United States
Environmental Protection
Agency

# Guidelines to Assessing Regional Vulnerabilities



RESEARCH AND DEVELOPMENT

EPA/600/R-08/xxx
April 2008
www.epa.gov

# Guidelines to Assessing Regional Vulnerabilities

by

Elizabeth R. Smith[1], Megan H. Mehaffey[1], Robert V. O'Neill[2], Timothy G. Wade[1], J. Vasu Kilaru[1], and Liem T. Tran[3]

[1]U.S. Environmental Protection Agency
Office of Research and Development
National Exposure Laboratory
Research Triangle Park, NC

[2]TN and Associates
Oak Ridge, TN

[3]University of Tennessee
Knoxville, TN



U.S. Environmental Protection Agency
Office of Research and Development
Washington, DC 20460

073cmb08

# Notice

The U.S. Environmental Protection Agency (U.S. EPA), through its Office of Research and Development (ORD), funded and managed parts of the research described here under contract EP-D-06-072 with TN and Associates, Inc. It has not yet been subjected to the Agency's peer and administrative review and has not yet been approved for publication as an EPA document.

# Acknowledgments

Many people contributed to the Regional Vulnerability Assessment methods development described in this report. Specifically, we would like to acknowledge the following:

Rochelle Araujo, EPA, National Exposure Research Laboratory
Tim Johnson, EPA, National Risk Management Research Laboratory
K. Bruce Jones, U.S. Geological Survey
Rick Linthurst, EPA, Office of Research and Development
Peter McKinnis, Waratah Corporation
Michael O'Connell, Waratah Corporation
Roger Tankersley, Tennessee Valley Authority
Paul Wagner, EPA, National Exposure Research Laboratory
Lisa Wainger, University of Maryland
Dennis Yankee, Tennessee Valley Authority



**Preferred Citation:**

**Smith, E.R., Mehaffey, M.H., O'Neill, R.V., Wade, T.G., Kilaru, J.V., and Tran, L.T. 2008. Guidelines to Assessing Regional Vulnerabilities.**

# Executive Summary

Environmental decision-makers today are faced with declining budgets, lack of problem-focused monitoring data, and issues that range from subtle and slow (such as changes in species composition) to conspicuous and immediate (e.g., catastrophic events). At the same time, there is greater recognition that environmental decisions that are made today are likely to impact human well-being in the future. Thus, there is a growing desire to evaluate potential decisions with regard to their future implications. Further, in attempting to reach a decision, an environmental decision-maker can quickly become overwhelmed by the huge amounts of disparate types of data and information that are available on resources, conditions, and stressors within a region. The EPA's Regional Vulnerability Assessment (ReVA) program was designed to deal with these problems: ReVA methodology establishes a platform that can help environmental decision-makers target limited resources and enable proactive decision-making.

ReVA has a broad spatial perspective, uses existing data, and applies an integrated approach to assessment; it can incorporate large, disparate sources of available spatial data on resources, environmental conditions, and stressors, and then visually express these conditions (or combinations of these conditions) in map form. ReVA methods also allow users to prepare "what if" scenarios; these scenarios permit inspection of likely future changes in environmental vulnerabilities, given user-determined inputs on anticipated regional changes in factors such as population growth, economic conditions, land use, transportation infrastructure, etc. ReVA can improve the environmental decision-making process by permitting more realistic inputs for environmental decision-making and by expressing results of multiple factors at a regional spatial scale.

Since 1998, much of the research effort within the ReVA program has focused on the mechanics of how data and model results can be integrated into meaningful indices designed to address specific assessment questions posed by environmental decision-makers. The approach developed by the ReVA program allows decision-makers to evaluate current conditions and vulnerabilities through the use of indices. This approach allows an evaluation of *net* change, so that the user can visualize how both positive and negative changes affect future conditions and vulnerabilities.

