landscape Pattern and Structure Metrics.* Landscape Ecology. (in press).

SAS Institute. 1982. User's Guide: Statistics. SAS Institute, Cary, NC.

Sember, P. R., F.P. Kapinos, and G.L. Knapp. 1987. United States Geological Survey Water Supply Paper 2294. U.S.G.S., Denver, CO.

Tennessee Valley Authority. 1992. Valley-wide Classification using Landsat TM Data. TVA memorandum.

Turner, M.G., R.H. Gardner, V.H. Dale, and R.V. O'Neill. 1988. *Landscape pattern and the spread of disturbance.* pp. 373-382 in Ruzicka, M., T. Hrnciarova, and L. Miklos (editors), Proc. VIIIth International Symposium on Problems of Landscape Ecological Research, Vol. 1. Institute of Experimental Biology and Ecology, CBES SAS, Bratislava, CSSR.

Turner, M.G. 1989. *Landscape Ecology: The Effect of Pattern on Process*. Annual Review of Ecological Systems. 20:171-197.

Turner, M.G., R.H. Gardner, V.H. Dale, R.V. O'Neill. 1989. *Predicting the spread of disturbance across heterogeneous landscapes*. Oikos. 55:121-129.

Turner, M.G. 1990. *Spatial and temporal analysis of landscape patterns*. Landscape Ecology. 4:21-30.

Turner, M.G. and R.H. Gardner (editors). 1991. Quantitative Methods in Landscape Ecology: The Analysis and Interpretation of Landscape Heterogeneity. Springer-Verlag, New York, NY.

United States Environmental Protection Agency. 1993. Discrimination of Transition Zone Sites in the Chesapeake Bay Preliminary Scene Accuracy Assessment. EPA Memorandum.

United States Environmental Protection Agency. 1994. Chesapeake Bay Watershed Classified Landsat/TM Image.

United States Environmental Protection Agency. 1994a. Landscape Monitoring and Assessment Research Plan. U.S. EPA 620/R-94-009, Office of Research and Development, Washington DC 20460.

Voss, R.F. 1988. *Fractals in nature: From characterization to simulation*. pp. 21-70 in Peitgen, H.O., and D. Saupe (editors), The Science of Fractal Images, Springer-Verlag, New York, NY.

Wickham, J.D., and D.J. Norton. 1994. *Mapping and analyzing landscape patterns*. Landscape Ecology. 9:7-23.

Wickham, J.D. 1995. Personal Interview.

Wickham, J.D., and K.H. Riitters. 1995. *Sensitivity of landscape metrics to pixel size*. International Journal of Remote Sensing (in press).

Wilson, E.O. (editor). 1988. Biodiversity. National Academy Press, Washington D.C.

APPENDICES

APPENDIX A

LANDSCAPE METRICS

APPENDIX A

LANDSCAPE METRICS

I selected landscape metrics for this study based upon past use in one or more studies. Some have been used extensively and others have only been used on one or a few occasions. At a minimum, all the metrics were used by Riitters *et al.* (1995). Magurran (1988) provided many of the formulas used for these metrics, but in a biological context of diversity and evenness. I used a program developed by Riitters *et al.* (1995) to process the land cover maps and calculate landscape metrics. This appendix lists the formulas used to compute these metrics. Calculation of the metrics involved several steps. First, per-patch measurements such as area, perimeter, *etc.*, the building blocks of many other metrics, were made. A simple count of the cells determined the frequencies of the different land cover classes. Next, an attribute (land cover class) adjacency matrix was constructed by counting edges between cells (using only cardinal directions).

Statistics and measurements were made for each patch in the map. Individual patches of contiguous cells of the same land cover were delineated, and patch area based on the number of cells in the patch was calculated. The program calculated patch perimeter in three ways: 1) the number of cells defining the outer limits of the patch (outside cells); 2) the number of edges enclosing the patch (outside edges); and 3) the sum of the number of edges between the patch and all inclusions (smaller patches) within the patch (total perimeter edge length). Patch centroids were then calculated based on the average row and column of the cells that were part of the patch. A bounding rectangle was determined for each patch, and the longest axis was determined as the greatest distance between any two cells in the patch. Finally, each patch was checked to see if it was adjacent to the map border. Metrics which use patch size as a part of their calculations included patches adjacent to the border, but ones that used patch shape excluded patches adjacent to the border due to the artificially straight boundary of the edge. The following abbreviations are used in several equations in this appendix:

A_{ij} is the frequency of land cover class i being located adjacent to attribute j in a cardinal direction (this is stored in the attribute adjacency matrix, A)
$(c_{\max,k}, c_{\min,k})$ are the maximum and minimum columns of patch k
CT_k is the centroid of patch k
F_i is the frequency of land cover i in a map
IE_k is the number of edges between patch k and all 'island' patches within patch k
LA_k is the longest axis across patch k
m is the number of patches in a map
n is the total number of cells in a map
OC_k is the number of cells enclosing patch k
OE_k is the number of edges enclosing patch k
$(r_{\max,k}, r_{\min,k})$ are the maximum and minimum rows of patch k
S_k is the area in numbers of cells of patch k
t is the number of land cover classes

TEXTURE/DIVERSITY METRICS

Texture and diversity describe land cover pattern, richness, and abundance. The following metrics measure different aspects of texture and diversity of land cover types in a map:

NTYP: Number of Types (Baker and Cai, 1992)

This is simply the number of types of land cover in a map (t). This measures the richness of different land cover types, but this alone is not much information. Abundance of a land cover is important as well, because one or a few land covers could dominate the landscape, lessening the importance of the remaining land covers.

PMAX: Maximum Attribute Class Proportion (Magurran, 1988)

This is the proportionality of the land cover, i, that has the highest proportion of cells in a map.

$$PMAX = \max\{p_i\}$$

where

$$p_i = \frac{F_i}{n}$$

PMAX is a measure of dominance, or the degree to which a single land cover dominates a landscape. While it ignores information about other land cover types, PMAX reveals the extent of influence of the most common cover. If a highly dominant land cover is a natural one, such as forest, then this metric could indicate a forest-biome landscape is in a healthy state. However, if urban cover exhibits similar dominance, this metric could indicate that ecosystems in the area are either in jeopardy or have already been severely damaged. Thus the metric is meaningful only when the cover type is noted. Also, because of its focus on one land cover type, PMAX should be used in conjunction with other texture/diversity metrics which capture information about the other land cover types.

SIDI: Simpson Diversity of Attribute Classes (Magurran, 1988, Baker and Cai, 1992)

$$SIDI = 1 - \sum_{i=1}^{I} p_i^2$$

This metric is a measure of land cover type richness (number of types) and abundance (evenness of representation of types). It is important to measure both aspects of diversity, because it is possible for two maps to have the same number of land covers (richness) yet different levels of abundance. In that situation, the map with more evenly distributed land cover abundance would be considered to be more diverse. SIDI is heavily weighted towards abundance, and is less sensitive to land cover richness (Magurran, 1988). A low value indicates dominance by one or a few land cover types.

SIEV: Simpson Evenness of Attribute Classes (Magurran, 1988)

$$SIEV = \frac{SIDI}{1-\frac{1}{t}}$$

SIEV isolates the abundance found in SIDI to measure how evenly distributed the land cover types are. This takes some of the influence of the dominant land covers away, but at the cost of losing the information about the total number or richness of land cover types. The purpose of determining evenness is so that relative abundance in maps with different numbers of land cover types can be compared.

SHDI: Shannon Diversity of Attribute Classes (Magurran, 1988, Baker and Cai, 1992, Turner, 1989)

$$SHDI = -\sum_{i=1}^{t} p_i \ln p_i$$

Like the SIDI metric, SHDI measures both richness and abundance of land cover types. The difference is that SHDI does not weight toward the most abundant type, so types with small proportions have more representation in the calculation of this metric.

SHEV: Shannon Evenness of Attribute Classes (Magurran, 1988)

$$SHEV = \frac{SHDI}{\ln t}$$

This metric compares the diversity of a landscape to the maximum possible diversity, $\ln t$. This removes the influence of the number of land cover types, leaving a measure of abundance or evenness.

SUMD: Sum of Diagonal Elements of Adjacency Matrix, A

$$SUMD = \sum_{i=1}^{t} v_{ii}$$

where

$$v_{ij} = \frac{A_{ij}}{\sum_{i=1}^{t}\sum_{j=1}^{t} A_{ij}}$$

SUMD measures contagion by calculating the proportion of cells adjacent to cells of the same land cover type. Contagion is an aspect of the texture of a map, describing how clustered cells of the same type are. High values of contagion indicate that large patches exist in a map, and low values indicate a fragmented landscape. The proportion of cells adjacent to cells of the same type is taken from the diagonal of the adjacency matrix. The adjacency matrix is tabulated by counting the times a particular cell type is adjacent (in cardinal directions only) to another cell type.

SIDA: Simpson Diversity of Adjacency Matrix

$$SIDA = 1 - \sum_{i=1}^{t}\sum_{j=1}^{t} v_{ij}^2$$

This metric measures the diversity of the adjacency matrix, and by doing so, it measures the diversity of relationships between patches of different land cover types. For example, SIDA measures the richness of different relationships, such as deciduous forest adjacent to pasture or urban adjacent to water, and combines this with the abundance of each of these different edge relationships. This is similar to SIDI, but it incorporates contagion into the calculation by using the adjacency matrix. A low value of SIDI often indicates high contagion.

SICO: Simpson Contagion

$$SICO = 1 - \frac{SIDA}{1 - \frac{1}{t^2}}$$

SICO is a measure of evenness of the adjacency matrix. Similar to SIEV, the influence of the richness of edge types is removed. This metric is describing evenness and contagion.

SHDA: Shannon Diversity of Adjacency Matrix

$$SHDA = -\sum_{i=1}^{t}\sum_{j=1}^{t} v_{ij} \ln\left(v_{ij}\right)$$

This measures diversity of the adjacency matrix using Shannon's equation instead of Simpson's. The same characteristics apply from the calculation of the diversity of land cover types.

SHCO: Shannon Contagion (Li and Reynolds, 1993)

$$SHCO = 1 - \frac{SHDA}{2\ln t}$$

SHCO removes the influence of the number of edge types to produce a measure of evenness and contagion based on Shannon's diversity index.

FDDA: Area-Weighted Average of Fractal Dimension from Density-area Scaling

This is a fractal estimator of attribute class mass from the scaling of attribute class density to the size of a neighborhood of an arbitrary cell in the class. It is a way of scaling contagion and texture. Let $Q_f(L)$ be the probability, for an arbitrary cell in an attribute class, of finding f other cells of the same class within a square of side length L centered on the arbitrary cell. In general,

$$Q_f(L) = \frac{Z_{f,L}}{H_i}$$

where the numerator is the number of cells (*Z*) for which f was realized for length L, and the denominator is the number of cells (*H*) sampled. Let $N(L)$ be the maximum number of cells of the same class observed in any square of size L.

Define

$$M(L) = \sum_{f=1}^{N(L)} f\, Q_f(L)$$

From each attribute class with > 500 cells, a random sample of at least 500 cells is selected. Square neighborhoods with a side length L = 5, 15, 25, 35, and 45 are placed around each sampled cell, and the occurrences of cells of the same class are counted for each neighborhood size. The values of $Q_f(L)$ and $M(L)$ are calculated after the counts are accumulated for all sampled cells. Let β_{1i} be the estimated slope from the regression of $\ln M(L)$ on $\ln L$ for the i-th attribute class. The fractal estimator is:

$$FDDA = \sum_{i=1}^{t} p_i \beta_{1i}$$

NPAT: Number of Patches

m is the number of patches. Patches are defined by connections of pixels of the same data code that share edges in common (not just vertices). This indirectly measures contagion in a landscape because a high number of patches indicates different land cover types are probably present, creating a more fragmented landscape, or low contagion. A low number of patches would inversely indicate high contagion.

LPAT: Largest Patch

The patch with the greatest number of cells.

$$LPAT = \max\{S_1, \ldots S_m\}$$

This indirectly measures contagion and dominance. If the largest patch covers a significant proportion of a map, then contagion should be high. If the largest patch is small, then all of the other patches will be smaller as well, reflecting a more fragmented landscape.

PSIZ: Average Patch Size

$$PSIZ = \frac{1}{m} \sum_{k=1}^{m} S_k$$

A small average patch size indicates a more fragmented landscape. This metric is, however, sensitive to very large or very small patches (outliers).

P005: Proportion of Area Consisting of Patches Containing Six or More Pixels (O'Neill *et al.*, 1994)

Weighted average proportion of cells contained in patches with an area > 5 cells (Jackson's contagion5 statistic)

$$P005 = \sum_{i=1}^{t} p_i \frac{\sum_{g=1}^{m_i} S_g^*}{\sum_{g=1}^{m_i} S_g}$$

where g subscripts patches, m_i is the total number of patches of type i,

$$S_g^* = 0 \text{ if } S_g < 6$$

$$= S_g \text{ otherwise}$$

This measures contagion at a certain level. If a low proportion of patches have an area > 5 cells, then the map is extremely fragmented. If a high proportion of patches have an area > 5 cells, then some contagion exists, but the metric does not reveal what degree of contagion. To determine this, different levels would need to be calculated (*i.e.* > 50 cells, > 500 cells, *etc.*).

PATCH SHAPE/COMPACTION METRICS

Simple shapes have low perimeter to area ratios, and are consequently more compact than complex shapes. O'Neill *et al.* (1988) have shown that simpler shapes on the landscape often are created by humans, demonstrating the utility of measuring the complexity of patch shape as a proxy for human disturbance. The following metrics measure the shape and compaction of patches:

PA-1: Average Perimeter-Area Ratio (Baker and Cai, 1992)

$$PA\text{-}1 = \frac{1}{m} \sum_{k=1}^{m} \frac{OE_k}{S_k}$$

This metric takes the total length of the outside edges of a patch divided by the patch area, giving a measure of how compact patches in a map are. It measures the complexity of the shape of a patch, with simpler patches having smaller perimeters than complex patches covering the same area. This metric is sensitive to the size of the patches.

PA-2: Average Adjusted Perimeter-Area Ratio (Baker and Cai, 1992)

$$PA\text{-}2 = \frac{1}{m}\sum_{k=1}^{m} 0.282\frac{OE_k}{S_k}$$

PA-2 measures the same aspect as PA-1, but attempts to lessen the sensitivity to differently sized patches by using the constant of 0.282.

DSTA: Average Adjusted Perimeter-Area Ratio using Gardner's D-statistic (Gardner *et al.*, 1987; Milne, 1991)

$$DSTA = \frac{1}{m}\sum_{k=1}^{m}\frac{\ln[4S_k]}{\ln[OE_k]}$$

This metric represents the perimeter-area relationship on a logarithmic scale, which tends to reduce the influence of extremely large numbers.

NASQ: Average Normalized Area, Square Model

$$NASQ = \frac{1}{m}\sum_{k=1}^{m} 16\frac{S_k}{OE_k^2}$$

This metric compares the shape of a patch to a square. NASQ has a value of zero for linear patches and one for square patches, and it is sensitive to patch size.

RGYR: Average Radius of Gyration (Pickover, 1990)

$$RGYR = \frac{1}{m}\sum_{k=1}^{m} RG_k$$

where, given patch centroid CT_k, the radius of gyration (RG) of patch k is

$$RG_k = \sqrt{\sum_{b=1}^{S_k} d_b^2}$$

and d_b is the distance from cell b to CT_k

This metric finds the average distance of cells within a patch to the center of that patch. It is a measure of overall compaction, independent of perimeter shape.

PTRD: Average Patch Topology Ratio Dimension

$$PTRD = \frac{1}{m}\sum_{k=1}^{m} 2 - \frac{OC_k - OC_{\min_k}}{OC_{\max_k} - OC_{\min_k}}$$

where

$$OC_{\min_k} = 4\sqrt{S_k}$$

which is the number of cells on the perimeter for a square, and

$$OC_{\max_k} = (2S_k) + 2$$

which is the number of cells on the perimeter of a linear patch.

This metric provides a measure of shape and compaction by placing a patch on a continuum from a square to a line.

ABRR: Average Bounding Rectangle Ratio

$$ABRR = \frac{1}{m}\sum_{k=1}^{m}\frac{S_k}{X_k}$$

where

$$X_k = \left(r_{\max_k} - r_{\min_k}\right) \times \left(c_{\max_k} - c_{\min_k}\right)$$

which is the area of the bounding rectangle

ABRR finds the ratio between the area of a patch and the smallest possible rectangle which encloses the patch. It is a measure of compaction.

ABSR: Average Area Boxside Ratio

$$ABSR = \frac{1}{m}\sum_{k=1}^{m}\frac{\ln S_k}{Y_k}$$

where

$$Y_k = \sup\left(r_{\max_k} - r_{\min_k}, c_{\max_k} - c_{\min_k}\right)$$

which is the side of the box with the greatest length.

ABSR measures the compaction of a patch by taking the longest side of a bounding rectangle and comparing this to the logarithm of the area of a patch.

ACCR: Average Circumscribing Circle Ratio (Baker and Cai, 1992)

$$ACCR = \frac{1}{m}\sum_{k=1}^{m}\frac{S_k}{\pi\left(\frac{LA_k}{2}\right)^2}$$

The denominator is the area of a circle based on the longest axis (*LA*) of patch *k*. ACCR measures the compaction of a patch by comparing it to a circle that encloses the entire patch.

ALAR: Average Area to Longest Axis Ratio

$$ALAR = \frac{1}{m}\sum_{k=1}^{m}\frac{\ln S_k}{\ln LA_k}$$

This is similar to ABSR in that it measures the compaction of a patch by comparing the area of a patch to a single axis, but instead of using the side of a box which encloses the patch, ALAR uses the longest axis, or a line connecting the two cells belonging to the patch that are the furthest distance from each other. The logarithm of the longest axis reduces the influence of distance in cases where patches might be extremely linear.

OCFC: Perimeter-Area Scaling—Pixels (Krummel *et al.*, 1987, O'Neill *et al.*, 1988, Wickham and Norton, 1994)

$$OCFC = 2\beta_2$$

OCFC measures the shape and compaction of patches using fractal dimensions. Shape complexity of patches is related to the fractal geometry of the landscape (Mandlebrot, 1983). Using calculations of patch statistics, the log of the patch perimeter (Y) is regressed against the log of the patch area (X) for each patch of a particular cover type. An estimate of complexity of shapes is given by 2.0 times the slope (β_2) of the regression line (Lovejoy, 1982). Larger values indicate more complex shapes. Small patches produce less reliable results because shape is determined more by the resolution of the grid than by the complexity of the patches (Milne, 1991). Therefore, only patches with an area greater than 3 pixels were included in this calculation. Perimeter is defined by the outside pixels of the patch.

OEFC: Perimeter-Area Scaling—Edges (Krummel *et al.*, 1987; Baker and Cai, 1992)

$$OEFC = 2\beta_3$$

This is similar to Perimeter-Area Scaling—Pixels (OCFC), except that perimeter is defined by outside plus inside edges. This may cause the OEFC dimension to differ because it takes fewer pixels than edges to enclose a given patch.

APPENDIX B

SUMMARY STATISTICS

Table B-1: Summary of Information for Tennessee River Watershed (TRW) and Chesapeake Bay Watershed (CBW) data sets.

Name	Abbreviation	Cell Size	Map Extent	Number of Maps	Notes
TRW original data set	T12_25	25 x 25 meters	1800 x 1100 cells	85	
TRW low-resolution majority-rule data set	T12_125M	125 x 125 meters	360 x 250 cells	85	5 x 5 window of the 25-meter map was compressed to 125 meters and given a land cover type based on the highest proportion in the window.
TRW high-resolution majority-rule data set	T12_25M	25 x 25 meters	1800 x 1100 cells	85	5 x 5 window of the 25-meter map was given a land cover type based on the highest proportion in the window, but not compressed to 125 meters
TRW low-resolution nearest-neighbor data set	T12_125N	125 x 125 meters	360 x 250 cells	85	5 x 5 window of the 25-meter map was compressed to 125 meters and given a land cover type based on the center cell in the window
TRW 5-class data set	T5_25	25 x 25 meters	1800 x 1100 cells	85	
CBW quadrants data set	C5_25Q	25 x 25 meters	1500 x 1900 cells	97	
CBW sub-watersheds data set	C5_25W	25 x 25 meters	varies	46	subset maps were cut out using 8-digit U.S.G.S. Hydrologic Units (sub-watersheds)

Table B-2: Summary Statistics for the Tennessee River Watershed original data set. See Appendix A for explanation of metric abbreviations and Table B-1 for description of data sets.

METRIC	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	MINIMUM	MAXIMUM	NORMAL DISTRIBUTION
NTYP	8.412	1.147	13.640	6.000	11.000	
PMAX	0.587	0.144	24.604	0.262	0.888	yes
SIDI	0.547	0.136	24.780	0.203	0.812	
SIEV	0.623	0.153	24.562	0.232	0.914	
SHDI	1.075	0.288	26.794	0.427	1.811	yes
SHEV	0.507	0.135	23.381	0.205	0.824	yes
SUMD	0.925	0.019	2.053	0.882	0.963	yes
SIDA	0.595	0.138	23.262	0.237	0.843	
SICO	0.396	0.140	35.407	0.146	0.760	
SHDA	1.396	0.357	25.554	0.583	2.221	yes
SHCO	0.670	0.084	12.573	0.495	0.860	yes
FDDA	1.820	0.052	2.845	1.739	1.928	
NPAT	11837.5	4222.0	35.7	4397.0	26654.0	
LPAT	759508.5	432003.3	56.9	111250.0	1670946.0	
PSIZ	180.865	66.560	36.801	71.958	436.200	
P005	0.995	0.002	0.188	0.989	0.999	
PA-1	1.261	0.058	4.631	1.121	1.367	
DSTA	1.372	0.011	0.767	1.356	1.397	
PA-2	1.657	0.027	1.646	1.580	1.713	yes
NASQ	0.568	0.010	1.759	0.552	0.604	yes
RGYR	3.675	0.428	11.659	2.890	4.792	
PTRD	1.964	0.007	0.357	1.946	1.979	yes
ABRR	0.635	0.010	1.575	0.618	0.670	
ABSR	1.574	0.011	0.673	1.546	1.598	yes
ACCR	0.697	0.021	3.033	0.641	0.736	
ALAR	1.670	0.010	0.593	1.646	1.698	yes
OCFC	1.190	0.015	1.265	1.156	1.227	yes
OEFC	1.296	0.019	1.501	1.253	1.341	yes

Table B-3: Summary Statistics for the Tennessee River Watershed low-resolution majority-rule data set. See Appendix A for explanation of metric abbreviations and Table B-1 for description of data sets.

