



# **Regional Vulnerability Assessment for the Mid-Atlantic Region: Evaluation of Integration Methods and Assessments Results**



EPA/600/R-03/082  
October 2003

# **Regional Vulnerability Assessment for the Mid-Atlantic Region: Evaluation of Integration Methods and Assessments Results**

by

Elizabeth R. Smith  
NERL

Liem T. Tran  
Florida Atlantic University

Robert V. O'Neill  
T N and Associates, Inc.

---

## Notice

The U.S. Environmental Protection Agency (U.S. EPA), through its Office of Research and Development (ORD), funded and managed the research described here under Interagency Agreement number DW1393920801-0 with the U.S. Department of Commerce. It has been subjected to the Agency's peer and administrative review and has been approved for publication as an EPA document.

## Acknowledgments

Many people contributed to this report, through analyses and interpretation, methods development, and other input to the ReVA program. Specifically we would like to acknowledge the following:

Rochelle Araujo, EPA, National Exposure Research Laboratory  
Michael Blum, EPA, National Exposure Research Laboratory  
Jeffrey Frithsen, EPA, National Center for Environmental Assessment  
Earl Greene, U.S. Geological Survey, Water Resources Division  
Laura Jackson, EPA, National Health and Environmental Effects Research Laboratory  
Kimberley Johnson, EPA, National Exposure Research Laboratory  
Bruce Jones, EPA, National Exposure Research Laboratory  
Vasu Kilaru, EPA, National Exposure Research Laboratory  
Daniel Kluza, EPA, National Center for Environmental Assessment  
Andrew La Motte, U.S. Geological Survey, Water Resources Division  
Joshua Lawler, National Research Council  
Rick Linthurst, EPA, Office of the Inspector General  
Nicholas Locantore, Waratah Corporation  
Peter McKinnis, Waratah Corporation  
Megan Mehaffey, EPA, National Exposure Research Laboratory  
Michael O'Connell, Waratah Corporation  
Esther Parrish, Tennessee Valley Authority  
Kurt Riitters, U.S. Forest Service  
Roger Tankersley, Tennessee Valley Authority  
Dennis Yankee, Tennessee Valley Authority  
Timothy Wade, EPA, National Exposure Research Laboratory  
Paul Wagner, EPA, National Exposure Research Laboratory  
Lisa Wainger, University of Maryland  
James Wickham, EPA, National Exposure Research Laboratory

---

## Executive Summary

Decision-makers need information on cumulative and aggregate stressors as well as clear information on where problems are likely to occur in the future in order to prioritize risk management actions. The most pervasive and difficult to assess changes are the result of regional-scale drivers of change that act simultaneously on a suite of resources that are important to society and to ecological sustainability. A great deal of data already exist that could potentially inform risk management decisions; however, there has been no effort previously to synthesize these data into meaningful assessment results that can inform the multiple criteria that go into any kind of decision-making. Methods to do this are critical to timely, responsive, and proactive decision-making.

The Regional Vulnerability Assessment (ReVA) Program has focused initially on the synthesis of existing data. We have used the same set of spatial data and synthesized these data using a total of 11 existing and newly developed integration methods. These methods were evaluated in terms of 1) how well each individual method performs given different data issues that are encountered with existing data, and 2) how effectively each method addresses different types of assessment questions.

Specific data issues that are addressed in our evaluation of integration methods include:

- Discontinuity – How are the methods affected by variables that (in raw form) are counts, such as number of aquatic species, versus having only continuous data?
- Imbalance – What effect does having too many variables of a particular type (e.g., representative of terrestrial conditions versus aquatic) have on the integration results from individual methods?
- Skewness – What effect does having variables with highly skewed distributions have on integration results? Many statistical methods are valid only for symmetrically distributed data or require transformation of the data.
- Interdependency – How are the methods affected by including variables that are highly correlated with one another?

Prioritization of risk management actions involves balancing many different factors that can be addressed through a series of assessment questions. ReVA's evaluation of integration methods considers which methods are most suitable to address questions such as:

- What is the overall environmental condition of the region?
- What is the relative condition of locations within a region?
- Where are the most vulnerable (i.e., both high stressor levels and high numbers of resources) locations in a region?
- How will conditions and vulnerabilities change in the future?
- How applicable are risk management options to other locations in the region?

Analysis results presented here should provide useful information to others involved in integrated risk assessment in that integration methods were tested and compared using the same set of regional spatial data. This should allow analysts to avoid problems presented by the use of existing data that may invalidate results from integration methods that are sensitive to inherent issues. Additionally, by comparing the results of each method, we have identified which integration methods are appropriate to address different types of assessment questions that contribute to informing decisions that require prioritization of areas for risk management actions. These results are directly transferable to other regions and can be applied to other scales of data.

Below are recommendations for the use of the integration methods described in this report:

- A. Use a suite of integration methods: There is no universal integration method that can cover all tasks of an integrated environmental assessment. An individual method has advantages in some aspects but is disadvantaged regarding others. The use of multiple methods in a complementary manner will help the user look at the problem from different angles/perspectives. It also gives the user a better chance to detect whether a pattern/abnormality on the map is a real environmental signal or just an arbitrary object created by some “strange” calculation.
- B. Start with the simple methods (*Simple Sum, Best/Worst Quantiles*) first and move to other complicated ones later. This will help the user to have a general picture of the study area before involving in more complicated and detailed calculations (i.e., see the forest first before get down to the tree).
- C. Keep it simple: If several methods provide similar patterns and/or results, stick with the simple methods and drop off the complicated ones.
- D. Pay proper care to data: The ways in which data are coded or transformed have a big influence on the integration results. Try to keep a balance between data transformation and data interpretation. It is because, while data transformation can reduce some particular problems (e.g., log transform to reduce skewness), it might cause difficulties in interpreting as well as in putting the transformed variable in the same calculation with other variables in the data set.

---

## Table of Contents

<b>Section 1 – Introduction .....</b>	<b>1</b>
Background .....	1
Regional Vulnerability .....	1
Purpose of this Report .....	2
Pilot Study Area .....	3
Data .....	3
<b>Section 2 – Integration Methods .....</b>	<b>9</b>
Best and Worst Quantiles .....	9
Results.....	10
Advantages & Disadvantages .....	11
Recommendations.....	11
Simple Sum .....	11
Advantages & Disadvantages .....	12
Recommendations.....	12
Principal Component Analysis.....	12
Advantages & Disadvantages .....	13
Recommendations.....	13
State Space Analysis .....	14
Tests .....	15
Advantages & Disadvantages .....	15
Recommendations.....	16
Criticality Analysis.....	16
Methodology .....	17
Preliminary Results.....	18
Testing the Sensitivity of the Method to Assumptions .....	19
Test.....	20
Advantages & Disadvantages .....	21
Recommendations.....	22
Analytic Hierarchy Process .....	22
Results.....	25
Tests .....	26
Advantages & Disadvantages .....	26
Recommendations.....	27
Clustering Analysis .....	27
Results.....	28
Tests .....	30
Advantages & Disadvantages .....	30
Recommendations.....	30
Self-organizing Map.....	31
Results.....	32
Tests & Sensitivity .....	35
Advantages & Disadvantages .....	36
Recommendation .....	37

Stressor-Resource Overlay .....	37
Advantages & Disadvantages .....	38
Recommendations.....	38
Change Analysis.....	39
Advantages & Disadvantages .....	40
Recommendation .....	40
Stressor-Resource Matrix Analysis .....	40
Correlation – Results and Tests .....	41
Stressors: Results and Tests .....	41
Vulnerable Resources: Results and Tests .....	42
Regression – Results and Tests .....	43
Stressors .....	43
Vulnerable Resources .....	43
Advantages & Disadvantages .....	44
Recommendations.....	45
<b>Section 3 – Sensitivity of Integration Methodology to Data .....</b>	<b>47</b>
<b>Section 4 – Discussion &amp; Recommendations .....</b>	<b>51</b>
Discussion .....	51
Classification/Ranking.....	51
Risk/Vulnerability Assessment.....	52
Planning/Restoration/Development .....	53
Subjective Judgments/Expert Knowledge .....	53
Combination of Integration Methods.....	54
Recommendations .....	54
<b>Appendix A – Data .....</b>	<b>55</b>
Table 1A. Correlation matrix for variables included in the evaluation of integration methods .....	56
Table 2A. Variables representing resources in analyses .....	61
Table 3A. Variables representing stressors in analyses.....	62
<b>Appendix B – Calculations .....</b>	<b>63</b>
Tran and Duckstein’s Fuzzy Ranking Method.....	63
Distance measure for interval numbers.....	63
Distance measure for fuzzy numbers .....	63
Table B1. Distance functions for some commonly used fuzzy numbers .....	65
Mechanics of Integration Methods and Supporting Software .....	66
Mechanics of Each Data Integration Method .....	66
Best Quintile .....	66
Worst Quintile.....	67
Simple Sum.....	67
PCA Distance (Euclidean) .....	67
State Space (zero reference location).....	68
Criticality Analysis .....	68
Stressor/Resource Overlay .....	68
References.....	71

---

## List of Tables

Table 1.	List of variables used in evaluating integration methods .....	6
Table 2.	Summary of data integration methods.....	9
Table 3.	Groups of variables for the <i>AHP</i> Method.....	24
Table 4.	Correlation matrix results: top four stressors .....	42
Table 5.	Correlation matrix results: top five vulnerable resources.....	42
Table 6.	Regression analysis results: top four stressors .....	43
Table 7.	Regression analysis results: top five vulnerable resources.....	44
Table 8.	Effects of various data issues on each integration method.....	48



## List of Figures

Figure 1.	Map displaying the Mid-Atlantic region watershed (8-digit Hydrologic Unit Code) and state boundaries. ....	4
Figure 2.	Land cover map of the Mid-Atlantic region.....	5
Figure 3.	Map displays results of the <i>Best Quantile</i> ranking method.....	10
Figure 4.	Map displays results of the <i>Worst Quantile</i> ranking method.....	10
Figure 5.	Map displays results of <i>Simple Sum</i> method.....	11
Figure 6.	The PCA results displayed in equal-interval septiles.....	13
Figure 7.	<i>State Space Analysis</i> results displayed in equal-size septiles with Appomattox as the “most vulnerable” watershed.....	14
Figure 8.	<i>State Space Analysis</i> results displayed in equal-size septiles with Lower York Shenandoah as the “most vulnerable” watershed.....	15
Figure 9.	Map shows <i>Criticality Analysis</i> results.....	19
Figure 10.	Results of the <i>Criticality Analysis</i> done with a second set of reference conditions that might be considered more conservative.....	21
Figure 11.	Diagram of the hierarchies in the <i>AHP</i> model for regional environmental assessment.....	23
Figure 12.	AHP-L <sub>1</sub> results displayed in equal-size septiles, where watersheds the furthest distance from ideal are shown in dark red.....	25
Figure 13.	AHP-L <sub>2</sub> results displayed in equal-size septiles, where watersheds the furthest distance from ideal are shown in dark red.....	26
Figure 14.	Results of K-means <i>Clustering</i> analysis.....	29
Figure 15.	Results of hierarchical <i>Clustering</i> analysis using within-group average linkage.....	29
Figure 16.	Results of hierarchical <i>Clustering</i> analysis using complete linkage and Ward linkage.....	30
Figure 17.	Diagram of a two-level self-organizing map model.....	32
Figure 18.	First-level 10×5 self-organizing map created by the U-matrix method.....	33
Figure 19.	Second-level 7×1 self-organizing map created by the U-matrix method.....	34
Figure 20.	Map shows geographic distribution of the two-level SOM’s seven clusters.....	35
Figure 21.	Map shows geographic distribution of the one-level SOM’s seven clusters.....	36
Figure 22.	Results of the <i>Stressor-Resource Overlay</i> method displayed in a 16-category map.....	38
Figure 23.	Graphic illustrates the <i>Change Analysis</i> method.....	39

---

# Section 1

## Introduction

### ***Background***

The U.S. EPA's Regional Vulnerability Assessment (ReVA) program is designed to develop and demonstrate approaches that address the latter phases of an integrated ecological risk assessment (U.S. EPA 1998), following development of specific assessment questions (problem formulation) and building on available monitoring data, with a focus on integrating and synthesizing information on the spatial patterns of multiple exposures to allow a comparison and prioritization of risks. ReVA is not designed to do complete regional assessments and assumes that assessment endpoints have already been identified and that monitoring data that represent these endpoints are available. ReVA will provide guidance and tools for this phase of the assessment process, but the full assessment of regional vulnerabilities is primarily the responsibility of regional decision-makers (Moss 2002), including those in EPA regional offices, as well as state and local administrators.

ReVA's strategic priorities include:

1. Focus on the synthesis of existing data. As the second phase in the assessment process, ReVA uses *available* monitoring data and model results to address current decision-making needs.
2. Expand the scope of research to include a full suite of stressors and ecological resources and refine the integration techniques through application to this broader array of information.
3. Develop the ReVA approaches and demonstrate their application at the regional, watershed, and local scale.
4. Initiate studies in other regions to test the applicability of the methods and approaches in other areas and repeat the process (Smith et al. 2002).

### ***Regional Vulnerability***

A *region* is defined as a large, multi-state geographic area such as the Mid-Atlantic, Northeast, Southeast, or Pacific Northwest regions within the United States. An EPA Region is a useful representation of a geographic region because it reflects the size of the geographic area initially considered in the ReVA program and because strategic planning and management decisions are made at this scale. This regional-scale information should prove valuable for decision-making at finer scales because it will provide context as well as insights into changes in regional-scale stresses. Integration methods presented here should be scale-independent.

*Vulnerability* has multiple elements in its definition but is most simply represented by the probability that future conditions will deteriorate or degrade. We see ecosystems as relatively stable configurations of a number of species with the ability to resist and/or recover from the normal array of disturbances such as

fire, flood, and drought that it has experienced over its evolutionary history. We assume stability, resiliency, adaptability, and resistance when we extract resources from the system, depend on it to purify wastes, or impose recreational impacts. However, these assumptions are no longer valid when the stresses we impose are outside the range that the organisms have evolved to resist and move that ecological system outside the normal range of variability. Thus, the vulnerability of an ecological system increases as the number, intensity, and frequency of stressors increases. Cumulative and aggregate stresses that have occurred over time may influence the prioritization of ecosystems based on vulnerability.

Regional vulnerability is many things. It is rarity, synergy, sensitivity, spatial context, and history. No single question or approach will suffice to encompass it. Likewise, decision-making based on ecological vulnerability is a complex prioritization process that includes evaluation of multiple criteria (Saaty and Vargas 1982, Ridgley and Rijsberman 1992). Decisions to implement risk management or risk reduction strategies can include evaluations of 1) current conditions, 2) risk of future harm, 3) feasibility of management options, and 4) value of the ecosystem at risk, all of which have different levels of importance for different decision-makers. Regional vulnerability analysis approaches developed by ReVA will thus draw on many sources of data, will explore many different assessment methods, and will enable decision-makers to ask many different questions.

## ***Purpose of this Report***

This report presents analyses supportive of our first strategic priority and evaluates existing and newly developed integration methods with regards to 1) how well each individual method performs given different data issues that are encountered with existing data, and 2) how effectively each method addresses different types of assessment questions. As our focus here is on the evaluation of integration methods, results of these integrations should not be considered to be a complete assessment of regional conditions or vulnerabilities. Those results will be presented in future ReVA reports which will focus on addressing a suite of assessment questions and future scenarios that might be used to target risk reduction actions and prioritize use of resources (ReVA's second strategic priority).

Limiting analyses to the use of existing data both poses both opportunities and problems. By constraining our analyses, we do not have the luxury of either collecting data specifically relevant to assessment questions or having data that are particularly suited for integration. However, use of existing data allows much more timely decision-making, avoids the expense of additional data collection, and to some degree reflects the regional issues of concern in that existing monitoring was initiated in response to some perceived need. Specific data issues that are being addressed in our evaluation of integration methods include:

- Discontinuity – How are the methods affected by variables that (in raw form) are counts, such as number of aquatic species, versus having only continuous data?
- Imbalance – What effect does having too many variables of a particular type (e.g., representative of terrestrial conditions versus aquatic) have on the integration results from individual methods?
- Skewness – What effect does having variables with highly skewed distributions have on integration results? Many statistical methods are valid only for symmetrically distributed data or require transformation of the data.
- Interdependency – How are the methods affected by including variables that are highly correlated with one another?

Prioritization of risk management actions involves balancing many different factors that can be addressed through a series of assessment questions. ReVA's evaluation of integration methods considers which methods are most suitable to address questions such as:

- What is the overall environmental condition of the region?
- What is the relative condition of locations within a region?
- Where are the most vulnerable (i.e., both high stressor levels and high numbers of resources) locations in a region?
- How will conditions and vulnerabilities change in the future?
- How applicable are risk management options to other locations in the region?

Analysis results presented here should provide useful information to others involved in integrated risk assessment in that a total of 11 different integration methods were tested and compared using the same set of regional spatial data. This should allow analysts to avoid problems presented by the use of existing data that may invalidate results from integration methods that are sensitive to inherent issues. Additionally, by comparing the results of each method, we have identified which integration methods are appropriate to address different types of assessment questions that contribute to informing decisions that require prioritization of areas for risk management actions. These results are directly transferable to other regions and can be applied to other scales of data.

## ***Pilot Study Area***

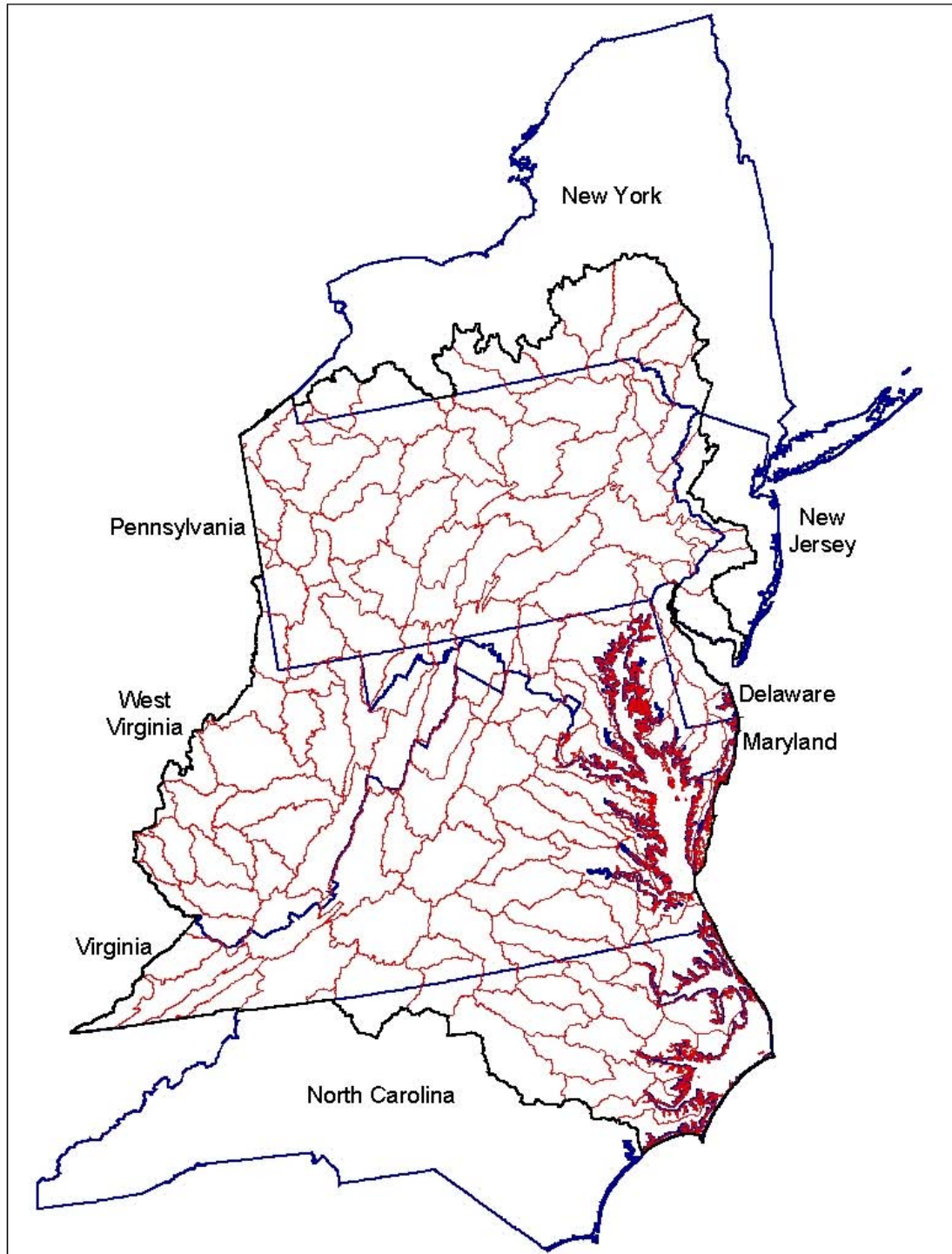
ReVA's pilot area is the Mid-Atlantic region as part of the Mid-Atlantic Integrated Assessment (MAIA) (Bradley and Landy 2000). The Mid-Atlantic encompasses portions of three physiographic provinces and eight states (see Figure 1), and it includes a wide range of ecological and environmental conditions. Human disturbance over the past 300 years has caused widespread changes. Some of these changes, such as the buildup of large urban areas, are obvious and easily interpreted; others are subtle synergistic effects that can only be examined by looking at groups of environmental measures.

Current land use patterns (see Figure 2) show predominantly forest with a long history of human disturbance. The coastal plain is dominated by urban development and most of the large cities in the region lie along the geologic boundary between the coastal plain and the piedmont. The piedmont includes most of the region's agricultural lands, with smaller cities and scattered forestland. The Appalachian highlands contain the world's largest remaining contiguous temperate forest (Riitters et al. 2000), interspersed with small- to medium-sized cities, some agriculture, and many mines. In essence, we have a gradient from the coastal plain (urban/agricultural matrix) to the piedmont (agriculture/forest matrix) to the Appalachians (forest matrix).

## ***Data***

For these analyses, our objective in data choice was to have regionally consistent spatial coverages for a reasonable number of reporting units. The choice of data was intended primarily to represent the range of data issues that might be encountered when doing a regional integrated assessment using existing data, rather than to be representative of data that would (or should) be used in the assessment of conditions or vulnerabilities for the region (these data will be used in ReVA's second report). The reporting units we used here include 8-digit hydrologic unit codes (HUCs), the only regionally consistent watershed delineation currently available. Beginning with a set of over 100 spatial coverages (variables)

for the region, we then eliminated variables with missing values and variables that were highly correlated (>98% correlation) (see data correlation Table 1A, Appendix). This left us with a total of 50 variables related to land cover, land use, aquatic life, terrestrial life, and economics over 141, 8-digit HUCs (see Table 1).



**Figure 1. Map displaying the Mid-Atlantic region watershed (8-digit Hydrologic Unit Code) and state boundaries.**

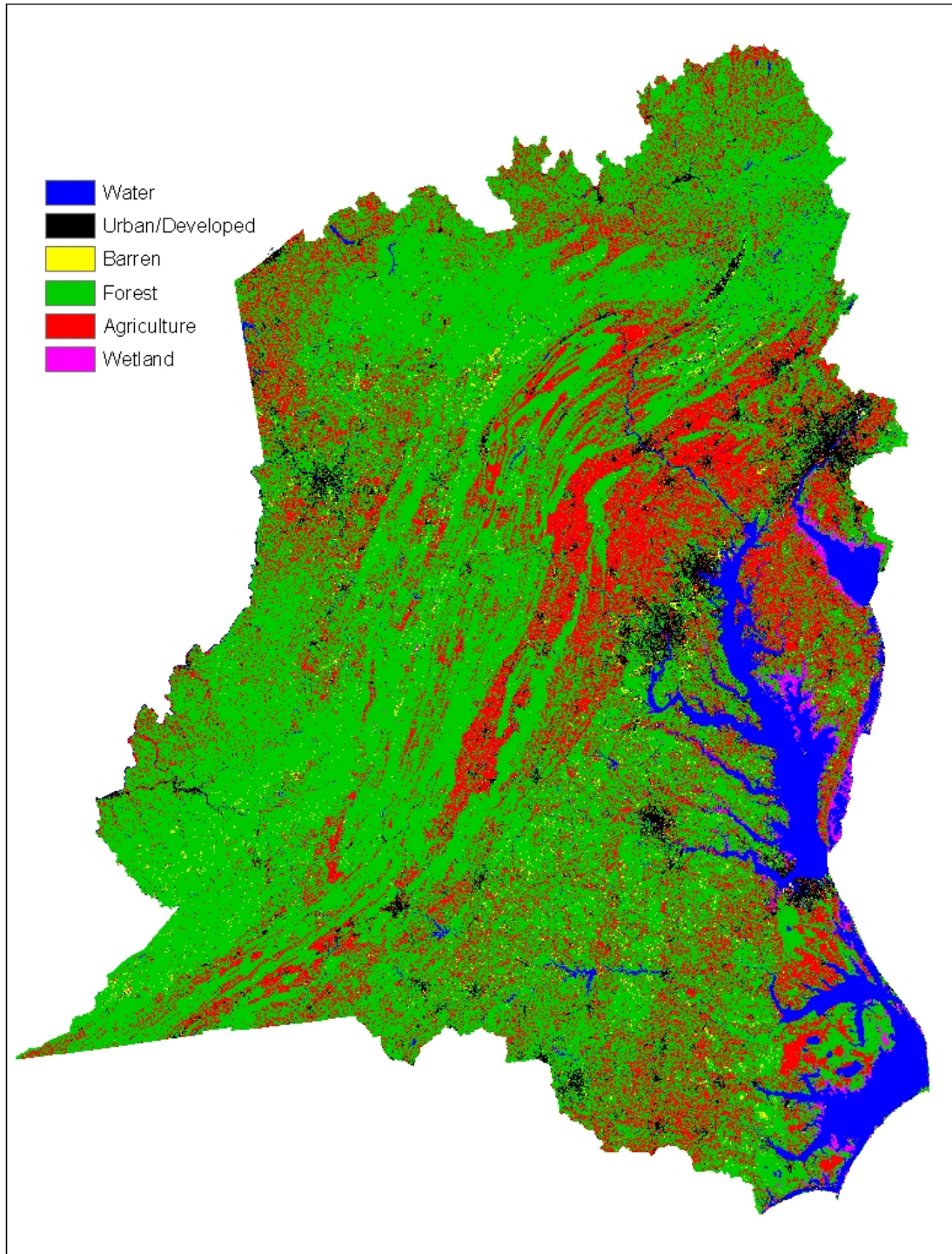


Figure 2. Land cover map of the Mid-Atlantic region (source: National Land Cover Database).



**Table 1. List of variables used in evaluating integration methods.**

<b>Abbreviation</b>	<b>Description</b>
AGSL	Proportion of watershed with agriculture land cover on slopes that are greater than three percent
AQUAEXOTIC	Count of exotic aquatic – fish and mussels – species
AQUANATIVE	Count of native aquatic – fish and mussels – species
AQUATE	Count of threatened and endangered aquatics – fish and mussels species
C5FS	Children (0-5) in families & subfamilies
CROPSL	Proportion of watershed with crop land cover on slopes that are greater than three percent
DAMS	Impoundment density (number of dams per 1,000 kilometers of stream length)
DISSOLVEDP	Estimated suspended sediment in streams modeled using land cover metrics
EDGE2	Percentage of forest habitat called edge (2 ha scale)
EDGE65	Percentage of forest habitat called edge (65 ha scale)
EMAGRIC	Employed persons by industry – agriculture, forestry, fisheries 1990
EMMINE	Employed persons by industry – mining 1990
FORCOVDEFOL	Percent of forest cover defoliated and with mortality as proportion of existing forest
FUNGICIDE	Annual fungicide loadings
HARDCHIPMIL	Estimate of increase (decrease) in chip mill for hardwoods capacity in tons, based on our regression, and assuming the Mid-Atlantic behaves like the South
HARDWOODINV	Index values for hardwood forest inventory. The index compares a baseline of most recently available FIA data against projections to 2020. Index values >1 are areas with increasing inventory
HARDWOODREM	Index values for hardwood removals. The index compares a baseline of most recently available FIA data against projections to 2020. Index values >1 are areas with increasing inventory
HERBICIDE	Annual atrazine loadings 1990-93
IMPLCPCT	Percent impervious surface by land cover
INDTHPTH	Infant deaths per 1,000 live births 1990
INSECTICIDE	Annual O-P Insecticides loadings 1990-93
INT2	Percentage of forest habitat called interior (2 ha scale)
INT65	Percentage of forest habitat called interior (65 ha scale)
MIGSCENARIO	Number of migratory scenarios for long-distance forest migrants that use a particular HUC or hexagon. Scenarios are defined by a combination of compass heading, landfall location along the gulf coast and southern Atlantic Coast, and nightly flight distance
NATCOVERPCT	Percent coverage with FOREST that matches potential vegetation in Kuchler
NBLDPM97	New private housing building permits 1997

**Table 1. (Continued)**

<b>Abbreviation</b>	<b>Description</b>
NO3DEPMODEL	Modeled annual wet deposition of nitrate based on averages from 1987-1999
NONCLIMAXPCT	Percent coverage with FOREST but the species are not the climax listed by Kuchler (1964)
NTCMPPLM	Incomplete plumbing
OZONE8HR	Ozone (8 hr max) is a human health indicator and is given in parts per billion (ppb)
POPDENS	Population density
POPGROWTH	Population growth rate from 1990-1995
POV65	65+ below poverty
PSOIL	Proportion of watershed with potential soil loss greater than 1 ton per acre per year; the percentage of HUC or hexagon area that is estimated to lose more than 1 ton/acre/year of soil due to erosion
RDDENS	The density numbers are meters of road per hectare of area
RIPAG	Proportion of total stream length with adjacent agriculture land cover; percent riparian buffer that is agricultural land
RIPFOR	Proportion of total stream length with adjacent forest land cover; percent riparian buffer that is forest
SO4DEPMODEL	Modeled annual wet deposition of sulfate based on averages from 1987-1999
SOFTCHIPMIL	Estimate of increase (decrease) in chip mill for softwoods capacity in tons, based on our regression, and assuming the Mid-Atlantic behaves like the South
SOFTWOODINV	Index values for softwood forest inventory. The index compares a baseline of most recently available FIA data against projections to 2020. Index values >1 are areas with increasing inventory
SOFTWOODREM	Index values for softwood removals. The index compares a baseline of most recently available FIA data against projections to 2020. Index values >1 are areas with increasing inventory
STRD	Number of road crossings per total stream length
SUM06	Cumulative sum of all hourly ozone concentrations equal to or above 0.06 ppm (or 60 ppb) for hours between 7 a.m. and 7 p.m. The SUM06 index is an indicator of ozone exposure that plants receive during daylight hours
TERREXOTIC	Count of exotic birds, mammals, butterflies, amphibians, and reptiles
TERRNATIVE	Count of native birds, mammals, butterflies, amphibians, and reptiles
TERRTE	Count of threatened and endangered birds, mammals, butterflies, amphibians, and reptiles
TOTALN	Estimated total nitrogen in streams modeled using land cover metrics
UINDEX	Human use index (proportion of watershed area with agriculture or urban land cover)
UVB	Mean Annual UV-B Irradiance
WETLNDSPCT	Percent of area classified as wetlands



All variables used in these analyses were normalized and inverted if necessary to make all indicators in the range of 0 to 1, where 0 and 1 represent environmentally desirable (good) and undesirable (poor) conditions, respectively. Values of zero represent the best conditions across the region, while those of one represent the worst conditions across the region. Normalization is used because it is a linear transformation that preserves the ranking and correlation structure of the variables, and it allows for variables with different scales to be used together (see Pielou 1984, p. 46-47).