ReVA's approach as presented in these guidelines includes the following steps:

- Acquisition of spatially explicit data
- Data processing
- Metric selection and integration
- Development/selection of spatially explicit models
- Creation of alternative scenarios
- Synthesis
- Results communication

# Table of Contents

# List of Figures

# List of Abbreviations and Acronyms

| | |
|---|---|
| ATtILA | Analytical Tools Interface for Landscape Assessments |
| BASINS | Better Assessment Science Integration Point and Nonpoint Sources |
| BIOCLIM | Bioclimatic prediction system |
| CMAQ | Community Multiscale Air Quality |
| DEMs | Digital Elevation Models |
| EDT | Environmental Decision Toolkit |
| FGDC | Federal Geographic Data Committee |
| FWS | Fish and Wildlife Service |
| GARP | Genetic Algorithm for Rule-set Production |
| GIS | Geographic Information System |
| HUC | Hydrologic Unit Code |
| ICLUS | Integrated Climate Land Use Scenarios |
| IDW | Inverse Distance Weighting |
| LANDSAT | Land Remote Sensing Satellite |
| LTHIA | Long-Term Hydrologic Impact Assessment |
| MAIA | Mid-Atlantic Integrated Assessment |
| NASS | National Agricultural Statistical Survey |
| NATA | National Air Toxics Assessment |
| NAWQA | National Water-Quality Assessment Program |
| NCDC | National Climatic Data Center |
| NED | National Elevation Dataset |
| NHD | National Hydrography Database |
| NIS | Non-Indigenous Species |
| NLCD | National Land Cover Dataset |
| NOAA | National Oceanic and Atmospheric Administration |
| NRCS | National Resources Conservation Service |
| NWI | National Wetlands Inventory |
| PCA | Principle Components Analysis |
| ReVA | Regional Vulnerability Assessment |
| RGDT | Regional Growth Decision Tool |
| RUSLE | Revised Universal Soil Loss Equation |
| SAB | Science Advisory Board |
| SEQL | Sustainable Environment for Quality of Life |

| | |
|---|---|
| SERGoM | Spatially Explicit Regional Growth Model |
| SETAC | Society of Environmental Toxicology and Chemistry |
| SLEUTH | Slope, Land use, Exclusion, Urban, Transportation, Hillshading model |
| SSURGO | Soil Survey Geographic database |
| STATSGO | State Soil Geographic database |
| TIGER | Topologically Integrated Geographic Encoding and Referencing |
| U.S. EPA | U.S. Environmental Protection Agency |
| USCB | U.S. Census Bureau |
| USDA | U.S. Department of Agriculture |
| USGS | U.S. Geological Survey |
| USLE | Universal Soil Loss Equation |
| XML | Extensible Markup Language |

# Section 1

# Introduction and Background

Decision-makers today face increasingly complex environmental problems that require integrative and innovative approaches for analyzing, modeling, and interpreting various types of information. ReVA acknowledges this need and is designed to evaluate methods and models for synthesizing diverse kinds of available information on the distribution of stressors and sensitive ecological resources. As with any study, the first and probably most important step is to establish a clear goal. For ReVA, the goal is to develop and demonstrate approaches that use existing data to evaluate current and future conditions and vulnerabilities of valued resources (native biodiversity, water quality, forest productivity, etc.) resulting from ecological drivers of change[1] and later, management alternatives.

## *Why Look at the Broad Scale?*

ReVA is designed to help decision-makers use existing data and model results at a broad scale, allowing insights into (1) where problems are likely to occur in the future, (2) what environmental stresses are likely to be of most concern, and (3) how alternative management decisions might play out in terms of trade-offs across the region. The broad-scale approach is important for several reasons. First, by stepping back and assessing landscape (regional scale) characteristics and the distribution of resources and stressors, spatial relationships become apparent. Over time, land use and invasive species may change and can be expected to move across the landscape. Identifying where these things are currently occurring can provide insights as to when and where these issues will occur in the future. Second, many of the drivers of ecological change occur at the regional scale over fairly long time periods (Figure 1). Thus, a broad-scale approach is necessary to capture these changes, for they could be easily overlooked at a finer scale. Third, evaluating projected changes at a broad scale enables strategic management responses by considering what is best for the region overall, even while managing finer-scale risks or problems that are unavoidable or are part of the trade-offs that come with any environmental decision.

[1]Drivers of ecological change are generally accepted as including land-use change, invasive non-indigenous species, resource extraction, pollution and pollutants, and climate change (Chambers et al., 2007).