METRIC	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	MINIMUM	MAXIMUM	NORMAL DISTRIBUTION
NTYP	8.318	1.147	13.784	6.000	11.000	
PMAX	0.599	0.147	24.605	0.262	0.903	yes
SIDI	0.533	0.141	26.499	0.178	0.812	
SIEV	0.607	0.159	26.166	0.203	0.914	
SHDI	1.041	0.296	28.420	0.376	1.809	yes
SHEV	0.493	0.137	27.716	0.181	0.823	yes
SUMD	0.784	0.058	7.454	0.672	0.908	
SIDA	0.658	0.151	22.978	0.262	0.898	
SICO	0.332	0.153	46.037	0.091	0.734	
SHDA	1.713	0.467	27.238	0.665	2.798	yes
SHCO	0.594	0.108	18.248	0.363	0.840	yes
FDDA	1.782	0.285	15.981	0.000	1.940	
NPAT	4004.2	1294.9	32.3	1607.0	7217.0	yes
LPAT	31179.3	18067.4	57.9	4015.0	68684.0	
PSIZ	21.473	8.282	38.571	10.684	47.849	
P005	0.934	0.021	2.242	0.880	0.974	yes
PA-1	1.762	0.060	3.433	1.612	1.910	
DSTA	1.275	0.008	0.650	1.255	1.296	yes
PA-2	1.867	0.059	3.185	1.745	2.026	yes
NASQ	0.483	0.014	2.954	0.445	0.519	yes
RGYR	2.560	0.282	10.988	1.998	3.193	
PTRD	1.823	0.018	0.998	1.741	1.859	
ABRR	0.579	0.013	2.315	0.540	0.602	
ABSR	1.451	0.018	1.206	1.390	1.487	
ACCR	0.656	0.025	3.864	0.596	0.713	yes
ALAR	1.572	0.018	1.161	1.513	1.611	yes
OCFC	1.279	0.022	1.704	1.221	1.328	yes
OEFC	1.480	0.027	1.833	1.418	1.543	yes

Table B-4: Summary Statistics for the Tennessee River Watershed high-resolution majority-rule data set. See Appendix A for explanation of metric abbreviations and Table B-1 for description of data sets.

METRIC	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	MINIMUM	MAXIMUM	NORMAL DISTRIBUTION
NTYP	8.318	1.147	13.784	6.000	11.000	
PMAX	0.599	0.147	24.605	0.262	0.903	yes
SIDI	0.533	0.141	26.515	0.177	0.812	
SIEV	0.607	0.159	26.182	0.203	0.914	
SHDI	1.040	0.296	28.437	0.375	1.810	yes
SHEV	0.493	0.137	27.734	0.180	0.824	yes
SUMD	0.957	0.012	1.221	0.935	0.982	
SIDA	0.562	0.144	25.671	0.195	0.833	
SICO	0.430	0.146	33.970	0.157	0.802	
SHDA	1.250	0.348	27.795	0.466	2.107	yes
SHCO	0.704	0.081	11.447	0.521	0.888	yes
FDDA	1.826	0.054	2.943	1.732	1.944	
NPAT	4004.8	1294.9	32.3	1607.0	7217.0	yes
LPAT	776783.5	449755.0	57.9	100375.0	1709122.0	
PSIZ	534.314	206.007	38.555	265.758	1192.850	
P005	1.000	0.000	0.000	1.000	1.000	
PA-1	0.648	0.019	2.869	0.605	0.686	
DSTA	1.538	0.002	0.098	1.533	1.541	yes
PA-2	1.322	0.029	2.166	1.258	1.393	yes
NASQ	0.838	0.018	2.134	0.792	0.877	yes
RGYR	4.803	0.496	10.324	3.874	6.339	yes
PTRD	1.975	0.004	0.184	1.965	1.982	yes
ABRR	0.886	0.014	1.552	0.855	0.918	yes
ABSR	1.859	0.010	0.557	1.830	1.878	
ACCR	0.757	0.021	2.767	0.702	0.802	
ALAR	1.766	0.007	0.394	1.746	1.780	yes
OCFC	1.307	0.011	0.869	1.277	1.334	yes
OEFC	1.351	0.012	0.926	1.318	1.383	yes

Table B-5: Summary Statistics for the Tennessee River Watershed low-resolution nearest-neighbor data set. See Appendix A for explanation of metric abbreviations and Table B-1 for description of data sets.

METRIC	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	MINIMUM	MAXIMUM	NORMAL DISTRIBUTION
NTYP	8.365	1.163	13.908	6.000	11.000	
PMAX	0.587	0.145	24.625	0.261	0.888	yes
SIDI	0.547	0.136	24.822	0.203	0.813	
SIEV	0.623	0.153	24.573	0.232	0.914	
SHDI	1.075	0.288	26.839	0.426	1.813	yes
SHEV	0.508	0.135	26.466	0.205	0.825	yes
SUMD	0.741	0.065	8.790	0.612	0.882	
SIDA	0.689	0.143	20.791	0.307	0.906	
SICO	0.300	0.145	48.259	0.082	0.688	
SHDA	1.833	0.467	25.470	0.778	2.880	yes
SHCO	0.566	0.110	19.411	0.337	0.813	yes
FDDA	1.786	0.285	15.944	0.000	1.939	
NPAT	5837.0	1800.1	30.8	2464.0	10274.0	yes
LPAT	28681.5	17438.8	60.8	3940.0	66621.0	
PSIZ	14.449	5.240	36.266	7.449	31.060	
P005	0.903	0.030	3.289	0.828	0.959	yes
PA-1	1.826	0.058	3.200	1.679	1.964	
DSTA	1.263	0.008	0.610	1.246	1.284	yes
PA-2	1.902	0.060	3.241	1.759	2.028	yes
NASQ	0.469	0.015	3.116	0.441	0.505	yes
RGYR	2.449	0.250	10.224	1.940	3.012	yes
PTRD	1.804	0.017	0.968	1.749	1.842	yes
ABRR	0.572	0.013	2.214	0.539	0.598	yes
ABSR	1.439	0.016	1.095	1.389	1.469	
ACCR	0.652	0.024	3.708	0.601	0.705	yes
ALAR	1.563	0.016	1.039	1.519	1.595	
OCFC	1.288	0.021	1.608	1.242	1.337	yes
OEFC	1.511	0.026	1.746	1.456	1.575	yes

Table B-6: Summary Statistics for the Tennessee River Watershed 5-class data set.
See Appendix A for explanation of metric abbreviations and Table B-1 for description of data sets.

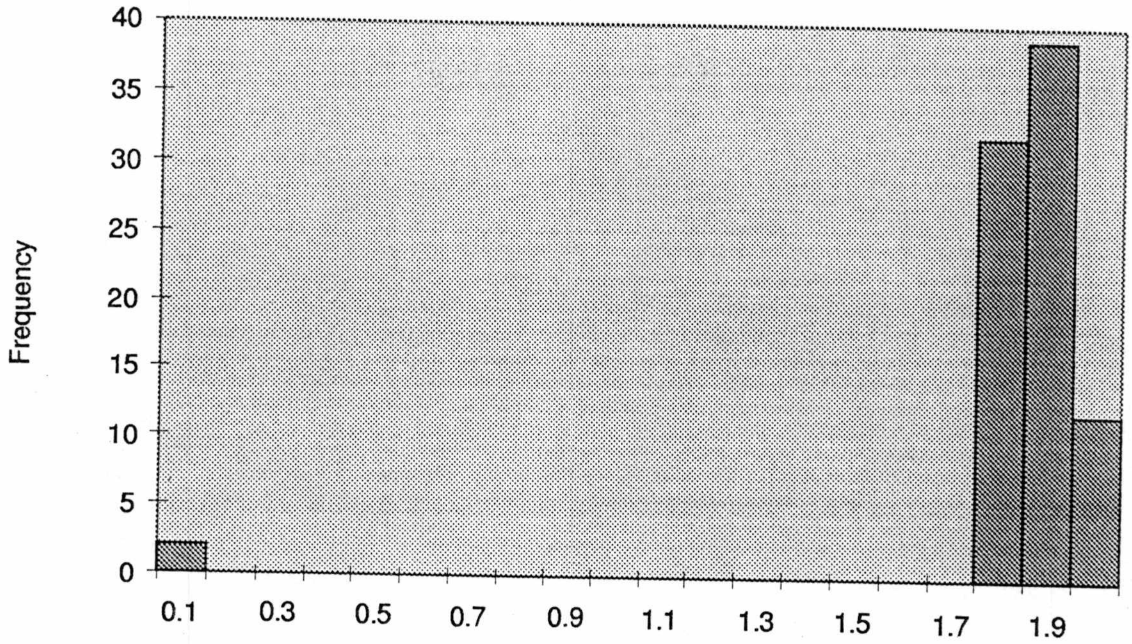
METRIC	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	MINIMUM	MAXIMUM	NORMAL DISTRIBUTION
NTYP	4.682	0.468	10.002	4.000	5.000	
PMAX	0.655	0.139	21.174	0.430	0.989	
SIDI	0.452	0.139	30.751	0.022	0.654	
SIEV	0.576	0.176	30.578	0.029	0.817	
SHDI	0.765	0.226	29.539	0.066	1.178	
SHEV	0.497	0.145	29.204	0.047	0.765	yes
SUMD	0.949	0.017	1.786	0.907	0.996	yes
SIDA	0.494	0.148	30.003	0.026	0.694	
SICO	0.482	0.155	32.140	0.278	0.973	
SHDA	0.978	0.286	29.209	0.086	1.479	
SHCO	0.682	0.092	13.478	0.503	0.969	yes
FDDA	1.879	0.042	2.212	1.810	1.992	
NPAT	5542.1	2481.3	44.8	794.0	16441.0	
LPAT	900319.5	458714.4	51.0	178570.0	1894339.0	
PSIZ	418.282	270.630	64.700	116.658	2413.307	
P005	0.998	0.001	0.118	0.992	1.000	
PA-1	1.226	0.054	4.380	1.103	1.350	yes
DSTA	1.374	0.010	0.745	1.349	1.394	yes
PA-2	1.720	0.035	2.037	1.597	1.800	yes
NASQ	0.551	0.014	2.487	0.525	0.599	
RGYR	4.558	0.469	10.297	3.505	5.873	yes
PTRD	1.963	0.009	0.449	1.943	1.978	
ABRR	0.622	0.013	2.036	0.598	0.666	yes
ABSR	1.574	0.014	0.888	1.540	1.603	yes
ACCR	0.675	0.023	3.441	0.612	0.748	yes
ALAR	1.664	0.014	0.840	1.634	1.697	yes
OCFC	1.205	0.017	1.386	1.161	1.247	yes
OEFC	1.300	0.017	1.318	1.258	1.351	yes

Table B-7: Summary Statistics for the Chesapeake Bay Watershed quadrants data set. See Appendix A for explanation of metric abbreviations and Table B-1 for description of data sets.

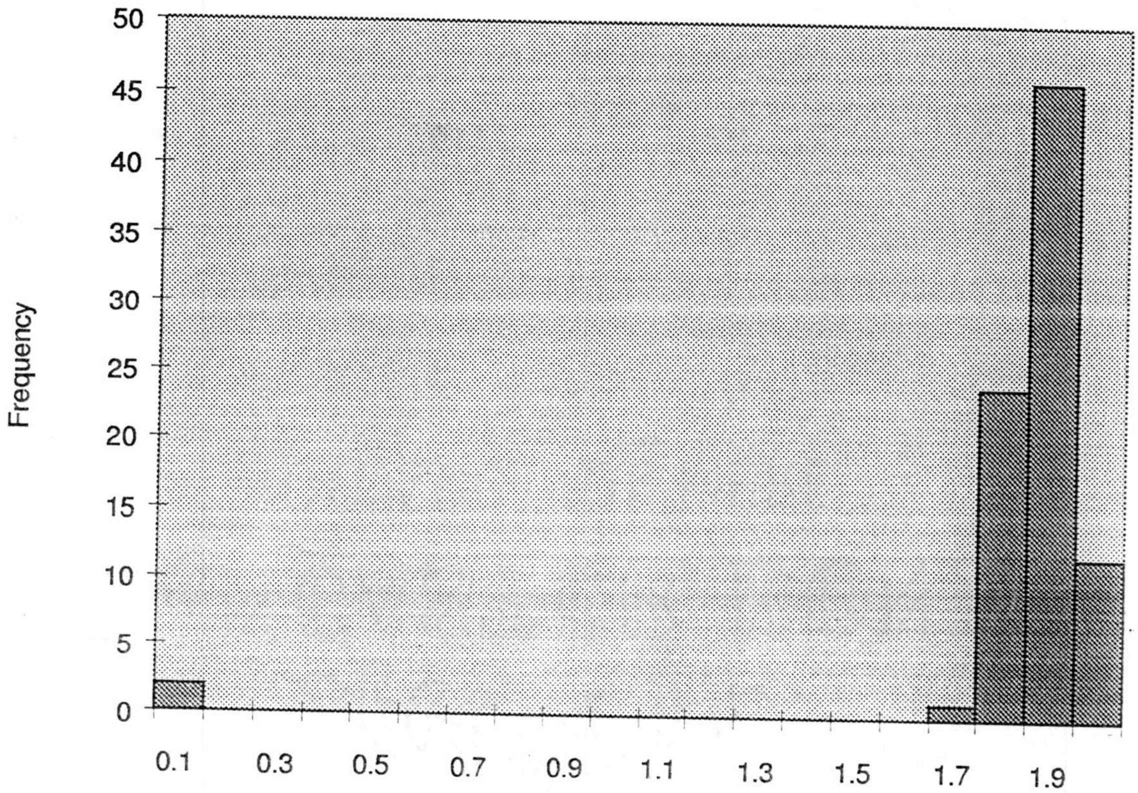
METRIC	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	MINIMUM	MAXIMUM	NORMAL DISTRIBUTION
NTYP	4.959	0.200	4.031	4.000	5.000	
PMAX	0.613	0.123	20.119	0.402	0.950	
SIDI	0.496	0.111	22.433	0.095	0.700	
SIEV	0.621	0.139	22.394	0.119	0.876	
SHDI	0.842	0.194	22.988	0.217	1.292	yes
SHEV	0.526	0.120	22.826	0.135	0.803	yes
SUMD	0.937	0.019	2.030	0.894	0.979	yes
SIDA	0.544	0.115	21.099	0.115	0.740	
SICO	0.433	0.120	27.649	0.229	0.880	
SHDA	1.098	0.236	21.479	0.306	1.630	yes
SHCO	0.657	0.073	11.170	0.494	0.905	yes
FDDA	1.873	0.037	1.981	1.807	1.962	
NPAT	12915.0	5443.2	42.1	2806.0	29285.0	
LPAT	828530.4	524927.5	63.4	154219.0	2315356.0	
PSIZ	274.404	154.616	56.346	98.559	920.177	
P005	0.996	0.002	0.185	0.991	0.999	
PA-1	1.303	0.094	7.218	1.051	1.504	yes
DSTA	1.329	0.016	1.174	1.298	1.368	
PA-2	2.048	0.037	1.792	1.963	2.165	yes
NASQ	0.403	0.007	1.674	0.388	0.419	yes
RGYR	5.271	0.924	17.523	3.740	8.593	
PTRD	1.874	0.018	0.964	1.818	1.905	
ABRR	0.517	0.008	1.612	0.496	0.536	yes
ABSR	1.509	0.022	1.474	1.448	1.557	yes
ACCR	0.508	0.018	3.591	0.456	0.546	yes
ALAR	1.551	0.017	1.103	1.506	1.583	
OCFC	1.224	0.017	1.410	1.173	1.260	yes
OEFC	1.332	0.027	2.002	1.253	1.373	

Table B-8: Summary Statistics for the Chesapeake Bay Watershed sub-watersheds data set. See Appendix A for explanation of metric abbreviations and Table B-1 for description of data sets.

METRIC	MEAN	STANDARD DEVIATION	COEFFICIENT OF VARIATION	MINIMUM	MAXIMUM	NORMAL DISTRIBUTION
NTYP	4.978	0.147	2.962	4.000	5.000	
PMAX	0.626	0.123	19.726	0.381	0.929	yes
SIDI	0.482	0.113	23.415	0.134	0.678	
SIEV	0.603	0.141	23.450	0.168	0.847	
SHDI	0.817	0.183	22.333	0.308	1.187	yes
SHEV	0.510	0.114	22.321	0.191	0.738	yes
SUMD	0.934	0.017	1.795	0.900	0.968	yes
SIDA	0.533	0.119	22.299	0.163	0.737	
SICO	0.444	0.124	27.905	0.233	0.830	
SHDA	1.083	0.236	21.777	0.441	1.584	yes
SHCO	0.662	0.073	11.068	0.508	0.863	yes
FDDA	1.868	0.036	1.901	1.813	1.949	
NPAT	25888.4	11891.1	45.9	5959.0	61826.0	yes
LPAT	1189443.0	825768.8	69.4	227226.0	3690582.0	
PSIZ	242.865	97.130	39.993	110.300	576.231	
P005	0.996	0.002	0.162	0.992	0.999	yes
PA-1	1.311	0.093	7.108	1.096	1.525	yes
DSTA	1.327	0.016	1.191	1.294	1.364	yes
PA-2	2.045	0.026	1.290	1.980	2.105	yes
NASQ	0.403	0.005	1.351	0.393	0.415	yes
RGYR	5.204	0.873	16.775	3.907	7.698	
PTRD	1.874	0.019	1.035	1.804	1.907	
ABRR	0.516	0.008	1.605	0.498	0.533	yes
ABSR	1.507	0.023	1.505	1.438	1.547	yes
ACCR	0.508	0.016	3.135	0.469	0.541	yes
ALAR	1.550	0.017	1.111	1.498	1.574	
OCFC	1.224	0.020	1.654	1.154	1.254	
OEFC	1.332	0.030	2.266	1.239	1.368	

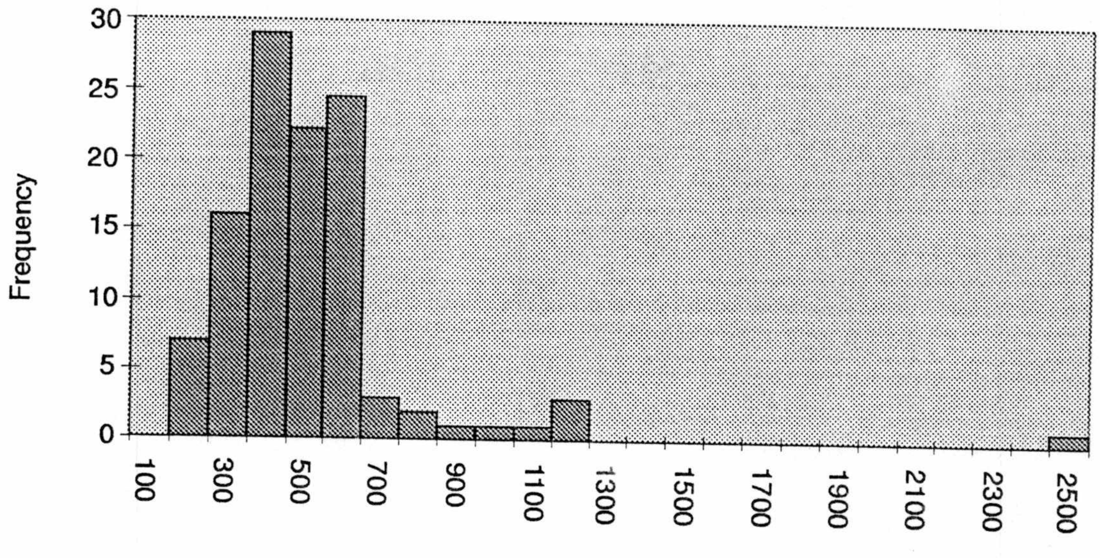


(a) Tennessee River Watershed low-resolution majority-rule data set

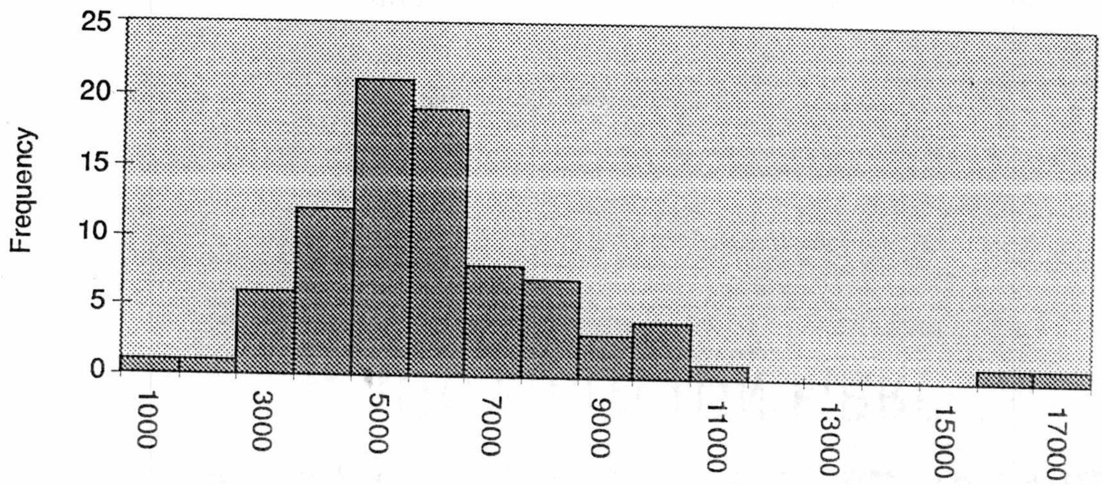


(b) Tennessee River Watershed low-resolution nearest-neighbor data set

Figure B-1: Histograms of Area-Weighted Average of Fractal Dimension from Density-Area Scaling (FDDA) showing extreme outliers.