Color-coded rankings presented here are designated based on equal intervals between values rather than having equal numbers of watersheds displayed for each class. This was an arbitrary choice and affects the resulting map. There is no apparent advantage to using either method. Ranking methods are displayed in this report in a red to green color-coding that reflects poor conditions to favorable. For the grouping methods, a palette that ranges from tan to brown was used as no ranking is intended. Watersheds with similar color-codings reflect similar characteristics.

---

## Section 2

### Integration Methods

The 11 methods presented in this report are outlined in Table 2. Each method will be described in detail.

**Table 2. Summary of data integration methods.**

Method	Description
Best/Worst Quantiles	Number of variables in the best/worst quantile
Simple Sum	Add the normalized values of all variables
Principal Component Analysis (PCA)	Transform variables and then calculate Euclidean distance from a reference
State Space Analysis	Mahalanobis distance from a reference
Criticality Analysis	Fuzzy distance to a hypothetical “natural” state
Analytical Hierarchy Process (AHP)	Multi-criteria tool that uses decision-maker preferences in the calculations
Cluster Analysis	Partitioning methods to group watersheds
Self-organizing Map (SOM)	Self-organizing map to group watersheds
Stressor-Resource Overlay	High-stress values with high-resource values
Change Analysis	Comparison of two regional maps
Stressor-Resource Matrix	Ranks stressors and resources

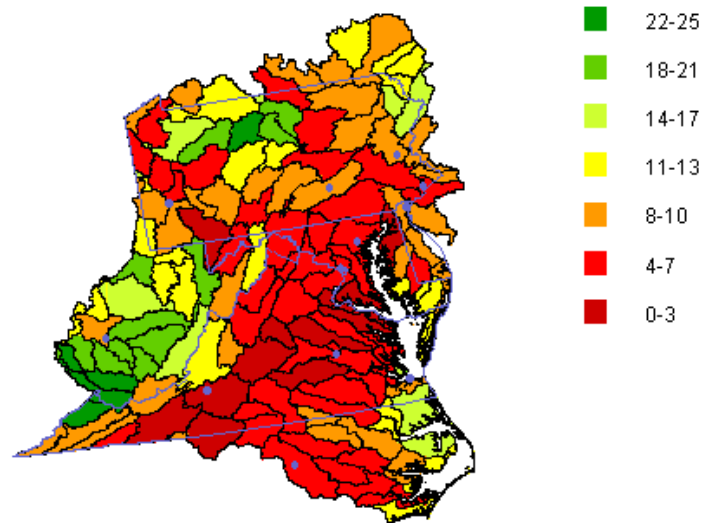
#### ***Best and Worst Quantiles***

One of the simplest ways to integrate environmental data is to compute a value for the phenomena of interest, rank the result, and create categories that represent an equal number of mapping units in each condition class. The resulting maps enable the user to see the relative difference between mapping units across the study area. A logical next step is to add the number of variables in the best class (or the worst class). This process creates easily interpretable maps that illustrate where favorable conditions cluster and where unfavorable conditions cluster.

The variables were ranked and subdivided into quantiles with the same number of watersheds. Each watershed was then evaluated in terms of the number of its scores that fell in the best and worst quantiles. The watersheds were depicted using the best quantile scores divided into seven equal intervals and then again using the worst quantile scores divided into seven equal intervals. The maps were colored such that green always implies best overall conditions and red always implies worst overall conditions.

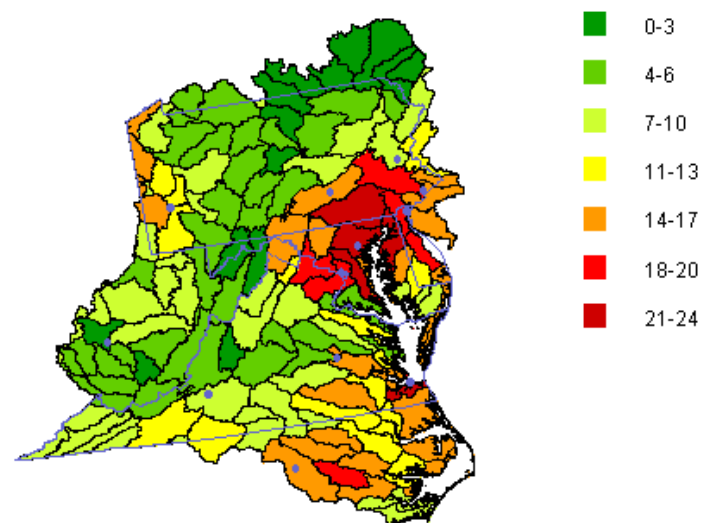
## Results

The *Best Quantile* map is shown in Figure 3. The map shows that the watersheds in the best environmental conditions (greens) are concentrated in the highlands along the western edge of the region. Areas of intermediate condition (yellows) are scattered over the region. Along the eastern edge of the map, these intermediate watersheds could be good candidates for preservation or remediation.



**Figure 3.** Map displays results of the *Best Quantile* ranking method. This method ranks watersheds based on the number of variables with the best values (closest to 0 on the normalized scale) within the region. Watersheds with a high number of variables in good condition are shown in green, while those with only few variables in good condition are shown in red.

The *Worst Quantile* map is shown in Figure 4. The map illustrates that the watersheds in the poorest environmental condition are concentrated in the urban areas along the eastern edge of the map.



**Figure 4.** Map displays results of the *Worst Quantile* ranking method. This method ranks watersheds based on the number of variables with the worst values (closest to 1 on the normalized scale) within the region.

## Advantages & Disadvantages

The advantage of the Quantile approach is ease of interpretation: The best and worst Quantile maps clearly present a landscape that separates high quality areas from the poor quality areas.

The Quantile method does not account for correlation between environmental measures. For example, if several variables measure forest condition in slightly different ways, a watershed with high percentage forest will have high scores for many forest variables and therefore have a high-best quantile score.

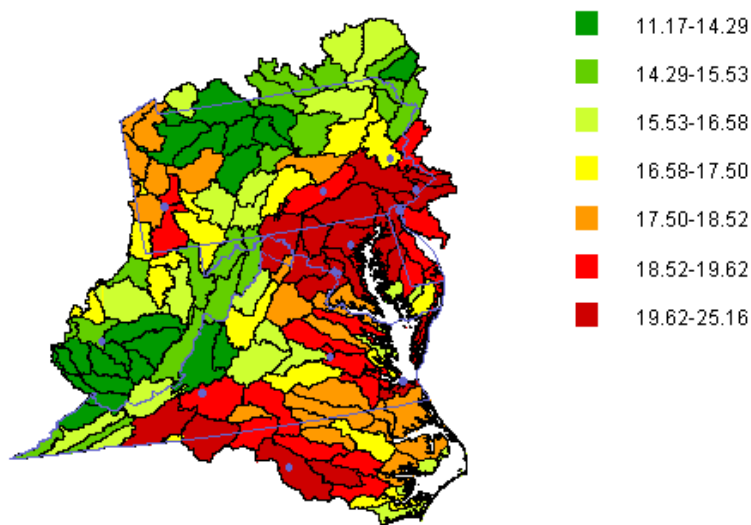
## Recommendations

The *Best and Worst Quantile* method is an unsophisticated approach to integration. It ignores the complex relationships among the variables. Any integration or assessment based on this method should be done with extreme caution and only accepted if supported with confirming independent evidence. Because these methods provide a quick overview of the region, they are best used as preliminary visualization methods.

To facilitate this application, it is recommended that the *Best Quantile* map show only the top three or four septiles in shades of green and leave the rest of the watersheds in gray. Likewise, the *Worst Quantile* map should show only the lowest three or four septiles in shades of red and leave the rest of the watersheds in gray.

## ***Simple Sum***

A straightforward approach to integrating environmental variables is to sum the normalized values that range from 0 (good) to 1 (bad). In this *Simple Sum* method, the smaller the sum, the better the overall environmental condition of the watershed. To display the results of the *Simple Sum* on the map, watersheds were grouped into seven equally sized groups or septiles (see Figure 5).



**Figure 5.** Map displays results of *Simple Sum* method. This method sums the normalized values of every variable, then it ranks the values based on equal intervals between values. Green indicates low values (good condition) and red indicates high values (bad conditions).

## Advantages & Disadvantages

The *Simple Sum* has the advantage of being intuitively simple and easily communicated. It makes no assumptions about the statistical distribution of the variables and therefore is not sensitive to discontinuities or non-normal distributions.

Using any single indicator always has the associated problem of occluding (Suter 1993a). This occurs when one or two variables that are clearly unsatisfactory cannot be detected because they occur on the same watershed with many variables indicating good conditions. The sum produces an intermediate value that occludes the danger. Although the problem is minimized in the ReVA analysis where the assessor has immediate access to all of the variables, nevertheless, any method as simple as summing the normalized variables incurs this problem.

A second disadvantage is that the *Simple Sum* does not account for the covariance structure of the data set. Thus, two highly correlated variables are added together as though they were independent of each other. The result is that a watershed may be overly penalized by a number of related stressors or appear better than it truly is because correlated resources are added as though they were independent.

Another disadvantage is that the *Simple Sum* only gives a relative ranking of the watersheds. There is no objective standard to which all the watersheds can be compared. Thus, the best watershed may not be objectively satisfactory and the worst watershed may not be objectively problematic. Therefore, the method should not be interpreted in isolation and should be used in conjunction with other methods that provide more objective standards.

## Recommendations

The *Simple Sum* method is an inadequate approach to integration. Because it ignores the complex relationships among the variables, any integration or assessment based on this method should be done with extreme caution and only accepted if backed up with confirming independent evidence. The *Simple Sum* method should not be used in isolation from other methods.

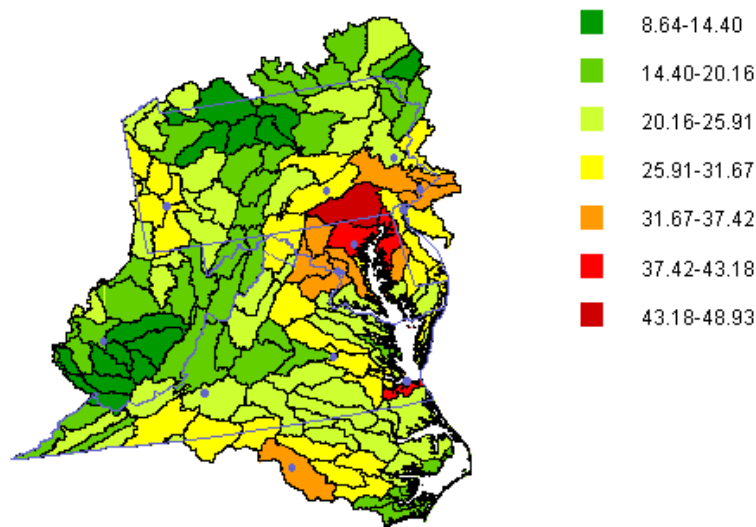
## ***Principal Component Analysis***

Principal Component Analysis (PCA), introduced by Pearson (1901) and independently by Hotelling (1933), is one of the oldest and most widely used statistical multivariate techniques. The basic idea is to describe the variation of a set of multivariate data with a new set of uncorrelated variables, each of which is a linear combination of the original variables, using the covariance (or correlation) matrix. PCA uses eigenvalues and their corresponding eigenvectors of the covariance (or correlation) matrix to derive the new variables in a decreasing order of importance in explaining variation of the original variables. Usually, if correlations among the original variables are large enough, the first few components will account for most of the variation in the original data. If that is the case, then a few can be used to represent the data with little loss of information, thus reducing the dimensionality of the data.

PCA has been applied in a wide array of studies in environmental sciences, especially for determining sources of some substances (e.g., Rachdawong and Christensen 1997; Statheropoulos and others 1998; Topalián and others 1999; Yu and Chang 2000) and revealing the relationships among different indicators (e.g., Calais and others 1996; Sjögren and others 1996; Yu and others 1998).

The PCA in this study was performed with varimax rotation to minimize the number of variables that have high loadings on each factor, simplifying the interpretation of the factors (Everitt and Dunn 1992). The use of the correlation matrix instead of the covariance matrix in the PCA was to assign equal weights

for all of the variables (Chatfield and Collins 1980). Then the eigenvectors (loadings) derived from the PCA was used to compute the so-called principal component (PC)-based indices. The PC-based indices were weighted sums of the 50 indicators where the weights were the loadings' absolute values (in other words, the PC-based indices were the principal component scores where all of the negative component score coefficients were converted to positive values). This was to make values of the PC-based indices represent environmental conditions similarly to those of the individual variables (i.e., small values for good conditions and large values for the opposite). Next, the averages of the 11 PC-based indices, calculated for all of the watersheds, were used as an integrated index for ranking and clustering purposes. Using 1.0 as the cut-off value for eigenvalues in the PCA, the first 11 PCs accounted for 80.49 percent of the total variation and were used to calculate the 11 PC-based indices mentioned above. The septile map of the PCA-based integrated index is displayed in Figure 6.



**Figure 6.** The PCA results displayed in equal-interval septiles. Values are principal component-based indices that reflect good (green) to bad (red) conditions.

### Advantages & Disadvantages

The main advantage of *PCA* is the replacement of a set of multivariate data with a new set of uncorrelated variables. By explicitly accounting for the correlation structure of the data sets, this method avoids a major disadvantage of the first two integration methods. However, this advantage comes with a price that it is often difficult to interpret environmental meanings of the new set of uncorrelated variables. Thus, while it will display a map of the region that is not distorted by correlations, it is more difficult to analyze the pattern in terms of the original variables. The main disadvantage of the *PCA* method is that it might be influenced by data abnormalities (e.g., non-normal distribution, discontinuities). Data abnormalities might have impacts to different extents on the correlation matrix, and subsequently, on the loadings of variables on various principal components. Details on possible impacts of data abnormalities on *PCA* and their treatment can be found in Jobson (1992) and Rencher (1995).

### Recommendations

The *Simple Sum* and the *PCA* methods should be used together because they are sensitive to different data problems. Where the methods agree on the spatial patterning of environmental quality, the results can reasonably be assumed to be free from data peculiarities.

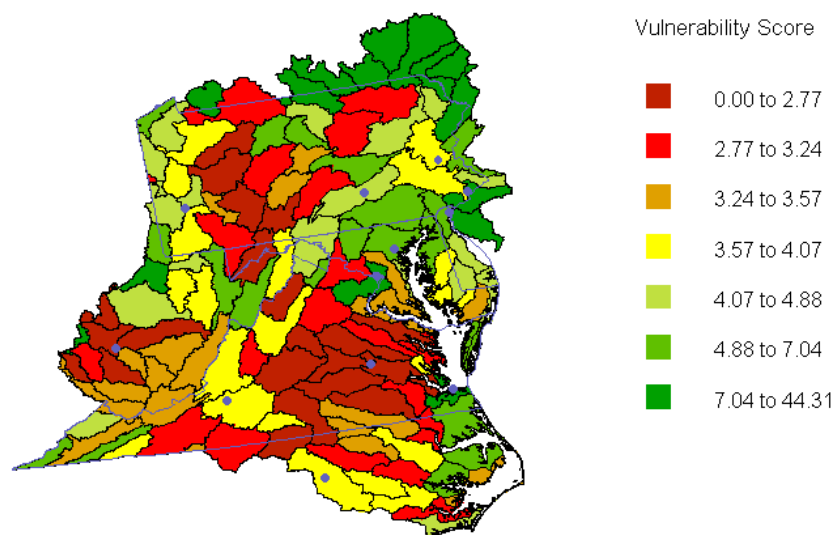
## State Space Analysis

Accurate assessment of the environmental quality and vulnerability of watersheds requires that we make maximum use of all of the available variables. A simple integration method, such as the *Simple Sum*, makes use of all the information available, but it takes no account of the correlations among the variables. A multivariate approach accounts for the correlation substructure of the variables, but it reduces the dimensionality of the analysis. Therefore, we considered a *State Space* approach (Johnson 1988) that calculates the Mahalanobis distance (Mahalanobis 1936). This approach uses all of the variables in calculating the distance between two watersheds, but it corrects the distances to account for covariances.

The objective of this approach is to determine the distance of each watershed from the most vulnerable watershed in the region. This would provide us with a different spatial pattern that might reveal new nuances about the regional environment. It remains a challenge to determine an objective means to specify the most vulnerable watershed. The concept is to choose a watershed that still has a reasonable amount of valued resources but is already under some stress.

To develop and test the method, it is sufficient to choose a test watershed that lies somewhere near the center of the distribution of variables, with both remaining resources and stressors acting on those resources. The coded variables range from 0 to 1. Therefore, a watershed with an average variable value near 0.5 would have both moderate resources and moderate stressors. Because the variables are not equally distributed between 0 and 1, the average is closer to 0.4. Using this arbitrary criterion, a watershed was chosen to serve as the surrogate “most vulnerable” watershed: Appomattox.

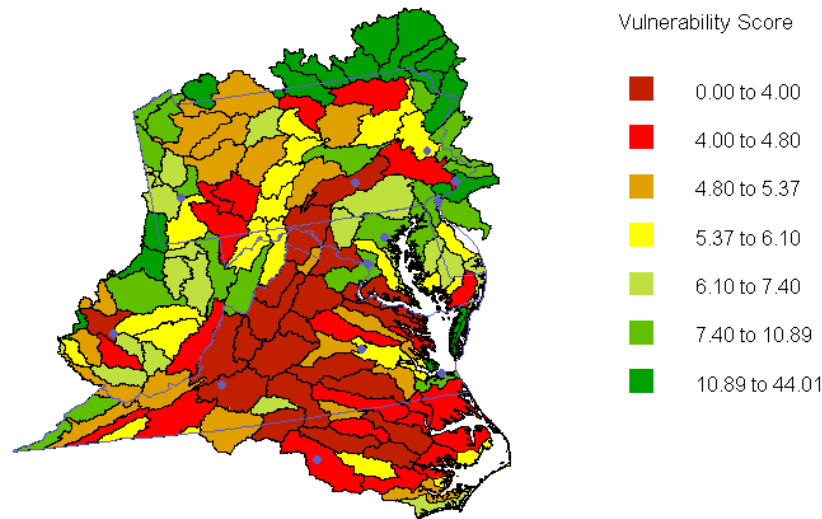
The resulting map is shown in Figure 7. The calculated distances were divided into septiles with dark red indicating the watershed closest to the “most vulnerable” watershed. The vulnerable watersheds are found on the coastal plain of Virginia, central Pennsylvania, western West Virginia, and along the North Carolina border. This pattern is different from those shown by other methods. It remains to be determined if the pattern is meaningful.



**Figure 7.** *State Space Analysis* results displayed in equal-size septiles with Appomattox as the “most vulnerable” watershed (average normalized variable value of 0.4). Watersheds with the darkest red colors are closest to this watershed in multivariate-state space; those in green are furthest away.

## Tests

The major weakness involved in this method is the choice of “most vulnerable” watershed. To test the sensitivity of the approach to this assumption, we chose a second watershed. In this case, we systematically went through the data set, eliminating watersheds with the lowest resources combined with the highest stressors. The logic was that these watersheds were already heavily impacted and had few resources left to be vulnerable to further stress. We also discarded watersheds with the highest resources combined with the lowest stressors. The logic was that these watersheds were relatively pristine and inaccessible. Therefore, they were less likely to be subjected to development. This systematic approach produced a second, test watershed, Lower York Shenandoah. The map is shown in Figure 8.



**Figure 8. *State Space Analysis* results displayed in equal-size septiles with Lower York Shenandoah as the “most vulnerable” watershed. Use of this watershed as reference reflects the removal of watersheds with few resources and high stress and high resources and low stress. Watersheds with dark red color-coding are closest to this watershed, while those in green are furthest away.**

The spatial pattern on Figure 8 is similar to Figure 7 with vulnerable watersheds closer to the coast. Most of the watersheds that have shifted only changed by one septile. Nevertheless, the test shows the approach is sensitive to the choice of “most vulnerable” watershed.

## Advantages & Disadvantages

The *State Space* method has the advantage of maintaining the full dimensionality of the regional database while modifying the calculation of distance in a manner that accounts for the covariance substructure of the data set. This is a property that we should take advantage of.

Calculating how far a given watershed is from the most vulnerable watershed in the region is a unique approach. The spatial pattern produced is quite different from the other forms of analysis in the regional assessment. If the results are meaningful, then it opens the possibility of another way to look at the environmental quality of the region.

The primary disadvantage of the *State Space* approach is not the method itself but the choice of the “most vulnerable” watershed. Presently, there is no objective definition of this reference point and the test showed the method to be sensitive to the reference. The sensitivity is likely to be due to the skewed distribution of many of the variables. Generally, the variables are not normally distributed over the range



from 0 to 1. Many variables have distributions weighted heavily toward 0 with a few outliers at the upper range. As a result, choosing a reference toward the center of the range may not be appropriate.

### Recommendations

The *State Space* approach should be retained as one of the ReVA integration methods. Some research may be needed to determine its sensitivity as a function of variable distributions. Nonetheless, the approach has advantages such as avoiding problems that occur with other methods for integrated assessment.

The challenge appears to be to choose an objective reference point. If the reference point is both objective and relevant, then the sensitivity problem is eliminated because the reference point is no longer an assumption. It may be possible, for example, to define a sufficient number of thresholds so that the Mahalanobis distance above or below the threshold references would be of interest.

### **Criticality Analysis**

*Criticality Analysis* calculates the Euclidean distance between the vector of variable values representing current conditions of each watershed and a vector representing a hypothetical “natural” state. This natural state is an attempt to reconstruct the set of conditions under which the ecological system evolved.

During the period of natural selection prior to human disturbance, ecological systems evolved a complex infrastructure of feedbacks that permitted recovery from natural disturbances and maintenance of relative stability. The further a system is moved from its natural state, the greater the probability that the system will be unable to respond stably to natural disturbances and normal variations in environmental conditions. Strangely missing from the assessment literature is a consideration of natural catastrophes that should be considered in the very definition of risk assessment (Suter 1993b) and can precipitate irreversible changes in ecological systems weakened by human stressors.

Nonlinear change is well documented in the ecological literature, including the fossil record (Crowley and North 1988, McGhee 1990). Sudden catastrophic changes have occurred in lake eutrophication, (Rosenzweig 1971), desertification (Schlesinger et al. 1990), forest pest outbreaks (Berryman et al. 1984), and fisheries collapses (Jones and Walters 1976). Nonlinear changes occur in soil forming processes (Phillips 1993), aquatic (Dubois 1979, Hughes 1994) and terrestrial systems (Gatto and Renaldi 1987) and in linked ecological-economic systems (Tainter 1988, Rosser 1991, Rosser et al. 1994).

We have known for a long time that even simple ecological systems can undergo catastrophic change (May 1977; O'Neill et al. 1982, 1989; Schaeffer and Kot 1986; Loehle 1989). We know that ecological systems are nonequilibrium (Kay 1991, Levin 1999, O'Neill 2001), can exist in multiple states (Peterman et al. 1979, Sutherland 1974), and that following a disturbance the system may not recover to the same state (Holling 1973, 1986; O'Neill 1999). Predictions of the effects of climate change on vegetation distributions (Bachelet et al. 2001, Iverson and Prasad 2001) show that even small incremental changes in environment can precipitate large spatial changes in the biotic system. We know that this type of radical change can occur in biological systems ranging from populations (Spromberg et al. 1998) to whole landscapes (Ingegnoli 1990). The challenge is estimating how tightly strung the rubber band is and the risk that it will snap with the next minor impact (Casti 1982).

ReVA attempts to estimate the relative risk of catastrophic nonlinear change. It is not even theoretically possible to predict the exact position of the bifurcation (i.e., the critical threshold beyond

which the ecological system will move to a new and undesired state). Therefore, we will focus on estimating how far ecological systems have moved from their natural state.

This approach assumes that systems in the natural state retain the feedback networks that permitted stable response to disturbances over the long period of evolutionary history. As human activities add stressors (e.g., chemical pollutants), extract resources (e.g., lumber), and change land cover (e.g., fragmentation), the natural feedbacks are disrupted and the system becomes more vulnerable to radical and potentially irreversible change.

## Methodology

The first step in applying the Criticality approach is to define the hypothetical “natural” state. The task is simple for some variables, e.g., human population and pollutants, which can be assumed to have been zero. The task is more arbitrary for other variables such as biodiversity. In addition, it cannot be assumed that biotic variables can be represented by a single value. The watersheds in the Mid-Atlantic Region range from highland forests in the Appalachians through the Ridge and Valley Province to the Coastal Plains. Even under pre-human conditions, it cannot be assumed that this diversity of systems was characterized by a single set of biotic variables.

To deal with the uncertainties involved in defining the natural state, the *Criticality Analysis* will be based on “fuzzy” values. A fuzzy value is expressed not as a single number but as a range of possible values plus an assumed distribution. The range of values is selected as the lowest and highest values that can be reasonably expected to have existed in the natural state. A triangular distribution is assumed if the most reasonable value would be expected to lie toward the center of this range. A flat or rectangular distribution is assumed if our ignorance only permits us to say that the value lies somewhere within the range.

For purposes of testing the method the following definition of the natural state was used:

- 1) Human population and activities were set equal to zero. This includes AGSL, CROPSL, DAMS, RIPAG, FUNGICIDE, HERBICIDE, INSECTICIDE, STRD, NTCMPLM, HARDCHIPMIL, SOFTCHIPMIL, EMMINE, IMPLCPCT, NBLDPM97, POPGROWTH, POPDENS, RDDENS, UINDEX, C5FS, EMAGRIC, INDTHPTH, and POV65.
- 2) Some pollutants are simply higher than normal values for natural phenomena (DISSOLVEDP, NO3DEPMODEL, OZONE8HR, SEDIMENT, SO4DEPMODEL, SUMO6, TOTALN, and UVB). For these we assumed a range from 0 to the upper bound of the second lowest quintile and a triangular distribution. In this and following definitions, the range of values for the variable (from the smallest value found on any watershed within the region to the largest value found on any watershed within the region) was divided into five intervals to define the quintiles.
- 3) Forest edge (EDGE2, EDGE65, and EDGE600) was assumed to range from 0 to the upper bound of the lowest quintile with a rectangular distribution. Forest interior (INT2, INT65, and INT600) as well as riparian forests (RIPFOR) and wetlands WETLANDSPECT) were assumed to range from the lower to the upper bound of the highest quintile.
- 4) Forest damage (FORDEFOL and FORMORT) is a natural phenomenon although the current state of the system is probably more damaged than the natural state due, for example, to introduced pests. For these we used a range of 0 to the upper bound of the second lowest quintile and a triangular distribution.
- 5) It is difficult to determine what the forest inventories (HARDWOODINV, HARDWOODREM, SOFTWOODINV, and SOFTWOODREM) would have been under natural conditions with no

harvesting. We assumed a range from 0 to the present value on a watershed with a triangular distribution.