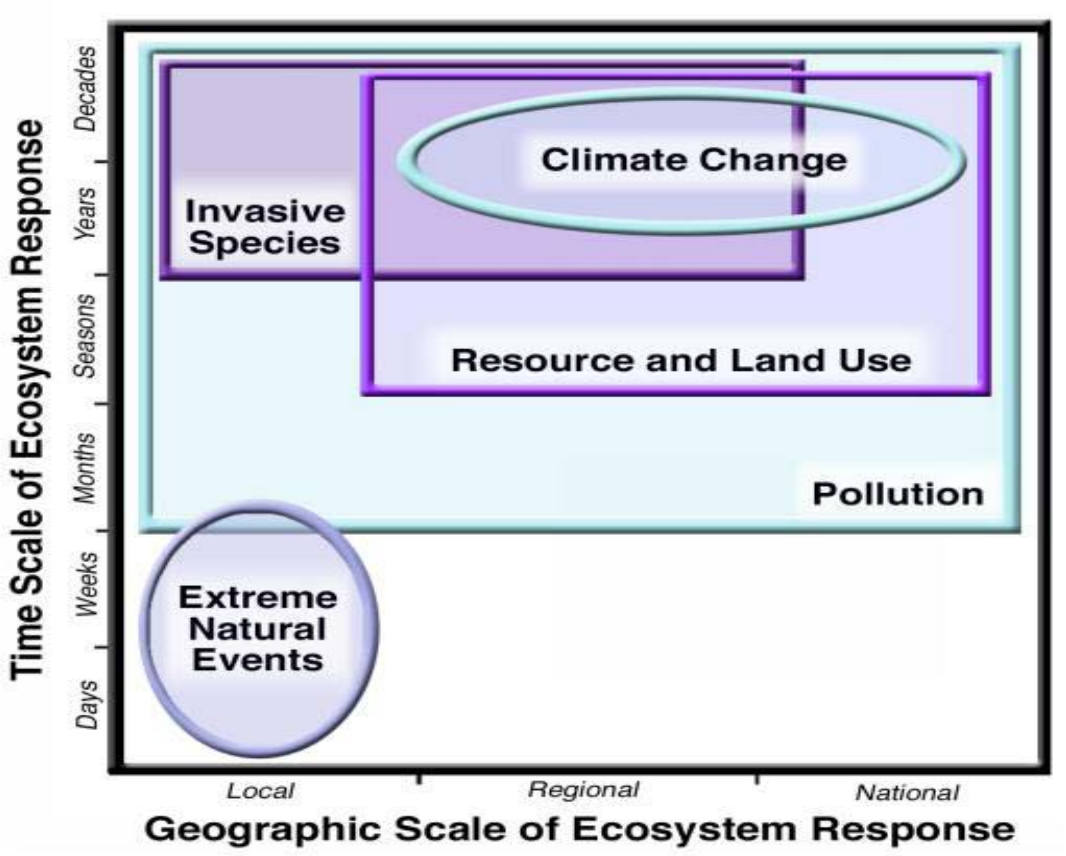


**Figure 1. Schematic depicting differences in the spatial (X axis) and temporal (Y axis) scales of ecosystem responses to various types of stressors.**

### *Vulnerability versus Risk*

As its name implies, ReVA is based on vulnerability assessment; it examines a broad range of information across a region and attempts to identify areas where as-yet-unidentified endpoints might be vulnerable. ReVA accomplishes this objective by applying environmental indicators (or descriptive metrics) to represent important endpoints and examines the co-occurrence of valued resources and stressors to represent vulnerability of sensitive endpoints to potential harm. The techniques used to examine how stressors and resources combine seek to reveal threats that are often not clearly identifiable or quantifiable, and allow the users of ReVA output to explore complex interdependencies of related issues (cf. Liotta, 2005; Liotta and Miskel, 2004).