(a) Average Patch Size (PSIZ)



(b) Number of Patches (NPAT)

Figure B-2: Histograms of metrics with extreme outliers for the Tennessee River Watershed 5-class data set.

Table B-9: Pearson Correlation Coefficients among metrics for the Tennessee River Watershed original data set.

	NTYP	PMAX	SIDI	SIEV	SHDI	SHEV	SUMD	SIDA	SICO	SHDA	SHCO	FDDA	NPAT	LPAT	PSIZ	P005	PA-1	DSTA	PA-2	NASQ	RGYR	PTRD	ABRR	ABSR	ACCR	ALAR	OCFC	OEFC
NTYP	1.000																											
PMAX	-.250	1.000																										
SIDI	.280	**-.969**	1.000																									
SIEV	.195	**-.965**	**.996**	1.000																								
SHDI	.285	**-.920**	**.970**	**.964**	1.000																							
SHEV	.011	**-.886**	**.934**	**.954**	**.958**	1.000																						
SUMD	-.137	**.742**	**-.779**	**-.788**	**-.760**	**-.773**	1.000																					
SIDA	.275	**-.967**	**.998**	**.995**	**.962**	**.929**	**-.812**	1.000																				
SICO	-.254	**.967**	**-.997**	**-.996**	**-.961**	**-.935**	**.815**	**-.999**	1.000																			
SHDA	.273	**-.919**	**.969**	**.965**	**.993**	**.959**	**-.829**	**.968**	**-.968**	1.000																		
SHCO	.020	**.874**	**-.921**	**-.945**	**-.939**	**-.993**	**.833**	**-.923**	**.929**	**-.953**	1.000																	
FDDA	-.228	**.878**	**-.890**	**-.890**	**-.887**	**-.867**	**.887**	**-.904**	**.905**	**-.919**	**.888**	1.000																
NPAT	.048	-.479	.546	.562	.574	**.609**	**-.881**	.576	-.581	**.650**	**-.680**	**-.642**	1.000															
LPAT	-.338	**.907**	**-.899**	**-.888**	**-.867**	**-.811**	**.661**	**-.892**	**.889**	**-.861**	**.793**	**.798**	-.470	1.000														
PSIZ	-.120	.523	-.586	-.593	**-.619**	**-.627**	**.854**	**-.616**	**.619**	**-.684**	**.686**	**.684**	**-.881**	.480	1.000													
P005	-.014	.362	-.433	-.450	-.463	-.506	**.804**	-.464	.469	-.540	.581	.524	**-.978**	.393	**.836**	1.000												
PA-1	-.209	.563	-.531	-.516	-.506	-.452	.201	-.515	.512	-.469	.406	.537	.191	.458	-.130	-.334	1.000											
DSTA	.184	-.485	.442	.428	.412	.360	-.110	.426	-.423	.371	-.312	-.466	-.275	-.381	.202	.408	**-.975**	1.000										
PA-2	.297	-.527	.553	.538	.530	.470	-.385	.549	-.546	.522	-.453	-.422	.220	-.583	-.159	-.182	-.328	.143	1.000									
NASQ	-.031	.239	-.298	-.307	-.336	-.353	.363	-.302	.304	-.353	.368	.211	-.437	.270	.356	.436	-.133	.344	**-.741**	1.000								
RGYR	.319	-.568	.545	.521	.502	.419	-.074	.518	-.512	.443	-.350	-.418	-.268	-.588	.248	.357	-.816	**.756**	**.647**	-.094	1.000							
PTRD	-.194	.448	-.512	-.502	**-.603**	-.559	.326	-.488	.485	-.581	.529	.453	-.326	.541	.264	.287	.264	-.154	-.550	.501	-.421	1.000						
ABRR	.168	-.039	.002	-.023	.003	-.071	.191	-.012	.019	-.029	.107	-.088	-.367	-.016	.293	.415	-.491	**.651**	-.423	**.847**	.272	.070	1.000					
ABSR	-.041	-.371	.298	.307	.213	.237	-.206	.307	-.309	.214	-.235	-.362	-.108	-.204	.065	.197	**-.642**	**.707**	-.125	.463	.339	.404	.456	1.000				
ACCR	-.207	.564	-.587	-.576	**-.620**	-.574	.292	-.566	.563	-.586	.531	.526	-.058	.546	.072	-.058	**.720**	-.580	**-.681**	.485	**-.743**	**.743**	.038	-.016	1.000			
ALAR	-.109	.288	-.351	-.348	-.429	-.408	.205	-.331	.330	-.409	.384	.264	-.231	.364	.209	.189	.145	.011	-.587	**.721**	-.314	**.852**	.403	.564	**.782**	1.000		
OCFC	.180	.151	-.127	-.150	-.091	-.161	-.024	-.117	.122	-.071	.144	.121	.090	.102	-.096	-.097	.231	-.349	.451	-.468	-.057	-.161	-.354	-.536	-.119	-.317	1.000	
OEFC	-.043	.362	-.339	-.341	-.318	-.317	.017	-.318	.318	-.275	.272	.340	.220	.305	-.180	-.289	**.673**	**-.759**	.180	-.444	-.493	.123	**-.603**	-.592	.321	**-.078**	**.822**	1.000

Notes: See Appendix A for explanation of metric abbreviations.

See Table B-1 for description of data set.

Absolute values greater than or equal to 0.6 are in boldface.

Table B-10: Pearson Correlation Coefficients among metrics for the Tennessee River Watershed low-resolution majority-rule data set.

	NTYP	PMAX	SIDI	SIEV	SHDI	SHEV	SUMD	SIDA	SICO	SHDA	SHCO	FDDA	NPAT	LPAT	PSIZ	P005	PA-1	DSTA	PA-2	NASQ	RGYR	PTRD	ABRR	ABSR	ACCR	ALAR	OCFC	OEFC
NTYP	1.000																											
PMAX	-.317	1.000																										
SIDI	.341	**-.969**	1.000																									
SIEV	.263	**-.965**	**.996**	1.000																								
SHDI	.337	**-.918**	**.968**	**.963**	1.000																							
SHEV	.081	**-.885**	**.935**	**.953**	**.963**	1.000																						
SUMD	-.283	**.851**	**-.870**	**-.871**	**-.848**	**-.833**	1.000																					
SIDA	.342	**-.962**	**.991**	**.988**	**.943**	**.913**	**-.915**	1.000																				
SICO	-.322	**.961**	**-.990**	**-.989**	**-.942**	**-.918**	**.916**	**-.999**	1.000																			
SHDA	.338	**-.922**	**.968**	**.963**	**.991**	**.957**	**-.909**	**.958**	**-.958**	1.000																		
SHCO	-.068	**.882**	**-.927**	**-.947**	**-.947**	**-.991**	**.888**	**-.921**	**.927**	**-.959**	1.000																	
FDDA	-.030	-.068	.090	.093	.062	.072	-.109	.109	-.111	.074	-.085	1.000																
NPAT	.217	**-.730**	**.776**	**.782**	**.802**	**.801**	**-.929**	**.806**	**-.809**	**.860**	**-.856**	.114	1.000															
LPAT	-.411	**.919**	**-.910**	**-.898**	**-.865**	**-.810**	**.770**	**-.897**	**.894**	**-.863**	**.800**	.012	**-.658**	1.000														
PSIZ	-.265	**.722**	**-.783**	**-.784**	**-.791**	**-.779**	**.914**	**-.827**	**.828**	**-.847**	**.830**	-.155	**-.924**	**.662**	1.000													
P005	-.207	**.694**	**-.744**	**-.749**	**-.772**	**-.774**	**.910**	**-.776**	**.778**	**-.833**	**.831**	-.099	**-.994**	**.624**	**.923**	1.000												
PA-1	-.368	.550	-.561	-.542	-.511	-.441	.339	-.538	.532	-.477	.400	.003	-.099	.596	.157	.038	1.000											
DSTA	.284	-.548	.565	.552	.543	.493	-.373	.538	-.534	.512	-.457	-.018	.171	-.537	-.215	-.118	**-.913**	1.000										
PA-2	.238	-.186	.155	.141	.041	-.010	-.031	.170	-.166	.032	.023	.007	-.156	-.324	.087	.184	-.433	.051	1.000									
NASQ	-.184	-.052	.060	.074	.107	.156	-.135	.058	-.062	.119	-.171	-.056	.201	.089	-.172	-.219	.170	.241	**-.856**	1.000								
RGYR	.186	-.156	.125	.114	.029	-.011	.166	.093	-.090	-.025	.068	-.056	-.331	-.326	.279	.366	-.558	.272	**.835**	**-.642**	1.000							
PTRD	-.148	-.287	.266	.289	.198	.268	-.253	.288	-.294	.206	-.277	-.074	.159	-.134	-.147	-.134	-.412	.508	-.076	.297	.046	1.000						
ABRR	.087	-.153	.194	.187	.263	.236	-.229	.184	-.182	.268	-.239	.012	.253	-.082	-.254	-.263	-.010	.343	**-.671**	**.805**	-.585	-.044	1.000					
ABSR	.124	-.552	.535	.538	.480	.481	-.493	.548	-.549	.487	-.485	-.106	.311	-.460	-.322	-.270	**-.741**	**.801**	.076	.216	.108	**.801**	.171	1.000				
ACCR	-.254	-.045	.037	.060	.053	.128	-.223	.069	-.076	.095	-.174	-.138	.299	.140	-.245	-.327	.293	.039	**-.718**	**.848**	**-.687**	.489	.593	.322	1.000			
ALAR	-.038	-.381	.375	.388	.349	.384	-.449	.403	-.406	.377	-.412	-.150	.373	-.221	-.349	-.364	-.302	.533	-.355	**.643**	-.315	**.770**	.499	**.792**	**.807**	1.000		
OCFC	.065	.094	-.173	-.188	-.207	-.251	.033	-.141	.145	-.170	.216	.186	-.084	.070	.085	.087	.208	-.386	.369	-.404	.080	-.409	-.234	-.299	-.278	-.330	1.000	
OEFC	-.056	.286	-.346	-.353	-.380	-.397	.134	-.297	.299	-.331	.347	.160	-.099	.269	.122	.081	.549	**-.694**	.214	-.332	-.152	-.466	-.239	-.494	-.126	-.374	**.844**	1.000

Notes: See Appendix A for explanation of metric abbreviations.

See Table B-1 for description of data set.

Absolute values greater than or equal to 0.6 are in boldface.

Table B-11: Pearson Correlation Coefficients among metrics for the Tennessee River Watershed high-resolution majority-rule data set.

	NTYP	PMAX	SIDI	SIEV	SHDI	SHEV	SUMD	SIDA	SICO	SHDA	SHCO	FDDA	NPAT	LPAT	PSIZ	P005	PA-1	DSTA	PA-2	NASQ	RGYR	PTRD	ABRR	ABSR	ACCR	ALAR	OCFC	OEFC
NTYP	1.000																											
PMAX	-.317	1.000																										
SIDI	.341	-.969	1.000																									
SIEV	.263	-.965	.996	1.000																								
SHDI	.337	-.918	.968	.963	1.000																							
SHEV	.081	-.885	.935	.953	.963	1.000																						
SUMD	-.283	.851	-.870	-.871	-.848	-.832	1.000																					
SIDA	.342	-.969	.999	.996	.963	.931	-.884	1.000																				
SICO	-.324	.969	-.999	-.997	-.963	-.936	.885	-.999	1.000																			
SHDA	.337	-.923	.971	.966	.998	.962	-.881	.968	-.968	1.000																		
SHCO	-.073	.886	-.934	-.953	-.957	-.998	.863	-.932	.937	-.961	1.000																	
FDDA	-.303	.868	-.882	-.880	-.881	-.858	.976	-.892	.892	-.908	.882	1.000																
NPAT	.217	-.730	.776	.782	.802	.801	-.930	.786	-.788	.833	-.831	-.892	1.000															
LPAT	-.411	.919	-.909	-.898	-.865	-.809	.769	-.908	.906	-.866	.806	.805	-.658	1.000														
PSIZ	-.265	.722	-.783	-.784	-.791	-.779	.915	-.796	.797	-.821	.806	.877	-.924	.661	1.000													
P005	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000												
PA-1	-.333	.677	-.677	-.660	-.658	-.594	.540	-.672	.669	-.652	.584	.632	-.346	.633	.339	.000	1.000											
DSTA	.025	-.240	.252	.254	.267	.272	-.228	.249	-.250	.268	-.272	-.287	.235	-.129	-.242	.000	-.478	1.000										
PA-2	.365	-.639	.621	.602	.558	.484	-.450	.620	-.615	.550	-.472	-.518	.188	-.673	-.202	.000	-.835	-.035	1.000									
NASQ	-.346	.649	-.647	-.629	-.626	-.556	.497	-.642	.638	-.618	.543	.582	-.282	.643	.269	.000	.915	-.095	-.948	1.000								
RGYR	.328	-.568	.547	.530	.480	.412	-.274	.539	-.535	.458	-.387	-.369	.026	-.642	-.042	.000	-.776	.009	.944	-.864	1.000							
PTRD	-.085	.188	-.197	-.192	-.201	-.180	.122	-.194	.193	-.192	.171	.157	-.026	.217	.021	.000	.188	.593	-.492	.532	-.404	1.000						
ABRR	-.286	.676	-.663	-.652	-.628	-.580	.528	-.660	.658	-.623	.571	.606	-.311	.651	.290	.000	.905	-.117	-.935	.977	-.863	.474	1.000					
ABSR	-.367	.508	-.521	-.496	-.533	-.445	.360	-.513	.507	-.519	.427	.456	-.190	.527	.183	.000	.803	-.029	-.821	.918	-.744	.633	.827	1.000				
ACCR	-.347	.630	-.633	-.614	-.625	-.551	.477	-.627	.622	-.615	.537	.573	-.279	.620	.269	.000	.927	-.176	-.897	.984	-.820	.518	.939	.958	1.000			
ALAR	-.323	.538	-.545	-.526	-.550	-.477	.390	-.538	.533	-.538	.461	.482	-.215	.547	.205	.000	.798	.009	-.835	.933	-.754	.687	.862	.989	.964	1.000		
OCFC	.139	.085	-.129	-.146	-.171	-.225	.033	-.120	.124	-.157	.212	.104	-.071	.014	.077	.000	.198	-.526	.096	.056	-.006	-.061	.113	.017	.098	.035	1.000	
OEFC	.158	.010	-.047	-.062	-.096	-.146	-.047	-.037	.041	-.080	.132	.033	.002	-.070	.002	.000	.181	-.555	.144	.027	.032	-.068	.067	.032	.090	.043	.980	1.000

Notes: See Appendix A for explanation of metric abbreviations.

See Table B-1 for description of data set.

Absolute values greater than or equal to 0.6 are in boldface.

Table B-12: Pearson Correlation Coefficients among metrics for the Tennessee River Watershed low-resolution nearest-neighbor data set.

	NTYP	PMAX	SIDI	SIEV	SHDI	SHEV	SUMD	SIDA	SICO	SHDA	SHCO	FDDA	NPAT	LPAT	PSIZ	P005	PA-1	DSTA	PA-2	NASQ	RGYR	PTRD	ABRR	ABSR	ACCR	ALAR	OCFC	OEFC
NTYP	1.000																											
PMAX	-.262	1.000																										
SIDI	.294	**-.969**	1.000																									
SIEV	.208	**-.965**	**.996**	1.000																								
SHDI	.298	**-.920**	**.970**	**.964**	1.000																							
SHEV	.023	**-.887**	**.933**	**.954**	**.958**	1.000																						
SUMD	-.206	**.831**	**-.853**	**-.857**	**-.825**	**-.819**	1.000																					
SIDA	.288	**-.959**	**.988**	**.985**	**.940**	**.908**	**-.908**	1.000																				
SICO	-.265	**.958**	**-.987**	**-.987**	**-.939**	**-.915**	**.910**	**-.999**	1.000																			
SHDA	.292	**-.923**	**.969**	**.964**	**.990**	**.953**	**-.895**	**.958**	**-.958**	1.000																		
SHCO	.001	**.879**	**-.922**	**-.945**	**-.937**	**-.989**	**.881**	**-.916**	**.923**	**-.953**	1.000																	
FDDA	-.017	-.085	.108	.112	.077	.085	-.138	.133	-.134	.093	-.101	1.000																
NPAT	.139	**-.699**	**.749**	**.759**	**.763**	**.776**	**-.931**	**.790**	**-.794**	**.833**	**-.839**	.144	1.000															
LPAT	-.375	**.907**	**-.895**	**-.882**	**-.851**	**-.790**	**.737**	**-.878**	**.875**	**-.849**	**.776**	-.015	**-.609**	1.000														
PSIZ	-.203	**.703**	**-.764**	**-.767**	**-.768**	**-.760**	**.918**	**-.819**	**.820**	**-.833**	**.818**	-.174	**-.924**	**.617**	1.000													
P005	-.134	**.670**	**-.722**	**-.732**	**-.741**	**-.754**	**.917**	**-.765**	**.768**	**-.813**	**.820**	-.123	**-.995**	.581	**.923**	1.000												
PA-1	-.343	.572	-.563	-.542	-.517	-.439	.274	-.522	.515	-.470	.383	.058	-.032	**.614**	.078	-.015	1.000											
DSTA	.252	-.554	.543	.528	.527	.472	-.287	.494	-.490	.483	-.420	-.076	.080	-.532	-.116	-.043	**-.897**	1.000										
PA-2	.228	-.204	.185	.170	.077	.026	-.049	.200	-.195	.067	-.011	-.005	-.112	-.358	.058	.140	-.448	.032	1.000									
NASQ	-.203	.001	-.015	.002	.040	.095	-.073	-.022	.018	.053	-.111	-.065	.139	.133	-.123	-.158	.215	.233	**-.861**	1.000								
RGYR	.191	-.163	.133	.119	.058	.010	.187	.087	-.083	-.005	.056	-.060	-.326	-.362	.286	.360	**-.602**	.301	**.819**	**-.622**	1.000							
PTRD	-.197	-.311	.278	.307	.193	.279	-.279	.305	-.312	.208	-.296	-.158	.224	-.168	-.187	-.216	-.258	.343	-.032	.247	.040	1.000						
ABRR	.081	-.108	.125	.116	.206	.176	-.126	.095	-.093	.202	-.169	-.040	.124	-.044	-.133	-.132	-.060	.430	**-.684**	**.808**	-.504	-.127	1.000					
ABSR	.070	**-.612**	.579	.585	.511	.523	-.494	.586	-.588	.515	-.521	-.167	.316	-.507	-.320	-.292	**-.697**	**.734**	.157	.140	.185	**.768**	.100	1.000				
ACCR	-.327	.000	-.021	.010	-.011	.090	-.206	.019	-.027	.041	-.147	-.129	.305	.189	-.257	-.336	.406	-.062	**-.687**	**.809**	**-.692**	.514	.478	.262	1.000			
ALAR	-.133	-.394	.362	.383	.323	.384	-.448	.392	-.398	.359	-.420	-.187	.389	-.217	-.359	-.392	-.210	.436	-.295	.575	-.299	**.776**	.391	**.782**	**.783**	1.000		
OCFC	.070	-.023	-.042	-.053	-.059	-.094	-.106	-.009	.012	-.018	.054	.110	.070	-.107	-.045	-.068	.108	-.271	.362	-.319	.090	-.307	-.195	-.142	-.157	-.134	1.000	
OEFC	-.066	.262	-.304	-.305	-.333	-.332	.041	-.241	.241	-.271	.268	.114	.016	.204	.025	-.034	.512	**-.675**	.251	-.322	-.141	-.310	-.333	-.376	-.004	-.204	**.823**	1.000

Notes: See Appendix A for explanation of metric abbreviations.

See Table B-1 for description of data set.

Absolute values greater than or equal to 0.6 are in boldface.

Table B-13: Pearson Correlation Coefficients among metrics for the Tennessee River Watershed 5-class data set.