- 6) For forest cover (LLSLPINE, MAPLEBEECH, OAKGUMCYPRESS, OAKHICKORY, OAKPINE, and SPRUCEFIR) we assumed that the natural state lay somewhere between the current cover and 100 percent cover with a rectangular distribution. For forest types that are largely planted (LOPSLPINE and WRJACKP), we assumed a range from 0 to the present cover on a watershed with a rectangular distribution.
- 7) For soil loss potential, we assumed a range from 0 to the upper bound of the lowest quintile and triangular distribution.
- 8) Our estimates for acid neutralizing capacity were based on underlying bedrock and the “natural” state was defined as the present value unchanged.
- 9) We assumed that exotic and rare/threatened species were not present in the natural state and these variables were set to 0: AMPHEXOTIC, BIRDEXOTIC, FISHEXOTIC, MAMMALEXOTIC, MUSSELEXOTIC, REPTILEEXOTIC, AMPHTE, BIRDTE, FISHTE, MAMMALTE, MUSSELTE, and REPTILETE.
- 10) For native biodiversity, we assumed that our ignorance is profound but that diversity was probably higher than under present conditions. We assumed a range from the lower bound of the second highest quintile to the largest current value with a triangular distribution for AMPHNATIVE, BIRDNATIVE, MAMMALNATIVE, MUSSELNATIVE, and REPTILENATIVE.

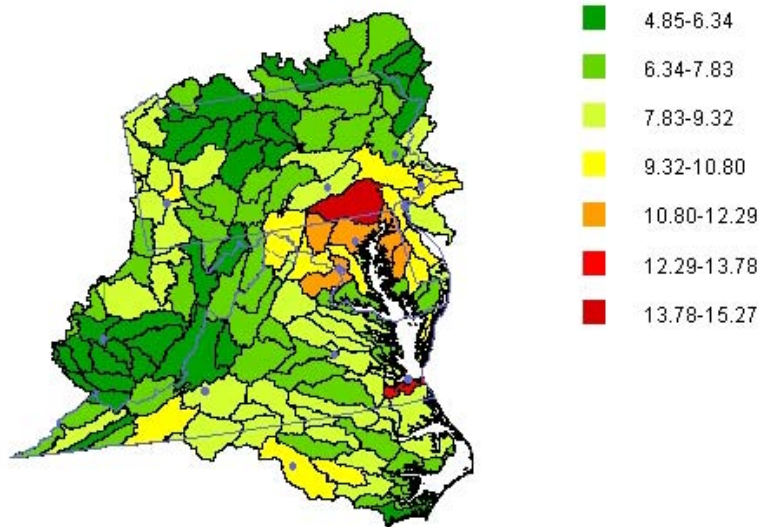
Once the definition of the “natural state” is established, it is possible to calculate a “fuzzy” distance between each watershed and the natural state. (See Appendix B; Tran and Duckstein 2002.)

### Preliminary Results

The relatively arbitrary definition of “natural state” given above produces the map shown in Figure 9. The map shows the regional pattern of vulnerability to passing through a critical threshold and moving to a new undesired state. Watersheds shown in green are closer to their natural state and less likely to experience radical changes due to natural processes and incremental stressors. Watersheds shown in yellow, orange and red are further from their natural state and more vulnerable, i.e., more likely to experience potentially irreversible changes.

In general, the map shows the typical pattern seen with other approaches. The greatest vulnerability is associated with watersheds subject to more intense human activity, particularly in the vicinity of the Philadelphia-Baltimore-Washington axis and Pittsburgh. The least vulnerable watersheds are in areas of high elevation and higher topographic gradients, i.e., less accessible to human activity.

Although the Criticality map shows the recognizable pattern, a comparison with the Simple Sum map (Figure 5) shows clear differences. In general, the *Criticality Analysis* provides a more conservative assessment. Because of our “fuzzy” definitions of the natural state, many (>20) more watersheds appear in the lowest two quintiles (darker shades of green). In other words, there are a number of watersheds that cannot be shown to be clearly vulnerable. Thus, the analysis appears to be biased in the direction of giving false negatives, i.e., possibly underestimating the risk of catastrophic change for a number of watersheds. This bias can be interpreted in either of two ways. First, it makes a strong case that the watersheds shown in yellow/orange/red are indeed vulnerable. Second, the results should not be interpreted as meaning that the watersheds shown in green are “safe” for further development.



**Figure 9. Map shows *Criticality Analysis* results. Watersheds in red are the furthest distance in multivariate space from natural conditions as specified in this example, while watersheds in green are the closest.**

A comparison between the Criticality map and the PCA/SUM map (Figure 6) shows similar results. The PCA/SUM method calculates the distance of a watershed from the best value of each variable found on any watershed in the region. So the similarity in the maps can be simply explained by the observation that the highest values in the region are close to our definition of the “natural” state. This is not surprising, given the relatively undisturbed nature of the Appalachian watersheds. However, the similarity in the results means that a decision will need to be made about whether the Criticality approach provides anything different from the PCA or if they should be combined.

### Testing the Sensitivity of the Method to Assumptions

The important assumption in the *Criticality Analysis* is the definition of the “natural” state. To test the sensitivity of the results to this assumption, we developed an alternative definition of the natural state. We chose to devise an even more conservative approach – accepting broader ranges that perhaps better represent our ignorance of the natural state. The alternative definition is:

- 1) Human population and activities remained equal to zero.
- 2) For naturally occurring pollutants (DISSOLVEDP, NO3DEPMODEL, OZONE8HR, SEDIMENT, SO4DEPMODEL, SUMO6, TOTALN, and UVB), we continued to assume a range from 0 to the upper bound of the second lowest quintile but used a more conservative rectangular distribution.
- 3) Forest edge was assumed to range from 0 to the upper bound of the second lowest quintile with a rectangular distribution. Forest interior, riparian forest, and wetlands were assumed to range from the lower to the upper bound of the second highest quintile with a triangular distribution.
- 4) Forest damage (FORDEFOL and FORMORT) continued to assume a range of 0 to the upper bound of the second lowest quintile but we chose a more conservative rectangular distribution.
- 5) Forest inventories (HARDWOODINV, HARDWOODREM, SOFTWOODINV, and SOFTWOODREM) were assumed to have been at their present values.
- 6) Forest cover variables were assumed to be at their present values.

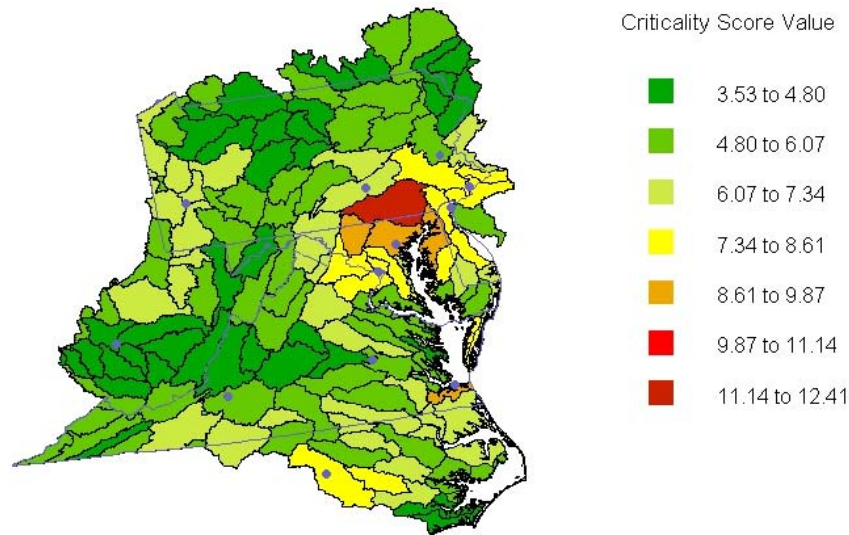
- 7) For soil loss potential we assumed a range from 0 to the upper bound of the second lowest quintile with a rectangular distribution.
- 8) We continued to assume that acid-neutralizing capacity was at the present value.
- 9) Since species invasion was a possible phenomenon in the natural state we assumed that exotic species and threatened/endangered species ranged from 0 to 1 with a rectangular distribution.
- 10) For native biodiversity, we assumed the same range from the lower bound of the second highest quintile to the largest current value but used a more conservative rectangular distribution.

## Test

The critical assumption in the *Criticality Analysis* is the definition of the “natural” state. To test the sensitivity of the results to this assumption, we developed an alternative definition of the natural state. We chose to devise an even more conservative approach – accepting broader ranges that perhaps better represent our ignorance of the natural state. The alternative definition is:

- 11) Human population and activities remained equal to zero.
- 12) For naturally occurring pollutants, we assumed the same range but used a more conservative rectangular distribution.
- 13) For forest edge, we increased the range to include the lowest two quintiles. For forest interior, riparian forest, and wetlands, we increased the range to include the highest two quintiles.
- 14) Forest defoliation was assumed to have a more conservative rectangular distribution.
- 15) For forest inventories, we assumed a more conservative rectangular distribution.
- 16) Our original assumptions for the forest cover variables were already conservative and were unchanged for the test.
- 17) For soil loss potential, we included the lowest two quintiles and a rectangular distribution.
- 18) Because species invasions were possible without human intervention, we assumed that exotic species and threatened/endangered species ranged from 0 to the present value on a watershed with a rectangular distribution.
- 19) For native biodiversity, we assumed the same range but used a more conservative rectangular distribution.

A comparison of the Criticality map using the original (see Figure 9) and alternative definitions (see Figure 10) shows the maps are very similar, with a few watersheds categorized one septile less vulnerable (green) in the alternative. This is the expected result of choosing more conservative values. Vulnerable watersheds remained vulnerable and intermediate watersheds were categorized as less vulnerable.



**Figure 10. Results of the *Criticality Analysis* done with a second set of reference conditions that might be considered more conservative. Watersheds that are the furthest distance from reference conditions are indicated in red.**

Given these results, the analysis appears relatively robust to the definition of the natural state. This insensitivity to the assumptions is most likely due to the fact that however one changes the assumptions about the biotic variables, setting human activities to zero is implicit in the definition of the natural state. Further, the values chosen for the biotic and land use variables are constrained because the natural state was less impacted than at present. Therefore, no matter what specific assumptions appeal to the assessor, the values are constrained. As a result, the general pattern of vulnerability (see Figure 9) is insensitive to the details of the definition of the natural state.

### Advantages & Disadvantages

The greatest strength of the *Criticality* approach is that it provides a unique perspective on watersheds in the Region. The phenomenon of catastrophic change in complex adaptive systems is a potentially important concept in large-scale assessment. The technique outlined above may provide a reasonable approach for bringing this concept to bear on assessment problems. Most importantly, the *Criticality* approach provides some perspective on the potentially irreversible impacts of further development on vulnerable watersheds.

Another strength of the *Criticality* approach is its apparent insensitivity to the assumptions involved in defining a “natural” state. By allowing the assessor to include uncertainty about the specifics of the natural state into the fuzzy definition of variables, a major impediment is overcome. In particular, the most vulnerable watersheds appear to emerge no matter how uncertain we are about the details of the pre-human system.

The calculation of criticality does not appear to be dependant on the distributions of the individual variables. As the results show, however, the distribution of calculated criticality values across watersheds is highly skewed, due in part, to the incorporation of uncertainty in the analysis through the use of fuzzy distance calculations.

The greatest weakness of the *Criticality* approach is our inability to predict where the threshold of criticality is. Although it appears reasonable to estimate relative risk, it is not possible to pinpoint exactly which watersheds will undergo radical change given a natural disturbance or further development. Thus, we cannot assess which watersheds are free of significant risk. Simply because a watershed is closer to

its natural state does not mean that it is “safe.” Because the analysis incorporates a large amount of uncertainty in our estimates of the natural conditions, it is relatively conservative. Therefore, the distribution of calculated criticality values are highly skewed with more watersheds classified as “not so vulnerable,” and a small number of watersheds classified as highly vulnerable (see Figure 9).

## Recommendations

*Criticality Analysis* is unique among the integration techniques presented here. It is the only method that compares the condition of a watershed to its assumed natural state. This comparison provides us with a better understanding of the vulnerability of a watershed than many of the other methods. Most of the other techniques rank the watersheds based on the condition of the “best” or “worst” watershed. This type of ranking is not very useful for determining the degree that a watershed has moved from its natural state toward a critical threshold beyond which irreversible change can occur. *Criticality* gives us some insight into how vulnerable a watershed is likely to be to catastrophic change in its ecological systems.

Despite the major advantage of the technique for addressing the issue of vulnerability, the method is biased toward identifying more watersheds that are closer to a natural state. Because we cannot identify the critical thresholds beyond which catastrophic changes may occur, we are unable to say with certainty that these watersheds are “safe,” i.e., not vulnerable. Therefore, it is not possible to make a quantitative statement about just how dangerous it is to make an incremental change in an individual watershed. The applicability of the approach in a decision-making context depends heavily on the ability of the decision-maker to evaluate and weigh the risk against the benefits involved in the decision. The only important contribution of this analysis is to inform the decision-makers that further changes on a vulnerable watershed might precipitate irreversible changes.

Because the approach is most robust for the most disturbed sites (the watersheds furthest from the natural state tend to remain unchanged as the assumptions about the natural state are changed), we suggest that the interpretation of the *Criticality Analysis* should focus on these sites. Instead of ranking all sites based on criticality, it might be wise to highlight only the most vulnerable sites. By showing only the worst watersheds, we indicate the watersheds that are at greatest risk of catastrophic change and our argument for robustness is strong. An objective approach for selecting the most vulnerable watersheds would be to construct a histogram of fuzzy distances. It might then be possible to argue that there is a breakpoint in the histogram and map only the watersheds above this breakpoint.

The *Criticality Analysis* could be improved by strengthening the definition of the natural state. Even though the analysis is relatively robust, the verisimilitude of the results is dependent on how well we can establish the natural state definitions.

## **Analytic Hierarchy Process**

Developed by Saaty in 1980 (Saaty 1980), Analytical Hierarchy Process (*AHP*) is one of the most widely used multi-criteria decision-making methods. One of the reasons for *AHP*’s popularity is that it derives preference information from the decision-makers in a manner that they find easy to understand. Beside the original version of Saaty, there have been several variants of *AHP* seen in the decision-making science literature. For instance, Lootsma (1997, 1999) modified the scale and aggregation procedure in the original *AHP* to come up with the Additive *AHP* and Multiplicative *AHP*. The *AHP*’s original version as well as its two variants developed by Lootsma has been altered to deal with fuzzy numbers (see Saaty (1977, 1978), Chen and Hwang (1992) for the original model, and Lootsma (1997, 1999) for the modified versions). *AHP* has been applied widely in different environmental problems (e.g., Saaty 1986, Lewis and Levy 1989, Varis 1989), especially in resources allocation and planning (e.g., Ramanathan and Ganesh 1995, Mummolo 1996, Alphonse 1997).

AHP is a systematic procedure to construct and represent the elements of a problem in a hierarchy. The basic rationale of *AHP* is organized by the breakdown of the problem into smaller constituent parts at different levels. Decision-makers are guided through a series of pair-wise comparison judgments to reveal the relative impacts, or the priorities of elements (e.g., criteria, alternatives) in the hierarchy. These judgments in turn are transformed to ratio-scale numbers representing relative local and global weights of the elements at a certain level of the hierarchy. The hierarchy in AHP is often constructed from the top (goal from management standpoint, e.g., environmentally-sound development), through intermediate levels (criteria on which subsequent levels depend, e.g., physical, chemical, biological, and socioeconomic criteria) to the lowest level (usually a set of alternatives, possible actions).

To perform a regional environmental assessment, we designed a four-level hierarchy *AHP* model displayed in Figure 11. The lowest level (the fourth level) is for the watersheds to be assessed and/or compared. The third level consists of all of the indicators used in the assessment. The second level has eight components representing eight groups of indicators at the third level (Table 3).

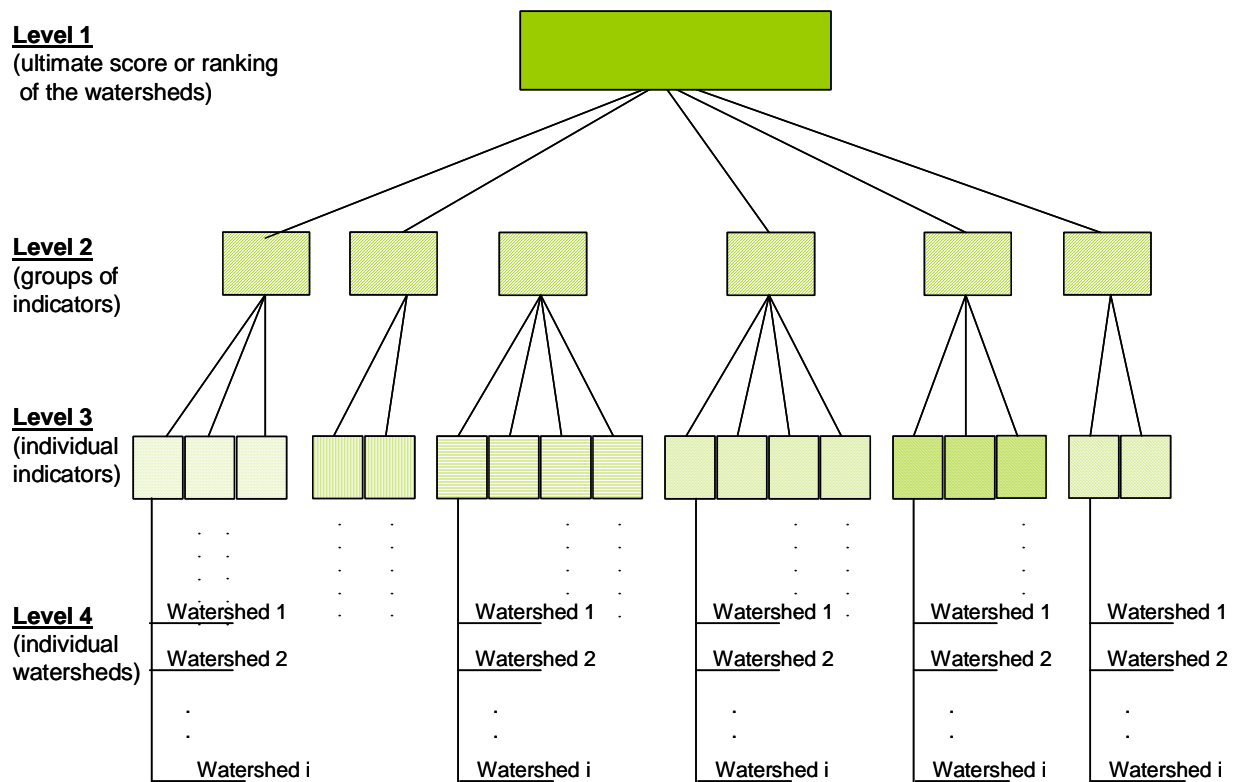


Figure 11. Diagram of the hierarchies in the *AHP* model for regional environmental assessment.



**Table 3. Groups of variables for the *AHP* Method.**

<b>Groups</b>	<b>Variables</b>
Aquatic Resources	AQUANATIVE, AQUATE, and PSOIL
Aquatic Stressors	AGSL, AQUAEXOTIC, CROPSL, DAMS, DISSOLVEDP, IMPLCPCT, NO3DEPMODEL, NTCMPPLM, RIPAG, SO4DEPMODEL, STRD, and TOTALN
Economics - Extractive Industries	EMAGRIC, EMMINE, HARDCHIPMIL, HARDWOODINV, HARDWOODREM, SOFTCHIPMIL, SOFTWOODINV, and SOFTWOODREM
Human Health - Sensitive Populations	C5FS, INDTHPTH, and POV65
Human Health - Stressors	OZONE8HR and UVB
Population Pressure	NBLDPM97, POPDENS, and POPGROWTH
Terrestrial Resources	EDGE2, EDGE65, INT2, INT65, MIGSCENARIO, NATCOVERPCT, NONCLIMAXPCT, RIPFOR, TERRNATIVE, TERRTE, and WETLNDSPCT
Terrestrial Stressors	FORCOVDEFOL, FUNGICIDE, HERBICIDE, INSECTICIDE, RDDENS, SUM06, TERREXOTIC, and UINDEX

Somewhat different from common *AHP* applications, we used measurement rather than pair-wise comparison at the lowest level of the hierarchy. Our aim is to rate the watersheds on a single-indicator basis. As measurement involves a measuring standard, the watersheds are rated against some reference points, namely some ideal and undesirable ecological states (conditions) of the indicator under study. In this analysis, we simply construct the ideal and undesirable states for a particular indicator by using its minimum and maximum values derived from the indicators' data from all of the watersheds. Then those single-indicator-based distances of a watershed are aggregated gradually from the bottom to the top of the hierarchy to come up with an ultimate score for that watershed. Conceptually, the ultimate score of a watershed represents the distance of the watershed to an arbitrary ideal watershed that has the ideal states for all of the indicators. Next, all of the ultimate scores are used to derive a relative ranking for the watersheds.

Note that grouping of indicators in the third level into groups at the second level can be done in various ways. For example, we can divide indicators into groups based on some subjective judgment/classification as presented above. Alternatively, we can group indicators in an objective manner with the use of multivariate statistical analysis (e.g., principal component analysis) (Tran et al. 2003). Furthermore, we can combine subjective judgment and objective analysis complementarily in grouping indicators.

Normally in *AHP*, the next step after constructing the hierarchy is to carry out pair-wise comparison judgments at different levels of the hierarchy to reveal the criteria's relative weights. However, to create the baseline model, we assigned equal weights for the eight components at the second level (i.e., equal local weights of 0.125), implying they were treated equally. In the same manner, weights at the third level for indicators in the same level-two component are equally assigned. When the model is used for a real ecological assessment with actual decision-maker(s) and stakeholder(s), pair-wise comparisons should be carried out thoroughly, following the common procedure of *AHP*, to determine the criteria's relative weights at all levels in the hierarchy (except the lowest level). Therefore, those potential real-world applications probably will have different sets of weights and consequently have different sets of

ranking which in turn might not be the same as those in this baseline analysis. Those differences reflect divergence in public values, preferences, and priorities of different decision-makers and stakeholders.

Commonly in *AHP*, local weights determined from pair-wise comparison are synthesized from the second level down to the very last level to derive the global weights for all criteria. However, we suggest the local weights to be used to compute scores of the watersheds at each criterion in the hierarchy. The reason is to make the scores computed at all criteria in all levels of the hierarchy be on the same 0-to-1 scale, conceptually representing the distances from the watersheds to the ideal states of the corresponding criteria. Note that the conversion between scores computed by local weights with those by global weights is trivial.

Within the hierarchical structure, the normalized values of watersheds regarding various indicators at the fourth level are aggregated to produce combined scores at other higher levels of the hierarchy. Scores at the third level are computed by two different methods:  $L_1$  norm (sum of the scores) and  $L_2$  norm (square root of sum of the squared scores) as follows:

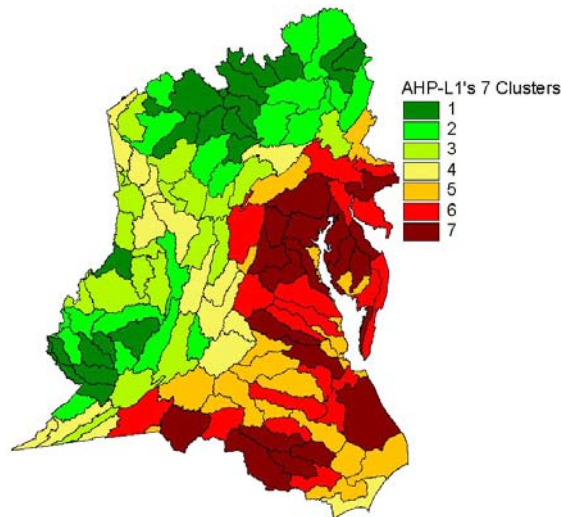
$$L_1: \quad D_{level\_j}^{criterion\_i} = \sum_{k=1}^m w_k \cdot D_{level\_j+1}^{criterion\_k}$$

$$L_2: \quad D_{level\_j}^{criterion\_i} = \sqrt{\sum_{k=1}^m w_k \cdot (D_{level\_j+1}^{criterion\_k})^2}$$

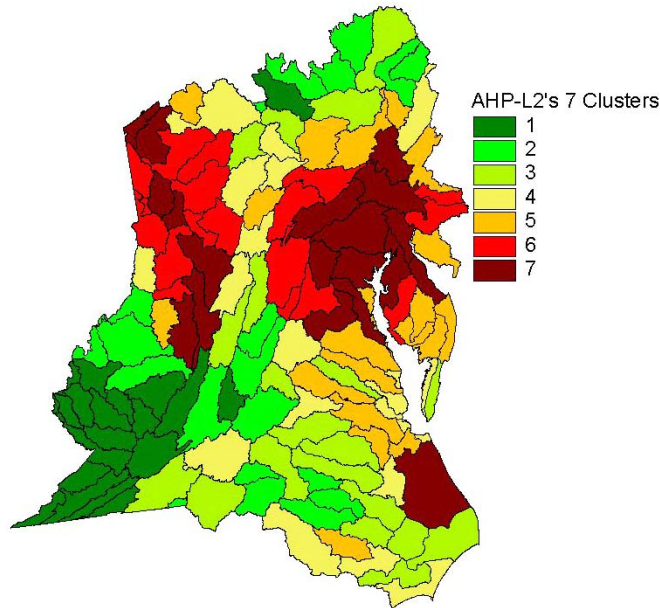
where  $D_{level\_j}^{criterion\_i}$  is the score at criterion  $i$  in the level  $j$ ;  $w_k$  is the local weight of criterion  $k$  in the level  $j+1$ ; and  $m$  is the number of indicators (criteria) in the level  $j+1$  associated with criterion  $i$ . Scores at the second and first levels are computed by the  $L_1$  norm only.

## Results

The ultimate scores for all of the watersheds and their rankings, derived from the two different methods (so-called AHP- $L_1$  and AHP- $L_2$ ), in turn are grouped into seven groups ranked from 1 (good condition) to 7 (bad condition) (see Figures 12 and 13).



**Figure 12. AHP- $L_1$  (sum of AHP scores) results displayed in equal-size septiles, where watersheds the furthest distance from ideal are shown in dark red.**



**Figure 13. AHP-L<sub>2</sub> (square root of the sum of the squared AHP scores) results displayed in equal-size septeles, where watersheds the furthest distance from ideal are shown in dark red.**

Some spatial patterns are revealed from AHP-based maps. In general, watersheds located near urban centers (e.g., Philadelphia; Washington, DC; Pittsburgh) have relative high ultimate impact scores (i.e., bad condition). There are several adjacent watersheds in the southwestern part of the study area (i.e., West Virginia) that were in good condition in comparison with the others in the region. However, there is no simple spatial transition from the bad watersheds to the good ones. Watersheds in one cluster are not clearly spatially contiguous, but rather, intermingled with watersheds in other clusters throughout the study area, showing that some relatively good watersheds are located right next to some bad watersheds. It is obvious that watersheds are not independent but rather interdependent in terms of ecological impacts. What happens in one watershed might have impacts on its neighboring watersheds to a certain extent. For example, a new transportation line is likely to cause some impacts (e.g., air pollution, changes in stream flow and sedimentation, etc.) not only on the watersheds that it goes through but also on watersheds nearby. Hence, even there are no direct risks within their boundaries; good watersheds are not completely safe from degradation due to interrelated impacts among all of the watersheds.

### Tests

Results of the two models AHP-L<sub>1</sub> and AHP-L<sub>2</sub> were different significantly from one to another. While AHP-L<sub>1</sub> showed patterns somewhat similar to *ecoregion* settings in the study area, AHP-L<sub>2</sub> revealed more rural-versus-urban arrangement. Note that the two models were the same in terms of structure and weights at every level except the fact that different norms (L<sub>1</sub> versus L<sub>2</sub>) were used at the lowest level of the hierarchy in each model. It suggested that distance measure (e.g., Euclidean and Mahalanobis distances) could play a role in shaping the classification/ranking pattern.

### Advantages & Disadvantages

The use of AHP for regional environmental assessment has several advantages. First, it helped to organize a complex problem into a well-structured hierarchy. AHP not only offers the ultimate scores at the highest level, it also provides the impact scores regarding a specific criterion at any level in the hierarchy, allowing a comprehensive multi-level assessment.

Second, AHP can deal with various types of data and information (e.g., qualitative and quantitative data, expert judgment) in a single framework without requiring data transformation. It is because AHP uses a ratio 1-9 scale to compare one variable/criterion/object to another. AHP uses different means to carry out the pair-wise comparison. Some of them include questionnaire, graphic, verbal, and matrix.

Third, AHP allows judgment from various experts to be included and integrated in a single model. This aspect is very attractive because integrated environmental assessment is often a process involving multiple stakeholders and/or decision-makers.

The main disadvantage of AHP is that the amount of pair-wise comparison might be enormous if there are many criteria and/or alternatives to be compared in the model. At one level, we were able to use measurements to avoid the problem. But the decision-maker would still need to pair-wise compare the groups of variables and determine relative ratings. So the disadvantage remains.

A second disadvantage of AHP is its sensitivity to the method of calculating the ultimate score. Some objective criterion must be developed to determine which method of calculating will be used in future assessments.