### *The Need for an Integrated Approach*

In addition to taking a broad spatial perspective, ReVA stresses an integrated approach to assessment. This emphasis imparts greater realism to environmental decision-making by presenting problems simultaneously to permit the decision-maker a broader perspective in identifying the most vulnerable resources within a region. In considering all resources,

conditions, and stressors, decision-makers typically confront huge amounts of data which results in a challenge to make the information meaningful. These difficulties are addressed by the ReVA approach (Figure 2), which allows decision-makers to evaluate current conditions and vulnerabilities using indices. The use of indices permits the decision-maker to evaluate how positive and negative changes affect future conditions and vulnerabilities.

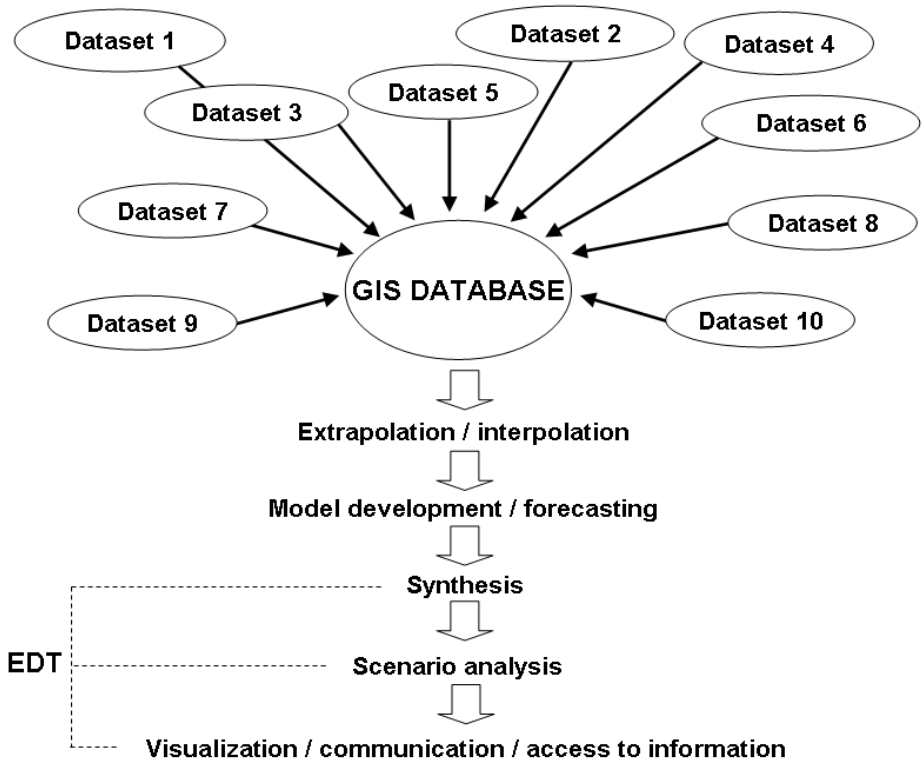


**Figure 2. Schematic of steps in the ReVA approach. ReVA's Environmental Decision Toolkit (EDT) is used for synthesis, scenario analysis and communication of results by visual representations.**

# Section 2

# Data Used in ReVA

## *Types of Spatial Data*

Geospatial data typically have two basic components: (1) the location or geographic context and (2) attributes of that location or area. The geographic context, in turn, falls into several categories: point, line, polygon, and grid. A point is simply a discrete location of an entity, designated by x and y coordinate values, such as an air monitoring station or a soil sampling site; a line abstraction is used to represent linear objects such as roads or streams; and polygons represent areas such as political borders or water bodies. A grid is a special raster-based (cell-based) geography where all the cells in the grid are square and equal in area and each cell contains only one value for the variable of concern. Typical gridded datasets are used for variables such as elevation (e.g., a digital elevation model) and land-cover classification. Points, lines, and polygons may have any number of attributes associated with a single element. For example, a polygon that represents a county may have attributes such as area, perimeter, population, and per capita income.