	NTYP	PMAX	SIDI	SIEV	SHDI	SHEV	SUMD	SIDA	SICO	SHDA	SHCO	FDDA	NPAT	LPAT	PSIZ	P005	PA-1	DSTA	PA-2	NASQ	RGYR	PTRD	ABRR	ABSR	ACCR	ALAR	OCFC	OEFC
NTYP	1.000																											
PMAX	-.159	1.000																										
SIDI	.213	**-.964**	1.000																									
SIEV	.120	**-.963**	**.995**	1.000																								
SHDI	.260	**-.918**	**.977**	**.967**	1.000																							
SHEV	.032	**-.911**	**.960**	**.974**	**.971**	1.000																						
SUMD	-.194	**.796**	**-.826**	**-.823**	**-.825**	**-.815**	1.000																					
SIDA	.212	**-.964**	**.999**	**.994**	**.972**	**.957**	**-.852**	1.000																				
SICO	-.178	**.964**	**-.998**	**-.997**	**-.969**	**-.963**	**.852**	**-.999**	1.000																			
SHDA	.259	**-.917**	**.973**	**.963**	**.994**	**.967**	**-.880**	**.974**	**-.972**	1.000																		
SHCO	-.028	**.909**	**-.954**	**-.968**	**-.963**	**-.994**	**.869**	**-.956**	**.963**	**-.970**	1.000																	
FDDA	-.277	**.855**	**-.871**	**-.858**	**-.864**	**-.831**	**.939**	**-.889**	**.885**	**-.901**	**.867**	1.000																
NPAT	.143	-.508	.579	.576	**.652**	**.644**	**-.830**	.599	-.598	**.705**	**-.698**	**-.652**	1.000															
LPAT	-.203	**.904**	**-.911**	**-.903**	**-.899**	**-.877**	**.744**	**-.906**	**.904**	**-.892**	**.869**	**.801**	-.548	1.000														
PSIZ	-.267	.547	**-.639**	**-.631**	**-.665**	**-.643**	**.724**	**-.657**	**.655**	**-.696**	**.673**	**.659**	**-.636**	.512	1.000													
P005	-.085	.402	-.471	-.474	-.550	-.556	**.740**	-.491	.493	**-.602**	**.609**	.532	**-.977**	.466	.572	1.000												
PA-1	-.051	.444	-.398	-.397	-.317	-.310	.212	-.396	.396	-.305	.297	.430	.253	.301	.164	-.400	1.000											
DSTA	.030	-.366	.306	.304	.220	.212	-.082	.299	-.299	.198	-.189	-.326	-.359	-.244	-.005	.489	**-.951**	1.000										
PA-2	.155	-.504	.556	.557	.547	.547	**-.611**	.571	-.571	.573	-.573	-.554	.465	-.498	-.594	-.428	-.091	-.165	1.000									
NASQ	-.046	.096	-.164	-.169	-.217	-.230	.329	-.176	.178	-.247	.261	.172	-.449	.085	.443	.455	-.239	.523	**-.753**	1.000								
RGYR	.134	-.502	.475	.470	.393	.377	-.130	.458	-.457	.350	-.332	-.362	-.250	-.515	.023	.317	-.575	.538	.397	.091	1.000							
PTRD	-.154	.187	-.271	-.259	-.375	-.347	.339	-.266	.262	-.384	.357	.276	-.489	.269	.253	.482	-.247	.393	-.475	**.605**	.027	1.000						
ABRR	.012	-.022	-.034	-.043	-.062	-.087	.175	-.046	.049	-.086	.112	.002	-.318	-.036	.359	.353	-.313	.573	**-.667**	**.948**	.109	.399	1.000					
ABSR	-.242	-.261	.171	.195	.088	.145	-.082	.175	-.184	.084	-.141	-.180	-.172	-.130	-.031	.222	-.564	**.656**	-.276	.519	.168	**.762**	.430	1.000				
ACCR	-.200	.345	-.395	-.384	-.404	-.376	.358	-.395	.392	-.410	.382	.412	-.158	.293	.413	.056	.455	-.207	**-.645**	**.627**	-.365	**.615**	.471	.361	1.000			
ALAR	-.214	.117	-.202	-.188	-.262	-.229	.243	-.201	.196	-.271	.238	.195	-.288	.154	.348	.255	-.050	.280	**-.604**	**.774**	-.078	**.847**	**.635**	**.749**	**.859**	1.000		
OCFC	.246	-.018	.067	.044	.074	.019	-.244	.084	-.076	.112	-.057	-.173	.210	-.028	-.311	-.161	.056	-.227	.576	-.469	-.007	-.374	-.367	-.501	-.452	-.515	1.000	
OEFC	.043	.161	-.126	-.133	-.100	-.113	-.164	-.105	.107	-.051	.063	.032	.327	.118	-.212	-.349	.455	**-.605**	.492	-.549	-.312	-.298	-.528	-.492	-.125	-.352	**.845**	1.000

Notes: See Appendix A for explanation of metric abbreviations.

See Table B-1 for description of data set.

Absolute values greater than or equal to 0.6 are in boldface.

Table B-14: Pearson Correlation Coefficients among metrics for the Chesapeake Bay Watershed quadrants data set.

	NTYP	PMAX	SIDI	SIEV	SHDI	SHEV	SUMD	SIDA	SICO	SHDA	SHCO	FDDA	NPAT	LPAT	PSIZ	P005	PA-1	DSTA	PA-2	NASQ	RGYR	PTRD	ABRR	ABSR	ACCR	ALAR	OCFC	OEFC
NTYP	1.000																											
PMAX	-.009	1.000																										
SIDI	.028	**-.947**	1.000																									
SIEV	-.030	**-.946**	**.998**	1.000																								
SHDI	.089	**-.856**	**.955**	**.950**	1.000																							
SHEV	-.037	**-.858**	**.954**	**.956**	**.992**	1.000																						
SUMD	.008	.525	-.487	-.489	-.431	-.435	1.000																					
SIDA	.014	**-.953**	**.991**	**.991**	**.930**	**.931**	-.590	1.000																				
SICO	.008	**.953**	**-.991**	**-.991**	**-.928**	**-.932**	.591	**-.999**	1.000																			
SHDA	.083	**-.871**	**.944**	**.939**	**.971**	**.963**	**-.634**	**.950**	**-.948**	1.000																		
SHCO	.054	**.871**	**-.941**	**-.945**	**-.959**	**-.970**	**.639**	**-.949**	**.951**	**-.990**	1.000																	
FDDA	.074	.564	-.571	-.576	-.529	-.542	**.798**	**-.643**	**.645**	**-.662**	**.676**	1.000																
NPAT	.108	-.304	.271	.266	.280	.267	**-.851**	.358	-.355	.469	-.456	-.425	1.000															
LPAT	-.052	.573	-.514	-.512	-.415	-.410	**.616**	-.578	.577	-.524	.518	.556	-.439	1.000														
PSIZ	-.085	.200	-.136	-.132	-.111	-.101	**.760**	-.233	.231	-.305	.295	.383	**-.826**	.511	1.000													
P005	-.076	.224	-.184	-.180	-.194	-.186	**.798**	-.271	.269	-.381	.373	.337	**-.984**	.419	**.824**	1.000												
PA-1	.059	.307	-.385	-.388	-.356	-.366	-.205	-.335	.337	-.245	.254	.337	.590	-.038	**-.622**	**-.671**	1.000											
DSTA	-.076	-.315	.391	.396	.364	.375	.200	.342	-.343	.252	-.263	-.345	-.576	.042	**.621**	**.657**	**-.993**	1.000										
PA-2	.077	-.138	.157	.154	.097	.090	.055	.155	-.153	.062	-.054	-.206	-.333	-.286	.317	.339	-.483	.404	1.000									
NASQ	-.169	-.234	.219	.229	.173	.193	-.348	.252	-.256	.237	-.260	-.350	.266	-.116	-.228	-.228	.005	.095	-.588	1.000								
RGYR	-.028	-.144	.205	.206	.176	.181	.411	.145	-.145	.036	-.040	-.099	**-.703**	-.014	.751	.730	**-.862**	**.840**	.710	-.282	1.000							
PTRD	-.097	-.374	.438	.444	.385	.398	.043	.405	-.407	.311	-.324	-.394	-.384	.049	.402	.482	**-.832**	**.849**	.179	.310	.565	1.000						
ABRR	-.076	-.248	.257	.263	.273	.285	-.465	.298	-.300	.352	-.366	-.467	.391	-.207	-.244	-.377	-.019	.087	-.265	**.666**	-.174	.022	1.000					
ABSR	-.079	-.428	.512	.516	.481	.494	-.009	.477	-.478	.406	-.418	-.463	-.330	.010	.408	.426	**-.866**	**.889**	.206	.322	.597	**.958**	.196	1.000				
ACCR	-.029	-.180	.167	.169	.161	.164	-.496	.214	-.214	.265	-.270	-.224	**.631**	-.139	-.537	**-.625**	.436	-.372	-.582	**.687**	**-.640**	.030	.490	.054	1.000			
ALAR	-.078	-.466	.539	.544	.508	.520	-.200	.526	-.528	.479	-.491	-.528	-.077	-.075	.197	.163	**-.661**	**.701**	.053	.506	.363	**.881**	.382	**.936**	.381	1.000		
OCFC	.085	.055	-.114	-.117	-.145	-.153	-.495	-.035	.037	.005	.003	-.174	.508	-.238	-.541	-.511	.484	-.541	.212	-.157	-.443	-.391	.005	-.443	.205	-.306	1.000	
OEFC	.115	.190	-.266	-.272	-.277	-.291	-.417	-.190	.192	-.125	.138	.082	**.625**	-.194	**-.667**	**-.673**	**.816**	**-.849**	-.102	-.088	**-.718**	**-.675**	-.004	**-.724**	.355	-.531	**.879**	1.000

Notes: See Appendix A for explanation of metric abbreviations.

See Table B-1 for description of data set.

Absolute values greater than or equal to 0.6 are in boldface.

Table B-15: Pearson Correlation Coefficients among metrics for the Chesapeake Bay Watershed sub-watersheds data set.

	NTYP	PMAX	SIDI	SIEV	SHDI	SHEV	SUMD	SIDA	SICO	SHDA	SHCO	FDDA	NPAT	LPAT	PSIZ	P005	PA-1	DSTA	PA-2	NASQ	RGYR	PTRD	ABRR	ABSR	ACCR	ALAR	OCFC	OEFC
NTYP	1.000																											
PMAX	.115	1.000																										
SIDI	-.046	**-.958**	1.000																									
SIEV	-.091	**-.961**	**.999**	1.000																								
SHDI	.024	**-.900**	**.966**	**.962**	1.000																							
SHEV	-.079	**-.909**	**.968**	**.969**	**.995**	1.000																						
SUMD	-.127	**.665**	**-.723**	**-.715**	**-.755**	**-.740**	1.000																					
SIDA	-.036	**-.953**	**.997**	**.996**	**.962**	**.962**	**-.765**	1.000																				
SICO	.052	**.954**	**-.997**	**-.997**	**-.960**	**-.963**	**.762**	**-.999**	1.000																			
SHDA	.049	**-.884**	**.952**	**.946**	**.990**	**.982**	**-.840**	**.957**	**-.955**	1.000																		
SHCO	.052	**.895**	**-.956**	**-.956**	**-.987**	**-.990**	**.827**	**-.960**	**.960**	**-.995**	1.000																	
FDDA	.040	**.640**	**-.730**	**-.730**	**-.731**	**-.733**	**.768**	**-.755**	**.755**	**-.766**	**.770**	1.000																
NPAT	.245	-.329	.320	.308	.377	.351	-.592	.348	-.344	.444	-.419	-.157	1.000															
LPAT	.091	**.620**	**-.613**	**-.615**	-.598	**-.606**	.472	**-.614**	**.615**	-.593	**.603**	.462	-.024	1.000														
PSIZ	-.251	.195	-.227	-.215	-.311	-.285	**.663**	-.265	.260	-.406	.381	.120	**-.706**	.196	1.000													
P005	-.204	.282	-.291	-.281	-.384	-.362	**.717**	-.326	.323	-.478	.458	.174	**-.702**	.254	**.908**	1.000												
PA-1	.240	.408	-.466	-.475	-.393	-.416	.065	-.450	.454	-.332	.356	**.616**	.349	.208	**-.612**	-.599	1.000											
DSTA	-.291	-.412	.462	.473	.388	.417	-.061	.446	-.450	.327	-.357	**-.620**	-.358	-.224	**.621**	.588	**-.994**	1.000										
PA-2	-.022	-.152	.228	.228	.142	.144	-.038	.233	-.233	.117	-.119	-.323	-.334	-.264	.355	.390	-.545	.487	1.000									
NASQ	-.346	-.386	.355	.370	.303	.338	-.403	.374	-.379	.336	-.371	-.475	.133	-.275	-.097	-.168	-.244	.322	-.358	1.000								
RGYR	-.351	-.215	.255	.270	.168	.204	.228	.228	-.234	.082	-.117	-.355	-.578	-.232	**.787**	**.711**	**-.871**	**.876**	**.657**	.058	1.000							
PTRD	-.127	-.513	.581	.585	.508	.520	-.268	.573	-.575	.472	-.485	**-.638**	-.164	-.190	.325	.354	**-.840**	**.831**	.315	.394	.568	1.000						
ABRR	-.306	-.243	.196	.209	.198	.229	-.433	.227	-.232	.257	-.288	-.421	.279	-.241	-.241	-.333	-.047	.113	-.204	**.677**	-.024	-.097	1.000					
ABSR	-.210	-.555	**.612**	**.619**	.549	.569	-.322	**.606**	**-.609**	.517	-.539	**-.736**	-.164	-.272	.344	.325	**-.890**	**.895**	.341	.458	**.644**	**.962**	.091	1.000				
ACCR	.024	-.254	.234	.232	.263	.260	-.567	.264	-.263	.343	-.341	-.195	.477	-.144	**-.683**	**-.707**	.387	-.349	-.534	.549	-.597	.062	.307	.059	1.000			
ALAR	-.198	-.599	**.652**	**.658**	**.604**	**.622**	-.490	**.657**	**-.660**	**.600**	**-.620**	**-.769**	-.004	-.310	.113	.079	**-.727**	**.744**	.192	.597	.443	**.908**	.220	**.953**	.346	1.000		
OCFC	.525	.150	-.145	-.168	-.130	-.183	-.319	-.104	.113	-.038	.091	.166	.478	.157	**-.632**	-.498	.585	**-.640**	-.073	-.225	**-.735**	-.324	-.134	-.450	.348	-.321	1.000	
OEFC	.465	.277	-.297	-.317	-.256	-.303	-.179	-.263	.270	-.169	.216	.401	.483	.163	**-.696**	**-.604**	**.836**	**-.869**	-.278	-.246	**-.866**	-.594	-.117	**-.702**	.400	-.543	**.922**	1.000

Notes: See Appendix A for explanation of metric abbreviations.

See Table B-1 for description of data set.

Absolute values greater than or equal to 0.6 are in boldface.

APPENDIX C

MULTIVARIATE STATISTICS

Table C-1: Results of principal components analysis and varimax rotation for the Tennessee River Watershed original data set. See Appendix A for explanation of metric abbreviations and Table B-1 for data set description.

Eigenvalues and cumulative proportion of variance explained by principal components analysis							
Eigenvalue	14.296	5.635	3.474	1.342	1.135	.720	
Cum. Variance	.511	.712	.836	.884	.924	.950	
METRIC	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	Communality
			Factor pattern after varimax rotation				
NTYP	-.160	.104	-.068	.092	.140	**.917**	.911
PMAX	**.889**	-.318	.054	.113	.113	-.144	.941
SIDI	**-.923**	.259	-.127	-.109	-.102	.153	.980
SIEV	**-.931**	.246	-.120	-.125	-.120	.072	.977
SHDI	**-.912**	.212	-.245	-.044	-.077	.128	.960
SHEV	**-.917**	.169	-.220	-.092	-.131	-.131	.960
SUMD	**.940**	.149	-.028	.002	-.190	.057	.945
SIDA	**-.938**	.236	-.097	-.106	-.076	.142	.982
SICO	**.941**	-.232	.095	.110	.080	-.122	.982
SHDA	**-.948**	.153	-.212	-.034	-.031	.102	.980
SHCO	**.943**	-.105	.185	.083	.089	.171	.979
TMAS	**.936**	-.201	.076	-.120	-.039	-.019	.939
NPAT	**-.783**	**-.535**	-.079	-.019	.172	-.085	.942
LPAT	**.809**	-.236	.173	.147	.142	-.301	.872
PSIZ	**.799**	.474	.038	-.053	-.184	.025	.902
P005	**.686**	**.643**	.071	.057	-.146	.068	.919
PA-1	.347	**-.896**	.056	-.193	.057	.001	.967
DSTA	-.270	**.877**	.034	.340	-.153	.010	.983
PA-2	-.397	.356	-.328	**-.618**	.338	.221	.937
NASQ	.283	.152	.467	**.723**	-.302	.075	.941
RGYR	-.254	**.855**	-.208	-.173	-.087	.264	.947
PTRD	.394	-.106	**.874**	.012	-.009	-.094	.940
ABRR	.066	.432	.142	**.820**	-.186	.168	.946
ABSR	-.303	**.551**	**.662**	.191	-.228	-.148	.944
ACCR	.383	**-.615**	**.633**	.157	-.083	.013	.958
ALAR	.219	-.069	**.904**	.285	-.110	.013	.963
OCFC	.115	-.089	-.222	-.217	**.904**	.144	.955
OEFC	.205	**-.521**	.014	-.353	**.715**	.021	.951
						Sum	26.603
			Variance explained by each factor after rotation				
	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	
	12.892	5.118	3.156	2.207	1.918	1.311	

Table C-2: Results of principal components analysis and varimax rotation for the Tennessee River Watershed low-resolution majority-rule data set. See Appendix A for explanation of metric abbreviations and Table B-1 for data set description.

Eigenvalues and cumulative proportion of variance explained by principle components analysis							
Eigenvalue	13.948	4.890	3.533	1.465	1.221	1.028	
Cum. Variance	.498	.673	.799	.851	.895	.932	
METRIC	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	Communality
			Factor pattern after varimax rotation				
NTYP	.251	-.119	-.086	.103	**.796**	-.150	.751
PMAX	**-.897**	.109	-.243	.087	-.186	-.041	.920
SIDI	**.935**	-.080	.188	-.156	.185	.056	.978
SIEV	**.940**	-.074	.200	-.171	.115	.066	.975
SHDI	**.934**	.006	.093	-.212	.187	.036	.963
SHEV	**.929**	.028	.130	-.256	-.039	.068	.952
SUMD	**-.936**	-.117	-.187	-.132	-.063	.003	.947
SIDA	**.944**	-.066	.218	-.082	.164	.053	.980
SICO	**-.946**	.064	-.221	.085	-.145	-.056	.979
SHDA	**.961**	.040	.106	-.135	.166	.025	.983
SHCO	**-.950**	-.065	-.143	.180	.073	-.059	.969
TMAS	.109	.003	-.085	.152	-.090	**.932**	.919
NPAT	**.920**	.256	-.002	.122	-.079	-.052	.936
LPAT	**-.844**	.251	-.124	.117	-.304	.038	.898
PSIZ	**-.903**	-.213	-.007	-.107	.014	.000	.872
P005	**-.902**	-.290	.036	-.136	.100	.080	.934
PA-1	-.306	.416	**-.538**	.380	-.487	-.148	.960
DSTA	.306	-.036	**.593**	**-.501**	.484	.168	.960
PA-2	.079	**-.923**	.088	.214	.141	-.018	.931
NASQ	.035	**.897**	.226	-.238	.009	.018	.914
RGYR	-.055	**-.885**	.140	-.198	.192	.011	.882
PTRD	.133	.060	**.885**	-.202	-.260	-.066	.918
ABRR	.147	**.819**	.037	-.166	.398	.126	.895
ABSR	.354	.005	**.871**	-.184	.158	-.030	.943
ACCR	.077	**.831**	.419	.064	-.230	-.173	.959
ALAR	.275	**.502**	**.791**	-.060	.011	-.122	.972
OCFC	-.080	-.238	-.133	**.854**	.141	.166	.858
OEFC	-.174	-.077	-.284	**.900**	-.084	.051	.936
						Sum	26.084

Variance explained by each factor after rotation

FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6
12.530	4.618	3.517	2.532	1.814	1.074

Table C-3: Results of principal components analysis and varimax rotation for the Tennessee River Watershed high-resolution majority-rule data set. See Appendix A for explanation of metric abbreviations and Table B-1 for data set description.

Eigenvalues and cumulative proportion of variance explained by principle components analysis							
Eigenvalue	16.340	4.414	2.540	1.285	.866	.690	
Cum. Variance	.605	.769	.863	.910	.942	.968	
METRIC	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	Communality
NTYP	.181	-.249	.113	.023	**.940**	.011	.993
PMAX	**-.866**	.384	-.005	-.048	-.071	-.204	.946
SIDI	**.905**	-.364	-.043	.022	.094	.162	.989
SIEV	**.915**	-.345	-.055	.019	.016	.166	.989
SHDI	**.905**	-.340	-.101	-.001	.092	.069	.958
SHEV	**.919**	-.274	-.139	-.008	-.159	.076	.970
SUMD	**-.924**	.193	-.079	-.056	-.071	.203	.946
SIDA	**.910**	-.357	-.032	.024	.096	.148	.988
SICO	**-.913**	.352	.034	-.024	-.077	-.149	.988
SHDA	**.923**	-.325	-.081	.007	.092	.032	.973
SHCO	**-.935**	.255	.119	.001	.165	-.038	.982
TMAS	**-.897**	.299	-.001	-.076	-.078	.189	.941
NPAT	**.929**	.039	.020	.031	.062	-.294	.956
LPAT	**-.793**	.397	-.069	.017	-.186	-.296	.914
PSIZ	**-.921**	-.050	-.010	-.032	-.127	.261	.936
P005	.000	.000	.000	.000	.000	.000	.000
PA-1	-.389	**.850**	.125	-.323	-.060	.015	.996
DSTA	.214	-.084	-.464	**.824**	.003	-.127	.964
PA-2	.298	**-.887**	.159	-.062	.077	.228	.962
NASQ	-.336	**.933**	.010	.054	-.055	-.013	.989
RGYR	.188	**-.847**	.042	-.019	.082	.452	.965
PTRD	-.050	.512	.012	**.834**	.029	.114	.974
ABRR	-.372	**.887**	.040	.016	.002	-.073	.933
ABSR	-.215	**.910**	.038	.183	-.120	.139	.943
ACCR	-.321	**.937**	.072	.022	-.069	.065	.996
ALAR	-.247	**.910**	.051	.239	-.070	.126	.970
OCFC	-.097	.011	**.978**	-.091	.052	-.034	.978
OEFC	-.006	.026	**.987**	-.116	.056	.030	.992
						Sum	26.134

Variance explained by each factor after rotation

FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6
12.283	8.050	2.281	1.616	1.099	.804

Table C-4: Results of principal components analysis and varimax rotation for the Tennessee River Watershed low-resolution nearest-neighbor data set. See Appendix A for explanation of metric abbreviations and Table B-1 for data set description.