Another disadvantage of our application is that our variables are not independent, as assumed in AHP. Several indicators in the same component at the second level describe more or less the same aspect of the ecosystem. As a consequence, it is likely that change in one indicator will accompany changes in other indicators in the same component. At this point, one might question why highly correlated indicators are not eliminated from the analysis, in general, and the AHP, in particular. Arguably, although some indicators are highly correlated, they still have their own signatures that are distinct from those of others to some extent. For example, although both of the indicators INT65 and INT2 – interior habitat in 65-ha and 2-ha windows, respectively – describe the same type of forest habitat condition, they are derived at two different scales, representing the picture of forest interior habitat from two different angles. Hence, choosing one while eliminating the other is not an easy decision to make. A better approach would be to use both of them but have some appropriate way to cope with the codependence problem. For example, the codependence problem among indicators can be solved by a careful weighting procedure, giving small weights to highly correlated indicators (e.g., inverse correlation coefficients as weights can be a feasible solution). However, equal weights at the second and third levels of the hierarchy as in this analysis is considered reasonable for a baseline model, when insights of the relationships among variables have not been verified by other careful judgments or analyses.

## Recommendations

Given the fact that the approach has real advantages in terms of ranking/classifying environmental conditions for integrated assessment in a user-friendly framework, the AHP approach should be used as one of the integration methods in ReVA. Some research may be needed to determine its appropriateness in dealing with the codependence problem (e.g., by an appropriate weighting scheme) or another more complicated model, the Analytical Network Process, might be needed.

## **Clustering Analysis**

*Clustering* is a very common approach used in a wide array of problems, including environmental studies. Its aim is to partition a data set into a set of clusters. However, “clustering” is a general term that embraces various approaches, such as crisp clustering, fuzzy clustering (Bezdek and Pal 1992), and mixture model-based clustering (McLahlan and Basford 1987). In this analysis, we focus only on: K-means cluster analysis and hierarchical cluster analysis. Although the general course of clustering is to maximize within-cluster similarity and/or between-cluster dissimilarity, various proximity measures (e.g.,

Euclidean, city-block, and Mahalanobis distances) and various distance criteria (within-cluster: average, nearest neighbor, and centroid distances; between-cluster: single, complete, average, and centroid linkages) exist, causing clustering results on the same data set to vary from one analysis to another. For illustration, we limit the scope of this analysis to the Euclidean distance measure and to a couple of common between-/within-cluster distance criteria only. A thorough discussion on proximity measures and clustering distance criteria can be found in various multivariate statistical textbooks, such as those of Jobson (1992) and Rencher (1995).

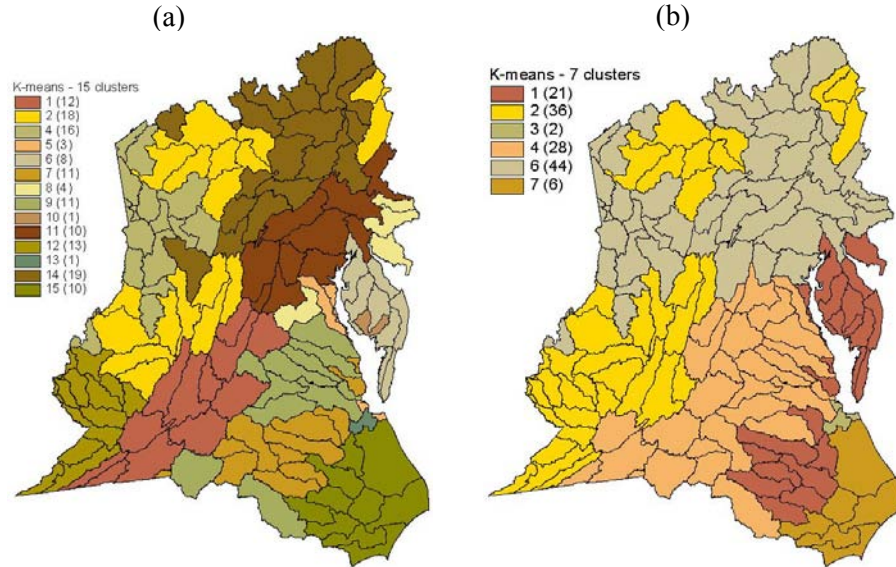
There are two main ways to cluster data: partitive and hierarchical approaches. K-means cluster analysis is a typical partitive clustering technique in which the data set is divided directly into a predefined number of clusters (e.g., the clustering process does not depend upon previously found clusters). This method implicitly assumes spherical shapes of the clusters. In the hierarchical clustering approach, the data set is organized into a hierarchical clustering tree (dendrogram) via either top-down (divisive) or bottom-up (agglomerative) algorithms. Between the two, agglomerative procedures are more commonly used than the divisive ones. The dendrogram can be used to study the data structure and to determine the number of clusters. With the dendrogram, it is guaranteed that a subcluster belongs completely to a larger cluster. This feature is not always true with the K-means clustering and other partitive approaches.

The “best” clustering (e.g., the number of clusters) among different clustering results can be selected by using some type of validity index such as those in Milligan and Cooper (1985) and Bezdek (1998). Some common validity indices include the Davies-Bouldin index (Davies and Bouldin 1979) and the average Silhouette width (Rousseeuw 1987). More on stopping rules and ways of finding out the “best” number of cluster can be found in McCune and Grace (2002).

In the ReVA context, clustering can be used to classify watersheds with similar environmental conditions into the same group (i.e., classification). However, clustering alone cannot tell which group of watershed is in good or bad condition or which group is better or worse than the others (i.e., ranking). To be able to do that, clustering needs to be combined with other ranking methods presented in this report (e.g., Simple Sum, PCA, AHP).

## Results

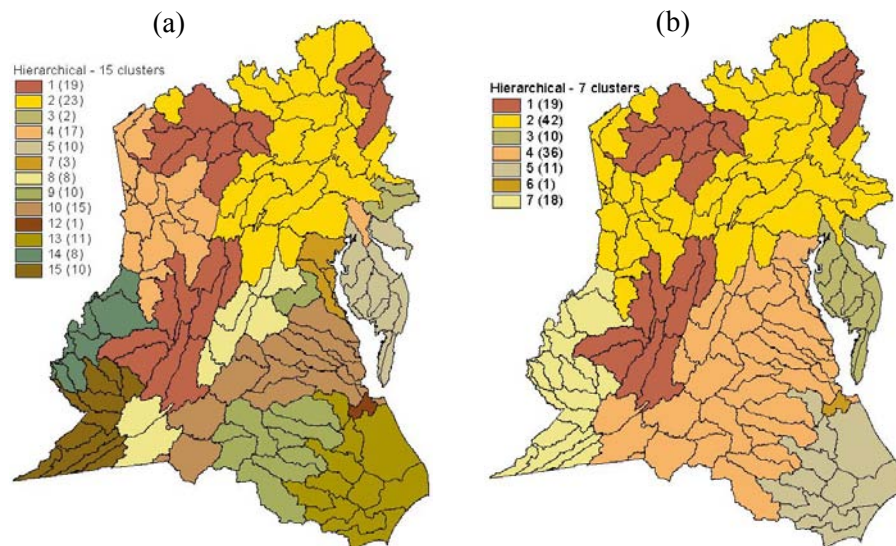
The results of the K-means clustering for two cases – 15 and 7 clusters – are shown on Figure 14. The spatial distribution of the K-means 15 clusters showed a combination of the urban-versus-rural patterns with the “ecoregion” settings of the Mid-Atlantic. However, the seven K-means clusters seemed too coarse to reveal any details. The number of watersheds associated with a cluster varied significantly from one cluster to another, especially in the case of the K-means 7 clusters. Some clusters had very few watersheds associated with them (e.g., clusters 13 and 3 of the K-means 15-cluster and 7-cluster analyses, respectively), showing that those watersheds were quite different from the others.



**Figure 14. Results of K-means *Clustering* analysis: (a) 15 clusters and (b) 7 clusters (numbers in parentheses are numbers of watersheds in each cluster). Watersheds with similar colors are in the same cluster.**

The results of the hierarchical clustering using the within-group average linkage for two cases – 15 and 7 clusters – are shown in Figure 15. Similar to the K-means clustering, the number of watersheds associated with a cluster in hierarchical clustering varied considerably from one cluster to another. Figures 14 and 15 show that the clustering patterns were very different from K-means clustering to hierarchical clustering.

Note that, using the average Silhouette width to compare various clustering results on the 50-variable data set, we could not determine the “best” clustering. It was because the differences of the Silhouette widths among different clusterings were insignificant. However, “best” clustering can be found in clusterings on some subsets of the data set (e.g., human health – stressors, terrestrial resources).

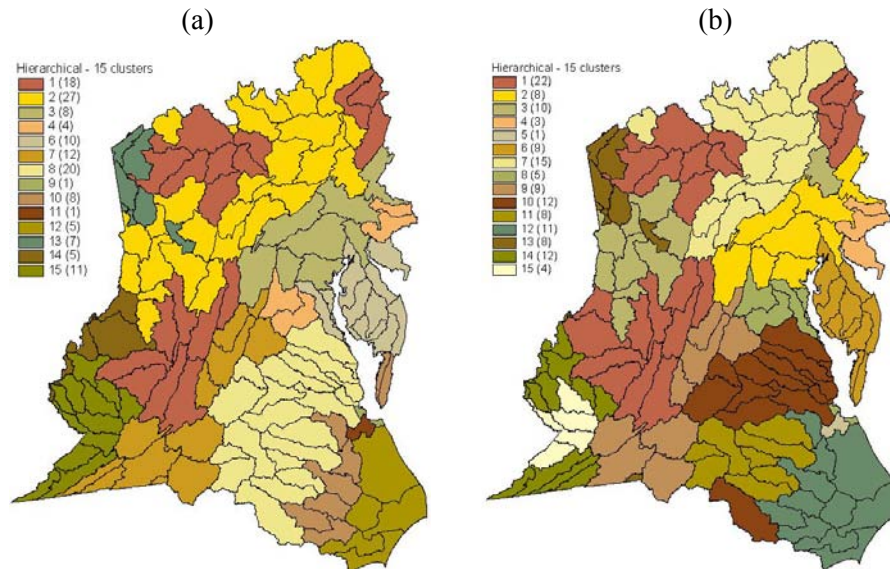


**Figure 15. Results of hierarchical *Clustering* analysis using within-group average linkage: (a) 15 clusters and (b) 7 clusters (numbers in parentheses are numbers of watersheds in each cluster). Watersheds with similar colors are in the same cluster.**



## Tests

To test the clustering approach, we carried out clustering analyses for a wide array of distance criteria. Some of the results are displayed in Figure 16. The main finding of these analyses was that the spatial clustering patterns were very diverse from one clustering analysis to another.



**Figure 16. Results of hierarchical *Clustering* analysis using (a) complete linkage and (b) Ward linkage (numbers in parentheses are numbers of watersheds in each cluster). Watersheds with similar colors are in the same cluster.**

## Advantages & Disadvantages

The *Clustering Analysis* has the advantage of being intuitively simple and easily communicated. It can be used to detect similarity and/or abnormality in environmental conditions. It makes no assumptions about the statistical distribution of the indicators. However, the *Clustering Analysis* may be influenced by the covariance structure of the data set, especially when the Euclidean distance is used.

The *Clustering Analysis* is only good at clustering and classification. It is impossible to do the ranking or vulnerability assessment with *Clustering Analysis* alone. However, it can be combined with other ranking and vulnerability integration methods in a complementary manner.

A crucial disadvantage of *Clustering Analysis*, which had been found from our analyses, is that the clustering patterns are quite diverse from one clustering technique to another, and from one distance criterion to another. It makes the task of interpreting the clustering results quite difficult.

## Recommendations

Given the fact that the approach has various advantages and disadvantages in terms of clustering/classification, the *Clustering Analysis* should be used in a complementary manner with other methods. The tests reveal that small changes in the methodology make significant changes in the spatial pattern of clusters. This sensitivity will make decision-makers wonder which approach is “correct.”

We recommend that further research be carried out with *Clustering* before it is used in regional assessments. It is possible, for example, to use the method for specific objectives, such as clustering

watersheds based on their potential for restoration. For a specific objective, it may be possible to specify the details of the methodology that are most appropriate for the purpose. It may also be possible to find that the clustering method is less sensitive for specific objectives.

## ***Self-organizing Map***

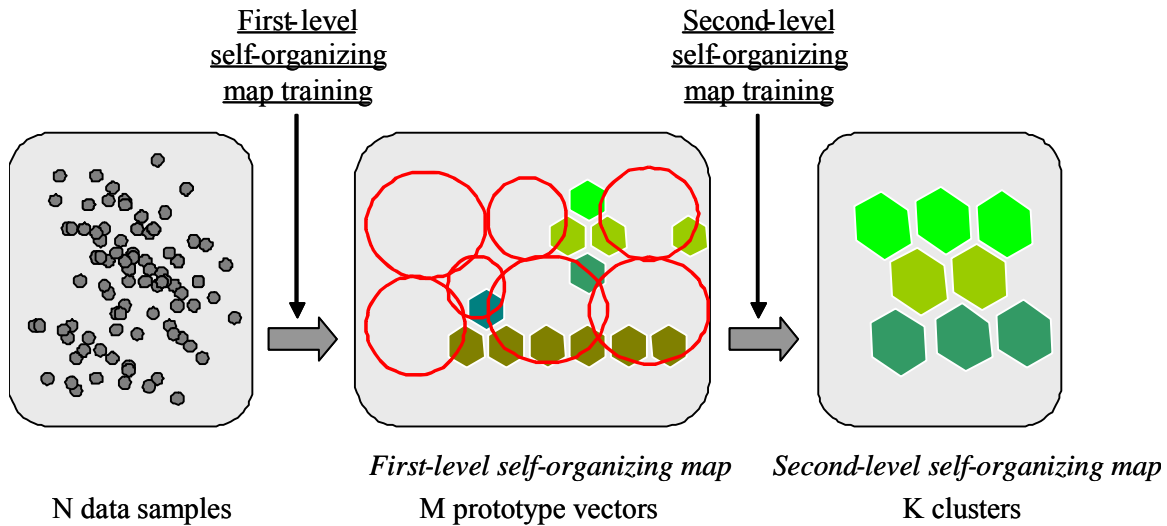
In a nutshell, the self-organizing map (*SOM*) is a clustering method with additional visualization capability to reveal the distribution of data under analysis. The self-organizing map (*SOM*) method is a neural network developed by Kohonen in the early 1980s (1982, 2001) (so-called Kohonen *SOM* hereafter). The Kohonen *SOM* is capable of learning from complex, multi-dimensional data without specification of what the outputs should be and generating a nonlinear classification of clusters. *SOM* has been applied in a wide array of classification problems. Kohonen (2001) identified more than 4,300 scientific papers related to the Kohonen *SOM* from 1981 to 2001. Recently, the Kohonen *SOM* has been used in several environmental studies including Trautmann and Denoeux (1995), Aguilera and others (2001), Brosse and others (2001), Cereghino and others (2001), Clare and Cohen (2001), Giraudel and Lek (2001), Obach and others (2001), and Walley and O'Connor (2001).

The Kohonen *SOMs* unsupervised learning algorithm involves a self-organizing process to identify the weight factors in the network, reflecting the main features of the input data as a whole. In that process, the input data is mapped onto a lower dimensional (usually two-dimensional) map of output nodes with little or no knowledge of the data structure being required. The output nodes, which associate with parametric-reference vectors having the same dimension as the input vectors (in this case, 50-dimensional space), represent groups of entities with similar properties, revealing possible clusters in the input data. Note that, although the Kohonen *SOM* is unsupervised in learning and determining the output nodes' parametric-reference vectors, the number of the output nodes and the output map's configuration need to be specified before the learning process. This aspect is similar to K-means clustering where the number of clusters is chosen beforehand.

The algorithm used in this analysis is the SOMPAK package, version 3.1 (Kohonen and others 1996). More specifically, the *vfind* program in SOMPAK was utilized to search for good mappings (i.e., low quantization error). The search in *vfind* is done by automatically repeating different initializing and training procedures (Kohonen and others 1996). Note that SOMPAK uses the Euclidean distance to measure dissimilarity between samples.

The clustering process used in this paper was a two-level *SOM* as illustrated in Figure 17. Note that similar multi-level *SOM* approaches have been explored in some papers with respect to their clustering capability (e.g., Lampinen and Oja 1992, Vesanto and Alhoniemi 2000). However, multi-level *SOM* has not been applied to environmental problems, in general, and environmental assessment, in particular. In this study, the first-level *SOM* was applied directly to the data to produce a 10x5 map (50 prototype vectors). The second-level *SOM* was then applied to the 50 prototype vectors in the 10x6 map to produce a more agglomerative 7x1 map.





**Figure 17. Diagram of a two-level self-organizing map model.**

## Results

The first-level SOM's 10x5 map is displayed in Figure 18. The map, created by the U-matrix method (Ultsch 1993), visualizes the distances between each prototype vector to each of its neighbors in a gray scale on the map. As density of the SOM vectors reflects the density of the data points, the distances among neighboring SOM vectors reflect the distribution of the data set (Kaski and others 2000). Lighter shades on the map indicate denser distribution of SOM vectors or clustering tendency whereas darker shades show big distances from one SOM vector to another or sparse area in between clusters.

Although there were 50 nodes on the 10x6 map, only 41 nodes were populated. The number of watersheds at each populated node ranged from 1 to 10. The second-level SOM's 7x1 map is displayed in Figure 19. All 7 nodes on the 7x1 map were occupied by several of the first-level SOM nodes. The nodes were represented by a set of 50-dimensional vectors that will be used as benchmark vectors in further calculations (e.g., environmental assessment and management).

The orders of the watersheds on the U-matrix displays of the first- and second-level SOMs (Figures 18 and 19) indicate the similarity of environmental conditions of the watersheds in the study area. Generally, watersheds within one node have more similarity in environmental conditions than with those associated with other nodes. Among all nodes on the map, neighboring nodes have more similarity in environmental conditions than nodes being far away on the SOMs. Among neighboring nodes, a lighter shade between two nodes indicates more similarity in environmental conditions between the two than between those with darker shades. In general, given a wide array of ecological indicators, the U-matrix displays of the SOMs were able to represent the relative arrangement of the watersheds under study in terms of overall environmental conditions.

Figure 20 geographically displays the watersheds associated with the seven second-level SOM nodes (or clusters) on the Mid-Atlantic region map. It shows that most of the watersheds associated with a second-level SOM cluster were spatially contiguous to some extent. It can be seen that two nodes that are faraway on the second-level SOM (Figure 19) can have their watersheds located adjacent to each other on the geographical map (Figure 20) (e.g., watersheds associated with clusters 1 and 7 in Virginia), showing that watersheds with distinctly different conditions (e.g., "good" versus "bad") can be spatially located next to each other.

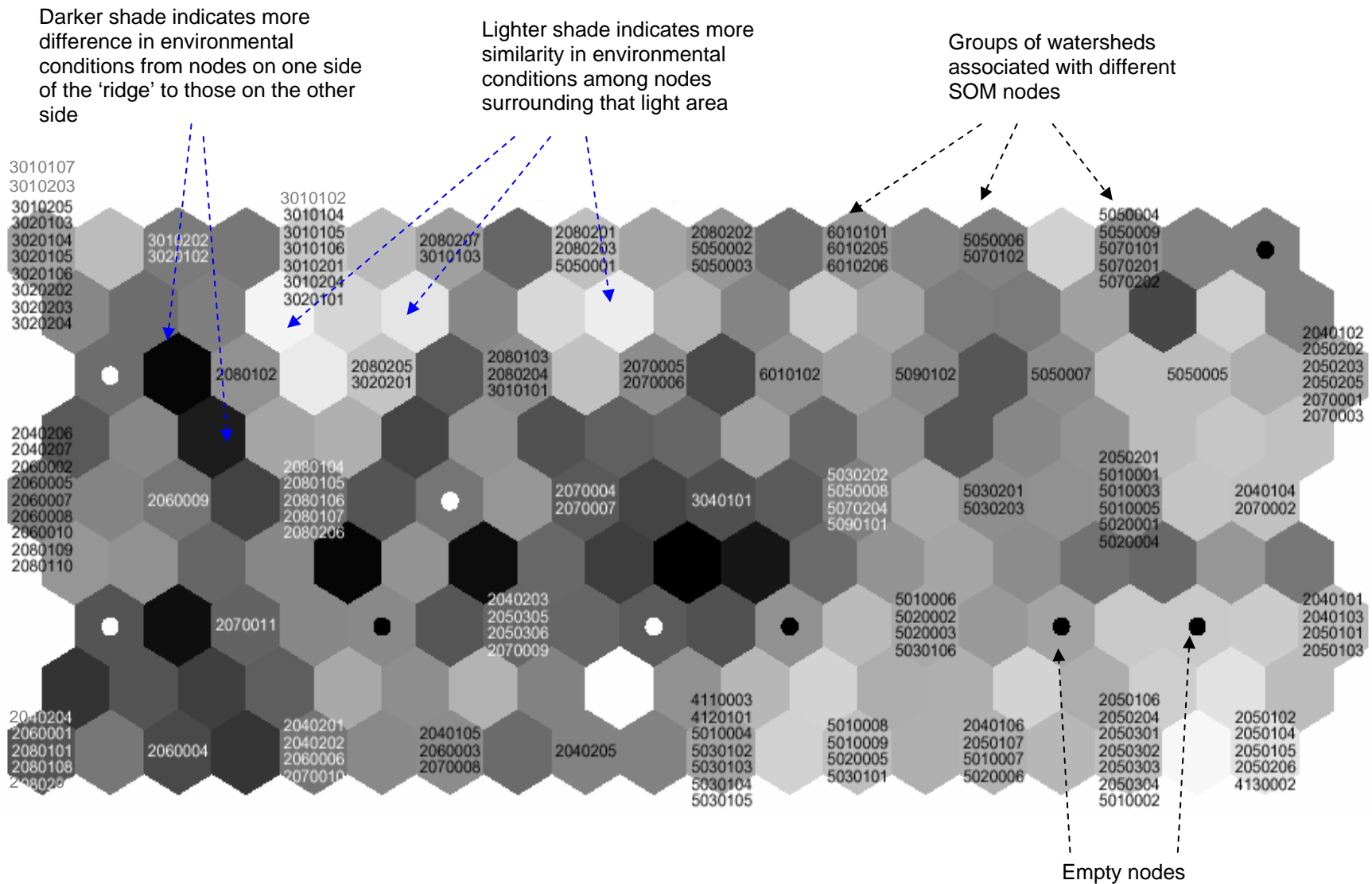


Figure 18. First-level 10x5 self-organizing map created by the U-matrix method.

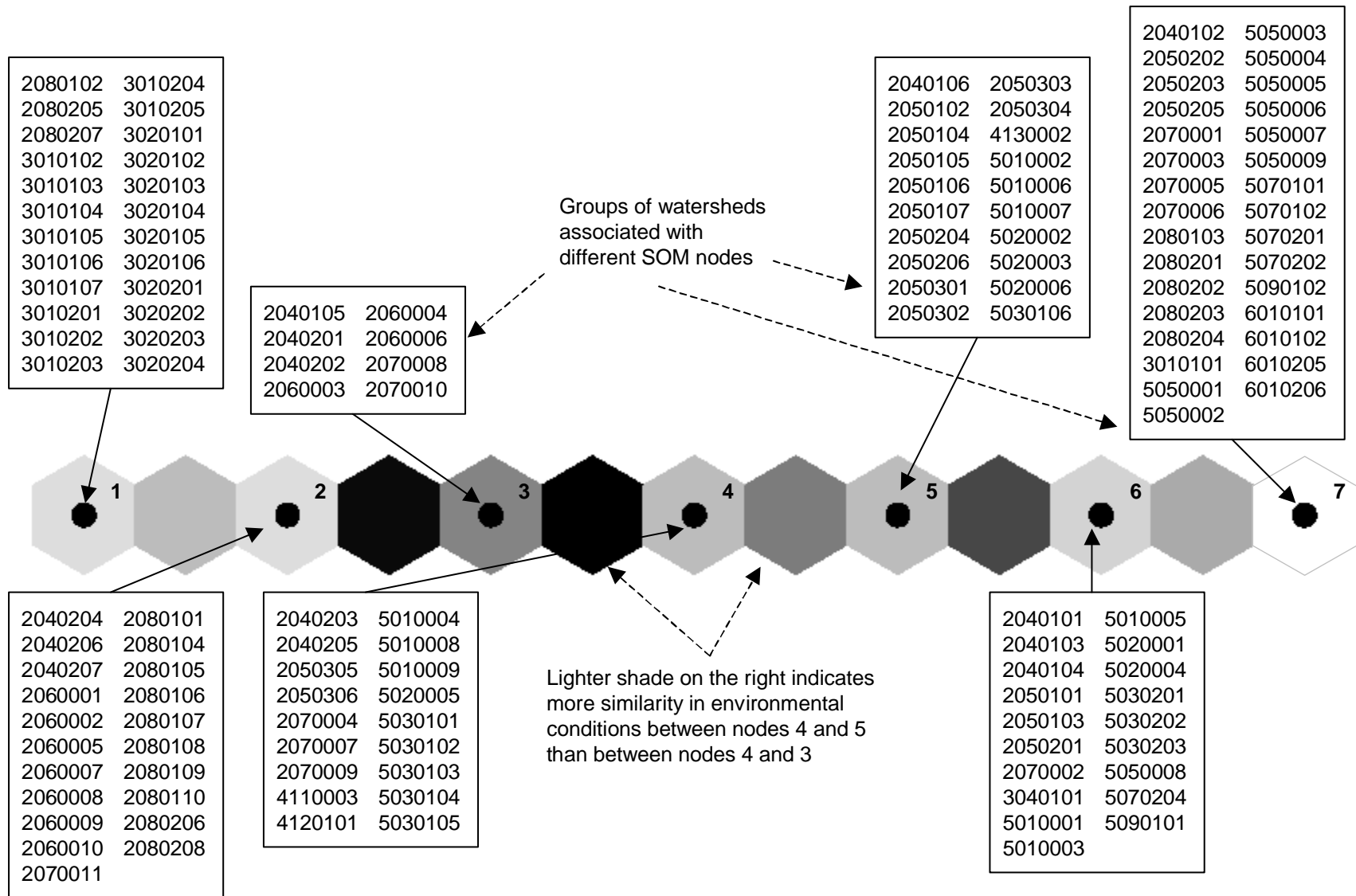
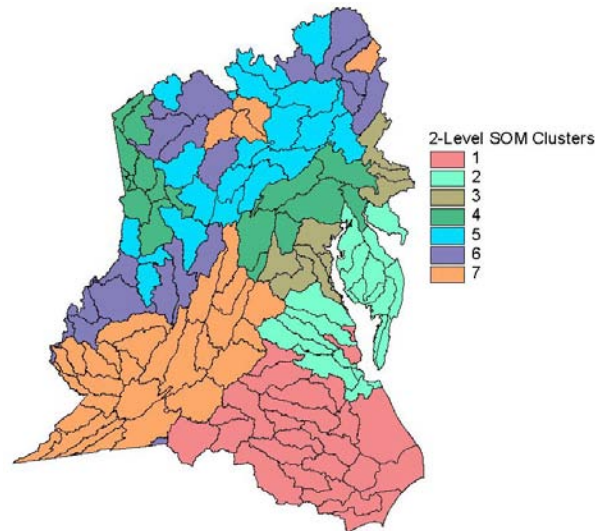


Figure 19. Second-level 7x1 self-organizing map created by the U-matrix method.

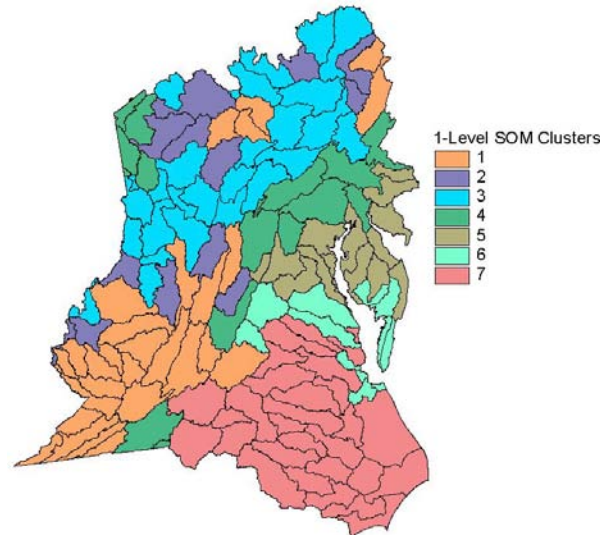


**Figure 20. Map shows geographic distribution of the two-level SOM's seven clusters. Watersheds with similar colors are in the same cluster.**

### Tests & Sensitivity

In general, the selection of the number of nodes in *SOM* is subjective. It can be decided via trial and error or by prior knowledge of the study problem. In this analysis, several factors were considered in deciding the size of the first- and second-level SOMs (10x5 and 7x1, respectively). They include data set size, possible clusters identified from the first-level SOM, and suitability for further analysis (i.e., a reasonable number of clusters, neither too many to fall short for generalization nor too few to dilute details). Regarding the relationship between the two SOMs, the 50 prototype vectors in the first-level SOM can be interpreted as subclusters that serve as ingredients to form larger clusters in the second-level SOM. Lampinen and Oja (1992) stated that the main benefit of the two-level SOM was noise reduction. Quantitatively, information from this analysis was insufficient to prove this. However, qualitatively, it can be seen that the within-cluster spatial adjoining of watersheds associated with cluster 5 in the two-level SOM (Figure 20) is better than those in the one-level SOM (Figure 21). This aspect can be explained because, as the prototype vectors are locally averaged from data associated with the subclusters, the impact of outliers on vector quantization would be reduced to some extent. Using the two-level approach, outliers associated with each subcluster are explored in detail by means of the first-level SOM. Whereas the first-level SOM showed a more detailed map of environmental conditions of single watersheds at a sub-regional level, the second-level SOM provided a more generalized picture of larger clusters of watersheds for the whole study area.