## *Data Inputs for ReVA*

The data required to perform ReVA-type analyses (see Figure 2) may be acquired from many sources. The area or region of concern, existing resources, types of stressors, and the questions and concerns about the region determine the data requirements. The main requirement is that the data used in any analysis must be collected consistently across the region of concern. Data are available from various sources at the national, regional, State, and local levels. National sources include many federal agencies such as the United States Geological Service (USGS) and the United States Environmental Protection Agency (U.S. EPA). An example of a regional source of information might be the U.S. EPA Chesapeake Bay Program Office. States, too, have geographic data holdings, but the extent and quality of these datasets can vary widely. Finally, at the finest scale, counties and local municipalities have geographic data at very local scales. These datasets can include land-parcel data and zoning information. The data in these local-scale datasets often vary greatly among local areas in terms of level of detail and quality, making combination across boundaries difficult.

### National Data Sources

ReVA uses a number of datasets that are available for at least the conterminous states. With these base layers, numerous landscape and environmental metrics can be computed for most areas in the nation. One of the most useful Web sites for obtaining data at a national scale is operated by the USGS. This Web site can be accessed at:

http://seamless.usgs.gov/website/seamless/viewer.php. The following are examples of datasets that can be downloaded from this site.

- NLCD 1992 and 2001 – The National Land Cover Dataset (NLCD) is a gridded dataset that contains consistently (within a year) collected and processed imagery with a land-cover classification scheme for the entire U.S. The 1992- and 2001-era data are nationally available and can be downloaded from the NLCD Web site: http://www.mrlc.gov/index.asp. Significant changes were made to the processing methodology of Land Remote Sensing Satellite (LANDSAT) imagery which makes direct comparison of NLCD 1992 and NLCD 2001 difficult.

- NED – The National Elevation Dataset (NED) is a gridded dataset that contains elevation values for each grid cell; such datasets are referred to as digital elevation models (DEMs). These data are available at several scales. Typically the 30-meter (or 1/3 arc-second) data are used; this dataset is available nationally. The 10-meter (or 1/9 arc-second) data also are available for some areas. Due to their fine scale, the 30- and 10-meter elevation datasets are large. For some applications, it may be acceptable to use a larger-scale grid, such as the 100-meter (1 arc-second) dataset.

- National Atlas – Data from the USGS National Atlas are also available at the USGS "seamless" server site. However, for better descriptions and access to metadata, it is helpful to visit the site at: http://nationalatlas.gov/pros.html. The National Atlas contains spatial datasets on diverse variables, including: the 2002 Census of Agriculture, breeding bird survey locations, invasive species, forest fragmentation estimates, vegetation growth, West Nile virus surveillance, wildlife mortality, and other variables, encompassing geology, climate, environment, transportation, and water.

- NHD – The National Hydrography Dataset (NHD) is a 1:100,000-scale digital representation of the nation's streams and rivers. NHD is very useful in many landscape analyses, especially in conjunction with DEMs and land cover. It also is useful for hydrologic modeling and is populated with various attributes that allow analysis of flow networks. The NHD is available at: http://nhd.usgs.gov/data.html.

- NWI – Maintained by the Fish and Wildlife Service (FWS), the National Wetlands Inventory (NWI) is a digital spatial dataset of the wetlands in the U.S. and is available at: http://wetlandsfws.er.usgs.gov/NWI/download.html.

- TIGER/Line 2000 – The Census 2000 TIGER/Line shapefiles were created from the Topologically Integrated Geographic Encoding and Referencing (TIGER) database of the United States Census Bureau (USCB). The shapefiles contain data on the following: line features such as roads, railroads, hydrography, and transportation and utility lines; boundary features such as statistical (e.g., census tracts and blocks), government (e.g., places and counties), and administrative (e.g., congressional and school districts); and boundaried and landmark features such as point (e.g., schools and churches), area (e.g., parks and cemeteries), and key geographic locations (e.g., apartment buildings and factories). A number of vendors offer value-added products that improve on the USCB's

version of the data. Freely available USCB data can be accessed at: http://arcdata.esri.com/data/tiger2000/tiger_download.cfm.

- Census 2000 – The USCB also administers the decadal census. While numerous products are available, the more detailed demographic data can provide useful information about housing, income, education, race, age, gender, and other socio-economic indicators. Like other U.S. government products, many vendors offer value-added products that build on the basic data collected by the USCB. For more information, visit: http://www.census.gov/main/www/cen2000.html.