Eigenvalues and cumulative proportion of variance explained by principal components analysis							
Eigenvalue	13.581	4.822	3.484	1.952	1.273	.952	
Cum. Variance	.485	.657	.782	.851	.897	.931	
METRIC	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	Communality
NTYP	.241	-.073	.351	-.457	.145	**-.522**	.689
PMAX	**-.890**	.082	-.340	-.129	.005	-.004	.931
SIDI	**.937**	-.072	.296	.068	-.075	.001	.981
SIEV	**.939**	-.071	.263	.114	-.094	.046	.980
SHDI	**.935**	.015	.267	-.020	-.116	-.023	.959
SHEV	**.918**	.025	.161	.127	-.171	.120	.929
SUMD	**-.944**	-.090	.011	-.143	-.141	.027	.940
SIDA	**.952**	-.071	.238	.107	-.018	.000	.979
SICO	**-.953**	.070	-.228	-.119	.023	-.013	.980
SHDA	**.968**	.041	.205	.007	-.054	-.031	.984
SHCO	**-.942**	-.053	-.091	-.161	.112	-.122	.951
TMAS	.152	.009	-.002	-.231	.133	**.829**	.781
NPAT	**.915**	.180	-.248	.089	.095	-.030	.949
LPAT	**-.817**	.216	-.395	.018	-.050	.104	.883
PSIZ	**-.910**	-.159	.188	-.041	-.077	.029	.897
P005	**-.899**	-.206	.285	-.089	-.100	.054	.952
PA-1	-.306	.319	**-.851**	-.111	.154	.061	.959
DSTA	.291	.105	**.881**	.183	-.254	-.037	.971
PA-2	.086	**-.905**	.204	-.006	.259	-.065	.940
NASQ	-.005	**.919**	.076	.233	-.161	.026	.931
RGYR	-.073	**-.802**	.471	-.013	-.062	-.021	.875
PTRD	.184	-.013	.086	**.909**	-.199	-.066	.911
ABRR	.088	**.878**	.375	-.143	-.057	-.032	.944
ABSR	.409	-.028	**.525**	**.694**	-.045	-.125	.943
ACCR	.066	**.741**	-.258	**.585**	.054	-.053	.968
ALAR	.303	.447	.206	**.786**	.065	-.122	.971
OCFC	.031	-.174	-.017	-.110	**.934**	.062	.920
OEFC	-.135	-.146	-.437	-.071	**.854**	.037	.966
						Sum	26.065

Variance explained by each factor after rotation					
FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6
12.532	4.185	3.479	2.808	1.998	1.062

Table C-5: Results of principal components analysis and varimax rotation for the Tennessee River Watershed 5-class data set. See Appendix A for explanation of metric abbreviations and Table B-1 for data set description.

Eigenvalues and cumulative proportion of variance explained by principle components analysis							
Eigenvalue	13.749	6.116	2.768	1.353	1.264	.931	
Cum. Variance	.491	.710	.808	.857	.902	.935	
METRIC	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	Communality
NTYP	.143	-.119	.054	.141	-.042	**.820**	.731
PMAX	**-.949**	.002	-.171	-.001	-.108	-.006	.941
SIDI	**.977**	-.081	.125	.009	.063	.049	.982
SIEV	**.979**	-.075	.122	-.002	.064	-.035	.984
SHDI	**.968**	-.148	.038	-.030	-.001	.102	.972
SHEV	**.969**	-.133	.029	-.056	-.001	-.102	.971
SUMD	**-.875**	.087	.020	-.211	.314	-.075	.923
SIDA	**.979**	-.075	.126	.035	.033	.047	.985
SICO	**-.981**	.073	-.125	-.031	-.033	-.016	.986
SHDA	**.975**	-.146	.027	.013	-.064	.106	.988
SHCO	**-.974**	.132	-.017	.012	.066	.100	.981
TMAS	**-.892**	.060	-.187	-.112	.152	-.179	.901
NPAT	**.682**	-.153	-.459	.139	-.477	.049	.948
LPAT	**-.926**	.022	-.005	.014	-.197	-.081	.904
PSIZ	**-.634**	.213	-.114	-.300	**.506**	-.053	.809
P005	**-.594**	.138	**.582**	-.122	.466	.013	.943
PA-1	-.294	-.034	**-.930**	.090	-.082	-.045	.970
DSTA	.220	.253	**.871**	-.252	.141	.122	.969
PA-2	**.534**	-.495	.022	**.605**	.089	-.156	.929
NASQ	-.148	**.760**	.186	-.429	.221	.280	.944
RGYR	.393	-.074	.453	.032	**.733**	-.012	.903
PTRD	-.265	**.805**	.371	.027	-.018	-.204	.898
ABRR	-.017	**.643**	.195	-.452	.198	.408	.861
ABSR	.197	**.695**	**.532**	-.169	-.159	-.314	.957
ACCR	-.279	**.832**	-.423	-.170	-.020	-.076	.984
ALAR	-.125	**.958**	.051	-.175	.007	-.124	.981
OCFC	.023	-.318	-.027	**.839**	.019	.293	.893
OEFC	-.095	-.180	-.392	**.856**	-.116	.027	.941
						Sum	26.181

Variance explained by each factor after rotation

FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6
12.977	4.395	3.353	2.560	1.592	1.304

Table C-6: Results of principal components analysis and varimax rotation for the Chesapeake Bay Watershed quadrants data set. See Appendix A for explanation of metric abbreviations and Table B-1 for data set description.

Eigenvalues and cumulative proportion of variance explained by principal components analysis							
Eigenvalue	11.522	8.040	2.889	1.645	1.016	.838	
Cum. Variance	.412	.699	.802	.861	.897	.927	
METRIC	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	Communality
NTYP	.020	-.079	-.092	.018	.001	**.989**	.994
PMAX	**-.922**	-.069	-.035	.154	-.030	.011	.881
SIDI	**.976**	.141	.042	-.104	.041	.021	.986
SIEV	**.974**	.145	.048	-.106	.042	-.036	.987
SHDI	**.966**	.125	.066	-.027	-.010	.094	.962
SHEV	**.966**	.136	.077	-.032	-.009	-.031	.959
SUMD	-.491	.464	-.240	**.612**	-.234	.025	.943
SIDA	**.964**	.073	.062	-.198	.071	.005	.982
SICO	**-.963**	-.075	-.064	.199	-.072	.018	.982
SHDA	**.957**	-.022	.119	-.184	.045	.081	.973
SHCO	**-.955**	.012	-.131	.190	-.046	.054	.971
TMAS	**-.500**	-.098	-.178	**.730**	-.235	.054	.883
NPAT	.350	**-.760**	.318	-.312	.101	.090	.918
LPAT	-.494	.236	.133	**.615**	.107	-.012	.708
PSIZ	-.208	**.775**	-.234	.338	-.114	-.039	.827
P005	-.272	**.811**	-.299	.294	-.045	-.056	.913
PA-1	-.252	**-.927**	.139	.149	-.100	.001	.975
DSTA	.252	**.944**	-.038	-.149	.064	-.009	.981
PA-2	.088	.327	**-.788**	-.389	.174	.047	.920
NASQ	.131	.073	**.884**	-.190	.046	-.106	.854
RGYR	.093	**.847**	-.444	-.136	-.078	.006	.948
PTRD	.310	**.787**	.190	-.006	.457	-.032	.962
ABRR	.174	.003	**.660**	**-.522**	-.201	.006	.778
ABSR	.378	**.811**	.238	-.094	.346	.006	.986
ACCR	.170	-.395	**.768**	-.019	.352	.017	.899
ALAR	.412	.613	.450	-.151	.437	.018	.962
OCFC	-.081	**-.660**	-.263	-.337	**.506**	.011	.880
OEFC	-.182	**-.895**	-.098	-.144	.269	.039	.937
						Sum	25.949

Variance explained by each factor after rotation

FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6
9.943	7.839	3.435	2.469	1.230	1.033

Table C-7: Results of principal components analysis and varimax rotation for the Chesapeake Bay Watershed sub-watersheds data set. See Appendix A for explanation of metric abbreviations and Table B-1 for data set description.

Eigenvalues and cumulative proportion of variance explained by principal components analysis						
Eigenvalue	13.144	7.444	2.336	1.522	1.018	.749
Cum. Variance	.469	.735	.819	.873	.909	.936

METRIC	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6	Communality
NTYP	-.002	-.490	.035	-.369	.275	.431	.639
PMAX	**-.927**	-.076	-.146	-.062	.048	.092	.901
SIDI	**.967**	.064	.204	.033	.032	-.037	.983
SIEV	**.964**	.086	.202	.050	.020	-.056	.983
SHDI	**.979**	.006	.143	.039	-.006	.027	.980
SHEV	**.976**	.056	.138	.076	-.034	-.017	.982
SUMD	**-.725**	.484	-.245	-.355	-.126	-.039	.963
SIDA	**.961**	.019	.222	.070	.057	-.033	.982
SICO	**-.960**	-.027	-.221	-.076	-.052	.041	.981
SHDA	**.965**	-.103	.166	.108	.015	.034	.983
SHCO	**-.965**	.053	-.162	-.146	.013	.010	.982
TMAS	**-.646**	-.076	-.490	-.398	-.292	-.070	.912
NPAT	.419	**-.578**	-.144	.213	-.205	.451	.821
LPAT	**-.624**	-.001	.045	-.145	-.235	**.606**	.834
PSIZ	-.334	**.851**	.155	-.172	.142	.018	.909
P005	-.412	**.773**	.219	-.260	.199	.040	.924
PA-1	-.319	**-.719**	**-.522**	-.047	-.292	-.099	.988
DSTA	.311	**.743**	**.522**	.112	.229	.065	.991
PA-2	.158	.304	.101	-.158	**.863**	-.117	.911
NASQ	.221	.031	.485	**.635**	-.397	-.197	.884
RGYR	.140	**.868**	.220	.009	.369	-.112	.969
PTRD	.409	.377	**.809**	-.134	.037	.028	.984
ABRR	.164	-.029	-.018	**.959**	-.069	-.008	.952
ABSR	.447	.452	**.759**	.053	.068	.006	.988
ACCR	.213	**-.656**	.362	.226	-.498	-.219	.953
ALAR	.492	.231	**.805**	.160	-.049	-.056	.974
OCFC	-.111	**-.891**	.005	-.147	.229	.144	.901
OEFC	-.208	**-.915**	-.239	-.125	.031	.029	.956
						Sum	26.212

Variance explained by each factor after rotation					
FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6
11.157	6.548	3.598	2.139	1.835	.936

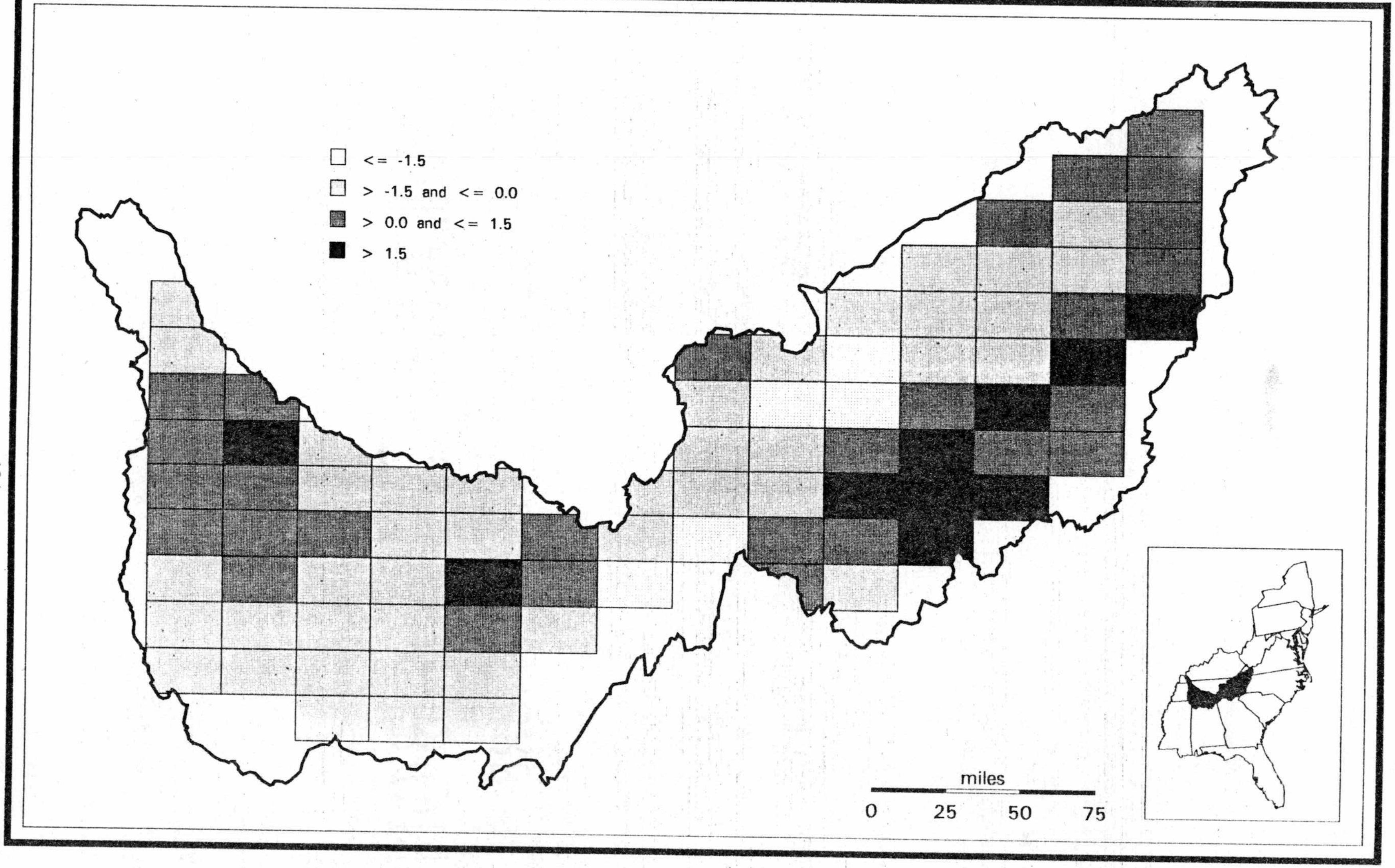


Figure C-1: Factor 1 Scores for the 12-Class, 25-Meter-Resolution Tennessee River Watershed Data Set.

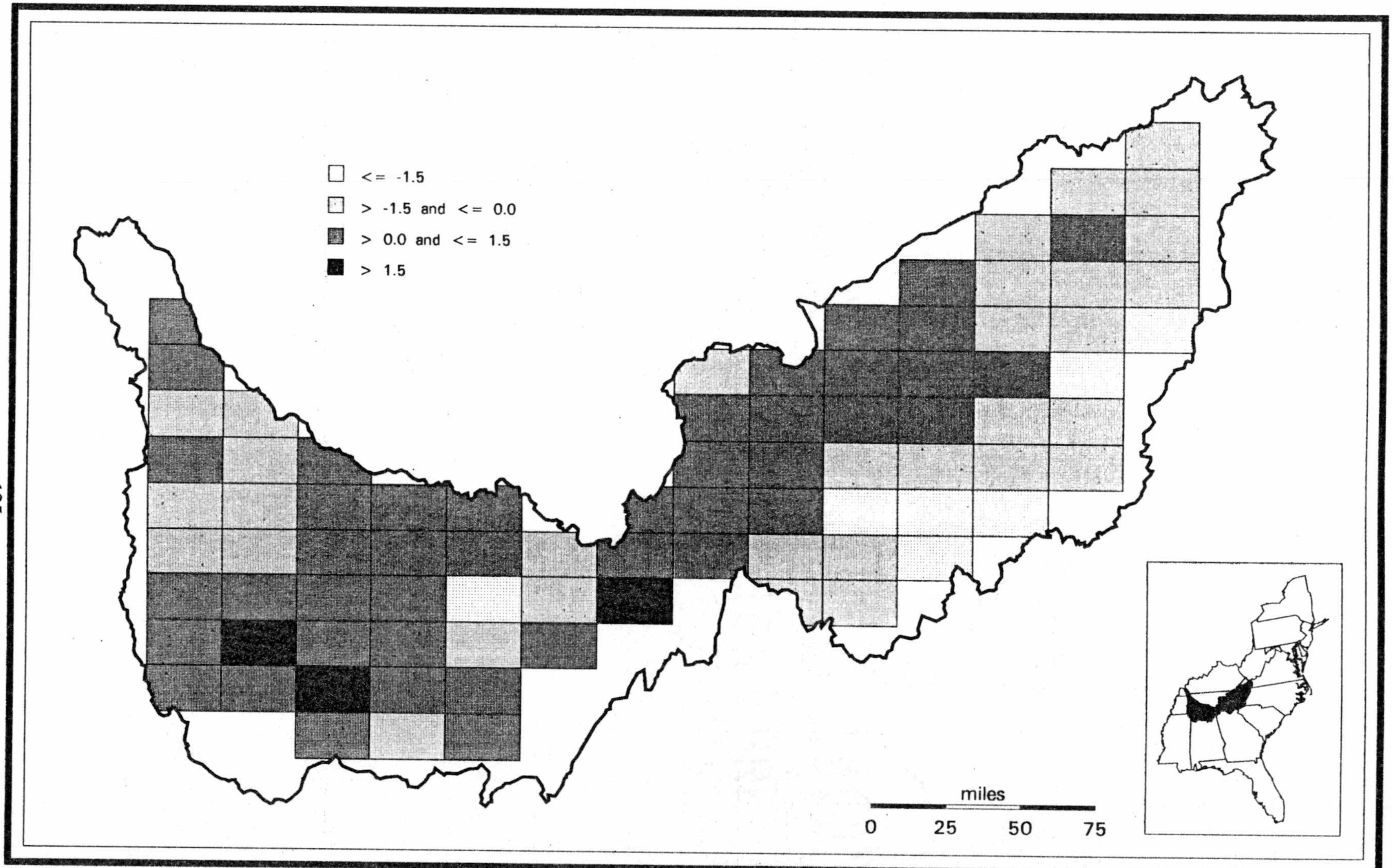


Figure C-2: Factor 1 Scores for the 12-Class, 125-Meter-Resolution, Majority-Rule-Resampled Tennessee River Watershed Data Set.

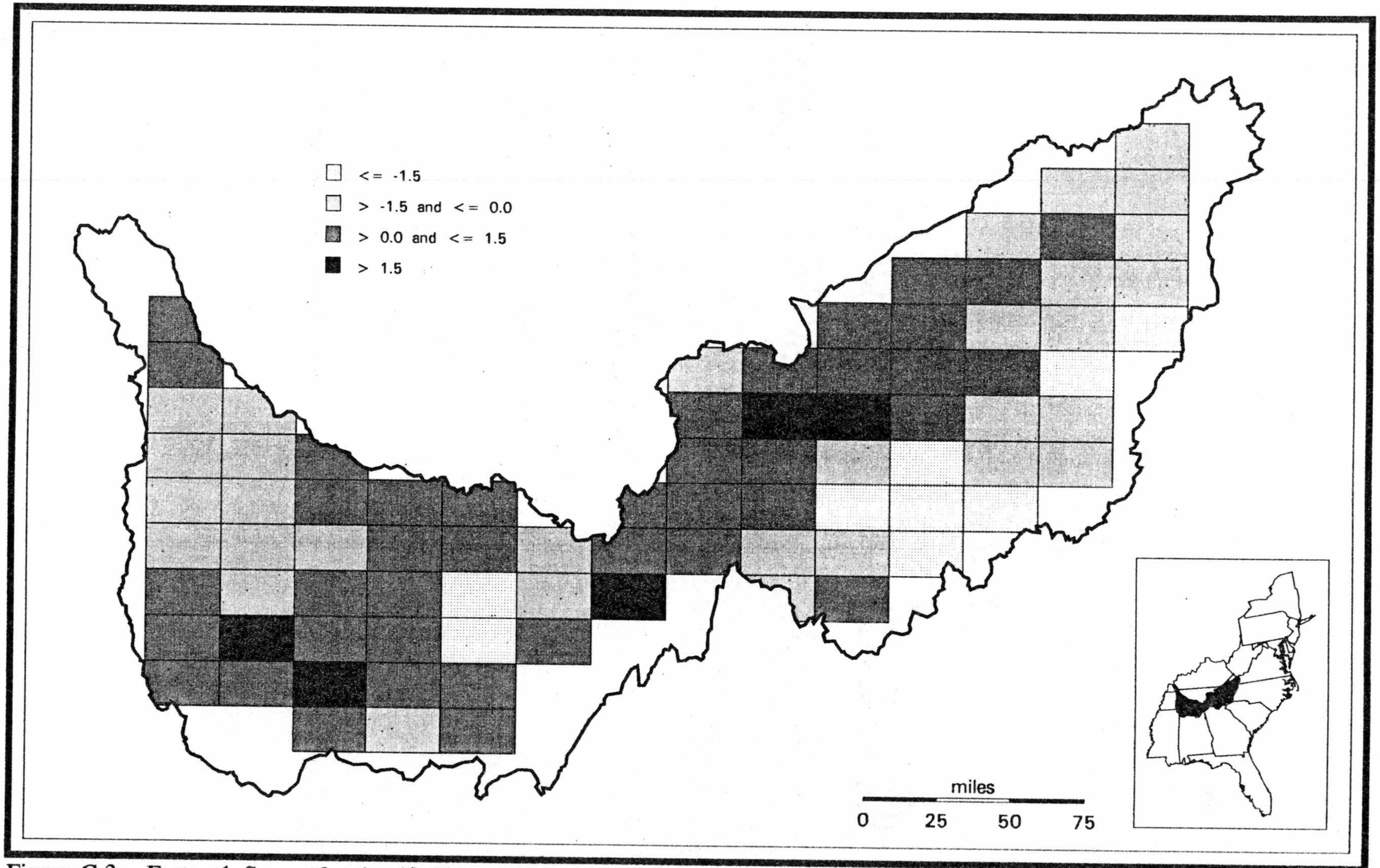


Figure C-3: Factor 1 Scores for the 12-Class, 25-Meter-Resolution, Majority-Rule-Resampled Tennessee River Watershed Data Set.