The use of the Euclidean distance to measure dissimilarity between samples in SOMPAC might have some implication on clustering results because it did not account for the interdependence among variables. Consequently, the clustering result might be biased toward indicators that are highly correlated and occupy a large portion in the data set. This problem can be addressed by using a more advanced distance measure, e.g., the Mahalanobis distance (Mahalanobis 1936). Another alternative is applying the Kohonen SOM to a set uncorrelated variables derived from PCA. However, this alternative has the tradeoff of being unable to identify the contribution of individual indicators to the clustering process.



**Figure 21. Map shows geographic distribution of the one-level SOM's seven clusters. Watersheds with similar colors are in the same cluster.**

### Advantages & Disadvantages

The main advantage of *SOM* for regional environmental assessment is in its nonlinear clustering ability and its capability in revealing the distribution of a multi-dimensional data set on a lower dimensional space (e.g., the U-matrix method). Another advantage of the two-level *SOM* used in this analysis is that it provided a unique means in exploring the region from different perspectives. Whereas the first-level *SOM* showed a detailed map of environmental conditions of sub-clusters at a sub-regional level, the second-level *SOM* provided a more generalized picture of larger clusters for the whole study area. Briefly, the two-level *SOM* in this study was proven a viable tool for clustering and classification of complex environmental data set.

Similar to other clustering methods, the primary disadvantage of the *SOM* approach is that it can be used for clustering/classification only but not for other common tasks of a regional environmental assessment, such as ranking or vulnerability analysis. However, this disadvantage can be overcome by combining *SOM* with other approaches to form a more comprehensive method. For example, Tran et al. (2003) combined *SOM* with PCA and various visualization tools in a method capable of classifying and ranking complex environmental conditions.

As a secondary disadvantage, while the ability of *SOM* in visualization is unquestionable (and was one of the main reasons that *SOM* was chosen in this analysis), there is no guarantee that *SOM* produces the best clusters within data (Kohonen 2001). However, this point was not so crucial because clustering was only one among various important aspects taken into consideration with *SOM* (e.g., visualization, transition among watersheds, and possibility of multi-level clustering).

## Recommendation

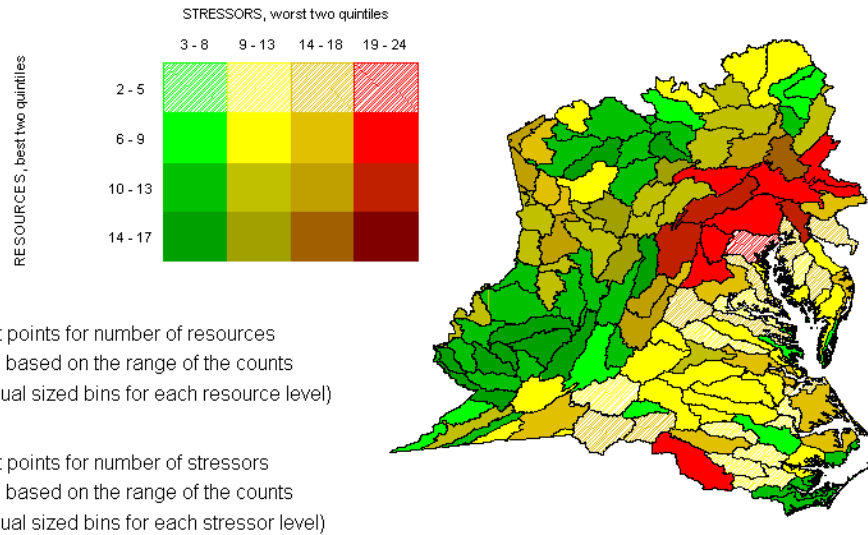
Given its unique capability in clustering and displaying distribution of the data set, SOM should be retained as a classification technique for regional environmental assessment. Furthermore, the strengths of the SOM's nonlinear clustering ability can be combined with the PCA's capability in reducing the dimensionality of multivariate data as well as with several constructive visualization tools (e.g., U-matrix map, component planes, parallel coordinate plots) to form a more comprehensive method which can perform a regional environmental assessment with the focus on cumulative impact of multiple stressors on a large area (Tran et al. 2003).

Despite its many advantages, the exact approach for incorporating SOM in regional assessments remains a research question. Certainly, further work is needed before *SOM* can be recommended for direct application by decision-makers without the guidance of an experienced analyst.

## ***Stressor-Resource Overlay***

The *Stressor-Resource Overlay* method attempts to locate watersheds where high levels of valued resources occur with high levels of stressors. For this analysis, the ReVA variables were first divided into stressors and resources. Stressor and resource variables were then divided into quintiles. Watersheds were scored on the number of stressor variables that fell into the worst two quintiles and also on the number of resource variables that fell into the best two quintiles. Results were displayed as a 16-category map in Figure 22. The solid red color showed that those watersheds are most vulnerable because they have valued resources endangered by multiple worst stressors. In contrast, the solid green color was for watersheds having good resources and no serious stressors.

The figure illustrates some of the subtle features of vulnerability analysis. For example, the watershed containing Baltimore is highly stressed, but it is not among the most vulnerable because few valued resources remain. If we focus on the most vulnerable watersheds (in red and dark tan), we find a dozen watersheds with similar properties in rural suburban areas or watersheds containing smaller cities such as Allentown and Raleigh. They have populations and/or population growth in the upper two quintiles but human resources, such as low poverty and low infant death, are still in the lower two quintiles of the region. These human resources are likely to be vulnerable to further urbanization and population growth. Many of these vulnerable watersheds have agriculture on steep slopes, in riparian zones, and on erodible soils. Combined with high levels of atmospheric deposition, agriculture has led to high levels of nitrogen in the aquatic systems and only one of the watersheds, Raleigh, retains native aquatic organisms in the highest two quintiles. While the aquatic resources have been largely lost, significant terrestrial resources remain. The forests are stressed by fragmentation, ozone, and exotic species but still have considerable forest resources in hardwoods and softwoods and five of the watersheds have native terrestrial fauna in the upper two quintiles of the region. Thus, the *Stressor-Resource Overlay* was able to identify a set of watersheds that have largely lost their aquatic resources but retained human and terrestrial resources that are vulnerable to further stresses.



**Figure 22. Results of the *Stressor-Resource Overlay* method displayed in a 16-category map. The two-dimensional matrix provides color-codings for different combinations of stressors and resources. Increasing stressors go from left to right across the top, while increasing resources go from top to bottom on the left. Areas with the highest resources and highest stressors are indicated by dark red, shown in the bottom right-hand corner.**

### Advantages & Disadvantages

The advantage of the *Stressor-Resource Overlay* is in its unique ability to identify areas that could lose valued resources with further stress by directly addressing the geographic distribution of vulnerability. A watershed could have many valued resources but still not be vulnerable because it is not experiencing stress. On the other hand, a highly stressed watershed may not be considered vulnerable because its valued resources have already been destroyed. The most vulnerable watersheds can be considered to be those with intermediate to high levels of stressors together with intermediate to high levels of resources.

As with many of the methods, the *Stressor-Resource Overlay* has the disadvantage that it does not account for correlation between variables. However, in this case, this is not a significant disadvantage. Two stressors may be highly correlated, but because of synergistic effects, the two stressors may represent more impact than each stressor independently. Similarly, two resources may be highly correlated but because each is valued the two resources may be more highly valued than each resource independently. As a result, the *Stressor-Resource Overlay* is not influenced by the correlation structure of the data as long as each resource is valued and stressors can interact to cause synergistic effects.

### Recommendations

The *Stressor-Resource Overlay* method is an effective approach to estimate vulnerability. It takes a unique view of vulnerability that does not overlap with the other integration methods and the simplicity of its logic makes it robust to the peculiarities of the data structure. As such it should be incorporated into any regional assessment.

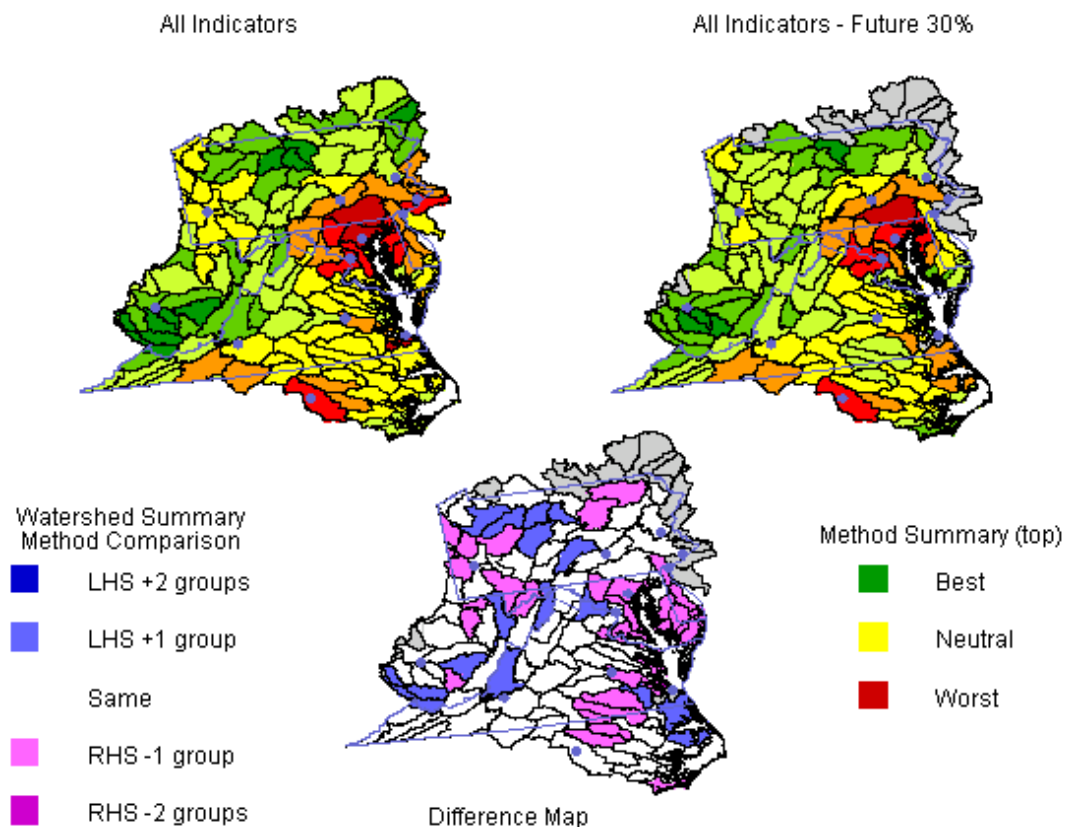


## Change Analysis

The *Change Analysis* method compares two regional maps. If a watershed falls into the same septile on both maps, the watershed is colored gray. If the watershed is improved by one septile, the watershed is colored light green. If the watershed is improved by two or more septiles, the watershed is colored dark green. If the watershed is degraded by one septile, it is colored light red. If the watershed is degraded by two or more septiles, it is colored dark red. Thus, the difference or change between the two maps is easily visualized.

The *Change Analysis* method can be used to compare any pair of maps. For example, it is possible to compare the results from different integration methods to determine differences in spatial patterns. To illustrate the method, we compared present conditions with a projected future scenario. The future scenario considered that forestland cover on each watershed was reduced by 30 percent. Riparian forest was changed to agriculture. Forest on slopes >3 percent was changed to crops. Interior forest was changed to human land use. All other variables were kept at their current values. The present and future scenarios were both evaluated using the *Simple Sum* method. This exercise was not intended to evaluate a realistic future scenario but simply to determine if the *Change Analysis* method would have the sensitivity to indicate the change.

The comparison map produced by the *Change Analysis* method test is shown in Figure 23. The comparison map indicates that a significant number of watersheds changed their rank by a septile. The map clearly shows where the changes occur and simplifies interpretation.



**Figure 23.** Graphic illustrates the *Change Analysis* method. The difference map at the bottom highlights areas where values are better in the map on the left-hand side (LHS) in blue and where values are better in the map on the right-hand side (RHS) in pink.



## Advantages & Disadvantages

The primary advantage of the *Change Analysis* method is its simplicity. As a result, the analysis is easily understood and easily communicated. It is immediately apparent to what degree a change has occurred and the change in overall regional pattern. The second advantage of the *Change Analysis* method is its sensitivity. The illustration, shown on upper right of Figure 23, reflected only a change in the land use variables but the spatial pattern of change was clearly indicated.

Because of the inherent simplicity of the approach, the *Change Analysis* method has no disadvantages or sensitivities to data distributions or other problems. Any disadvantages come directly from the choice of the two maps being compared.

## Recommendation

The *Change Analysis* method should prove useful in future assessments. The method is simple and intuitive and should appeal to a broad audience. The method is flexible and can be used to compare present and future scenarios as well as testing the effectiveness of restoration strategies over a region or subregion. It shows the spatial pattern of change that can occur across the region in response to any postulated change.

It should be remembered, however, that the *Change Analysis* method is only useful in comparing two maps. It is not, in itself, an integration method. It is rather a visualization technique that allows the assessor to determine where change has occurred.

## ***Stressor-Resource Matrix Analysis***

As an approach to deal with multiple stressors, the assessment community developed a matrix procedure (Foran and Ferenc 1999, Ferenc and Foran 2000). The matrix represents stressors as the rows and the endpoints as the columns. Using a matrix format simply to organize complex assessment information has a long history (Phillips Brandt Reddick, McDonald and Grefe, Inc. 1978; Lumb 1982a, 1982b; Witmer et al. 1985; Clark 1986; Emery 1986; Risser 1988). Leopold et al. (1971) originally proposed the approach and Canter (1977) reviewed a number of variations.

In addition to organizing information, matrices have been used for diverse applications, each application involving significant variations. For example, Cada and McLean (1985) used a definition matrix to associate rankings with quantitative ranges of potential impact (sedimentation, cover loss, restriction of fish movement, and loss of food base) and weighted means to compare impacts across 12 projects. Bain et al. (1986) used a computer-assisted matrix method to analyze the impact of multiple human developments on multiple resources. All possible combinations of stressors are considered with the impact of each combination computed as the sum of all project-specific impacts, adjusted for the effect of interactions among projects. This results in a matrix representing the relative impact of every possible combination of stressors on each endpoint. The matrix is then searched for combinations that minimize the impact summed across all endpoints. Stull et al. (1987) used a matrix approach to evaluate multiple hydropower impacts on elk and salmon. The approach involves multiplying interaction and impact matrices to determine cumulative impact. Landis and Wiegers (1997) used matrices to find the overlap of wastewater inputs with crab and clam endpoints across different spatial areas. Foran and Ferenc (1999) use a matrix to compare qualitative estimates of likelihood and consequences across a set of potential scenarios. After eliminating scenarios of low probability and impact, further analyses might involve quantitative estimates of risk curves (i.e., probability of different levels of impact) and possibilities of remedial actions. A number of additional refinements using probability theory and fuzzy set theory are available (e.g., Ferson and Kuhn 1992, Jooste 2000).

In more recent applications of the approach, the emphasis has been on ranking stressors (Cormier et al. 2000). Harris et al. (1994) developed an impact matrix of stressors and endpoints in the form of impaired use criteria. Experts then filled in the matrix with values from 0 (no impact) to 3 (major impact). The stressors can then be ranked by looking at the row sums of the matrix. The row sums point out the stressor with the greatest impacts summed across the suite of endpoints.

In the typical application of the matrix approach, quantitative information is not available for the individual cells of the matrix. A panel of experts is asked to assess the individual impacts and supply a qualitative value, for example, 1 for a minor impact, 2 for a moderate impact, and 3 for a major impact. When the matrix is complete, the values in each row are summed and taken as the total impact of each stressor across all endpoints. The row sums can then be ranked to indicate which stressors are having the greatest impact and, in a decision-making context, are in greatest need of control.

Data are available for stressors and resources and a unique opportunity exists to apply the matrix approach quantitatively. The rows contain both traditional stressors, such as nitrate and sulfate deposition, and also some conditioning variables, such as soil loss potential, which might modify stressor-resource relationships across watersheds. The columns contain both traditional endpoints, such as native aquatic biota, and also assets such as interior forest (see Table 2A, Appendix). The values in each column can also be summed and taken as an indicator of vulnerability, in the sense of how much stress is already being imposed on the assets across all stressors. In a decision-making context, these assets might be considered the most vulnerable to further stress and most in need of protection or remediation.

The regional data can be analyzed in either of two ways. A correlation matrix expresses how the variability in a stressor is related to variability in a resource. This provides insight into the overall cumulative relationship among all stressors and resources across the watersheds. A second approach involves regression analysis in which each coefficient expresses how a small increase in a stressor results in a small increase or decrease in a resource. The two approaches provide different insights into the relationships between stressors and resources and both approaches were used.

### Correlation – Results and Tests

In the first approach, a correlation matrix is calculated for the entire data set over all watersheds and each row of correlation coefficients is summed. The resultant row sum is then an indicator that includes all of the direct and indirect interactions among the variables. Thus, indirect effects (both positive and negative) with other stressors are included along with the effects of stressors on resources.

### Stressors: Results and Tests

The first column of Table 4 gives the top four stressors: human land cover (UINDEX), dissolved phosphorus (DISSOLVEDP), Nitrogen in the aquatic system (TOTALN), and small-scale forest fragmentation (EDGE2). The second column gives their row sums. The importance of the land cover and fragmentation stressors is consistent with previous large-scale studies. The importance of aquatic nutrients is also reasonable in light of our understanding of nutrient dynamics. The dominance of the aquatic nutrient variables is somewhat surprising, given that many of the resource variables are terrestrial. The explanation is probably that the aquatic nutrients are leached from the watershed and therefore nutrient status in the freshwater system reflects nutrient status in the soils on the watershed as well.

**Table 4. Correlation matrix results: top four stressors.**

	Full	>  0.1	>  0.2	>  0.3
UINDEX	12.3	12.1	12.2	12.1
DISSOLVEDP	12.0	11.9	11.5	10.5
TOTALN	11.2	11.2	10.5	11.0
EDGE2	9.6	9.6	8.7	7.5

The correlation analysis involves 50x50 interactions and therefore probably includes a number of spurious correlations. Small, spurious coefficients might accumulate over the row sum and alter the results. To test the sensitivity of the method, we systematically removed smaller coefficients and recalculated the row sums. Columns 3 to 5 in the table represent the row sums with all coefficients smaller than +/- 0.1, +/- 0.2, and +/- 0.3 set to 0. The results in the table indicate that the ranking is insensitive. The likely explanation is that small coefficients are equally likely to be positive or negative and only have a minor influence on the row sum. Beyond the first four stressors, the row sums become more sensitive and the ordering of the rest of the stressors tends to change as the smaller coefficients are removed from the matrix. It appears, therefore, that the correlation matrix approach should be limited to designating the top three or four stressors.

#### Vulnerable Resources: Results and Tests

The correlation matrix can also be used to indicate the resources experiencing the greatest stress by ranking the column sums associated with the resources instead of the stressors. The most vulnerable resources are given in the first column of Table 5: intact forest at a small scale (INT2) and habitat for migratory birds (MIGSCENARIO). The vulnerability of habitat assets seems reasonable.

**Table 5. Correlation matrix results: top five vulnerable resources.**

	Full	>  0.1	>  0.2	>  0.3
INT2	10.9	10.8	10	10.2
MIGSCENARIO	10.3	10.1	9.6	8.5
EMAGRIC	8.4	8.2		
SOFTWOODREM			8.4	
RIPFOR				6.8

As with the stressor analysis, we tested the sensitivity of these results by removing small correlation coefficients from the matrix. Because of the number of variables in the analysis, small and potentially spurious values could be summed to change the rankings of a stressor or vulnerable asset. Columns 2 to 4 in the table represent the row sums with all coefficients smaller than +/- 0.1, +/- 0.2, and +/- 0.3 set to 0. The top two resources do not change. Once again, the matrix approach is robust to this test primarily because both positive and negative values are eliminated. As a result, the sums are only minimally influenced and the rankings are unchanged. However, it should be noted that beyond the first two variables, the rankings of the remaining variables change significantly as the smaller coefficients are removed.

The question arises as to whether removing the small coefficients is a better approach, rather than just a test. In typical applications of a correlation matrix, significant causes or correlates are sought. Within a stated probability (alpha), only values greater than a certain value are significant. Thus, the typical test guards against false positives: falsely identifying a variable when, in fact, the correlation is spurious. The test does not guard against false negatives: falsely eliminating a small but real correlation. Therefore, eliminating small correlations is not necessarily justified. Alternatively, there is a good reason to retain them: these small correlations may quantify the subtle cumulative effects of large numbers of stressors.

## Regression – Results and Tests

The regression analysis involves a univariate regression of each resource as a function of all the stressors. The regression coefficients associated with a stressor are then summed across all resources. This method will be less sensitive to *indirect* effects than the correlation approach, but it should be more useful for determining the stressors that dominate *direct* effects. To emphasize the direct effects, insignificant regression coefficients (less than  $|0.25|$ ) and negative coefficients were dropped from the matrix before doing the row sums. Because the regressions were done with the coded variables, only a significant, positive regression coefficient indicates a direct effect.

## Stressors

The first column of Table 6 shows the top four stressors determined by the regression matrix row sums. Three of the four are identical to those determined by the correlation approach (Table 4) and road density replaces EDGE2. Because the insignificant coefficients were already removed from the matrix before summing the rows, the removal of small coefficients could not be used here to test the sensitivity of the method. Instead, another aspect of the approach was tested.

**Table 6. Regression analysis results: top four stressors.**

	Full	WS
UINDEX	9.3	1.5
DISSOLVEDP	14.2	2.4
TOTALN	6.8	1.4
RDDENS	9.1	2.1

Taking the row sum across all resources gives equal weight to each resource. However, the ReVA data set does not have the same number of variables in each resource categories or families. There are six socio-economic resources, nine forest variables, four aquatic and three terrestrial population resources. With the possibility that stressor effects may dominate the row sum on forest alone, we performed a weighted sum to test for the influence of this imbalance among resources. The coefficients within each family of variables were averaged and then the four averages were summed. The second column of Table 2 shows the top four stressors using the weighted sums of the regression coefficients. Both the weighted and unweighted sums indicate the same top stressors. Therefore, the regression matrix approach does not appear to be particularly sensitive to the imbalance among resources.

## Vulnerable Resources

Regression analysis can also be applied to determine the most stressed resources. The approach would be to sum the regression coefficients across all the stressors. However, there would then be a problem comparing these sums because each univariate regression considers only a single resource. Therefore, the sum of the regression coefficients may have more to do with the goodness of fit of the

individual univariate regressions than with the overall relative stress on the resource. As a result, one might be comparing goodness of fit rather than levels of stress when one univariate row sum is compared to another univariate row sum.

Table 7 illustrates the problems with the regression approach. The first column shows the sum of all coefficients for the top three positive values. Remember that with the coded variables, a positive sum indicates a negative impact on the resource. The resources are reasonable as the most stressed: employment in agriculture (EMAGRIC), habitat for migratory birds (MIGSCENARIO), and intact forest (INT65). However, only one of these variables, MIGSCENARIO, was also determined to be a highly stressed resource in the correlation analysis. In comparing the correlation and regression approaches for the stressors, three of the top four stressors were the same. In comparing the correlation and regression approach for the resources, only one variable was common to both.

**Table 7. Regression analysis results: top five vulnerable resources.**

	Full	>  0.25	> 0.25
EMAGRIC	2.4		
MIGSCENARIO	1.6		
INT65	1.6	1.5	5.3
NONCLIMAXPCT		1.6	
TERRTE		1	
NATCOVERPCT			4.3
AQUANATIVE			3.4

The problem is more evident in the second column of Table 4, which removes the insignificant regression coefficients. Two new variables enter the top three and only the INT65 remains. In column three, only positive values are included in the sum and again two new variables enter. So the rankings based on the row sums do not appear to be very stable to these manipulations. Because of this instability and the lack of agreement between the regression analysis and the correlation analysis, this analysis should probably be dropped from future work. There really is no theoretical foundation for comparing coefficient sums for unrelated univariate regressions.

### Advantages & Disadvantages

The primary advantage of the *Matrix* approach to integration is that it takes a completely different view of the region. Instead of comparing watersheds, it compares stressors and resources across the entire region. Thus, it provides a viewpoint that is complementary to the other integration methods.

The primary disadvantage of this approach to integration is the difficulty of interpreting and communicating the results. This approach has been used in the literature to rank direct impacts. In the current analysis, correlations are not direct impacts. Even large negative values are a complex combination of direct and indirect impacts as well as synergistic and cumulative effects. The correlations will be strongly influenced by spatial patterns of co-occurrence. In some cases, the present condition of a resource may be strongly influenced by past stressors not captured in the present values used in the analysis.

The important point is that the correlations express the resultant of past and present stressors, spatial pattern, and synergistic and cumulative effects. Therefore, the results of the correlation analysis will have to be carefully stated with sufficient caveats warning the reader against the natural tendency to interpret

the results as direct impacts. In the usual application, the top ranked stressors are identified as the one most in need of immediate managerial action. That conclusion is not necessarily valid in our case and elimination of the top stressors might well result in little, or long-delayed, responses.

A similar problem exists with spatial autocorrelation. Spatial autocorrelation occurs when data from two adjacent points in space show a higher correlation than data from two randomly chosen points. Such proximity relationships clearly occur in the ReVA data set. Spatial autocorrelation causes a problem, for example, in regression analyses that test for significant relationships between variables. The variables may simply co-occur in space rather than being causally related.

Spatial autocorrelation does not pose a serious problem for the *Stressor-Resource Matrix* analysis because no hypotheses are tested and no attempt is made to establish causal relationships between variables. However, spatial autocorrelation does restrict the interpretation that can be applied to the results.

In regression analysis, negative coefficients are more closely related to direct effects. However, the measurements are not consecutive in time on a single watershed. Rather, they are at a single time across many watersheds. Therefore, reducing a top stressor can be expected to have an effect across the region. However, reducing a top stressor will not necessarily be an effective strategy on any particular watershed. So the interpretation is still different from the smaller scaled applications in the literature. Once again this makes it more difficult to communicate the results.

## Recommendations

We should use this approach in the assessment – but we will need to carefully argue each step so that the results are not misinterpreted. The results of the correlation analysis seem to be reasonable for both stressors and resources. The results of the regression analysis should be limited to the stressors and seem to reinforce the results obtained from the correlation analysis. If future applications continue to show that the regression results for stressors do not differ from the correlation results, then the regression approach can be dropped.

The technique was originally designed to identify the top few stressors within a site or watershed. The tests performed here reinforce the idea of limiting the results to two or three stressors and vulnerable assets. Once the analysis gets beyond about three, the row or column sums tend to converge and the exact ranking of a stressor or asset becomes very sensitive to small changes in the sums. The first two or three are robust to the assumptions. But the rest of the rankings are less reliable and more dependent on assumptions.

---

## Section 3

### Sensitivity of Integration Methodology to Data

We examined the methods described in previous sections for their sensitivity to various data problems. We identified various data issues that might influence the performance of the integration methods. They include:

- Discontinuity – There are variables that took on only integer values, e.g., presence/absence of a species. This might influence methods using statistics that assume continuous variables. This problem was solved in the Mid-Atlantic Region by redefining variables so that the discontinuities were eliminated. However, care must be taken in future regional applications to eliminate discontinuous variables or test explicitly for their impact on the methodology.
- Skewness – Many variables have distributions that are highly skewed toward zero. This is the natural result of collecting data across a diverse region. Many watersheds had little or no problems, only a few had high values. Some methods had no problem with this (e.g., SUM) but other methods did (e.g., regression in the *Stressor-Resource Matrix* method).
- Imbalance – The variables are not equally distributed across families (e.g., terrestrial biodiversity, human variables, etc.). This is a problem with methods that sum a score/rank across variables (e.g., SUM). The obvious solution is to average within families of similar variables and then sum across the families.
- Interdependency – The variables are correlated and methods (e.g., SUM) that do not account for this interdependency can give misleading results. The bias introduced is similar to the Imbalance problem because two highly correlated variables are added in twice when their information is not actually independent. Some of the statistically based methods automatically account for interdependency. For other methods the problem must be resolved individually.
- Auto-correlation – Although the variables are auto-correlated to various extents, it is not a problem for any of the methods except the *Stressor-Resource Matrix*. This is because the methods in this report do not seek to explain the pattern-process relationships among variables but focus on the overall environmental condition on the individual watershed basis.

On the one hand, we identified possible alternatives to handle each problem properly. For example, the skewness problem can be solved by appropriate data transformation (e.g., log transform) or dropping outliers; the imbalance can be dealt by reducing the between-family imbalance or averaging the within-family values first. On the other hand, we examined the methods to find out how sensitive each of them was to the data used in the analysis. Table 8 displays information on sensitivity of the integration methods regarding various data problems listed above.