- Soil data – The State Soil Geographic database/Soil Survey Geographic database (STATSGO/SSURGO) are geographic databases maintained by the Natural Resources Conservation Service (NRCS) that contain generalized soil types. The datasets were created by generalizing more detailed soil survey maps. For STATSGO, where more detailed soil survey maps were not available, data on geology, topography, vegetation, and climate were assembled, together with LANDSAT images. Soils of like areas were studied, and the probable classification and extent of the soils were determined. Map unit composition was determined by transecting or sampling areas on the more detailed maps and expanding the data statistically to characterize the whole map unit.

The STATSGO dataset consists of geo-referenced vector digital data and tabular digital data. The map data were collected in 1- by 2-degree topographic quadrangle units and merged into a seamless national dataset. It is distributed in state/territory and national extents. The soil map units are linked to attributes in the tabular data, which give the proportionate extent of the component soils and their properties.

The tabular data contain estimated and measured data on the physical and chemical soil properties, soil interpretations, and static and dynamic metadata. Most tabular data exist in the database as a range of soil properties, depicting the range for the geographic extent of the map unit. In addition to low and high values for most data, a representative value is also included for these soil properties. For more information, see: http://www.ncgc.nrcs.usda.gov/products/datasets/statsgo/data/index.html.

The STATSGO database is being updated and renamed to the Digital General Soil Map of the United States. The updated version will be available for download from the Soil Data Mart: http://soildatamart.nrcs.usda.gov/.

The STATSGO database is designed primarily for regional, multistate, river basin, state, and multicounty resource planning, management, and monitoring. It is not detailed enough for analyses at the county level or finer-scale. The SSURGO dataset is much more detailed than STATSGO. It is designed primarily for farm and ranch, landowner/user, township, county, or parish natural resource planning and management.[2]

[2]Pennsylvania State University Cooperative Agriculture Extension, November 2007. http://lal.cas.psu.edu/software/tutorials/soils/st_diff.html

Currently, plans are for the digital data for SSURGO to be completed in 2008. For more information on SSURGO, see: http://soildatamart.nrcs.usda.gov/SSURGOMetadata.aspx.

- Omernik Ecoregions – Ecoregions are areas of the landscape that are classified into regions on the basis of geology, physiography, vegetation, climate, soils, land use, wildlife, and hydrology. For more information and access to data that can be downloaded, visit: http://www.epa.gov/wed/pages/ecoregions/level_iii.htm.

- Climate Data – The National Oceanographic and Atmospheric Administration's (NOAA's) National Climatic Data Center (NCDC) collects and disseminates climate data that includes such parameters as temperature, precipitation, and wind speeds. These data are available for download at: http://lwf.ncdc.noaa.gov/oa/climate/climatedata.html.

At least three other national-scale sources for environmental data can be accessed for use in ReVA:

- NOAA Geophysical Data – NOAA provides access to a wide variety of geophysical data. These can be accessed at: http://www.ngdc.noaa.gov/ngdcinfo/onlineaccess.html.

- A site operated by Collins Software (Houston, Texas) contains links to various GIS data: http://www.collinssoftware.com/freegis_by_region.htm

- Digital Watershed – This site, maintained by Michigan State University, includes spatial data and models similar to those found in EPA's Better Assessment Science Integrating Point & Nonpoint Sources (BASINS) program (see: http://www.epa.gov/waterscience/basins/). The Digital Watershed can be found at: http://www.iwr.msu.edu/dw/.

Regional Data Sources

Because ReVA focuses on regions, regional data sources can be well-suited for ReVA applications. The types of spatial data available at regional scales obviously vary with the region of interest. To date, ReVA has used regional datasets from the following sources:

- The Chesapeake Bay watershed has long been an area of concern and widely studied. The Chesapeake Bay Program databases can be queried based upon user-defined inputs such as geographic region and date range. Each query results in a downloadable, tab- or comma-delimited text file that can be imported to programs such as SAS, Excel, or Access for further analysis. Chesapeake Bay Program databases can be found at: http://www.chesapeakebay.net/data/. GIS data for the Chesapeake Bay monitoring program are available at: http://www.chesapeakebay.net/data/data_desc.cfm?DB=CBP_GIS

- The Mid-Atlantic Integrated Assessment (MAIA) encompasses the Chesapeake Bay watersheds, but extends farther, including the Mid-Atlantic states. Due to its high population density and rapid growth in population, the Mid-Atlantic region has been

studied intensively. Data for this region can be obtained at: http://www.epa.gov/maia/html/data.html

- The Southeastern Ecological Framework is a comprehensive set of spatial data on ecological resources and habitat for the Southeastern United States (U.S. EPA Region 4). These data can be found at: http://www.geoplan.ufl.edu/epa/connectivity.