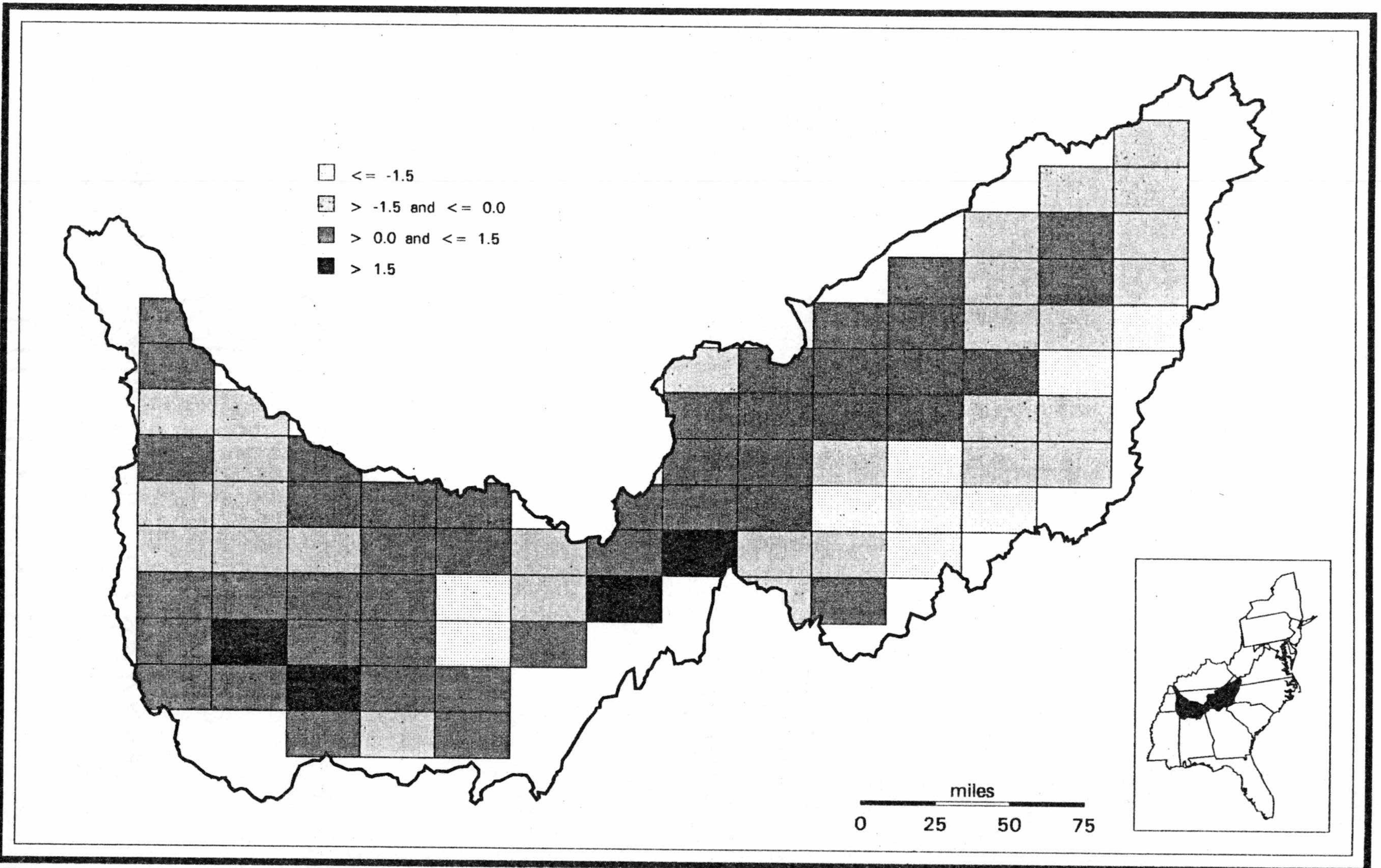


Figure C-4: Factor 1 Scores for the 12-Class, 125-Meter-Resolution, Nearest-Neighbor-Resampled Tennessee River Watershed Data Set.

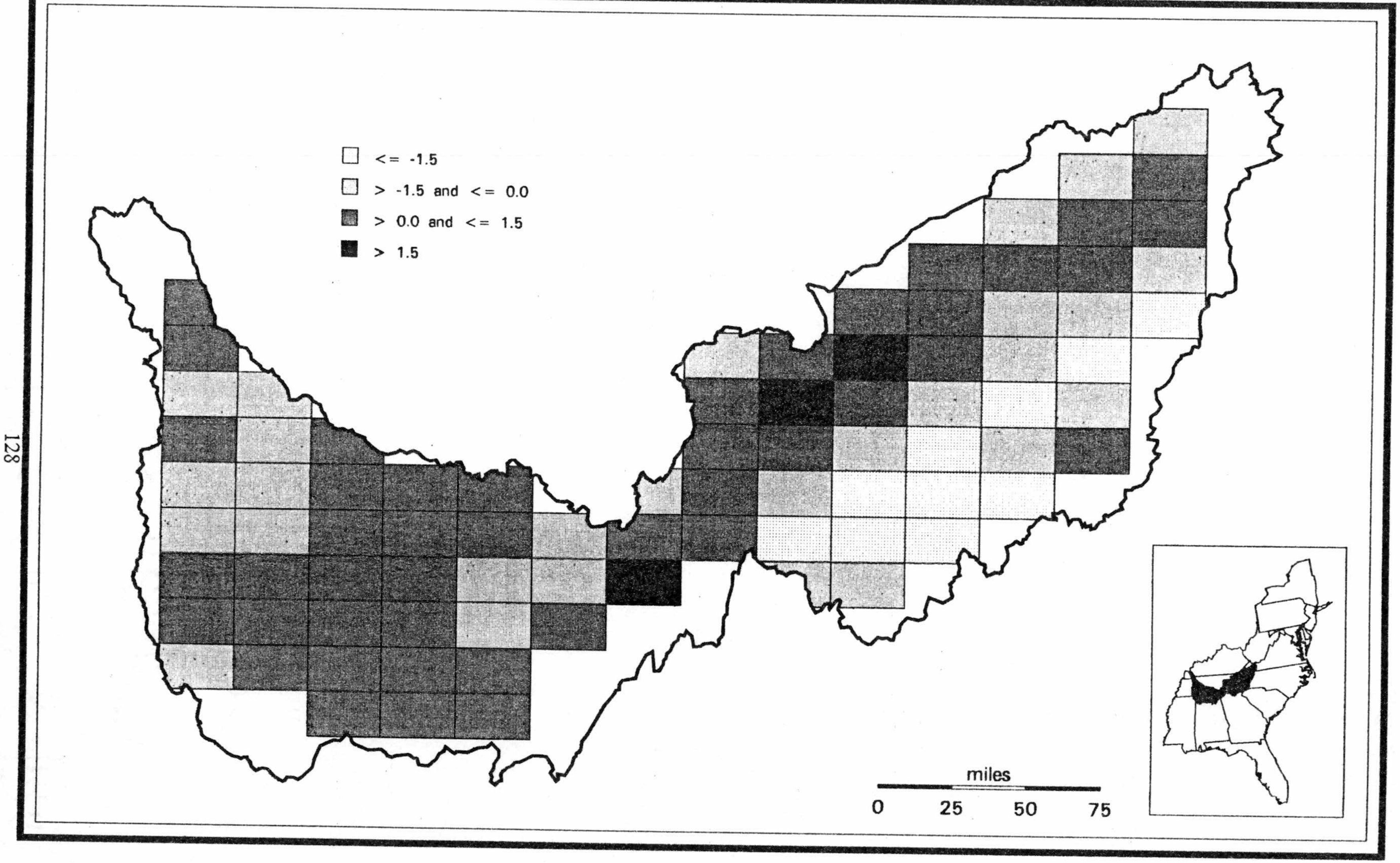


Figure C-5: Factor 1 Scores for the 5-Class, 25-Meter-Resolution Tennessee River Watershed Data Set.

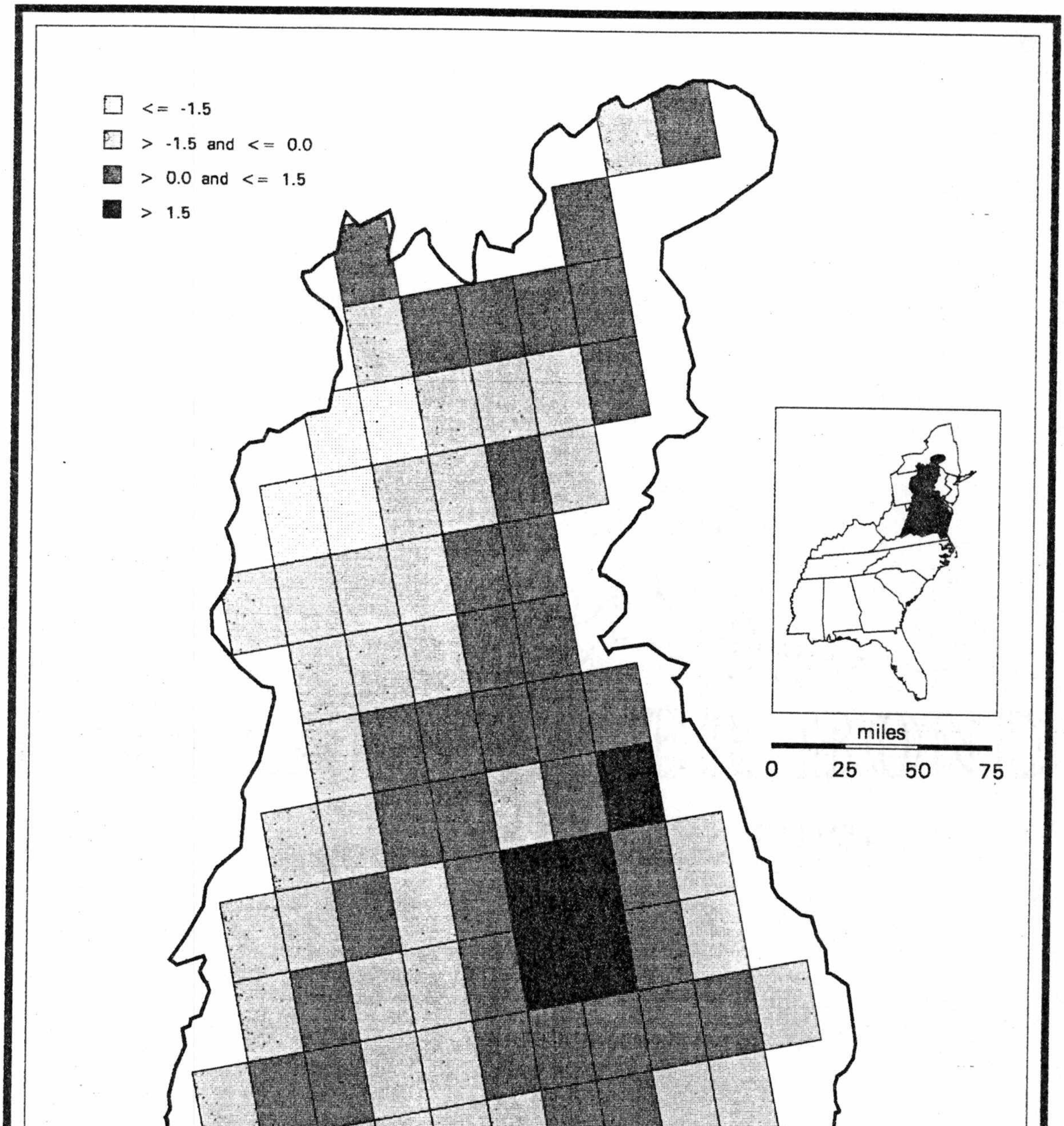


Figure C-6: Factor 1 Scores for the 5-Class, 25-Meter-Resolution, Quadrant-Split Chesapeake Bay Watershed Data Set.

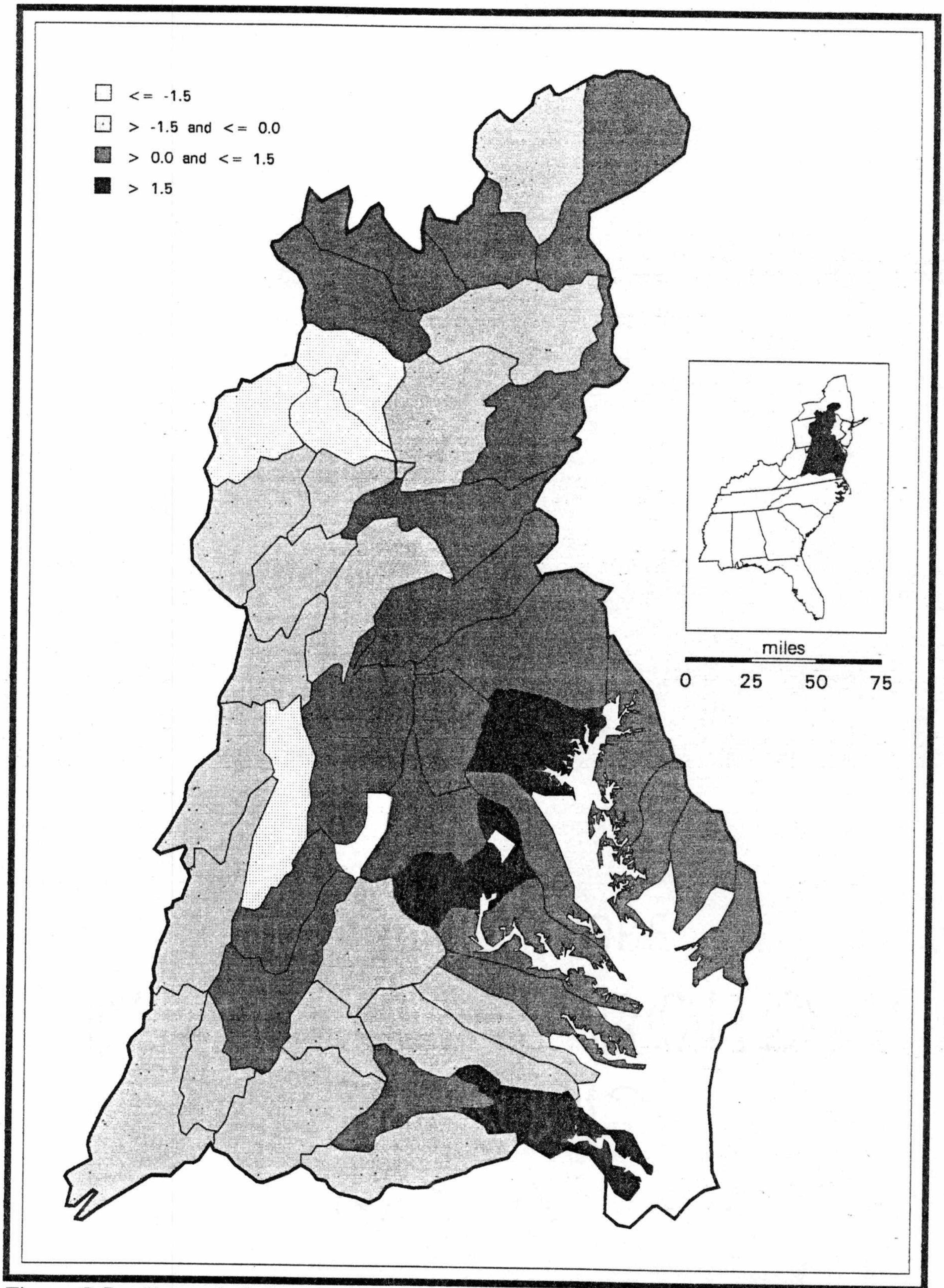


Figure C-7: Factor 1 Scores for the 5-Class, 25-Meter-Resolution, Sub-Watershed-Split Chesapeake Bay Watershed Data Set.

Table C-8: Results of Ward's minimum variance cluster analysis for the Tennessee River Watershed original data set.

Number of Clusters	--Clusters	Joined--	Frequency of New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
84	10000001	10000011	2	0.00018		0.99983	
83	10100100	10110000	2	0.00045	0.00027	0.99938	
82	01101001	01101101	2	0.00047	0.00003	0.99891	
81	00100011	00101001	2	0.00049	0.00002	0.99842	
80	01101111	01111000	2	0.00054	0.00005	0.99787	
79	10100001	10110010	2	0.00057	0.00003	0.99730	
78	10001101	10100011	2	0.00060	0.00002	0.99670	
77	10000101	10010010	2	0.00062	0.00002	0.99608	
76	01101011	01101110	2	0.00064	0.00002	0.99544	
75	10101101	10111001	2	0.00066	0.00002	0.99478	
74	01111010	11000110	2	0.00069	0.00003	0.99409	
73	CL79	CL83	4	0.00071	0.00002	0.99339	
72	01011001	01011010	2	0.00079	0.00008	0.99260	
71	11001001	11010000	2	0.00079	0.00001	0.99181	
70	11001011	11001110	2	0.00081	0.00001	0.99100	
69	10001110	10001111	2	0.00084	0.00003	0.99016	
68	00101011	10000110	2	0.00088	0.00004	0.98929	
67	10011000	10110100	2	0.00095	0.00007	0.98834	
66	01101010	01110011	2	0.00119	0.00024	0.98716	
65	10001011	10001100	2	0.00121	0.00002	0.98595	
64	01111001	01111011	2	0.00123	0.00003	0.98471	
63	10010011	CL73	5	0.00124	0.00000	0.98348	
62	CL77	10000111	3	0.00134	0.00010	0.98214	
61	10011011	10110001	2	0.00140	0.00006	0.98074	
60	01100111	01110010	2	0.00144	0.00004	0.97930	
59	01011100	01011110	2	0.00151	0.00007	0.97779	
58	01100110	11000001	2	0.00157	0.00006	0.97621	
57	CL68	00101110	3	0.00176	0.00019	0.97445	
56	10011010	10100101	2	0.00177	0.00001	0.97268	
55	10011101	11000010	2	0.00182	0.00005	0.97087	
54	CL67	10100110	3	0.00183	0.00001	0.96904	
53	10001001	CL65	3	0.00215	0.00032	0.96688	
52	CL63	10100111	6	0.00222	0.00006	0.96467	
51	CL57	10000100	4	0.00234	0.00013	0.96232	
50	01100101	01110000	2	0.00237	0.00003	0.95995	
49	10011001	10110011	2	0.00238	0.00000	0.95757	
48	01010110	CL72	3	0.00239	0.00001	0.95518	
47	01011011	10010111	2	0.00255	0.00016	0.95264	
46	CL84	11010001	3	0.00273	0.00018	0.94991	
45	CL64	11001100	3	0.00283	0.00010	0.94708	
44	CL55	11001000	3	0.00287	0.00004	0.94421	
43	CL82	CL80	4	0.00295	0.00008	0.94127	
42	CL66	11000000	3	0.00305	0.00011	0.93821	
41	CL47	01110001	3	0.00306	0.00000	0.93516	
40	CL76	01101100	3	0.00324	0.00018	0.93192	
39	01110110	11010010	2	0.00325	0.00001	0.92868	
38	CL74	11000111	3	0.00337	0.00013	0.92530	
37	CL59	01110100	3	0.00347	0.00009	0.92184	
36	10011100	10011110	2	0.00357	0.00011	0.91827	
35	01101000	CL78	3	0.00417	0.00060	0.91410	
34	11000011	CL71	3	0.00429	0.00013	0.90981	

Table C-8 (continued)

Number of Clusters	--Clusters	Joined--	Frequency of New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
33	11000100	11000101	2	0.00446	0.00017	0.90534	
32	CL75	10111000	3	0.00455	0.00008	0.90080	
31	CL81	CL46	5	0.00495	0.00040	0.89585	
30	CL48	CL41	6	0.00551	0.00056	0.89033	
29	CL54	CL49	5	0.00579	0.00028	0.88455	
28	CL37	CL50	5	0.00622	0.00043	0.87833	
27	CL39	11001101	3	0.00667	0.00045	0.87166	
26	CL44	10011111	4	0.00673	0.00007	0.86493	
25	CL35	CL53	6	0.00716	0.00042	0.85778	
24	CL25	CL34	9	0.00737	0.00022	0.85040	
23	CL52	CL56	8	0.00760	0.00023	0.84281	
22	CL62	CL70	5	0.00855	0.00095	0.83426	
21	CL51	CL36	6	0.00869	0.00014	0.82558	
20	CL58	CL40	5	0.01041	0.00173	0.81516	
19	CL27	CL38	6	0.01110	0.00069	0.80407	
18	CL31	CL21	11	0.01195	0.00085	0.79212	
17	CL24	CL69	11	0.01203	0.00008	0.78009	
16	CL28	CL60	7	0.01226	0.00024	0.76783	
.15	CL29	CL32	8	0.01270	0.00043	0.75514	
14	CL45	CL33	5	0.01270	0.00001	0.74243	
13	CL17	CL42	14	0.01336	0.00065	0.72908	
12	CL19	CL14	11	0.02346	0.01011	0.70562	
11	CL22	CL15	13	0.02406	0.00060	0.68156	
10	CL20	CL43	9	0.02463	0.00057	0.65692	
9	CL23	CL61	10	0.02988	0.00525	0.62704	
8	CL30	CL13	20	0.03499	0.00511	0.59206	yes
7	CL8	CL26	24	0.05070	**0.01571**	0.54136	
6	CL18	CL9	21	0.06533	0.01463	0.47603	
5	CL16	CL10	16	0.07356	0.00823	0.40248	
4	CL6	CL11	34	0.08405	0.01050	0.31843	
3	CL7	CL12	35	0.09141	0.00736	0.22702	
2	CL4	CL3	69	0.11051	0.01910	0.11651	
1	CL2	CL5	85	0.11651	0.00600	0.00000	

Note: Clustering was stopped at 8 clusters because of a large increase in the semi-partial r-squared value at Cluster 7 surrounded by smaller values. The larger increase at Cluster 2 was not used because more than 3 clusters were desired.