**Table 8. Effects of various data issues on each integration method.**

	<b>Discontinuity</b>	<b>Skewness</b>	<b>Imbalance</b>	<b>Correlation</b>
Quantiles	Not sensitive	Not sensitive	Sensitive	Sensitive
Simple Sum	Not sensitive	Not sensitive	Sensitive	Sensitive
PCA	Sensitive	Sensitive	Not sensitive	Not sensitive
State Space	Sensitive	Sensitive	Not sensitive	Not sensitive
Criticality	Not sensitive	Not sensitive	Sensitive	Sensitive
AHP	Not sensitive	Not sensitive	Not sensitive	Sensitive
Cluster	Not sensitive	Sensitive	Sensitive	Sensitive
SOM	Not sensitive	Sensitive	Sensitive	Sensitive
Overlay	Not sensitive	Not sensitive	Not sensitive	Not sensitive
Matrix	Sensitive	Sensitive	Sensitive	Not Sensitive

Discontinuity and skewness do not affect the *Best/Worst Quantiles* but imbalance and interdependency do when they coexist. In such cases, the results might be biased toward a large family of highly correlated variables.

With the *Simple Sum*, discontinuity is not a problem as there is no assumption about continuous variables. Skewness is not a problem for the *Simple Sum*, either. Since with *Simple Sum*, we are only interested in the relative environmental quality within the region. Outliers will stand out and that is appropriate. But it also means that we cannot depend on this method to finely discriminate among the intermediate watersheds. There is a problem with imbalance so we should always average families of variables and then sum. There is a potential problem with interdependency (i.e., double-counting related variables) for the *Simple Sum*. Hence, this method should be used in a complementary manner with the *PCA* and other distance-based methods.

As the *PCA* uses the covariance (or correlation) matrix in its calculation, discontinuity and skewness might affect the *PCA*. Regardless, the *PCA* is an effective way to handle imbalance and interdependency.

Using the Mahalanobis distance in its calculation, the State Space Analysis handled the problems of imbalance and interdependency quite well. However, abnormal values might occur if collinearity exists (due to problem with the inversed covariance matrix in case of collinearity).

In this study, we evaluated the Criticality Analysis using fuzzy distances. Other methods of calculating the distances from the natural state would change the sensitivities. However, using the fuzzy distances seems most appropriate for regional application given the difficulty in precisely defining the natural state.

The *AHP* does not make any assumption on discontinuity and skewness but assumes independence among variables. If interdependence exists, another more comprehensive model named the Analytical Network Process (*ANP*) (Saaty 2001) may be more appropriate and this alternative should be explored in future research.

Although there is no assumption about data in *Cluster Analysis*, clustering result is highly dependent upon the distance criteria used in the partitioning algorithm, which in turn is strongly influenced by the data distributions. For example, the clustering results of single linkage (or nearest neighbor) and the complete linkage (or furthest neighbor), which are based on a pair of two objects to measure the distance



between two groups, are sensitive to outliers. Interdependency and imbalance also affect the *Cluster Analysis* whose results might be biased toward the pattern of big family of highly-correlated variables, masking the signals of other important but stand-alone variables. In such a case, the problem can be overcome by the use of the Mahalanobis distance instead of the common Euclidean distance. The *SOM*'s sensitivity to data problems is similar to the *Cluster Analysis*.

The *Stressor-Resource Overlay* is the only method that is not sensitive to data peculiarities. However, this conclusion depends on two assumptions. First, multiple stressors impose synergistic effects. Therefore, if two stressors are present at high levels on a watershed, then the effect is at least twice as great. This will be true whether the stressors are or are not correlated across the region. Second, each of the resources is valued independently whether or not the resources are correlated across the region. Generally, these assumptions will hold as long as the variable list is carefully scrutinized. For example, two variables, which are really different measures of the same stressor or resource, should not be used together.

The *Stressor-Resource Matrix* uses either correlation or regression coefficients and has a problem with integer variables. Skewness is likely to cause problems with the correlations and regressions. Log transformations are inappropriate here because the method directly interprets the regression coefficients. Dropping outliers will not work, either. This is because we are looking for the outlier stressors. In addition, dropping variables will not work because the worst stressors are associated with the most skewed variables. So, skewness is a substantial problem here. The analyses will have to be carefully interpreted and checks provided for the influence of skewness. Imbalance is also a problem with this method since correlation/regression coefficients are summed across families. This will be addressed in the following two ways: (1) the analysis will be done for each family separately and (2) the sum will be taken over the averages of each family.

Generally, integration methodology is sensitive to some or all of the data problems. There is no universal solution for all methods and/or all problems. Hence, we need to be aware of the potential pitfall(s) of each method regarding data problems. Furthermore, we need to deal with each method regarding each problem individually.

---

## Section 4

### Discussion & Recommendations

#### ***Discussion***

In the following section, we discuss the integration methods regarding various key components of a regional environmental assessment, such as classification/ranking, risk/vulnerability assessment, and planning/restoration/development, as well as other related issues (e.g., data, uncertainty).

#### **Classification/Ranking**

Generally, the aim of classification in environmental assessment is to group watersheds into homogeneous classes or to rank watersheds from the best to the worst with regard to a set of criteria. In more detail, this type of problem can be referred to as classification or sorting. Although these two terms describe the overall objective of assigning watersheds into groups, they represent two different situations. While classification does not need an order among groups (nominal measurements), sorting requires groups to be ordered from the best to the worst one (ordinal measurements). For example, in an ecoregion-based classification scheme, the identification of a watershed according to its physical and ecological attributes is a classification problem because it is nonsense to establish a preference ordering among ecoregions and eco-subregions. However, the vulnerability assessment of watersheds in ReVA is a sorting problem because vulnerability can be ordered into different levels ranging from high to low. Both classification and sorting problems have been encountered extensively in environmental studies and they have practically motivated researchers to develop models based on various approaches to achieve higher accuracy in classification and prediction. For several decades, the source-based approaches such as those focused on a specific chemical contaminant (e.g., dose-response curves) have been applied extensively in risk assessment and related sorting and classification problems. Multivariate statistical analyses (e.g., various clustering techniques, linear, and quadratic discriminant analysis) have become more prevalent recently. While multivariate analysis can be used to analyze and present associations between multiple stressors and multiple endpoints/impacts, the parametric nature as well as the statistical assumptions and restrictions of these techniques have been a major issue in their practical applicability and usefulness. Recently, classification and sorting models have been developed based on techniques in decision-making science and artificial intelligence.

The integration methods presented in previous sections cover various approaches ranging from simple ones (e.g., *Simple Sum*, *Stressor-Resource Overlay*) to conventional statistics (*PCA*, *Cluster Analysis* with K-means clustering and hierarchical clustering), to artificial neural network (e.g., multi-level *SOM*) and decision-making methods (*AHP*). These methods are capable of classifying/ranking at various levels. For example, while the clustering methods (*Cluster Analysis*, *SOM*) are mainly for classification, their use for ranking purposes is absolutely possible by combining them with another method (e.g., *Simple Sum*, *PCA*). Alternatively, the distance-based methods (e.g., *Simple Sum*, *PCA*, *State Space Analysis*, *Criticality Analysis*, *AHP*) are mainly for ranking purpose. However, they should be used in parallel with a clustering method to find out whether or not their score-/rank-based groups reflect any real environmental pattern within and/or between groups.

Common integrated assessment questions regarding classification and ranking include (but are not limited to): What is the overall environmental condition of the region? What is relative condition of locations within a region? What is the pattern of the overall condition? What are the patterns of abiotic/biotic factors and/or resources associated with environmental condition of the region? What are the relative rankings of regional stressors/resources? What are the differences between areas with good versus poor condition?

Depending upon the question(s) and study phase, integration methods can be used individually or in combination. For example, for preliminary exploratory purposes, *Best/Worst Quantiles* and *Simple Sum* are very suitable due to their simplicity and visual effectiveness. Similarly, preliminary classification can be done with *Clustering Analysis* and/or *SOM* to explore spatial pattern of the overall environmental condition. At a later phase of the assessment, pattern classification can also be carried out with *Clustering Analysis* and/or *SOM* but in a more complex manner, such as multiple-level clustering (e.g., hierarchical clustering and multi-level *SOM*) on different subsets or the whole set of variables to explore spatial patterns of various environmental dimensions (e.g., stressors, resources, human health pressure, etc.). Questions regarding overall or specific environmental conditions of the region as well as relative condition of locations can be revealed in a distance-based measuring/ranking method, from a simple method like *Simple Sum* to more complicated ones, such as *PCA*, *State Space Analysis*, *Criticality Analysis*, and *AHP*, which can be applied on different subsets or the whole set of variables.

### Risk/Vulnerability Assessment

Conventional ecological risk/vulnerability assessment is mainly based on a “source-based” approach (single stressor on single resource) in which the concept of probability is dominant. However, it is not easy (or almost impossible) to derive a probabilistic risk/vulnerability in a “place-based” method as in ReVA where data are multiple stressors and resources collected from various sources with different types of uncertainty (or no information on uncertainty at all). In that context, the integration methods in ReVA portray a risk/vulnerability concept in a “qualitative” and “relative” context (e.g., relatively low risk, high risk, extremely high risk, etc.). This risk/vulnerability concept is based on a relative comparison and spatial relationships among watersheds. For example, the closer a watershed with high resources is to a watershed with high levels of stressors, the higher risk the former watershed is facing. And if there is no plan of protection/preservation/adaptation for those resources, the more vulnerable the watershed is.

Common risk/vulnerability assessment questions include (but are not limited to): *Where are the “hot spots” of poor condition (i.e., places with high stressors and poor resources) or the “most vulnerable” (i.e., both high stressor levels and high numbers of resources) locations in the region? Which resources are at risk? What are the socioeconomic factors contributing to stressors and condition? How will conditions and vulnerabilities change in the future? Which areas will change in a negative direction environmentally because of future conditions? Which environmental problems are putting the Region at greatest risk?*

As mentioned in the Introduction section, regional vulnerability includes many aspects: rarity, synergy, sensitivity, spatial context, and history. In that context, the questions listed above are not separate but interwoven. There is no single method that will suffice to answer those questions. However, the combination of both clustering- and ranking-oriented methods described in this report is capable of answering several aspects of regional vulnerability. *Criticality Analysis*, *Stressor-Resource Overlay*, *Stressor-Resource Matrix*, and *State Space Analysis* are methods that adequately demonstrate the concept of relative risk/vulnerability stated above. Those methods can be used in combination with future scenario analysis to explore conditions and vulnerabilities that change in the future. Furthermore, *Clustering Analysis* and/or *SOM* can assist in explaining some spatial aspects of regional vulnerability (e.g., whether vulnerability is locally limited or spatially widespread). Note that the integration methods do not handle time series data (because no time series data are available). Additional analyses (e.g., time

series data analysis if data are available, experts' judgment) are needed to account for the temporal aspects of vulnerability (e.g., cumulative and aggregate stresses occurring over time can influence the prioritization of ecosystems regarding vulnerability).

### Planning/Restoration/Development

The assessment questions listed in the two sections above are also common questions being asked in environmental planning, restoration, and development. Other common questions include: *How applicable are risk management options to other locations in the region? What are socioeconomic and environmental costs and benefits associated with alternative management programs? What are the tradeoffs associated with alternative management programs?*

All of the integration methods in this report can be used in various tasks of planning, restoration, and development. For example, both clustering- and ranking-oriented methods are capable of pointing out which watersheds are at risk or in bad shape for protection or restoration purposes. Future scenario analysis is an important task in environmental planning and policy development and each method in this report can facilitate this task in its own fashion. For example, the *Change Analysis* directly serves scenario analysis for either a single variables or a whole data set. For the clustering-oriented methods (*Cluster Analysis*, *SOM*), scenario analysis can be carried out by comparing changes with the benchmark clustering pattern to see if a watershed is moving from the current cluster to a better one or the reverse in terms of environmental conditions. For the ranking-oriented and distance-based methods (e.g., *Simple Sum*, *PCA*, *State Space Analysis*, *Criticality Analysis*, *AHP*), rank or distance to some reference points (e.g., a pristine watershed) can be used to measure the magnitude of change from a future scenario in comparison with the status quo (i.e., the current ranking or distances). Among the integration methods, *AHP* is an explicitly designated multi-criteria decision-making model. Understandably, *AHP* has several advantages compared with other methods in exploring *socioeconomic and environmental tradeoffs associated with alternative management programs*. In addition to being able to organize a complex problem into a well-structured hierarchy, *AHP* can be expanded to include other social, cultural, and economic components (e.g., putting the hierarchy in this analysis into another larger hierarchy), moving from ecological assessment to social-economic-environmental-policy evaluation.

### Subjective Judgments/Expert Knowledge

An integrated assessment is often put in the context of a larger decision-making problem where variables or indicators are seldom weighted equally. For example, resources providing goods and services or directly benefiting human health might be considered by some to be more important than non-monetized resources, such as native biodiversity. Similarly, some stressors have larger impacts or affect a greater number of resources than others and thus should be given greater weights. User-specified weightings offer a means to communicate results of the assessment in terms of what is important to different groups with different values and a way to explore how weighting groups of stressors or resources more heavily than others affects overall patterns of vulnerability. Varying weights allows the user to explore the data by focusing on subsets of information to evaluate the contribution of these subsets to overall patterns. Comparison of resulting maps based on different weightings allows an assessment of tradeoffs associated with different management priorities. Within that context, the *AHP* provides a powerful and flexible framework to facilitate such a multiple-criteria decision-making process with multiple stakeholders. Weights for different variables or criteria can be derived and/or put into the model via pair-wise comparison (by ways of graphic, questionnaire, verbal, or matrix) or absolute measurements (direct data entry). The model can incorporate group judgments and derive combined weights from multiple stakeholders. Furthermore, the *AHP* helps to organize a complex problem into a well-structured hierarchy. It also can be expanded in the future to include other social, cultural, and economic components (e.g., putting the hierarchy in this analysis into another larger hierarchy), moving from ecological assessment to social-economic-environmental-policy evaluation.

## Combination of Integration Methods

Although the integration methods are presented as separate methods in this report, their combination to create a more comprehensive integration model is feasible and promising. For example, Tran et al. (2002) developed a fuzzy decision analysis method for integrating ecological indicators. This was a combination of the fuzzy ranking method (Section 13) with the *AHP* and the *PCA*. The method was capable of ranking ecosystems in terms of environmental conditions and suggesting cumulative impacts across a large region. Using a data set on land cover, population, roads, streams, air pollution, and topography of the Mid-Atlantic region which is very similar to the one in this report, the authors were able to point out areas which were in relatively poor condition and/or vulnerable to future deterioration. The method offered an easy and comprehensive way to combine the strengths of fuzzy set theory and the *AHP* for ecological assessment. Furthermore, the suggested method can serve as a building block for the evaluation of environmental policies. In another example, Tran et al. (2003) created another assessment model by combining the *SOM* and the *PCA*. The method is capable of clustering ecosystems in terms of environmental conditions and suggesting relative cumulative environmental impacts of multiple factors across a large region. We expect to develop other comprehensive models in the next phase of ReVA.

Clearly, a regional vulnerability assessment involves various aspects with many interwoven questions. Furthermore, there is no single method appropriate for a specific question. The set of questions should be examined with a set of methods in an integrated fashion. In that context, a classification of “method versus question” or the question of “what is the best method?” can be misleading and it is not recommended in this report.

## **Recommendations**

Following are recommendations for the use of the integration methods described in this report:

- Use a suite of integration methods: There is no universal integration method that can cover all tasks of an integrated environmental assessment. A method has advantages on some aspects but is disadvantaged on others. The use of multiple methods in a complementary manner will help a user look at the problem from different angles/perspectives. It also gives the user a better chance to detect whether a pattern/abnormality on the map is a real environmental signal or just an arbitrary view created by some “strange” calculation.
- Start with the simple methods (*Simple Sum*, *Best/Worst Quantiles*) first and move to other complicated ones later. It will help the user to have a general picture of the study area before involving in more complicated and detailed calculations (i.e., see the forest first before viewing the tree).
- Keep it simple: If several methods provide similar patterns and/or results, stick with the simple methods and drop off the complicated ones.
- Pay proper care to data: How data are coded or transformed has a big influence on the integration results. Try to balance between data transformation and data interpretation because data transformation can reduce some particular problems (e.g., log transform to reduce skewness), while causing difficulties in interpreting due to the combination of transformed variables used in combination with non-transformed ones.
- Note that the methods described in this report are products of the ReVA’s first phase only. ReVA will expand its focus to include more products that have broader application and will work directly with clients to develop tools that will support environmental decision-making.

---

## **Appendix A**

### **Data**

Table 1A. Correlation matrix for variables included in the evaluation of integration methods.

Huc.dims	SOFTWOODINV	SOFTWOODREM	HARDWOODINV	HARDWOODREM	IMPLCPCT	RDDENS	DISSOLVEDP	UINDEX	TOTALN	PSOIL
SOFTWOODINV	1.000	0.980	0.615	0.773	0.402	0.290	0.330	0.302	0.123	0.019
SOFTWOODREM	0.980	1.000	0.521	0.729	0.413	0.318	0.379	0.351	0.182	0.073
HARDWOODINV	0.615	0.521	1.000	0.722	0.199	0.082	0.123	0.142	0.017	-0.010
HARDWOODREM	0.773	0.729	0.722	1.000	0.411	0.304	0.343	0.328	0.138	0.040
IMPLCPCT	0.402	0.413	0.199	0.411	1.000	0.961	0.744	0.546	0.283	-0.125
RDDENS	0.290	0.318	0.082	0.304	0.961	1.000	0.826	0.670	0.475	0.097
DISSOLVEDP	0.330	0.379	0.123	0.343	0.744	0.826	1.000	0.958	0.847	0.582
UINDEX	0.302	0.351	0.142	0.328	0.546	0.670	0.958	1.000	0.923	0.766
TOTALN	0.123	0.182	0.017	0.138	0.283	0.475	0.847	0.923	1.000	0.897
PSOIL	0.019	0.073	-0.010	0.040	-0.125	0.097	0.582	0.766	0.897	1.000
EDGE2	0.379	0.408	0.081	0.216	0.164	0.284	0.517	0.567	0.537	0.519
INT2	0.488	0.526	0.181	0.378	0.526	0.579	0.746	0.752	0.565	0.454
INT65	0.368	0.387	0.058	0.169	0.245	0.328	0.434	0.448	0.376	0.317
MIGSCENARIO	0.171	0.180	0.253	0.309	0.263	0.399	0.654	0.704	0.680	0.614
N03DEPMODEL	-0.549	-0.495	-0.436	-0.487	-0.039	0.091	0.005	-0.015	0.060	0.048
S04DEPMODEL	-0.478	-0.434	-0.435	-0.510	-0.032	0.105	-0.002	-0.027	0.043	0.026
UVB	0.616	0.554	0.474	0.513	0.071	-0.012	0.051	0.072	-0.011	-0.017
NONCLIMAXPCT	0.316	0.280	0.260	0.434	-0.037	-0.042	0.067	0.087	0.085	0.081
NATCOVERPCT	0.410	0.380	0.308	0.460	0.119	0.004	0.134	0.115	0.043	-0.016
AGSL	-0.365	-0.306	-0.267	-0.451	-0.115	0.050	0.175	0.212	0.373	0.398
CROPSL	-0.341	-0.300	-0.140	-0.269	-0.085	0.075	0.166	0.204	0.332	0.343
RIPAG	-0.202	-0.152	-0.138	-0.259	-0.185	-0.047	0.295	0.411	0.581	0.672
EDGE65	-0.309	-0.343	-0.215	-0.375	-0.431	-0.365	-0.540	-0.568	-0.326	-0.291
RIPFOR	0.392	0.418	0.201	0.408	0.401	0.355	0.539	0.567	0.340	0.309
EMAGRIC	-0.040	0.023	-0.202	0.078	0.227	0.338	0.493	0.481	0.520	0.418
FUNGICIDE	0.188	0.226	0.053	0.312	0.017	0.061	0.323	0.430	0.398	0.466
HERBICIDE	0.264	0.282	0.216	0.356	-0.073	0.014	0.408	0.571	0.556	0.664
INSECTICIDE	0.265	0.250	0.312	0.401	-0.103	-0.065	0.241	0.363	0.357	0.447
NBLDPM97	0.291	0.267	0.247	0.303	-0.022	-0.003	0.232	0.301	0.302	0.335
OZONE8HR	0.417	0.439	0.008	0.260	0.233	0.284	0.432	0.465	0.357	0.328
POPGROWTH	0.280	0.249	0.217	0.281	0.097	0.102	0.281	0.292	0.317	0.265
SOFTCHIPMIL	0.203	0.142	0.366	0.265	-0.108	-0.157	-0.089	-0.077	-0.053	-0.032
TERRTE	0.496	0.497	0.375	0.616	0.279	0.182	0.313	0.322	0.141	0.102
WETLNDSPCT	-0.384	-0.360	-0.350	-0.514	0.056	0.125	-0.056	-0.138	-0.003	-0.126
SUM06	-0.175	-0.192	-0.081	-0.164	-0.173	-0.160	-0.134	-0.119	-0.048	0.002
HARDCHIPMIL	-0.217	-0.255	-0.058	-0.211	-0.226	-0.254	-0.346	-0.371	-0.274	-0.241
NTCMPPLM	-0.063	-0.111	0.078	-0.032	-0.320	-0.394	-0.459	-0.415	-0.378	-0.263
POV65	0.117	0.063	0.162	0.083	-0.287	-0.296	-0.322	-0.260	-0.223	-0.113
EMMINE	-0.175	-0.151	-0.260	-0.287	-0.057	-0.015	-0.257	-0.293	-0.290	-0.297
STRD	-0.155	-0.115	-0.273	-0.355	0.122	0.158	-0.087	-0.160	-0.147	-0.226
AQUAEXOTIC	0.011	0.014	-0.024	0.019	0.019	0.000	0.072	0.053	0.114	0.077
AQUANATIVE	-0.114	-0.114	0.144	0.190	0.092	0.102	0.133	0.173	0.102	0.115
AQUATE	0.184	0.213	-0.002	0.031	-0.066	-0.088	-0.102	-0.115	-0.087	-0.076
C5FS	0.344	0.354	0.314	0.477	0.390	0.302	0.445	0.426	0.312	0.206
DAMS	-0.188	-0.174	-0.188	-0.183	0.153	0.243	0.107	0.024	0.061	-0.056
FORCOVDEFOL	-0.040	0.007	-0.401	0.007	0.280	0.280	0.237	0.157	0.126	-0.010
INDTHPTH	0.258	0.211	0.262	0.329	0.150	0.066	0.051	0.033	-0.059	-0.093
POPDENS	0.207	0.227	0.032	0.203	0.494	0.476	0.496	0.398	0.338	0.120
TERREXOTIC	-0.131	-0.053	-0.324	-0.036	0.080	0.168	0.341	0.369	0.344	0.351
TERRNATIVE	-0.061	-0.066	-0.133	-0.274	-0.107	-0.045	-0.172	-0.154	-0.128	-0.088

Table 1A. Continued

Huc.dims	EDGE2	INT2	INT65	MIGSCENARIO	NO3DEPMODEL	SO4DEPMODEL	UVB	NONCLIMAXPCT	NATCOVERPCT	AGSL
SOFTWOODINV	0.379	0.488	0.368	0.171	-0.549	-0.478	0.616	0.316	0.410	-0.365
SOFTWOODREM	0.408	0.526	0.387	0.180	-0.495	-0.434	0.554	0.280	0.380	-0.306
HARDWOODINV	0.081	0.181	0.058	0.253	-0.436	-0.435	0.474	0.260	0.308	-0.267
HARDWOODREM	0.216	0.378	0.169	0.309	-0.487	-0.510	0.513	0.434	0.460	-0.451
IMPLCPCT	0.164	0.526	0.245	0.263	-0.039	-0.032	0.071	-0.037	0.119	-0.115
RDDENS	0.284	0.579	0.328	0.399	0.091	0.105	-0.012	-0.042	0.004	0.050
DISSOLVEDP	0.517	0.746	0.434	0.654	0.005	-0.002	0.051	0.067	0.134	0.175
UINDEX	0.567	0.752	0.448	0.704	-0.015	-0.027	0.072	0.087	0.115	0.212
TOTALN	0.537	0.565	0.376	0.680	0.060	0.043	-0.011	0.085	0.043	0.373
PSOIL	0.519	0.454	0.317	0.614	0.048	0.026	-0.017	0.081	-0.016	0.398
EDGE2	1.000	0.854	0.846	0.614	-0.065	-0.007	0.185	0.243	0.243	0.052
INT2	0.854	1.000	0.760	0.640	-0.060	-0.031	0.156	0.135	0.224	0.011
INT65	0.846	0.760	1.000	0.417	-0.118	-0.052	0.213	0.210	0.215	-0.016
MIGSCENARIO	0.614	0.640	0.417	1.000	0.073	0.034	0.022	0.271	0.297	0.137
NO3DEPMODEL	-0.065	-0.060	-0.118	0.073	1.000	0.955	-0.878	-0.618	-0.541	0.420
SO4DEPMODEL	-0.007	-0.031	-0.052	0.034	0.955	1.000	-0.773	-0.610	-0.536	0.443
UVB	0.185	0.156	0.213	0.022	-0.878	-0.773	1.000	0.710	0.623	-0.467
NONCLIMAXPCT	0.243	0.135	0.210	0.271	-0.618	-0.610	0.710	1.000	0.787	-0.443
NATCOVERPCT	0.243	0.224	0.215	0.297	-0.541	-0.536	0.623	0.787	1.000	-0.464
AGSL	0.052	0.011	-0.016	0.137	0.420	0.443	-0.467	-0.443	-0.464	1.000
CROPSL	0.115	0.063	0.087	0.254	0.436	0.400	-0.431	-0.245	-0.288	0.711
RIPAG	0.070	0.052	-0.019	0.128	0.166	0.169	-0.209	-0.272	-0.313	0.700
EDGE65	-0.220	-0.592	0.007	-0.391	-0.004	0.044	0.008	0.091	-0.131	0.101
RIPFOR	0.352	0.634	0.303	0.353	-0.124	-0.148	0.118	0.013	0.210	-0.092
EMAGRIC	0.250	0.301	0.142	0.369	0.107	0.018	-0.127	0.141	0.086	0.237
FUNGICIDE	0.352	0.353	0.261	0.359	-0.185	-0.227	0.225	0.246	0.297	-0.280
HERBICIDE	0.433	0.443	0.290	0.483	-0.192	-0.235	0.253	0.289	0.322	-0.305
INSECTICIDE	0.318	0.284	0.189	0.390	-0.422	-0.435	0.540	0.519	0.514	-0.388
NBLDPM97	0.231	0.219	0.131	0.437	-0.448	-0.458	0.411	0.451	0.369	-0.180
OZONE8HR	0.595	0.626	0.523	0.346	-0.045	0.033	0.171	0.083	0.196	-0.182
POPGROWTH	0.161	0.142	0.105	0.413	-0.376	-0.391	0.301	0.341	0.269	-0.052
SOFTCHIPMIL	0.104	-0.005	0.085	0.276	-0.537	-0.539	0.542	0.608	0.498	-0.258
TERRTE	0.157	0.357	0.068	0.249	-0.348	-0.381	0.434	0.416	0.486	-0.485
WETLNDSPCT	-0.257	-0.286	-0.147	-0.221	0.406	0.412	-0.533	-0.530	-0.520	0.588
SUM06	-0.254	-0.254	-0.221	-0.085	0.014	0.021	-0.027	-0.060	-0.050	0.270
HARDCHIPMIL	-0.431	-0.477	-0.296	-0.241	-0.053	-0.049	0.002	-0.004	-0.091	0.133
NTCMPPLM	-0.333	-0.395	-0.344	-0.391	-0.246	-0.225	0.185	0.099	0.035	-0.237
POV65	-0.066	-0.201	-0.009	-0.296	-0.417	-0.341	0.485	0.324	0.209	-0.214
EMMINE	-0.194	-0.240	-0.034	-0.470	0.201	0.274	-0.144	-0.291	-0.290	0.117
STRD	-0.271	-0.213	-0.082	-0.540	0.190	0.234	-0.218	-0.461	-0.472	0.279
AQUAEXOTIC	-0.162	-0.134	-0.186	0.077	-0.228	-0.290	0.064	0.078	0.023	0.305
AQUANATIVE	-0.076	0.041	-0.231	0.371	0.052	-0.049	-0.157	-0.020	0.048	0.001
AQUATE	-0.010	-0.047	-0.021	-0.295	-0.158	-0.124	0.182	-0.009	-0.016	0.057
C5FS	0.232	0.387	0.203	0.446	-0.099	-0.173	0.081	0.153	0.388	-0.217
DAMS	0.078	0.038	0.066	0.097	0.416	0.429	-0.349	-0.269	-0.143	0.235
FORCOVDEFOL	0.100	0.159	0.063	-0.130	0.074	0.065	-0.111	0.029	0.007	-0.089
INDTHPTH	0.103	0.109	0.147	0.125	-0.281	-0.247	0.378	0.349	0.390	-0.468
POPDENS	0.215	0.301	0.148	0.284	-0.082	-0.090	0.053	0.109	0.131	0.089
TERREXOTIC	0.175	0.286	0.017	0.320	0.402	0.304	-0.420	-0.131	-0.102	0.063
TERRNATIVE	0.253	0.083	0.308	-0.158	0.194	0.311	-0.157	-0.280	-0.230	0.128