Table C-9: Results of Ward's minimum variance cluster analysis for the Tennessee River Watershed low-resolution majority-rule data set.

Number of Clusters	--Clusters	Joined--	Frequency of New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
84	10000001	10000011	2	0.00034		0.99966	
83	10000111	10011000	2	0.00041	0.00008	0.99925	
82	10000110	10001100	2	0.00047	0.00006	0.99878	
81	10011010	10101101	2	0.00048	0.00000	0.99830	
80	01011001	01011010	2	0.00049	0.00001	0.99781	
79	10100100	10100101	2	0.00055	0.00006	0.99726	
78	01101111	01111000	2	0.00063	0.00008	0.99663	
77	10100001	10100111	2	0.00071	0.00008	0.99593	
76	10011001	10111001	2	0.00074	0.00004	0.99519	
75	01111001	01111010	2	0.00083	0.00008	0.99436	
74	01101100	01101110	2	0.00083	0.00000	0.99353	
73	10001110	10100011	2	0.00086	0.00003	0.99268	
72	10010011	10110000	2	0.00086	0.00001	0.99182	
71	01011110	10011110	2	0.00087	0.00001	0.99094	
70	01011011	01100101	2	0.00091	0.00004	0.99003	
69	10100110	10110100	2	0.00099	0.00008	0.98904	
68	00100011	CL82	3	0.00100	0.00000	0.98804	
67	CL83	10001111	3	0.00122	0.00022	0.98682	
66	CL80	10001001	3	0.00129	0.00007	0.98553	
65	11000111	11001101	2	0.00130	0.00001	0.98424	
64	CL79	10110010	3	0.00131	0.00001	0.98293	
63	CL84	10001011	3	0.00133	0.00002	0.98159	
62	01110000	01110001	2	0.00135	0.00001	0.98025	
61	01100110	01101011	2	0.00135	0.00000	0.97890	
60	01100111	10010111	2	0.00135	0.00000	0.97755	
59	10010010	11000011	2	0.00156	0.00021	0.97599	
58	CL76	10111000	3	0.00159	0.00003	0.97440	
57	11001001	11010001	2	0.00170	0.00011	0.97269	
56	CL68	CL73	5	0.00172	0.00001	0.97098	
55	CL69	11000010	3	0.00172	0.00001	0.96926	
54	01111011	10001101	2	0.00174	0.00002	0.96752	
53	01101010	11000000	2	0.00184	0.00010	0.96568	
52	00101001	00101011	2	0.00187	0.00003	0.96382	
51	10000101	11001011	2	0.00188	0.00002	0.96193	
50	10011111	11001000	2	0.00194	0.00005	0.95999	
49	CL72	CL81	4	0.00239	0.00046	0.95760	
48	01101101	CL78	3	0.00242	0.00002	0.95519	
47	10011011	10110001	2	0.00248	0.00006	0.95271	
46	01101000	01110010	2	0.00257	0.00009	0.95013	
45	01011100	CL62	3	0.00259	0.00002	0.94754	
44	CL75	11001100	3	0.00307	0.00048	0.94447	
43	10000100	11000101	2	0.00312	0.00005	0.94135	
42	11000110	CL65	3	0.00324	0.00012	0.93810	
41	CL60	10110011	3	0.00331	0.00006	0.93480	
40	CL48	01110011	4	0.00331	0.00000	0.93149	
39	CL61	01101001	3	0.00332	0.00001	0.92817	
38	CL57	11001110	3	0.00345	0.00013	0.92472	
37	01010110	CL45	4	0.00348	0.00003	0.92124	
36	10011100	11010010	2	0.00356	0.00008	0.91768	
35	CL67	CL55	6	0.00367	0.00011	0.91401	
34	CL59	10011101	3	0.00384	0.00018	0.91017	

Table C-9 (continued)

Number of Clusters	--Clusters	Joined--	Frequency of New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
33	CL66	CL70	5	0.00413	0.00028	0.90604	
32	CL52	CL63	5	0.00444	0.00031	0.90160	
31	CL46	CL51	4	0.00453	0.00009	0.89707	
30	CL34	CL38	6	0.00467	0.00014	0.89241	
29	CL77	CL64	5	0.00476	0.00009	0.88765	
28	CL40	11000100	5	0.00482	0.00007	0.88283	
27	CL54	11010000	3	0.00490	0.00007	0.87793	
26	CL71	CL36	4	0.00558	0.00068	0.87236	
25	CL56	CL49	9	0.00574	0.00017	0.86661	
24	CL35	CL58	9	0.00611	0.00037	0.86050	
23	CL26	CL47	6	0.00629	0.00018	0.85421	
22	CL74	11000001	3	0.00719	0.00089	0.84702	
21	CL44	CL43	5	0.00752	0.00034	0.83950	
20	CL25	CL41	12	0.00787	0.00035	0.83163	
19	CL39	CL53	5	0.00821	0.00034	0.82343	
18	01110100	01110110	2	0.00893	0.00073	0.81450	
17	CL37	CL33	9	0.00952	0.00059	0.80498	
16	CL32	00101110	6	0.01225	0.00273	0.79273	
15	CL17	CL27	12	0.01357	0.00132	0.77916	
14	CL24	CL30	15	0.01369	0.00013	0.76547	
13	CL28	CL21	10	0.01579	0.00209	0.74968	
12	CL31	CL22	7	0.01864	0.00286	0.73104	
11	CL20	CL29	17	0.02072	0.00208	0.71032	
10	CL14	CL50	17	0.02401	0.00329	0.68631	
9	CL13	CL42	13	0.02575	0.00174	0.66056	
8	CL11	CL16	23	0.04201	0.01627	0.61855	
7	CL19	CL12	12	0.04232	0.00030	0.57623	
6	CL15	CL23	18	0.05385	0.01154	0.52238	yes
5	CL7	CL9	25	0.07174	**0.01788**	0.45064	
4	CL8	CL10	40	0.07552	0.00378	0.37512	
3	CL4	CL6	58	0.09892	0.02340	0.27621	
2	CL3	CL5	83	0.12274	0.02383	0.15347	
1	CL2	CL18	85	0.15347	0.03073	0.00000	

Note: Clustering was stopped at 6 clusters because of a large increase in the semi-partial r-squared value at Cluster 5 surrounded by smaller increases.

Table C-10: Results of Ward's minimum variance cluster analysis for the Tennessee River Watershed high-resolution majority-rule data set.

Number of Clusters	--Clusters	Joined--	Frequency of New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
84	10100100	10100111	2	0.00045		0.99956	
83	00101011	10000011	2	0.00049	0.00004	0.99907	
82	01101111	01111000	2	0.00052	0.00004	0.99855	
81	10011001	10110000	2	0.00065	0.00013	0.99790	
80	00101110	01011010	2	0.00085	0.00020	0.99705	
79	10010011	10110010	2	0.00090	0.00005	0.99615	
78	01110001	10001011	2	0.00092	0.00002	0.99523	
77	10000111	11010001	2	0.00099	0.00007	0.99424	
76	10011000	11000011	2	0.00099	0.00001	0.99325	
75	10001100	10100011	2	0.00100	0.00001	0.99224	
74	11001001	11001011	2	0.00118	0.00018	0.99106	
73	10011111	11001000	2	0.00148	0.00030	0.98958	
72	01011001	10001001	2	0.00153	0.00004	0.98805	
71	00101001	10010111	2	0.00156	0.00003	0.98650	
70	CL78	10000001	3	0.00156	0.00001	0.98493	
69	00100011	10110100	2	0.00170	0.00014	0.98323	
68	01111010	11000110	2	0.00174	0.00004	0.98149	
67	10000110	10111000	2	0.00179	0.00005	0.97970	
66	CL75	10101101	3	0.00184	0.00004	0.97786	
65	10100001	10111001	2	0.00186	0.00002	0.97600	
64	01101011	01101100	2	0.00207	0.00021	0.97394	
63	CL79	10100110	3	0.00209	0.00003	0.97184	
62	11000100	11001110	2	0.00212	0.00003	0.96972	
61	01010110	01011011	2	0.00220	0.00008	0.96752	
60	CL83	CL77	4	0.00221	0.00001	0.96531	
59	10001110	10001111	2	0.00224	0.00003	0.96307	
58	CL82	01110011	3	0.00230	0.00006	0.96077	
57	01100111	01110010	2	0.00233	0.00003	0.95844	
56	11000111	11001101	2	0.00235	0.00002	0.95609	
55	01011110	10011110	2	0.00239	0.00004	0.95370	
54	10010010	CL76	3	0.00244	0.00005	0.95126	
53	01110110	10000100	2	0.00252	0.00008	0.94874	
52	01111011	10001101	2	0.00256	0.00004	0.94618	
51	10000101	CL74	3	0.00269	0.00013	0.94350	
50	01011100	01100101	2	0.00276	0.00007	0.94074	
49	01100110	01101010	2	0.00277	0.00002	0.93797	
48	CL59	CL81	4	0.00280	0.00003	0.93516	
47	CL69	11000010	3	0.00330	0.00050	0.93187	
46	01101101	CL65	3	0.00338	0.00008	0.92849	
45	CL52	11010000	3	0.00363	0.00026	0.92486	
44	01111001	11001100	2	0.00382	0.00019	0.92105	
43	CL60	CL70	7	0.00384	0.00002	0.91721	
42	01101000	10011100	2	0.00395	0.00011	0.91326	
41	CL50	01110000	3	0.00399	0.00004	0.90928	
40	CL63	CL84	5	0.00445	0.00046	0.90483	
39	CL68	11010010	3	0.00456	0.00011	0.90027	
38	CL49	11000000	3	0.00491	0.00035	0.89537	
37	CL67	CL54	5	0.00515	0.00024	0.89021	
36	CL80	CL61	4	0.00533	0.00018	0.88489	
35	CL71	10011101	3	0.00540	0.00007	0.87949	
34	CL55	10110011	3	0.00547	0.00007	0.87402	

Table C-10 (continued)

Number of Clusters	--Clusters	Joined--	Frequency of New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
33	CL38	01101110	4	0.00596	0.00050	0.86806	
32	CL48	10011010	5	0.00629	0.00033	0.86177	
31	CL42	CL56	4	0.00661	0.00032	0.85516	
30	CL37	CL66	8	0.00676	0.00015	0.84840	
29	CL45	11000101	4	0.00689	0.00013	0.84151	
28	CL47	CL43	10	0.00710	0.00021	0.83441	
27	10011011	10110001	2	0.00774	0.00064	0.82667	
26	CL32	10100101	6	0.00780	0.00006	0.81887	
25	CL35	CL72	5	0.00803	0.00023	0.81084	
24	01101001	CL46	4	0.00807	0.00004	0.80277	
23	CL41	01110100	4	0.00847	0.00040	0.79430	
22	CL64	11000001	3	0.00956	0.00109	0.78474	
21	CL53	CL39	5	0.01074	0.00118	0.77400	
20	CL58	CL62	5	0.01163	0.00090	0.76237	
19	CL28	CL30	18	0.01212	0.00048	0.75025	
18	CL57	CL51	5	0.01294	0.00082	0.73730	
17	CL31	CL21	9	0.01409	0.00115	0.72321	
16	CL23	CL34	7	0.01443	0.00034	0.70879	
15	CL26	CL40	11	0.01473	0.00030	0.69406	
14	CL33	CL20	9	0.01788	0.00315	0.67619	
13	CL25	CL24	9	0.01915	0.00128	0.65703	
12	CL36	CL29	8	0.02350	0.00435	0.63353	
11	CL19	CL73	20	0.02414	0.00064	0.60939	
10	CL14	CL22	12	0.02429	0.00015	0.58510	
9	CL17	CL44	11	0.02785	0.00357	0.55725	
8	CL15	CL27	13	0.04139	0.01353	0.51586	
7	CL11	CL18	25	0.04158	0.00020	0.47428	
6	CL13	CL16	16	0.04954	0.00796	0.42474	
5	CL12	CL9	19	0.05859	0.00905	0.36614	yes
4	CL7	CL8	38	0.07783	**0.01924**	0.28831	
3	CL5	CL10	31	0.08777	0.00994	0.20054	
2	CL4	CL6	54	0.09196	0.00419	0.10858	
1	CL2	CL3	85	0.10858	0.01661	0.00000	

Note: Clustering was stopped at 5 clusters because the largest increase in the semi-partial r-squared value at Cluster 4.

Table C-11: Results of Ward's minimum variance cluster analysis for the Tennessee River Watershed low-resolution nearest-neighbor data set.

Number of Clusters	--Clusters	Joined--	Frequency of New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
84	10000011	10100011	2	0.00036		0.99964	
83	10100001	10100111	2	0.00041	0.00006	0.99923	
82	10000101	11000011	2	0.00062	0.00021	0.99861	
81	10011001	10111001	2	0.00068	0.00005	0.99793	
80	CL84	10000110	3	0.00073	0.00006	0.99720	
79	01111001	11000111	2	0.00077	0.00003	0.99643	
78	10100101	10110000	2	0.00079	0.00003	0.99564	
77	01011011	01100101	2	0.00086	0.00007	0.99478	
76	00101110	10000100	2	0.00089	0.00003	0.99390	
75	01101100	01101110	2	0.00091	0.00003	0.99298	
74	01101111	01110011	2	0.00093	0.00001	0.99206	
73	10010011	10101101	2	0.00096	0.00003	0.99110	
72	10100100	10110010	2	0.00097	0.00002	0.99013	
71	00100011	10001110	2	0.00100	0.00002	0.98913	
70	00101001	10001011	2	0.00112	0.00013	0.98801	
69	11000010	11001000	2	0.00122	0.00009	0.98679	
68	10001001	10001111	2	0.00124	0.00003	0.98555	
67	01100111	01110010	2	0.00127	0.00003	0.98428	
66	10000111	11010001	2	0.00127	0.00000	0.98301	
65	10011010	10100110	2	0.00135	0.00008	0.98166	
64	10011000	10110100	2	0.00137	0.00001	0.98029	
63	01101101	CL82	3	0.00138	0.00001	0.97891	
62	01110000	10011110	2	0.00146	0.00008	0.97745	
61	00101011	CL76	3	0.00148	0.00002	0.97598	
60	01100110	01101011	2	0.00154	0.00007	0.97444	
59	01010110	01011001	2	0.00156	0.00002	0.97288	
58	CL80	10001100	4	0.00159	0.00003	0.97129	
57	01111011	11010000	2	0.00165	0.00006	0.96964	
56	11001011	11001110	2	0.00168	0.00004	0.96796	
55	CL72	CL78	4	0.00177	0.00009	0.96619	
54	01011110	01101010	2	0.00188	0.00010	0.96431	
53	10110011	10111000	2	0.00202	0.00015	0.96229	
52	CL71	CL64	4	0.00222	0.00020	0.96007	
51	CL79	11000110	3	0.00225	0.00002	0.95782	
50	11000100	CL56	3	0.00226	0.00001	0.95556	
49	CL77	01110001	3	0.00227	0.00001	0.95329	
48	01111010	11001100	2	0.00231	0.00004	0.95099	
47	10010010	11001001	2	0.00238	0.00007	0.94861	
46	CL66	10001101	3	0.00238	0.00000	0.94623	
45	10011011	10110001	2	0.00265	0.00027	0.94359	
44	10000001	CL68	3	0.00276	0.00012	0.94082	
43	01111000	CL50	4	0.00279	0.00003	0.93803	
42	01011100	CL62	3	0.00289	0.00010	0.93514	
41	CL73	CL81	4	0.00294	0.00005	0.93219	
40	CL70	10010111	3	0.00313	0.00019	0.92906	
39	CL59	01011010	3	0.00316	0.00003	0.92590	
38	01101000	11000000	2	0.00316	0.00000	0.92274	
37	CL41	CL65	6	0.00324	0.00008	0.91950	
36	CL74	CL48	4	0.00391	0.00067	0.91559	
35	10011101	CL69	3	0.00416	0.00025	0.91143	
34	CL35	10011111	4	0.00428	0.00012	0.90715	

Table C-11 (continued)

Number of Clusters	--Clusters	Joined--	Frequency of New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
33	CL60	01101001	3	0.00464	0.00037	0.90251	
32	CL52	CL44	7	0.00505	0.00041	0.89746	
31	CL40	CL58	7	0.00515	0.00010	0.89231	
30	CL42	CL53	5	0.00523	0.00008	0.88708	
29	CL51	11001101	4	0.00600	0.00077	0.88108	
28	CL38	11000001	3	0.00603	0.00003	0.87506	
27	CL63	CL47	5	0.00608	0.00005	0.86898	
26	CL67	10011100	3	0.00618	0.00010	0.86280	
25	CL57	CL46	5	0.00732	0.00114	0.85548	
24	CL39	11010010	4	0.00748	0.00016	0.84800	
23	CL36	CL43	8	0.00764	0.00016	0.84036	
22	CL32	CL49	10	0.00779	0.00016	0.83257	
21	CL37	CL55	10	0.00818	0.00039	0.82439	
20	CL31	CL83	9	0.00871	0.00053	0.81569	
19	CL33	CL75	5	0.00884	0.00013	0.80685	
18	CL26	CL28	6	0.01040	0.00156	0.79645	
17	01110100	01110110	2	0.01206	0.00166	0.78439	
16	CL24	11000101	5	0.01300	0.00094	0.77140	
15	CL30	CL54	7	0.01504	0.00204	0.75636	
14	CL27	CL23	13	0.01659	0.00155	0.73977	
13	CL20	CL61	12	0.01929	0.00270	0.72048	
12	CL22	CL25	15	0.02212	0.00283	0.69836	
11	CL15	CL45	9	0.02319	0.00108	0.67517	
10	CL12	CL34	19	0.02942	0.00622	0.64576	
9	CL14	CL29	17	0.03066	0.00124	0.61510	
8	CL11	CL18	15	0.03240	0.00174	0.58270	
7	CL13	CL16	17	0.04416	0.01176	0.53855	
6	CL19	CL9	22	0.05690	0.01274	0.48164	
5	CL8	CL21	25	0.05822	0.00132	0.42342	yes
4	CL10	CL7	36	0.08299	**0.02477**	0.34043	
3	CL4	CL6	58	0.10020	0.01721	0.24023	
2	CL3	CL5	83	0.10821	0.00801	0.13202	
1	CL2	CL17	85	0.13202	0.02381	0.00000	

Note: Clustering was stopped at 5 clusters because of the largest increase in the semi-partial r-squared value at Cluster 4.

Table C-12: Results of Ward's minimum variance cluster analysis for the Tennessee River Watershed 5-class data set.

Number of Clusters	--Clusters	Joined--	Frequency New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
84	10010011	10011001	2	0.00037		0.99963	
83	10100110	10101101	2	0.00052	0.00015	0.99911	
82	10001011	10100011	2	0.00060	0.00009	0.99851	
81	00101001	10110010	2	0.00062	0.00001	0.99789	
80	00100011	10100001	2	0.00073	0.00011	0.99717	
79	01010110	11001110	2	0.00077	0.00005	0.99640	
78	10000011	CL82	3	0.00087	0.00010	0.99553	
77	01011001	01011010	2	0.00090	0.00003	0.99463	
76	10110100	11000000	2	0.00092	0.00002	0.99372	
75	01110011	01110100	2	0.00092	0.00000	0.99280	
74	10000111	10011000	2	0.00097	0.00006	0.99183	
73	10100100	10100111	2	0.00127	0.00030	0.99056	
72	01101011	01101110	2	0.00135	0.00008	0.98921	
71	CL75	11000011	3	0.00137	0.00003	0.98784	
70	01011100	11000010	2	0.00137	0.00000	0.98647	
69	01100111	01110010	2	0.00138	0.00000	0.98509	
68	11001011	11001101	2	0.00139	0.00002	0.98370	
67	10001101	10001111	2	0.00140	0.00001	0.98230	
66	01101010	10010111	2	0.00141	0.00001	0.98089	
65	10000001	10001001	2	0.00143	0.00002	0.97946	
64	CL78	10000110	4	0.00163	0.00020	0.97783	
63	10011101	11010001	2	0.00166	0.00003	0.97617	
62	01101000	10010010	2	0.00167	0.00001	0.97450	
61	CL76	10111001	3	0.00173	0.00006	0.97276	
60	CL73	10110000	3	0.00175	0.00001	0.97101	
59	11000111	11010010	2	0.00184	0.00009	0.96918	
58	CL79	10000101	3	0.00187	0.00004	0.96730	
57	CL80	CL81	4	0.00205	0.00018	0.96525	
56	01111001	01111010	2	0.00213	0.00009	0.96312	
55	CL66	CL61	5	0.00219	0.00006	0.96093	
54	01100110	01101101	2	0.00232	0.00013	0.95861	
53	10001100	10011010	2	0.00237	0.00005	0.95624	
52	01100101	CL67	3	0.00243	0.00006	0.95381	
51	CL71	11000100	4	0.00243	0.00000	0.95138	
50	00101011	10011110	2	0.00244	0.00001	0.94894	
49	CL65	CL64	6	0.00262	0.00018	0.94633	
48	CL74	10011100	3	0.00287	0.00025	0.94346	
47	01101111	CL59	3	0.00301	0.00015	0.94045	
46	01011011	01101001	2	0.00308	0.00007	0.93737	
45	00101110	10000100	2	0.00330	0.00022	0.93407	
44	01110001	11010000	2	0.00335	0.00005	0.93072	
43	CL77	CL52	5	0.00360	0.00025	0.92712	
42	CL70	CL44	4	0.00363	0.00003	0.92349	
41	01011110	CL54	3	0.00375	0.00012	0.91974	
40	01110110	CL68	3	0.00376	0.00002	0.91598	
39	10011011	10110011	2	0.00382	0.00006	0.91216	
38	11000110	11001100	2	0.00423	0.00041	0.90793	
37	CL55	10111000	6	0.00439	0.00016	0.90354	
36	CL57	CL50	6	0.00453	0.00014	0.89900	
35	CL58	CL42	7	0.00456	0.00003	0.89444	
34	CL62	CL83	4	0.00482	0.00026	0.88963	

Table C-12 (continued)

Number of Clusters	--Clusters	Joined--	Frequency New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
33	CL69	11000001	3	0.00503	0.00022	0.88459	
32	CL53	10001110	3	0.00507	0.00003	0.87953	
31	CL63	10011111	3	0.00517	0.00011	0.87436	
30	01111000	10100101	2	0.00550	0.00033	0.86886	
29	CL34	CL84	6	0.00637	0.00087	0.86249	
28	CL31	11001000	4	0.00652	0.00016	0.85597	
27	CL41	CL72	5	0.00665	0.00013	0.84932	
26	01110000	CL30	3	0.00712	0.00047	0.84220	
25	CL56	01111011	3	0.00770	0.00059	0.83450	
24	CL37	CL48	9	0.00849	0.00078	0.82602	
23	CL40	11001001	4	0.00852	0.00004	0.81750	
22	CL45	CL49	8	0.00959	0.00107	0.80790	
21	CL39	10110001	3	0.00966	0.00007	0.79824	
20	CL33	CL51	7	0.01162	0.00196	0.78662	
19	CL46	CL28	6	0.01219	0.00057	0.77443	
18	CL47	CL38	5	0.01274	0.00055	0.76169	
17	CL36	CL60	9	0.01463	0.00189	0.74706	
16	CL27	01101100	6	0.01830	0.00367	0.72876	
15	CL43	CL25	8	0.01836	0.00006	0.71040	
14	CL24	CL21	12	0.01979	0.00142	0.69062	
13	CL35	CL29	13	0.02182	0.00203	0.66880	
12	CL22	CL32	11	0.02362	0.00180	0.64518	
11	CL18	CL23	9	0.02413	0.00052	0.62105	
10	CL20	CL26	10	0.02575	0.00162	0.59530	
9	CL17	CL12	20	0.03127	0.00551	0.56403	
8	CL16	CL10	16	0.03214	0.00087	0.53189	
7	CL13	CL14	25	0.03590	0.00376	0.49599	
6	CL11	11000101	10	0.05153	0.01564	0.44446	
5	CL7	CL15	33	0.05881	0.00728	0.38565	
4	CL5	CL19	39	0.06959	0.01078	0.31606	yes
3	CL4	CL6	49	0.09688	**0.02729**	0.21918	
2	CL9	CL3	69	0.10102	0.00414	0.11816	
1	CL2	CL8	85	0.11816	0.01714	0.00000	

Note: Clustering was stopped at 4 clusters because of the largest increase in the semi-partial r-squared value at Cluster 3.