Table 1A. Continued

Huc.dims	CROPSL	RIPAG	EDGE65	RIPFOR	EMAGRIC	FUNGICIDE	HERBICIDE	INSECTICIDE	NBLDPM97	OZONE8HR	POPGROWTH
SOFTWOODINV	-0.341	-0.202	-0.309	0.392	-0.040	0.188	0.264	0.265	0.291	0.417	0.280
SOFTWOODREM	-0.300	-0.152	-0.343	0.418	0.023	0.226	0.282	0.250	0.267	0.439	0.249
HARDWOODINV	-0.140	-0.138	-0.215	0.201	-0.202	0.053	0.216	0.312	0.247	0.008	0.217
HARDWOODREM	-0.269	-0.259	-0.375	0.408	0.078	0.312	0.356	0.401	0.303	0.260	0.281
IMPLCPCT	-0.085	-0.185	-0.431	0.401	0.227	0.017	-0.073	-0.103	-0.022	0.233	0.097
RDDENS	0.075	-0.047	-0.365	0.355	0.338	0.061	0.014	-0.065	-0.003	0.284	0.102
DISSOLVEDP	0.166	0.295	-0.540	0.539	0.493	0.323	0.408	0.241	0.232	0.432	0.281
UINDEX	0.204	0.411	-0.568	0.567	0.481	0.430	0.571	0.363	0.301	0.465	0.292
TOTALN	0.332	0.581	-0.326	0.340	0.520	0.398	0.556	0.357	0.302	0.357	0.317
PSOIL	0.343	0.672	-0.291	0.309	0.418	0.466	0.664	0.447	0.335	0.328	0.265
EDGE2	0.115	0.070	-0.220	0.352	0.250	0.352	0.433	0.318	0.231	0.595	0.161
INT2	0.063	0.052	-0.592	0.634	0.301	0.353	0.443	0.284	0.219	0.626	0.142
INT65	0.087	-0.019	0.007	0.303	0.142	0.261	0.290	0.189	0.131	0.523	0.105
MIGSCENARIO	0.254	0.128	-0.391	0.353	0.369	0.359	0.483	0.390	0.437	0.346	0.413
N03DEPMODEL	0.436	0.166	-0.004	-0.124	0.107	-0.185	-0.192	-0.422	-0.448	-0.045	-0.376
S04DEPMODEL	0.400	0.169	0.044	-0.148	0.018	-0.227	-0.235	-0.435	-0.458	0.033	-0.391
UVB	-0.431	-0.209	0.008	0.118	-0.127	0.225	0.253	0.540	0.411	0.171	0.301
NONCLIMAXPCT	-0.245	-0.272	0.091	0.013	0.141	0.246	0.289	0.519	0.451	0.083	0.341
NATCOVERPCT	-0.288	-0.313	-0.131	0.210	0.086	0.297	0.322	0.514	0.369	0.196	0.269
AGSL	0.711	0.700	0.101	-0.092	0.237	-0.280	-0.305	-0.388	-0.180	-0.182	-0.052
CROPSL	1.000	0.421	0.113	-0.095	0.316	-0.193	-0.117	-0.248	-0.069	-0.065	-0.001
RIPAG	0.421	1.000	-0.071	0.203	0.268	0.058	0.181	0.056	0.042	-0.077	0.045
EDGE65	0.113	-0.071	1.000	-0.734	-0.211	-0.323	-0.467	-0.278	-0.175	-0.400	-0.064
RIPFOR	-0.095	0.203	-0.734	1.000	0.166	0.329	0.518	0.249	0.270	0.466	0.127
EMAGRIC	0.316	0.268	-0.211	0.166	1.000	0.163	0.234	0.146	0.199	0.206	0.381
FUNGICIDE	-0.193	0.058	-0.323	0.329	0.163	1.000	0.614	0.684	0.160	0.329	0.094
HERBICIDE	-0.117	0.181	-0.467	0.518	0.234	0.614	1.000	0.742	0.418	0.495	0.198
INSECTICIDE	-0.248	0.056	-0.278	0.249	0.146	0.684	0.742	1.000	0.323	0.165	0.175
NBLDPM97	-0.069	0.042	-0.175	0.270	0.199	0.160	0.418	0.323	1.000	0.292	0.700
OZONE8HR	-0.065	-0.077	-0.400	0.466	0.206	0.329	0.495	0.165	0.292	1.000	0.248
POPGROWTH	-0.001	0.045	-0.064	0.127	0.381	0.094	0.198	0.175	0.700	0.248	1.000
SOFTCHIPMIL	-0.129	-0.270	0.168	-0.205	-0.007	0.068	0.042	0.305	0.408	-0.110	0.396
TERRTE	-0.321	-0.175	-0.553	0.609	0.151	0.368	0.513	0.464	0.358	0.325	0.210
WETLNDSPCT	0.457	0.186	0.373	-0.500	0.040	-0.453	-0.610	-0.680	-0.317	-0.225	-0.089
SUM06	0.241	0.133	0.099	-0.234	0.374	-0.105	-0.190	-0.105	0.017	-0.078	0.249
HARDCHIPMIL	0.153	-0.002	0.382	-0.416	0.066	-0.300	-0.369	-0.198	-0.066	-0.453	0.127
NTCMPPLM	-0.278	-0.215	0.159	-0.301	-0.354	0.098	-0.118	0.085	-0.074	-0.209	-0.127
POV65	-0.240	-0.122	0.284	-0.219	-0.346	0.152	0.016	0.300	-0.131	-0.130	-0.227
EMMINE	0.104	-0.129	0.304	-0.319	-0.006	-0.283	-0.333	-0.333	-0.399	-0.065	-0.316
STRD	0.131	0.155	0.243	-0.224	-0.046	-0.313	-0.402	-0.386	-0.469	-0.205	-0.406
AQUAEXOTIC	0.197	0.145	0.047	-0.112	0.336	-0.127	-0.225	-0.185	0.223	-0.112	0.417
AQUANATIVE	0.109	-0.047	-0.345	0.291	0.029	0.137	0.205	0.082	0.160	-0.051	0.066
AQUATE	-0.075	0.000	0.087	-0.132	-0.014	-0.070	-0.110	-0.063	-0.029	0.015	-0.092
C5FS	-0.079	-0.049	-0.400	0.412	0.301	0.316	0.369	0.300	0.332	0.276	0.349
DAMS	0.275	0.029	0.119	-0.083	0.209	-0.142	-0.196	-0.209	-0.221	0.036	-0.077
FORCOVDEFOL	-0.109	0.001	-0.133	0.081	0.261	0.084	-0.040	-0.071	-0.083	0.195	-0.050
INDTHPTH	-0.295	-0.346	-0.037	-0.021	-0.059	0.284	0.198	0.347	-0.010	0.138	0.082
POPDENS	0.032	0.045	-0.212	0.173	0.581	0.042	0.003	-0.043	0.125	0.246	0.491
TERREXOTIC	0.060	0.146	-0.475	0.417	0.492	0.184	0.387	0.020	0.119	0.406	0.134
TERRNATIVE	0.133	-0.032	0.264	-0.140	-0.378	-0.133	-0.163	-0.279	-0.231	0.171	-0.277

Table 1A. Continued

Huc.dims	SOFTCHIPMIL	TERRTE	WETLNDSPCT	SUM06	HARDCHIPMIL	NTCMPPLM	POV65	EMMINE	STRD	AQUAEXOTIC
SOFTWOODINV	0.203	0.496	-0.384	-0.175	-0.217	-0.063	0.117	-0.175	-0.155	0.011
SOFTWOODREM	0.142	0.497	-0.360	-0.192	-0.255	-0.111	0.063	-0.151	-0.115	0.014
HARDWOODINV	0.366	0.375	-0.350	-0.081	-0.058	0.078	0.162	-0.260	-0.273	-0.024
HARDWOODREM	0.265	0.616	-0.514	-0.164	-0.211	-0.032	0.083	-0.287	-0.355	0.019
IMPLCPCT	-0.108	0.279	0.056	-0.173	-0.226	-0.320	-0.287	-0.057	0.122	0.019
RDDENS	-0.157	0.182	0.125	-0.160	-0.254	-0.394	-0.296	-0.015	0.158	0.000
DISSOLVEDP	-0.089	0.313	-0.056	-0.134	-0.346	-0.459	-0.322	-0.257	-0.087	0.072
UINDEX	-0.077	0.322	-0.138	-0.119	-0.371	-0.415	-0.260	-0.293	-0.160	0.053
TOTALN	-0.053	0.141	-0.003	-0.048	-0.274	-0.378	-0.223	-0.290	-0.147	0.114
PSOIL	-0.032	0.102	-0.126	0.002	-0.241	-0.263	-0.113	-0.297	-0.226	0.077
EDGE2	0.104	0.157	-0.257	-0.254	-0.431	-0.333	-0.066	-0.194	-0.271	-0.162
INT2	-0.005	0.357	-0.286	-0.254	-0.477	-0.395	-0.201	-0.240	-0.213	-0.134
INT65	0.085	0.068	-0.147	-0.221	-0.296	-0.344	-0.009	-0.034	-0.082	-0.186
MIGSCENARIO	0.276	0.249	-0.221	-0.085	-0.241	-0.391	-0.296	-0.470	-0.540	0.077
N03DEPMODEL	-0.537	-0.348	0.406	0.014	-0.053	-0.246	-0.417	0.201	0.190	-0.228
S04DEPMODEL	-0.539	-0.381	0.412	0.021	-0.049	-0.225	-0.341	0.274	0.234	-0.290
UVB	0.542	0.434	-0.533	-0.027	0.002	0.185	0.485	-0.144	-0.218	0.064
NONCLIMAXPCT	0.608	0.416	-0.530	-0.060	-0.004	0.099	0.324	-0.291	-0.461	0.078
NATCOVERPCT	0.498	0.486	-0.520	-0.050	-0.091	0.035	0.209	-0.290	-0.472	0.023
AGSL	-0.258	-0.485	0.588	0.270	0.133	-0.237	-0.214	0.117	0.279	0.305
CROPSL	-0.129	-0.321	0.457	0.241	0.153	-0.278	-0.240	0.104	0.131	0.197
RIPAG	-0.270	-0.175	0.186	0.133	-0.002	-0.215	-0.122	-0.129	0.155	0.145
EDGE65	0.168	-0.553	0.373	0.099	0.382	0.159	0.284	0.304	0.243	0.047
RIPFOR	-0.205	0.609	-0.500	-0.234	-0.416	-0.301	-0.219	-0.319	-0.224	-0.112
EMAGRIC	-0.007	0.151	0.040	0.374	0.066	-0.354	-0.346	-0.006	-0.046	0.336
FUNGICIDE	0.068	0.368	-0.453	-0.105	-0.300	0.098	0.152	-0.283	-0.313	-0.127
HERBICIDE	0.042	0.513	-0.610	-0.190	-0.369	-0.118	0.016	-0.333	-0.402	-0.225
INSECTICIDE	0.305	0.464	-0.680	-0.105	-0.198	0.085	0.300	-0.333	-0.386	-0.185
NBLDPM97	0.408	0.358	-0.317	0.017	-0.066	-0.074	-0.131	-0.399	-0.469	0.223
OZONE8HR	-0.110	0.325	-0.225	-0.078	-0.453	-0.209	-0.130	-0.065	-0.205	-0.112
POPGROWTH	0.396	0.210	-0.089	0.249	0.127	-0.127	-0.227	-0.316	-0.406	0.417
SOFTCHIPMIL	1.000	0.102	-0.216	0.186	0.299	0.100	0.230	-0.300	-0.417	0.216
TERRTE	0.102	1.000	-0.722	-0.109	-0.218	0.008	0.013	-0.317	-0.422	-0.024
WETLNDSPCT	-0.216	-0.722	1.000	0.275	0.280	0.019	-0.142	0.316	0.452	0.311
SUM06	0.186	-0.109	0.275	1.000	0.758	0.132	0.097	0.247	-0.015	0.546
HARDCHIPMIL	0.299	-0.218	0.280	0.758	1.000	0.146	0.162	0.187	0.023	0.416
NTCMPPLM	0.100	0.008	0.019	0.132	0.146	1.000	0.671	-0.068	-0.137	0.017
POV65	0.230	0.013	-0.142	0.097	0.162	0.671	1.000	-0.032	-0.056	-0.034
EMMINE	-0.300	-0.317	0.316	0.247	0.187	-0.068	-0.032	1.000	0.630	-0.072
STRD	-0.417	-0.422	0.452	-0.015	0.023	-0.137	-0.056	0.630	1.000	-0.125
AQUAEXOTIC	0.216	-0.024	0.311	0.546	0.416	0.017	-0.034	-0.072	-0.125	1.000
AQUANATIVE	-0.044	0.286	-0.139	-0.126	-0.167	-0.031	-0.172	-0.205	-0.341	0.088
AQUATE	-0.030	-0.063	0.080	0.060	0.024	0.138	0.177	0.213	0.260	0.112
C5FS	0.080	0.450	-0.373	-0.047	-0.166	-0.266	-0.365	-0.301	-0.358	-0.057
DAMS	-0.168	-0.233	0.255	0.062	-0.042	-0.234	-0.207	0.268	0.207	-0.064
FORCOVDEFOL	-0.182	0.063	0.035	-0.128	-0.193	-0.104	-0.133	-0.012	0.120	-0.017
INDTHPTH	0.316	0.251	-0.372	-0.084	-0.042	0.092	0.172	-0.081	-0.241	-0.093
POPDENS	-0.023	0.134	0.086	0.193	-0.087	-0.276	-0.303	-0.014	-0.022	0.379
TERREXOTIC	-0.371	0.316	-0.152	0.048	-0.259	-0.299	-0.474	-0.086	-0.224	0.067
TERRNATIVE	-0.240	-0.513	0.302	-0.351	-0.241	0.012	0.069	0.192	0.270	-0.416

Table 1A. Continued

Huc.dims	AQUANATIVE	AQUATE	C5FS	DAMS	FORCOVDEFOL	INDTHPTH	POPDENS	TERREXOTIC	TERRNATIVE
SOFTWOODINV	-0.114	0.184	0.344	-0.188	-0.040	0.258	0.207	-0.131	-0.061
SOFTWOODREM	-0.114	0.213	0.354	-0.174	0.007	0.211	0.227	-0.053	-0.066
HARDWOODINV	0.144	-0.002	0.314	-0.188	-0.401	0.262	0.032	-0.324	-0.133
HARDWOODREM	0.190	0.031	0.477	-0.183	0.007	0.329	0.203	-0.036	-0.274
IMPLCPCT	0.092	-0.066	0.390	0.153	0.280	0.150	0.494	0.080	-0.107
RDDENS	0.102	-0.088	0.302	0.243	0.280	0.066	0.476	0.168	-0.045
DISSOLVEDP	0.133	-0.102	0.445	0.107	0.237	0.051	0.496	0.341	-0.172
UINDEX	0.173	-0.115	0.426	0.024	0.157	0.033	0.398	0.369	-0.154
TOTALN	0.102	-0.087	0.312	0.061	0.126	-0.059	0.338	0.344	-0.128
PSOIL	0.115	-0.076	0.206	-0.056	-0.010	-0.093	0.120	0.351	-0.088
EDGE2	-0.076	-0.010	0.232	0.078	0.100	0.103	0.215	0.175	0.253
INT2	0.041	-0.047	0.387	0.038	0.159	0.109	0.301	0.286	0.083
INT65	-0.231	-0.021	0.203	0.066	0.063	0.147	0.148	0.017	0.308
MIGSCENARIO	0.371	-0.295	0.446	0.097	-0.130	0.125	0.284	0.320	-0.158
N03DEPMODEL	0.052	-0.158	-0.099	0.416	0.074	-0.281	-0.082	0.402	0.194
S04DEPMODEL	-0.049	-0.124	-0.173	0.429	0.065	-0.247	-0.090	0.304	0.311
UVB	-0.157	0.182	0.081	-0.349	-0.111	0.378	0.053	-0.420	-0.157
NONCLIMAXPCT	-0.020	-0.009	0.153	-0.269	0.029	0.349	0.109	-0.131	-0.280
NATCOVERPCT	0.048	-0.016	0.388	-0.143	0.007	0.390	0.131	-0.102	-0.230
AGSL	0.001	0.057	-0.217	0.235	-0.089	-0.468	0.089	0.063	0.128
CROPSL	0.109	-0.075	-0.079	0.275	-0.109	-0.295	0.032	0.060	0.133
RIPAG	-0.047	0.000	-0.049	0.029	0.001	-0.346	0.045	0.146	-0.032
EDGE65	-0.345	0.087	-0.400	0.119	-0.133	-0.037	-0.212	-0.475	0.264
RIPFOR	0.291	-0.132	0.412	-0.083	0.081	-0.021	0.173	0.417	-0.140
EMAGRIC	0.029	-0.014	0.301	0.209	0.261	-0.059	0.581	0.492	-0.378
FUNGICIDE	0.137	-0.070	0.316	-0.142	0.084	0.284	0.042	0.184	-0.133
HERBICIDE	0.205	-0.110	0.369	-0.196	-0.040	0.198	0.003	0.387	-0.163
INSECTICIDE	0.082	-0.063	0.300	-0.209	-0.071	0.347	-0.043	0.020	-0.279
NBLDPM97	0.160	-0.029	0.332	-0.221	-0.083	-0.010	0.125	0.119	-0.231
OZONE8HR	-0.051	0.015	0.276	0.036	0.195	0.138	0.246	0.406	0.171
POPGROWTH	0.066	-0.092	0.349	-0.077	-0.050	0.082	0.491	0.134	-0.277
SOFTCHIPMIL	-0.044	-0.030	0.080	-0.168	-0.182	0.316	-0.023	-0.371	-0.240
TERRTE	0.286	-0.063	0.450	-0.233	0.063	0.251	0.134	0.316	-0.513
WETLNDSPCT	-0.139	0.080	-0.373	0.255	0.035	-0.372	0.086	-0.152	0.302
SUM06	-0.126	0.060	-0.047	0.062	-0.128	-0.084	0.193	0.048	-0.351
HARDCHIPMIL	-0.167	0.024	-0.166	-0.042	-0.193	-0.042	-0.087	-0.259	-0.241
NTCMPPLM	-0.031	0.138	-0.266	-0.234	-0.104	0.092	-0.276	-0.299	0.012
POV65	-0.172	0.177	-0.365	-0.207	-0.133	0.172	-0.303	-0.474	0.069
EMMINE	-0.205	0.213	-0.301	0.268	-0.012	-0.081	-0.014	-0.086	0.192
STRD	-0.341	0.260	-0.358	0.207	0.120	-0.241	-0.022	-0.224	0.270
AQUAEXOTIC	0.088	0.112	-0.057	-0.064	-0.017	-0.093	0.379	0.067	-0.416
AQUANATIVE	1.000	-0.539	0.187	-0.047	-0.137	-0.073	0.025	0.317	-0.154
AQUATE	-0.539	1.000	-0.140	-0.092	0.013	-0.093	-0.035	-0.196	0.001
C5FS	0.187	-0.140	1.000	0.027	0.019	0.197	0.234	0.292	-0.289
DAMS	-0.047	-0.092	0.027	1.000	0.141	-0.040	0.110	0.127	0.083
FORCOVDEFOL	-0.137	0.013	0.019	0.141	1.000	-0.023	0.218	0.224	-0.028
INDTHPTH	-0.073	-0.093	0.197	-0.040	-0.023	1.000	0.102	-0.144	-0.139
POPDENS	0.025	-0.035	0.234	0.110	0.218	0.102	1.000	0.278	-0.233
TERREXOTIC	0.317	-0.196	0.292	0.127	0.224	-0.144	0.278	1.000	-0.332
TERRNATIVE	-0.154	0.001	-0.289	0.083	-0.028	-0.139	-0.233	-0.332	1.000

**Table 2A. Variables representing resources in analyses.**

<b>NAME</b>	<b>Display Name</b>	<b>Description</b>
<b>C5FS</b>	Children (0-5) in families & subfamilies	Children (0-5) in families & subfamilies
<b>EMAGRIC</b>	Employed persons in agriculture, forestry, fisheries	Employed persons by industry - agriculture, forestry, fisheries 1990
<b>EMMINE</b>	Employed persons in mining 1990	Employed persons by industry – mining 1990
<b>HARDWOODINV</b>	Index values for hardwood inventory	Index values for hardwood forest inventory. The index compares a baseline of most recently available FIA data against projections to 2020. Index values > 1 are areas with increasing inventory.
<b>HARDWOODREM</b>	Index values for hardwood removals	Index values for hardwood removals. The index compares a baseline of most recently available FIA data against projections to 2020. Index values > 1 are areas with increasing inventory.
<b>INDTHPTH</b>	Infant deaths per 1,000 live births 1990	Infant deaths per 1,000 live births 1990
<b>INT2</b>	Forest interior habitat at the 2 ha scale	Percentage of forest habitat called interior (2 ha scale)
<b>INT65</b>	Forest interior habitat at the 65 ha scale	Percentage of forest habitat called interior (65 ha scale)
<b>MIGSCENARIO</b>	Migratory scenarios that use area	The number of migratory scenarios for long-distance forest migrants that use a particular HUC or hexagon. Scenarios are defined by a combination of compass heading, landfall location along the gulf coast and southern Atlantic Coast, and nightly flight distance
<b>NTCMPPLM</b>	Incomplete plumbing	Incomplete plumbing
<b>POV65</b>	65+ below poverty	65+ below poverty
<b>PSOIL</b>	Soil loss potential	Proportion of watershed with potential soil loss greater than 1 ton per acre per year; the percentage of HUC or hexagon area that is estimated to lose more than 1 ton/acre/year of soil due to erosion
<b>RIPFOR</b>	Forest land cover along streams	Proportion of total stream length with adjacent forest land cover; % riparian buffer that is forest
<b>SOFTWOODINV</b>	Index values for softwood inventory	Index values for softwood forest inventory. The index compares a baseline of most recently available FIA data against projections to 2020. Index values > 1 are areas with increasing inventory.
<b>SOFTWOODREM</b>	Index values for softwood removals	Index values for softwood removals. The index compares a baseline of most recently available FIA data against projections to 2020. Index values > 1 are areas with increasing inventory.
<b>WETLNDSPCT</b>	Percent wetlands land cover	Percent of area classified as wetlands
<b>AQUANATIVE</b>	Native aquatic species	Count of native aquatic – fish and mussels – species
<b>AQUATE</b>	Threatened and endangered aquatic species	Count of threatened and endangered aquatics – fish and mussels species
<b>TERRNATIVE</b>	Native terrestrial species	Count of native birds, mammals, butterflies, amphibians, and reptiles
<b>TERRTE</b>	Threatened and endangered terrestrial species	Count of threatened and endangered birds, mammals, butterflies, amphibians, and reptiles
<b>NONCLIMAXPCT</b>	Percent coverage of non-climax forest	Percent coverage with FOREST but the species are not the climax listed by Kuchler
<b>NATCOVERPCT</b>	Percent coverage of natural forest	Percent coverage with FOREST that matches potential vegetation in Kuchler

**Table 3A. Variables representing stressors in analyses.**

NAME	Display Name	Description
AGSL	Agriculture land on steep slopes	Proportion of watershed with agriculture land cover on slopes that are greater than 3%
CROPSL	Crop land on steep slopes	Proportion of watershed with crop land cover on slopes that are greater than 3%
DAMS	Impoundment density	Impoundment density (number of dams per 1,000 kilometers of stream length)
DISSOLVEDP	Dissolved phosphorus	Estimated suspended sediment in streams modelled using land cover metrics
EDGE2	Forest edge habitat at the 2 ha scale	Percentage of forest habitat called edge (2 ha scale)
EDGE65	Forest edge habitat at the 65 ha scale	Percentage of forest habitat called edge (65 ha scale)
FUNGICIDE	Annual fungicide loadings	Annual fungicide loadings
HARDCHIPMIL	Chip mill capacity for hardwoods	Estimate of increase (decrease) in chip mill for hardwoods capacity in tons, based on our regression, and assuming the Mid-Atlantic behaves like the South
HERBICIDE	Annual atrazine loadings 1990-93	Annual atrazine loadings 1990-93
IMPLCPCT	Percent impervious land cover	Percent impervious surface by land cover
INSECTICIDE	Annual O-P insecticides loadings 1990-93	Annual O-P Insecticides loadings 1990-93
NBLDPM97	New private housing building permits 1997	New private housing building permits 1997
NO3DEPMODEL	Nitrate wet deposition – modeled	Modeled annual wet deposition of nitrate based on averages from 1987-1999
OZONE8HR	Ozone - 8 hr max	Ozone (8 hr max) is a human health indicator and is given in parts per billion (ppb)
POPDENS	Population density – 1995	Population density
POPGROWTH	Annual population growth rate 1990-1995	Population growth rate from 1990-1995
RDDENS	Road density	The density numbers are meters of road per hectare of area
RIPAG	Agriculture land cover along streams	Proportion of total stream length with adjacent agriculture land cover; % riparian buffer that is agricultural land
SO4DEPMODEL	Sulfate wet deposition – modeled	Modeled annual wet deposition of sulfate based on averages from 1987-1999
SOFTCHIPMIL	Chip mill capacity for softwoods	Estimate of increase (decrease) in chip mill for softwoods capacity in tons, based on our regression, and assuming the Mid-Atlantic behaves like the South
STRD	Roads crossing streams	Number of road crossings per total stream length
SUM06	Ozone – sum 06	Cumulative sum of all hourly ozone concentrations equal to or above 0.06 ppm (or 60 ppb) for hours between 7 a.m. and 7 p.m. The SUM06 index is an indicator of ozone exposure that plants receive during daylight hours.
TOTALN	Nitrogen in surface water	Estimated total nitrogen in streams modelled using land cover metrics
UINDEX	Human use index	Human use index (proportion of watershed area with agriculture or urban land cover)
UVB	Mean annual UV-B irradiance	Mean Annual UV-B Irradiance
FORCOVDEFOL	Pct forest cover defoliated as pct of existing forest	Percent of forest cover defoliated and with mortality as proportion of existing forest
AQUAEXOTIC	Introduced (exotic) aquatic species	Count of exotic aquatic – fish and mussels – species.
TERREXOTIC	Introduced (exotic) terrestrial species	Count of exotic birds, mammals, butterflies, amphibians, and reptiles

---

## Appendix B

### Calculations

#### *Tran and Duckstein's Fuzzy Ranking Method*

The fuzzy ranking method developed by Tran and Duckstein is based on a distance measure for fuzzy numbers (FNs), which in turn is established on a distance measure for interval numbers (INs) as follows:

##### Distance measure for interval numbers

Let  $F(\mathbf{R})$  be the set of INs in  $\mathbf{R}$  and the distance between two INs  $A(a_1, a_2)$  and  $B(b_1, b_2)$  be defined as (Tran and Duckstein, in press):

$$\begin{aligned} D^2(A, B) &= \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \left\{ \left[ \left( \frac{a_1 + a_2}{2} \right) + x(a_2 - a_1) \right] - \left[ \left( \frac{b_1 + b_2}{2} \right) + y(b_2 - b_1) \right] \right\}^2 dx dy \\ &= \left[ \left( \frac{a_1 + a_2}{2} \right) - \left( \frac{b_1 + b_2}{2} \right) \right]^2 + \frac{1}{3} \left[ \left( \frac{a_2 - a_1}{2} \right)^2 + \left( \frac{b_2 - b_1}{2} \right)^2 \right] \end{aligned}$$

##### Distance measure for fuzzy numbers

To be able to deal with curvilinear membership functions, generalized left right fuzzy numbers (GLRFN) of Dubois and Prade (1980) as described by Bárdossy and Duckstein (1995) are defined first. A fuzzy set  $A = (a_1, a_2, a_3, a_4)$  is called a GLRFN if its membership function satisfies the following:

$$\mu(x) = \begin{cases} L\left(\frac{a_2 - x}{a_2 - a_1}\right) & \text{for } a_1 \leq x \leq a_2 \\ 1 & \text{for } a_2 \leq x \leq a_3 \\ R\left(\frac{x - a_3}{a_4 - a_3}\right) & \text{for } a_3 \leq x \leq a_4 \\ 0 & \text{else} \end{cases}$$

where  $L$  and  $R$  are strictly decreasing functions defined on  $[0, 1]$  and satisfying the conditions:

$$L(x) = R(x) = 1 \quad \text{if } x \leq 0 \quad \text{and} \quad L(x) = R(x) = 0 \quad \text{if } x \geq 1$$

For  $a_2 = a_3$ , we have the classical definition of left right fuzzy numbers (LRFN) of Dubois and Prade (1980). Trapezoidal fuzzy numbers (TrFN) are special cases of GLRFN with  $L(x) = R(x) = 1 - x$ . Triangular fuzzy numbers (TFN) are also special cases of GLRFN with  $L(x) = R(x) = 1 - x$  and  $a_2 = a_3$ .

A GLRFN  $A$  is denoted as:

$$A = (a_1, a_2, a_3, a_4)_{L-A-R-A}$$

and an  $\alpha$ -level interval of fuzzy number  $A$  as:

$$A(\alpha) = (A_L(\alpha), A_U(\alpha)) = \left( a_2 - (a_2 - a_1)L_A^{-1}(\alpha), a_3 + (a_4 - a_3)R_A^{-1}(\alpha) \right)$$

Let  $F(\mathbf{R})$  be the set of GLRFNs in  $\mathbf{R}$ . Using the distance measure for interval numbers defined above, a distance between two GLRFNs  $A$  and  $B$  can be defined as:

$$D^2(A, B, f) = \left\langle \int_0^1 \left\{ \left[ \left( \frac{A_L(\alpha) + A_U(\alpha)}{2} \right) - \left( \frac{B_L(\alpha) + B_U(\alpha)}{2} \right) \right]^2 + \right. \right. \\ \left. \left. \frac{1}{3} \left[ \left( \frac{A_U(\alpha) - A_L(\alpha)}{2} \right)^2 + \left( \frac{B_U(\alpha) - B_L(\alpha)}{2} \right)^2 \right] \right\} f(\alpha) d\alpha \right\rangle / \int_0^1 f(\alpha) d\alpha$$

Here  $f$ , which serves as a weighting function, is a continuous positive function defined on  $[0, 1]$ . The distance is a weighted sum (integral) of the distances between two intervals at all  $\alpha$  levels from 0 to 1. It is reasonable to choose  $f$  as an increasing function, indicating greater weight assigned to the distance between two intervals at a higher  $\alpha$  level. The equations to compute distance for some of the commonly used fuzzy numbers with two different weighting functions ( $f(\alpha)=1$  representing equal weights for intervals at different  $\alpha$  levels and  $f(\alpha)=\alpha$  indicating more weight given to intervals at higher  $\alpha$  level) are presented in Table B1.