Table C-13: Results of Ward's minimum variance cluster analysis for the Chesapeake Bay Watershed quadrants data set.

Number of Clusters	--Clusters	Joined--	Frequency of New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
96	01011000	01011010	2	0.00013		0.99987	
95	10010101	11000010	2	0.00014	0.00001	0.99974	
94	11000110	11001001	2	0.00021	0.00007	0.99953	
93	11000011	11001000	2	0.00021	0.00000	0.99932	
92	10110001	10110010	2	0.00028	0.00008	0.99904	
91	01100101	10110011	2	0.00035	0.00007	0.99868	
90	01101010	11001010	2	0.00036	0.00000	0.99833	
89	10011000	10100100	2	0.00038	0.00003	0.99794	
88	10011001	10110000	2	0.00047	0.00009	0.99747	
87	01100011	01100110	2	0.00047	0.00000	0.99700	
86	01001111	10011110	2	0.00048	0.00001	0.99652	
85	10011101	10110110	2	0.00050	0.00002	0.99601	
84	01101011	01101110	2	0.00051	0.00000	0.99551	
83	10011100	10100110	2	0.00053	0.00002	0.99498	
82	01001100	11000001	2	0.00053	0.00000	0.99445	
81	01001101	10110111	2	0.00056	0.00003	0.99389	
80	01001000	01100001	2	0.00058	0.00002	0.99332	
79	00111101	01100000	2	0.00061	0.00003	0.99271	
78	10010001	10010100	2	0.00062	0.00001	0.99209	
77	01010001	01010100	2	0.00066	0.00004	0.99142	
76	11100000	11100011	2	0.00067	0.00001	0.99075	
75	10010010	10010110	2	0.00072	0.00004	0.99003	
74	01010010	CL85	3	0.00072	0.00001	0.98931	
73	10011010	11001111	2	0.00074	0.00002	0.98857	
72	01001011	01001110	2	0.00076	0.00002	0.98780	
71	CL92	10110100	3	0.00086	0.00010	0.98695	
70	10011111	CL76	3	0.00087	0.00001	0.98608	
69	01001001	CL95	3	0.00097	0.00011	0.98511	
68	CL84	01101111	3	0.00099	0.00002	0.98412	
67	11000101	11001100	2	0.00099	0.00000	0.98313	
66	01100010	CL88	3	0.00101	0.00002	0.98213	
65	CL93	11100010	3	0.00113	0.00012	0.98100	
64	10100101	11110000	2	0.00113	0.00001	0.97987	
63	CL81	CL96	4	0.00117	0.00003	0.97870	
62	CL90	11000000	3	0.00129	0.00012	0.97741	
61	CL79	01101000	3	0.00145	0.00016	0.97596	
60	00111100	00111111	2	0.00145	0.00000	0.97450	
59	10010111	11000100	2	0.00148	0.00002	0.97303	
58	CL66	10011011	4	0.00158	0.00011	0.97145	
57	11010010	11011000	2	0.00162	0.00004	0.96982	
56	01000010	CL80	3	0.00163	0.00000	0.96820	
55	00011111	00110101	2	0.00163	0.00001	0.96657	
54	00110110	00110111	2	0.00176	0.00013	0.96480	
53	11001101	11100101	2	0.00179	0.00002	0.96302	
52	CL87	01100100	3	0.00186	0.00008	0.96116	
51	CL53	11100111	3	0.00196	0.00010	0.95920	
50	CL74	CL59	5	0.00200	0.00004	0.95721	
49	01001010	CL72	3	0.00209	0.00009	0.95512	
48	11001110	11100100	2	0.00226	0.00017	0.95287	
47	11000111	11101100	2	0.00232	0.00006	0.95055	
46	CL62	CL75	5	0.00248	0.00016	0.94807	

Table C-13 (continued)

Number of Clusters	--Clusters	Joined--	Frequency of New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
45	CL51	11100110	4	0.00255	0.00007	0.94552	
44	CL86	10110101	3	0.00283	0.00027	0.94269	
43	CL89	CL64	4	0.00291	0.00008	0.93979	
42	CL65	CL67	5	0.00312	0.00021	0.93666	
41	CL47	11001011	3	0.00315	0.00003	0.93352	
40	CL56	CL82	5	0.00316	0.00001	0.93036	
39	00110011	00110100	2	0.00316	0.00000	0.92720	
38	10010011	CL71	4	0.00321	0.00005	0.92399	
37	CL49	CL91	5	0.00327	0.00005	0.92072	
36	CL61	CL83	5	0.00336	0.00009	0.91736	
35	CL52	01101001	4	0.00364	0.00029	0.91372	
34	CL58	01101100	5	0.00423	0.00059	0.90949	
33	CL40	CL69	8	0.00449	0.00026	0.90500	
32	CL63	CL44	7	0.00501	0.00052	0.89999	
31	00111110	CL68	4	0.00515	0.00014	0.89484	
30	CL34	CL73	7	0.00546	0.00031	0.88938	
29	CL78	CL38	6	0.00601	0.00055	0.88337	
28	CL45	11110010	5	0.00606	0.00005	0.87731	
27	CL60	CL31	6	0.00661	0.00055	0.87069	
26	CL37	CL29	11	0.00679	0.00018	0.86390	
25	CL35	11011010	5	0.00691	0.00011	0.85699	
24	CL70	CL57	5	0.00750	0.00059	0.84949	
23	CL36	CL46	10	0.00789	0.00039	0.84160	
22	CL42	CL94	7	0.00823	0.00034	0.83337	
21	10001111	11100001	2	0.00840	0.00017	0.82498	
20	CL30	CL41	10	0.00926	0.00087	0.81571	
19	CL55	CL54	4	0.01031	0.00105	0.80540	
18	CL32	CL50	12	0.01044	0.00013	0.79496	
17	CL25	CL48	7	0.01081	0.00037	0.78415	
16	CL19	CL39	6	0.01403	0.00322	0.77012	
15	CL77	CL21	4	0.01488	0.00085	0.75524	
14	CL20	11101101	11	0.01529	0.00041	0.73995	
13	CL27	CL33	14	0.01575	0.00046	0.72420	
12	CL23	CL43	14	0.01719	0.00144	0.70701	
11	CL24	CL28	10	0.01978	0.00259	0.68724	
10	CL11	11011011	11	0.02184	0.00206	0.66540	
9	CL12	CL26	25	0.02273	0.00090	0.64267	
8	CL13	CL17	21	0.03466	0.01192	0.60801	
7	CL18	CL22	19	0.03910	0.00444	0.56892	yes
6	CL16	CL9	31	0.06513	**0.02603**	0.50379	
5	CL6	CL14	42	0.06863	0.00351	0.43516	
4	CL8	CL7	40	0.08477	0.01613	0.35039	
3	CL4	CL10	51	0.08593	0.00116	0.26446	
2	CL5	CL3	93	0.09886	0.01294	0.16560	
1	CL2	CL15	97	0.16560	0.06674	0.00000	

Note: Clustering was stopped at 7 clusters because of the largest increase in the semi-partial r-squared value at Cluster 6 surrounded by smaller increases.

Table C-14: Results of Ward's minimum variance cluster analysis for the Chesapeake Bay Watershed sub-watersheds data set.

Number of Clusters	--Clusters	Joined--	Frequency of New Cluster	Semipartial R-Squared	Increase in Semipartial R-Squared	R-Squared	Clustering Stopped
45	2070001	2080201	2	0.00098		0.99902	
44	2050107	2070011	2	0.00102	0.00004	0.99801	
43	2060003	2070010	2	0.00108	0.00006	0.99693	
42	2050106	2080207	2	0.00114	0.00006	0.99579	
41	2080105	2080106	2	0.00200	0.00086	0.99379	
40	2050206	2050306	2	0.00208	0.00009	0.99171	
39	2050203	2050205	2	0.00250	0.00041	0.98921	
38	2070002	2080203	2	0.00259	0.00010	0.98662	
37	2050103	2060006	2	0.00265	0.00006	0.98397	
36	2050101	CL42	3	0.00285	0.00020	0.98112	
35	2070006	2080202	2	0.00301	0.00016	0.97810	
34	2050303	2050304	2	0.00304	0.00002	0.97506	
33	2050104	2050105	2	0.00347	0.00044	0.97159	
32	2050204	2050302	2	0.00362	0.00015	0.96797	
31	CL43	2080206	3	0.00384	0.00021	0.96413	
30	2060002	2080104	2	0.00396	0.00013	0.96017	
29	2070008	2080205	2	0.00436	0.00040	0.95581	
28	2050305	2070004	2	0.00453	0.00017	0.95128	
27	CL44	CL40	4	0.00462	0.00009	0.94666	
26	2060005	2060008	2	0.00549	0.00087	0.94117	
25	CL37	2070009	3	0.00594	0.00045	0.93523	
24	2080103	CL41	3	0.00679	0.00085	0.92844	
23	CL36	2050102	4	0.00687	0.00007	0.92157	
22	CL39	2070003	3	0.00697	0.00010	0.91460	
21	CL27	CL28	6	0.01111	0.00414	0.90349	
20	2050301	CL30	3	0.01158	0.00047	0.89191	
19	2070005	CL35	3	0.01194	0.00035	0.87997	
18	CL25	CL29	5	0.01268	0.00075	0.86729	
17	CL32	CL19	5	0.01314	0.00045	0.85415	
16	CL22	CL38	5	0.01556	0.00242	0.83859	
15	CL20	CL26	5	0.01860	0.00304	0.81999	
14	CL18	2080204	6	0.01920	0.00060	0.80079	
13	CL33	CL34	4	0.01950	0.00030	0.78129	
12	2050201	2050202	2	0.02155	0.00205	0.75974	
11	CL23	CL24	7	0.02329	0.00174	0.73645	
10	CL21	CL31	9	0.02745	0.00416	0.70899	
9	CL16	CL45	7	0.02878	0.00132	0.68022	yes
8	CL14	CL13	10	0.04827	**0.01950**	0.63194	
7	CL9	CL17	12	0.05426	0.00599	0.57768	
6	CL8	CL10	19	0.07117	0.01691	0.50651	
5	CL12	2060009	3	0.08435	0.01317	0.42216	
4	CL11	CL15	12	0.08658	0.00224	0.33558	
3	CL4	CL6	31	0.09029	0.00370	0.24529	
2	CL5	CL7	15	0.11451	0.02423	0.13078	
1	CL3	CL2	46	0.13078	0.01626	0.00000	

Note: Clustering was stopped at 9 clusters because of a large increase in the semi-partial r-squared value at Cluster 8 surrounded by two very small increases. The larger increase at Cluster 2 was not used because more than 3 clusters were desired.

Table C-15: Results of discriminant analysis of the Tennessee River Watershed original data set. See Table B-1 for data set description.

		Number of Observations classified into each cluster								
		1	2	3	4	5	6	7	8	Total
		Five Representative Metrics								
Original Cluster	1	10	0	0	0	0	1	0	0	11
	2	0	10	0	0	0	0	0	0	10
	3	2	0	7	0	0	0	0	0	9
	4	3	1	0	12	1	2	0	1	20
	5	2	1	0	0	9	1	0	0	13
	6	0	0	0	0	0	11	0	0	11
	7	0	0	0	0	0	0	7	0	7
	8	0	0	0	0	0	0	0	4	4
Total Observations										85
Observations Correctly Classified		10	10	7	12	9	11	7	4	70
Percentage Correctly Classified										82.35%
		Six Representative Metrics								
Original Cluster	1	10	0	0	0	0	1	0	0	11
	2	0	10	0	0	0	0	0	0	10
	3	2	0	7	0	0	0	0	0	9
	4	1	1	0	15	1	1	0	1	20
	5	3	1	0	0	9	0	0	0	13
	6	0	0	0	0	0	11	0	0	11
	7	0	0	0	0	0	0	7	0	7
	8	0	0	0	0	0	0	0	4	4
Total Observations										85
Observations Correctly Classified		10	10	7	15	9	11	7	4	73
Percentage Correctly Classified										85.88%

Table C-16: Results of discriminant analysis of the Tennessee River Watershed low-resolution majority-rule data set. See Table B-1 for data set description.

		Number of Observations classified into each cluster						
		1	2	3	4	5	6	Total
		Five Representative Metrics						
Original Cluster	1	14	4	1	1	3	0	23
	2	1	13	0	2	1	0	17
	3	2	3	12	0	0	1	18
	4	0	0	0	10	1	2	13
	5	1	0	0	1	10	0	12
	6	0	0	0	0	0	2	2
Total Observations								85
Observations Correctly Classified		14	13	12	10	10	2	61
Percentage Correctly Classified								71.76%
		Six Representative Metrics						
Original Cluster	1	18	4	1	0	0	0	23
	2	1	13	1	1	1	0	17
	3	0	1	15	0	0	2	18
	4	0	0	0	10	1	2	13
	5	1	0	0	1	9	1	12
	6	0	0	0	0	0	2	2
Total Observations								85
Observations Correctly Classified		18	13	15	10	9	2	67
Percentage Correctly Classified								78.82%

Table C-17: Results of discriminant analysis of the Tennessee River Watershed high-resolution majority-rule data set. See Table B-1 for data set description.

		Number of Observations classified into each cluster						
		1	2	3	4	5	6	Total
		Five Representative Metrics						
Original Cluster	1	13	0	0	0	0	0	13
	2	1	16	3	3	0	2	25
	3	0	1	10	0	1	0	12
	4	0	0	0	8	0	0	8
	5	1	3	1	1	10	0	16
	6	0	1	0	1	0	9	11
Total Observations								85
Observations Correctly Classified		13	16	10	8	10	9	66
Percentage Correctly Classified								77.65%
		Six Representative Metrics						
Original Cluster	1	13	0	0	0	0	0	13
	2	2	21	0	0	0	2	25
	3	0	1	11	0	0	0	12
	4	0	0	0	8	0	0	8
	5	1	2	0	1	12	0	16
	6	0	0	0	1	0	10	11
Total Observations								85
Observations Correctly Classified		13	21	11	8	12	10	75
Percentage Correctly Classified								88.24%

Table C-18: Results of discriminant analysis of the Tennessee River Watershed low-resolution nearest-neighbor data set. See Table B-1 for data set description.

		Number of Observations classified into each cluster					
		1	2	3	4	5	Total
		Five Representative Metrics					
Original Cluster	1	13	0	2	0	2	17
	2	1	18	0	1	2	22
	3	3	0	20	1	1	25
	4	0	1	2	16	0	19
	5	1	0	0	0	1	2
Total Observations							85
Observations Correctly Classified		13	18	20	16	1	68
Percentage Correctly Classified							80.00%
		Six Representative Metrics					
Original Cluster	1	12	1	2	0	2	17
	2	1	18	0	0	3	22
	3	2	1	21	0	1	25
	4	0	0	2	17	0	19
	5	1	0	0	0	1	2
Total Observations							85
Observations Correctly Classified		12	18	21	17	1	69
Percentage Correctly Classified							81.18%

Table C-19: Results of discriminant analysis of the Tennessee River Watershed 5-class data set. See Table B-1 for data set description.

		Number of Observations classified into each cluster				
		1	2	3	4	Total
		Five Representative Metrics				
Original Cluster	1	26	7	1	5	39
	2	2	13	5	0	20
	3	1	4	10	1	16
	4	0	0	0	10	10
Total Observations						85
Observations Correctly Classified		26	13	10	10	59
Percentage Correctly Classified						69.41%
		Six Representative Metrics				
Original Cluster	1	27	9	0	3	39
	2	1	19	0	0	20
	3	0	2	13	1	16
	4	0	0	0	10	10
Total Observations						85
Observations Correctly Classified		27	19	13	10	69
Percentage Correctly Classified						81.18%

Table C-20: Results of discriminant analysis of the Chesapeake Bay Watershed quadrants data set. See Table B-1 for data set description.

		Number of Observations classified into each cluster							
		1	2	3	4	5	6	7	Total
		Five Representative Metrics							
Original Cluster	1	13	0	2	0	3	1	0	19
	2	1	11	5	1	1	1	1	21
	3	1	4	16	1	1	0	2	25
	4	0	2	1	7	1	0	0	11
	5	2	0	0	0	1	1	0	4
	6	0	0	0	1	0	10	0	11
	7	0	0	0	0	0	0	6	6
Total Observations									97
Observations Correctly Classified		13	11	16	7	1	10	6	64
Percentage Correctly Classified									65.98%
		Six Representative Metrics							
Original Cluster	1	17	0	1	0	0	1	0	19
	2	2	11	5	1	0	1	1	21
	3	1	5	15	2	0	0	2	25
	4	1	2	1	7	0	0	0	11
	5	0	0	0	0	4	0	0	4
	6	0	0	0	1	0	10	0	11
	7	0	0	0	0	0	0	6	6
Total Observations									97
Observations Correctly Classified		17	11	15	7	4	10	6	70
Percentage Correctly Classified									72.16%

Table C-21: Results of discriminant analysis of the Chesapeake Bay Watershed sub-watersheds data set. See Table B-1 for data set description.

		Number of Observations classified into each cluster									
		1	2	3	4	5	6	7	8	9	Total
		Five Representative Metrics									
Original Cluster	1	7	0	0	0	0	0	0	0	0	7
	2	0	7	0	1	0	1	0	0	0	9
	3	0	0	7	0	0	0	0	0	0	7
	4	0	1	0	4	0	1	0	0	0	6
	5	0	0	0	0	5	0	0	0	0	5
	6	0	1	0	0	2	1	0	0	0	4
	7	0	0	0	0	0	0	5	0	0	5
	8	0	0	0	0	0	0	0	2	0	2
	9	0	0	0	0	0	0	0	0	1	1
Total Observations											46
Observations Correctly Classified		7	7	7	4	5	1	5	2	1	39
Percentage Correctly Classified											84.78%
		Six Representative Metrics									
Original Cluster	1	7	0	0	0	0	0	0	0	0	7
	2	0	7	0	1	0	1	0	0	0	9
	3	0	0	7	0	0	0	0	0	0	7
	4	0	1	0	4	0	1	0	0	0	6
	5	0	0	0	0	5	0	0	0	0	5
	6	0	1	0	0	2	1	0	0	0	4
	7	0	0	0	0	0	0	5	0	0	5
	8	0	0	0	0	0	0	0	2	0	2
	9	0	0	0	0	0	0	0	0	1	1
Total Observations											46
Observations Correctly Classified		7	7	7	4	5	1	5	2	1	39
Percentage Correctly Classified											84.78%

VITA

Douglas Hamilton Cain was born in Chattanooga, Tennessee on July 18, 1967. He attended schools in the public system of Hamilton County, Tennessee, where he graduated from East Ridge High School in May, 1985. He entered the University of Tennessee at Knoxville during September of 1985 where in May, 1989, he received the Bachelor of Arts in History. After serving three years as a commissioned officer in the United States Army, he entered the Master's program in Geography at the University of Tennessee at Knoxville in January of 1993, officially receiving the Master's degree in August, 1995.

He is presently working as an analyst/programmer for the Sidwell Company in Chicago, Illinois.