**Table B1. Distance functions for some commonly used fuzzy numbers.**

Fuzzy numbers	$f(\alpha)$	$D_T^2(A, B, f)$
<b>Trapezoidal fuzzy numbers</b> $A = (a_1, a_2, a_3, a_4)_{Tr}$ $B = (b_1, b_2, b_3, b_4)_{Tr}$	$\alpha$	$\left(\frac{a_2 + a_3}{2} - \frac{b_2 + b_3}{2}\right)^2 + \frac{1}{3}\left(\frac{a_2 + a_3}{2} - \frac{b_2 + b_3}{2}\right)[(a_4 - a_3) - (a_2 - a_1) - (b_4 - b_3) + (b_2 - b_1)]$ $+ \frac{2}{3}\left(\frac{a_3 - a_2}{2}\right)^2 + \frac{1}{9}\left(\frac{a_3 - a_2}{2}\right)[(a_4 - a_3) + (a_2 - a_1)] + \frac{2}{3}\left(\frac{b_3 - b_2}{2}\right)^2 + \frac{1}{9}\left(\frac{b_3 - b_2}{2}\right)[(b_4 - b_3) + (b_2 - b_1)] +$ $\frac{1}{18}[(a_4 - a_3)^2 + (a_2 - a_1)^2 + (b_4 - b_3)^2 + (b_2 - b_1)^2] - \frac{1}{18}[(a_2 - a_1)(a_4 - a_3) + (b_2 - b_1)(b_4 - b_3)] +$ $\frac{1}{12}[(a_4 - a_3)(b_2 - b_1) + (a_2 - a_1)(b_4 - b_3) - (a_4 - a_3)(b_4 - b_3) - (a_2 - a_1)(b_2 - b_1)]$
1		$\left(\frac{a_2 + a_3}{2} - \frac{b_2 + b_3}{2}\right)^2 + \frac{1}{2}\left(\frac{a_2 + a_3}{2} - \frac{b_2 + b_3}{2}\right)[(a_4 - a_3) - (a_2 - a_1) - (b_4 - b_3) + (b_2 - b_1)]$ $+ \frac{1}{3}\left(\frac{a_3 - a_2}{2}\right)^2 + \frac{1}{6}\left(\frac{a_3 - a_2}{2}\right)[(a_4 - a_3) + (a_2 - a_1)] + \frac{1}{3}\left(\frac{b_3 - b_2}{2}\right)^2 + \frac{1}{6}\left(\frac{b_3 - b_2}{2}\right)[(b_4 - b_3) + (b_2 - b_1)] +$ $\frac{1}{9}[(a_4 - a_3)^2 + (a_2 - a_1)^2 + (b_4 - b_3)^2 + (b_2 - b_1)^2] - \frac{1}{9}[(a_2 - a_1)(a_4 - a_3) + (b_2 - b_1)(b_4 - b_3)] +$ $\frac{1}{6}[(a_4 - a_3)(b_2 - b_1) + (a_2 - a_1)(b_4 - b_3) - (a_4 - a_3)(b_4 - b_3) - (a_2 - a_1)(b_2 - b_1)]$
<b>Triangular fuzzy numbers</b> $A = (a_1, a_2, a_3)_T$ $B = (b_1, b_2, b_3)_T$	$\alpha$	$(a_2 - b_2)^2 + \frac{1}{3}(a_2 - b_2)[(a_3 + a_1 - 2a_2) - (b_3 + b_1 - 2b_2)] +$ $\frac{1}{18}[(a_3 - a_2)^2 + (a_2 - a_1)^2 + (b_3 - b_2)^2 + (b_2 - b_1)^2] -$ $\frac{1}{18}[(a_2 - a_1)(a_3 - a_2) + (b_2 - b_1)(b_3 - b_2)] - \frac{1}{12}(2a_2 - a_1 - a_3)(2b_2 - b_1 - b_3)$
		$(a_2 - b_2)^2 + \frac{1}{2}(a_2 - b_2)[(a_3 + a_1 - 2a_2) - (b_3 + b_1 - 2b_2)] + \frac{1}{9}[(a_3 - a_2)^2 + (a_2 - a_1)^2 + (b_3 - b_2)^2 + (b_2 - b_1)^2] -$ $\frac{1}{9}[(a_2 - a_1)(a_3 - a_2) + (b_2 - b_1)(b_3 - b_2)] - \frac{1}{6}(2a_2 - a_1 - a_3)(2b_2 - b_1 - b_3)$



## ***Mechanics of Integration Methods and Supporting Software***

The majority of the landscape metrics used in this report were calculated with the Analytical Tools Interface for Landscape Assessments (ATtILA), an ArcView extension developed by the EPA Landscape Ecology Branch. ATtILA is available free of charge via email at [ebert.donald@epa.gov](mailto:ebert.donald@epa.gov). Using ATtILA requires ArcView software and the Spatial Analyst extension; both are commercial products available from Environmental Systems Research Institute (ESRI; [www.esri.com](http://www.esri.com)).

An interactive web-based application was built specifically to allow the comparison and evaluation of each of the integration methods that were tested for this report. This web application was made available to the ReVA scientists for this work and results of the evaluation will be incorporated into a new version of the tool that will be released later as a decision-support toolkit.

The web-based application is a statistical framework that uses S-Plus Stat Server Software (Insightful Corporation; [www.insightful.com](http://www.insightful.com)). ArcView (version 3.2) shape files are read in S-PLUS using standard read file functions. The shape file provides attribute information (metric or variable values), which is read into various data frames and used to produce the graphs. The polygon information includes the x and y points that are passed to the S-PLUS polygon function to draw the maps.

Integration of data to produce the final maps follows a series of steps. The first step is to take the raw data and run the specific calculation required by that method. This calculation creates a list of values for each reporting unit (watershed in this case). These values are then classified either by putting an equal number of watersheds in each bin (quantile) or by creating equal size bins (equal interval). There are seven bins that are represented by unique colors on the map.

## ***Mechanics of Each Data Integration Method***

*Note: Number of variables used in report is 50.*

*N = number of watersheds (141 in region)*

*W = watershed*

### **Best Quintile**

For each variable:

- Rank the watersheds from best (1) to worst (N)
- Identify watersheds in the best quintile [ $\text{rank}(W)/N < 0.20$ ]

For each watershed:

- Count the number of variables in the best quintile
- Use the count as the value for the watershed
- Use this value to bin and color the watershed

### Worst Quintile

For each variable:

- Rank the data from best (1) to worst (N)
- Identify watersheds in the worst quintile [ $\text{rank}(W)/N > 0.80$ ]

For each watershed:

- Count the number of variables in the worst quintile
- Plot watersheds using an equal-interval classification
- Use the count as the value for the watershed
- Use this value to bin and color the watershed

### Simple Sum

*Note: This is a special case of the “Weighted Sum” method. For the Simple Sum, all variables are used and have equal weight.*

For each variable:

- Raw data are converted to normalized data
- “Best” variable value is assigned to a value of 0
- “Worst” variable value is assigned to a value of 1

For each watershed:

- Add the normalized values for each variable to get the “Simple Sum” for that watershed.
- Use this value to bin and color the watershed

### PCA Distance (Euclidean)

Calculate the correlation matrix ( $\Sigma$ ) for the normalized data (X)

*Note: The correlation values are identical for raw data and normalized data except possibly a sign difference (+/-).*

Find the principal components for the correlation matrix ( $\Sigma = Q\Lambda Q^T$ )

Use the vector of first five principal components ( $Q_5$ )

Take the absolute value of the “loadings” of the first five principal components,  $\text{abs}(Q_5) = Q_5^*$

“Rotate” the normalized data by post-multiplying the data by the principal components: ( $Y = X Q_5^*$ )

For each watershed:

- Calculate the Euclidean distance from ideal (0). [*NOTE: This is equivalent to calculating the sum of squares of each resulting row (watershed)*]
- Use this value to bin and color the watershed

### State Space (zero reference location)

Reference (ideal) watershed has zero value for all variables

Calculate the Mahalanobis distance from the current watershed to the ideal watershed

The Mahalanobis formula is:  $x\Sigma^{-1}x^T$ , where x is the vector of a watershed's variable values

Plot watersheds using an equal-interval classification

*Note: Hampton Roads (value = 152.20) was a drastic outlier, and it was assigned the next highest value (85.63).*

- Using the smallest and largest value, create seven equally spaced *bins* to color-code the watersheds
- Assign the watershed to a color based on which bin the simple sum falls in

### Criticality Analysis

Reference values for all variables based on a “natural state” (see Section 2 for further details)

Use fuzzy distance measure to get each watershed's “distance from natural state”

Plot watersheds using an equal-interval classification

- Using the smallest and largest value, create seven equally spaced *bins* to color-code the watersheds
- Assign the watershed to a color based on which bin the simple sum falls in

## ***Stressor/Resource Overlay***

For each watershed:

- Count the number of stressor variables that are in the worst **two** quintiles (worst 40 percent of watersheds)
- Count the number of resource variables that are in the best **two** quintiles (best 40 percent of watersheds)

Determine stressors and resource thresholds separately using equal interval and quantile classifications

- Equal bin sizes (for individual category, not combined category)
- Bin by quantile (for individual category, not combined category)

### Color scheme for map

- Lighter to darker with increasing resources (top to bottom on tables)
- Green to red with increasing stressors (left to right on tables)

---

## References

- Aguilera, P.A., A.G. Frenich, J.A. Torres, H. Castro, J.L.M. Vidal, and M. Canton. 2001. Application of the Kohonen neural network in coastal water management: methodological development for the assessment and prediction of water quality. *Water Research* 35:4053-4062.
- Alphonse, C.B. 1997. Application of the Analytic Hierarchy Process in agriculture in developing countries. *Agricultural systems* 53(1):97-112.
- Bachelet, D., R. P. Neilson, J. M. Lenihan, and R. J. Drapek. 2001. Climate change effects on vegetation distribution and carbon budget in the United States. *Ecosystems* 4:164-185.
- Bain, M. B., J. S. Irving, R. D. Olsen, E. A. Stull, and G. W. Witmer. 1986. *Cumulative impact assessment: evaluating the environmental effects of multiple human developments*. Argonne Nat. Lab. ANL/EES-TM-309. Argonne, IL. 71pp.
- Berryman, A. A., N. C. Stenseth, and D. J. Wollkin 1984. Metastability of forest ecosystems infested by bark beetles. *Res. Popul. Ecol.* 26:1329-1340.
- Bezdek, J. C., and Pal, S. K. (Eds.) 1992. *Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data*. New York: IEEE.
- Bezdek, J. C. 1998. Some new indexes of cluster validity. *IEEE Trans. Syst., Man, Cybern. B*, vol. 28, pp. 301-315.
- Bradley, M.P., and R.B. Landy. 2000. The Mid-Atlantic Integrated Assessment (MAIA). *Environmental Monitoring and Assessment* 63:1-13.
- Brosse, S., J.L. Giraudel, and S. Lek. 2001. Utilisation of non-supervised neural networks and principal component analysis to study fish assemblages. *Ecological Modelling* 146(1):159-166.
- Cada, G. F., and R. B. McLean. 1985. An approach for assessing the impacts on fisheries of basin-wide hydropower and development. Pp. 367-372. IN F. W. Olson, R. G. White, and R. H. Hamre (eds). *Proceedings of the symposium on small hydropower and fisheries*. American Fisheries Society.
- Calais, M.D., R.G. Kerzee, J. Bing-Canar, E.K. Mensah, K.G. Croke, and R.S. Swger. 1996. An indicator of solid waste generation potential for Illinois using principal component analysis and geographic information system. *Journal of the Air and Waste Management Association* 46: 414-419.
- Canter, L. W. 1977. *Environmental impact assessment*. McGraw-Hill, NY. 331pp.
- Casti, J. 1982. Catastrophes, control, and the inevitability of spruce budworm outbreaks. *Ecological Modelling* 14:293-300.

- Cereghino, R., J.L. Giraudel, and A. Compin. 2001. Spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen self organizing maps. *Ecological Modelling* 146:167-180.
- Chatfield, C., and A.J. Collins. 1980. *Introduction to Multivariate Analysis*. Chapman and Hall, London, 246 pp.
- Chen, S.-J., and C.-L. Hwang. 1992. *Fuzzy Multiple Attribute Decision Making*. Springer-Verlag, Berlin, 536 pp.
- Clare A.P., and D.R. Cohen. 2001. A comparison of unsupervised neural networks and K-means clustering in the analysis of multi-element stream sediment data. *Geochemistry: Exploration, Environment, Analysis* 1:119-134.
- Clark, W. C. 1986. The cumulative impacts of human activities on the atmosphere. Pp. 113-124. In *Cumulative environmental effects: a binational perspective*. Minister of Supply and Services, Canada.
- Cormier, S., S. B. Norton, G. Suter III, and D. Reed-Judkins. 2000. *Stressor identification guidance document*. EPA-822-B-00-025.
- Crowley, T. J., and G. R. North. 1988. Abrupt climate change and extinction events in earth history. *Science* 240:996-1002.
- Davies, D. L., and Bouldin, D.W. 1979. A cluster separation measure. *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-1, pp. 224-227.
- Dubois, D. M. 1979. Catastrophe theory applied to water quality regulation of rivers. Pp 751-758. In S. E. Jorgensen (ed.), *State of the art of Ecological Modelling*. International Society for Ecological Modelling. Copenhagen, Denmark.
- Emery, R. M. 1986. Impact interaction potential: a basin-wide algorithm for assessing cumulative impacts from hydropower projects. *J. Environ. Manage.* 23:341-360.
- Everitt, B.S., and G. Dunn. 1992. *Applied Multivariate Data Analysis*. Oxford University Press, New York, 304 pp.
- Ferenc, S. A., and J. A. Foran (eds.). 2000. *Multiple stressors in ecological risk and impact assessment: approaches to risk estimation*. SETAC Press, Pensacola, FL.
- Ferson, S., and R. Kuhn. 1992. Propagating uncertainty in ecological risk analysis using interval and fuzzy arithmetic. Pp 387-401. IN P. Zanetti (ed.). *Computer techniques in environmental studies IV*. Elsevier Applied Science, London.
- Foran, J. A., and S. A. Ferenc (eds.). 1999. *Multiple stressors in ecological risk and impact assessment*. SETAC Press, Pensacola, FL.
- Gatto, M., and S. Renaldi. 1987. Some models of catastrophic behavior in exploited forests. *Vegetatio* 69:213-222.
- Giraudel, J. L., and S. Lek. 2001. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. *Ecological Modelling* 146:329-339.

- Harris, H.J., Wenger, R.B., Harris, V.A., Devault, D.S. 1994. A method for assessing environmental risk: a case study of Green Bay, Lake Michigan, USA. *Environ Manage.* 18(2):295-306.
- Holling, C. S. 1986. The resilience of terrestrial ecosystems: local surprise and global change. Pp. 292-317. IN W. C. Clark and R. E. Munn (eds.). *Sustainable development of the biosphere*. International Institute of Applied Systems Analysis, Luxemburg, Austria.
- Holling, C. S. 1973. Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics* 4:1-24.
- Hotelling, H. 1933. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24:417-41.
- Hughes, T. P. 1994. Catastrophes, phase shifts, and large scale degradation of a Caribbean Coral Reef. *Science* 265:1547-1551.
- Ingegnoli, V. 1990. Human influences in landscape change: thresholds of metastability. Pp 303-309. IN O. Ravera (ed.). *Terrestrial and Aquatic Ecosystems: perturbation and recovery*. Ellis Howard, London, England.
- Iverson, L. R., and A. M. Prasad. 2001. Potential changes in tree species richness and forest community types following climate change. *Ecosystems* 4:186-199.
- Jobson, J.D. 1992. *Applied multivariate data analysis, volume II: categorical and multivariate methods*. Springer-Verlag, New York, 731 pp.
- Johnson, A.R. 1988. Evaluating ecosystem response to toxicant stress: a state space approach. Pages 275-285 in W.J. Adams, G.A. Chapman, and W.G. Landis (eds.), *Aquatic Toxicology and Hazard Assessment: 10th Volume*. ASTM STP 971, American Society for Testing and Materials, Philadelphia.
- Jones, D. D., and C. J. Walters. 1976. Catastrophe theory and fisheries regulation. *J. Fish Res. Board Can.* 33:2829-2833.
- Jooste, S. 2000. A model to estimate the total ecological risk in the management of water resources subject to multiple stressors. *Water SA* 26:159-166.
- Kay, J. J. 1991. A nonequilibrium thermodynamic framework for discussing ecosystem integrity. *Environmental Management* 15:483-495.
- Kohonen, T. 1982. Analysis of a simple self-organizing process. *Biological Cybernetics* 44:135-140.
- Kohonen, T. 2001. *Self-Organizing Maps* (3<sup>rd</sup> edition). Springer, Berlin, 501pp.
- Kohonen, T., J. Hynninen, J. Kangas, and J. Laaksonen. 1996. *SOM-PAK: The Self-Organizing Map Program Package*. Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, FIN-02150 Espoo, Finland, 1996.
- Kuchler, A.H. 1964. Potential natural vegetation of the conterminous United States. American Geogr. Soc. Spec. Publ. No. 36, Washington, D.C.
- Lampinen, J., and E. Oja. 1992. Clustering properties of hierarchical self-organizing maps. *Journal of Mathematical Imaging and Vision* 2:261-272.

- Landis, W.G., and Wiegers, J.A. 1997. Design considerations and a suggested approach for regional and comparative risk assessment. *Human Ecol. Risk Assessment* 3:287-297.
- Leopold, L. B., F. E. Clarke, B. B. Hanshaw, and J. R. Balsley. 1971. *A procedure for evaluating environmental impact*. Geological Survey Circular 645. U.S. Government Printing Office, Washington, D.C.
- Levin, S. A. 1999. *Fragile Dominion*. Perseus Books, Reading, MA.
- Lewis, R., and D.E. Levy. 1989. Predicting a national acid rain policy. Pages 155-170 in B.L. Golden, E.A. Wasil, and P.T. Harker (eds.). *Application of the Analytic Hierarchy Process*. Springer-Verlag, New York.
- Loehle, C. 1989. Catastrophe theory in ecology: a critical review and an example of the butterfly catastrophe. *Ecological Modelling* 49:125-152.
- Lootsma, F.A. 1997. *Fuzzy Logic for Planning and Decision Making*. Kluwer Academic Publishers, Dordrecht, 195 pp.
- Lootsma, F.A. 1999. *Multi-criteria Decision Analysis via Ratio and Difference Judgment*. Kluwer Academic Publishers, Dordrecht, 283 pp.
- Lumb, A. M. 1982b. *Procedures for assessment of cumulative impacts of surface mining on the hydrologic balance*. U.S. Geol. Surv. Open-File Rep. 82-334. 50pp.
- Lumb, A. M. 1982a. Cumulative impact assessment of surface mining. Pp 145-150. In F. Kilpatrick and D. Matchett (eds.). *Proceedings of the eastern conference on water and energy: technical and policy issues*. Amer. Soc. Of Civil Engineers, NY.
- Mahalanobis, P. C. 1936. On the generalized distance in statistics. *Proceedings of the National Institute of Science of India* 12:49-55.
- May, R. M. 1977. Thresholds and breakpoints in ecosystems with a multiplicity of stable states. *Nature* 269:471-477.
- McGhee, G. R. 1990. Catastrophes in the history of life. Pp 26-500. IN K. C. Allen and D. E. G. Briggs (eds.). *Evolution and the Fossil Record*. Smithsonian Institution Press, Washington, D.C.
- McLahlan, G. J., and Basford, K. E. 1987. *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker, vol. 84.
- Milligan, G.W., and Cooper, M. C. 1985. An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2):159-179.
- Moss, D. A. 2002. When all else fails – Government as the ultimate risk manager. Harvard University Press.
- Mummolo, G. 1996. An Analytic Hierarchy Process Model for Landfill Site Selection. *Journal of environmental systems* 24(4):445-465.
- Obach, M., R. Wagner, H. Werner, and H.-H. Schmidt. 2001. Modelling population dynamics of aquatic insects with artificial neural networks. *Ecological Modelling* 146:207-217.



- O'Neill, R. V. 1999. Recovery in complex ecosystems. *Journal of Aquatic Ecosystem Stress and Recovery* 6:181-187.
- O'Neill, R. V. 2001. Is it time to bury the ecosystem concept? *Ecology* 82:3275-3284.
- O'Neill, R. V., A. R. Johnson, and A. W. King. 1989. A hierarchical framework for the analysis of scale. *Landscape Ecology* 3:193-205.
- O'Neill, R. V., R. H. Gardner, and D. E. Weller. 1982. Chaotic models as representations of ecological systems. *American Naturalist* 120:259-263.
- Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2:559-72.
- Peterman, R. M., W. C. Clark, and C. S. Holling. 1979. The dynamics of resilience: shifting stability domains in fish and insect systems. Pp 321-342. In R. M. Anderson, B. D. Turner, and L. R. Taylor (eds.). *Population Dynamics*. Blackwell, London, England.
- Phillips Brandt Reddick, McDonald and Grefe, Inc. 1978. *The cumulative impacts of shorezone development at Lake Tahoe*. Report prepared for California State Lands Commission, State of Nevada, Tahoe Regional Planning Agency and U.S. Army Corps of Engineers.
- Phillips, J. D. 1993. Spatial domain chaos in landscapes. *Geogr. Anal.* 25:101-117.
- Pielou, E.C. 1984. *The Interpretation of Ecological Data: a Primer on Classification and Ordination*. John Wiley and Sons, New York.
- Rachdawong, P., and E.R. Christensen. 1997. Determination of PCB sources by principal component method with nonnegative constraints. *Environmental Science and Technology* 31:2686-2691.
- Ramanathan, R., and L.S. Ganesh. 1995. Energy resource allocation incorporating qualitative and quantitative criteria: an integrated model using goal programming and AHP. *Socio-Economic Planning Sciences* 29(3):197-218.
- Rencher, A.C. 1995. *Methods of multivariate analysis*. John Wiley and Sons, New York, 627 pp.
- Ridgley, M. A., and F. R. Rijsberman. 1992. Multicriteria evaluation in a policy analysis of a Rhine estuary. *Water Resources Bulletin* 28:1095-1110.
- Riitters, K., J. Wickham, R. O'Neill, B. Jones, and E. Smith. 2000. Global-scale patterns of forest fragmentation. *Conservation Ecology* 4(2): 3.
- Risser, P. G. 1988. General concepts for measuring cumulative impacts on wetland ecosystems. *Environ. Manage.* 12:585-590.
- Rosenzweig, M. 1971. Paradox of enrichment: destabilization of enrichment systems in ecological time. *Science* 171:385-387.
- Rosser, J. B. 1991. *From catastrophe to chaos: a general theory of economic discontinuities*. Kluwer Academic Publishers, Boston, MA.
- Rosser, J. B., C. Folke, F. Gunther, H. Isomaki, C. Perrings, and T. Puu. 1994. Discontinuous change in multilevel hierarchical systems. *Systems Research* 11:77-94.

- Rousseeuw, P.J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational Mathematics* 20:53-65.
- Saaty, T.L. 1977. A scaling method for priorities in hierarchical structure. *Journal of Mathematical Psychology* 15:234-281.
- Saaty, T.L. 1978. Exploring the interface between hierarchies, multiple objectives, and fuzzy sets. *Fuzzy Sets and Systems* 1(1):57-68.
- Saaty, T.L. 1980. *The Analytic Hierarchy Process, Planning, Priority Setting, and Resource Allocation*. RWS Publications, Pittsburgh, 376 pp.
- Saaty, T.L. 1986. Absolute and relative measurement with the AHP: the most livable cities in the US. *Socio-Economic Planning Sciences* 20(6):327-331.
- Saaty, T.L., and Luis G. Vargas. 1982. The logic of priorities: applications in business, energy, health and transportation. Kluwer-Nijhoff Publishing, Boston (reprinted 1991, RWS Publications).
- Saaty, T.L. 2001. *The Analytic Network Process*. McGraw-Hill, New York, 287 pp.
- Schaeffer, W. M., and M. Kot. 1986. Chaos in ecological systems: the coals that Newcastle forgot. *Trends Ecol. Evol.* 1:58-63.
- Schlesinger, W. H., J. F. Reynolds, G. L. Cunningham, L. F. Huenneke, W. M. Jarrell, R. A. Virginia, and W. G. Whitford. 1990. Biological feedback in global desertification. *Science* 247:1043, 1048.
- Sjögren, M., H. Li, and R. Westerholm. 1996. Multivariate analysis of exhaust emissions from heavy-duty diesel fuels. *Environmental Science and Technology* 30: 38-49.
- Smith, E.R., O'Neill, R.V., Wickham, J.D., Jones, K.B., Jackson, L., Kilaru, J.K., and Reuter, R. 2002. *The U.S. EPA's Regional Vulnerability Assessment Program: a Research Strategy for 2001-2006*. EPA/600/R-01/008. <http://epa.gov/reva/reva-strategy.pdf>
- Spromberg, J. A., B. M. John, and W. G. Landis. 1998. Metapopulation dynamics: indirect effects and multiple distinct outcomes in ecological risk assessment. *Environmental Toxicology and Chemistry* 17:1640-1649.
- Statheropoulos, M., N. Vassiliadis, and A. Pappa. 1998. Principal component and canonical correlation analysis for examining air pollution and meteorological data. *Atmospheric Environment* 32(6): 1087-1095.
- Stull, E. A., K. E. LaGory, and W. S. Vinikour. 1987. Methodologies for assessing the cumulative environmental effects of hydroelectric development on fish and wildlife in the Columbia River Basin. Volume 2: Example and procedural guidelines. Argonne Nat. Lab Report. 92pp.
- Suter, G. W. 1993a. A critique of ecosystem health concepts and indexes. *Environmental Toxicology and Chemistry* 12:1533-1539.
- Suter, G. W. 1993b. *Ecological Risk Assessment*. Lewis Publishers, Ann Arbor, MI.
- Sutherland, J. P. 1974. Multiple stable points in natural communities. *American Naturalist* 108:859-873.
- Tainter, J. A. 1988. *The Collapse of Complex Societies*. Cambridge University Press, NY.

- Topalián, M.L., P.M. Castañé, M.G. Rovedatti, and A. Salibián. 1999. Principal component analysis of dissolved heavy metals in water of the Reconquista River (Buenos Aires, Argentina). *Bulletin of Environmental Contamination and Toxicology* 63: 484-490.
- Tran, L., and L. Duckstein, L. 2002. Comparison of fuzzy numbers using a fuzzy distance measure. *Fuzzy Sets and Systems* 130:331-341.
- Tran, L.T., C.G. Knight, R.V. O'Neill, E.R. Smith, K.H. Riitters, and J. Wickham. 2002. Fuzzy decision analysis for integrated environmental vulnerability assessment of the mid-Atlantic region. *Environmental Management* 29:845-859.
- Tran, L.T., C.G. Knight, R.V. O'Neill, E.R. Smith, and J. M. O'Connell. 2003. Self-organizing maps for integrated environmental assessment of the mid-Atlantic region. *Environmental Management*. In press.
- Trautmann, T., and T. Denoeux. 1995. Comparison of dynamic feature map models for environmental monitoring. In *Proceedings of ICNN'95, IEEE International Conference on Neural Networks*, volume I, pages 73-78, Piscataway, NJ, 1995. IEEE Service Center.
- USEPA. 1998. Guidelines for Ecological Risk Assessment. Office of Research and Development, Washington, DC, EPA/630/R-95/002F.
- Ultsch, A. 1993. Self-organizing neural networks for visualization and classification. Pages 307-313 in Opitz, O., B. Lausen, and R. Klar (eds.): *Information and Classification*, London, UK. Springer.
- Varis, O. 1989. The Analysis of preferences in complex environmental judgments-A focus on the Analytic Hierarchy Process. *Journal of Environmental Management* 28(4):283-294.
- Vesanto, J., and E. Alhoniemi. 2000. Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks* 11(3):586-597.
- Walley, W.J., and M.A. O'Connor. 2001. Unsupervised pattern recognition for the interpretation of ecological data. *Ecological Modelling* 146:219-230.
- Witmer, G., J. S. Irving, and M. Bain. 1985. *A review and evaluation of cumulative impact assessment techniques and methodologies*. Argonne Nat. Lab. Report.
- Yu, C.-C., J.T. Quinn, C.M. Dufournaud, J.J. Harrington, P.P. Rogers, and B. Lohani. 1998. Effective dimensionality of environmental indicators: a principal component analysis with bootstrap confidence intervals. *Journal of Environmental Management* 53: 101-119.
- Yu, T.-Y., and L.-F. Chang. 2000. Selection of the scenarios of ozone pollution at southern Taiwan area utilizing principal component analysis. *Atmospheric Environment* 34: 4499-4509.