Local Data Sources

With the spread of GIS technology for integrating and managing municipal functions, many cities, towns, and counties now generate and manage spatial data at the local scale. Local datasets include information on variables such as school and fire district maps, and zoning and land-use maps. Examples of local datasets for Wake County, North Carolina, can be reviewed at: http://www.lib.ncsu.edu/gis/wake.html#layers.

### *Data Quality Considerations*

The usefulness of any data depends upon their quality. For geospatial data, the Federal Geographic Data Committee (FGDC) has established metadata standards. Many spatial datasets now come with metadata files in text, html, or XML formats. These metadata describe the nature of the data, its lineage, the procedures that were used in processing and generating the data, and the potential uses and limitations of the data.

Two main data-quality elements are of concern when using spatial data: locational accuracy and attribute accuracy. Locational accuracy refers to the accuracy of information about the spatial location. For example, if the location of a soil-sampling site is given in latitude and longitude, the associated metadata should reflect the accuracy of that measurement (i.e., within 10 meters). Attribute accuracy refers to the accuracy measurement of the variable of interest at the location. Again, using the same soil-sampling example, the accuracy of a measured constituent in the soil (such as cadmium) at the location of interest might be plus or minus 5 parts per million. Frequently, further processing or generalization of the data may introduce additional uncertainties that should also be documented and considered.

# Section 3

# Data Processing in ReVA

Once all of the individual core datasets are acquired, they must be assembled into a single GIS database and one or more spatial units must be selected for reporting final results (Figure 2). Some data may need to be reapportioned if their collecting or enumeration unit differs from those of the reporting unit. Additionally, missing data may need to be estimated by interpolation or extrapolation to complete a dataset. Then, metrics can be calculated or modeled (modeling is discussed in the next chapter). Finally, appropriate metrics can be identified from the full suite of variables for integration using ReVA integration tools.

## *Database Management*

It is good practice to choose a single projection and datum (reference point) for storing all spatial data before generating metrics. In the Mid-Atlantic study, for example, two raster datasets (NED and NLCD) were used to prepare several metrics. The native projection for both of these datasets was standard U.S. Albers, NAD83. Projecting raster data requires resampling the data, and should be avoided if possible. Therefore, U.S. Albers, NAD83 was chosen for the Mid-Atlantic study, and all data in other projections were converted to this projection before further processing.

## *Reporting Units*

Descriptive metrics must be summarized and reported for specific areas. These areas, called reporting units, need to be of appropriate scale and relevant to the study. Some examples of commonly used reporting units are political boundaries (such as counties), naturally-defined areas (such as watersheds or ecoregions), or equal-sized cells in a square or hexagonal grid. Each type of reporting unit has advantages and disadvantages.

Watersheds are a good choice for reporting units for water-quality studies: for many watersheds, data are available for various factors at the watershed outlet. Further, most stressors and resources that can affect the sample data are contained within the reporting unit. In the ReVA Mid-Atlantic study, the 8-digit Hydrologic Unit Code (HUC) was one of the reporting units used. HUCs are advantageous in that they can be scaled in size, from 2-digit (the largest) to 12-digit (the smallest). Currently, HUCs that are smaller than 8 digits are not available for the entire United States, but are available for some areas.

An advantage of counties as a reporting unit is they often represent the decision-maker's area of interest. Further, information in one of the core datasets, census data, is collected by county, or even by smaller units that nest within county boundaries. Disadvantages of

counties as reporting units are that county boundaries may not correlate well with natural boundaries, and they cannot be scaled.

Grid cells as reporting units have several advantages. They are all the same size, which can facilitate comparisons between areas. Grid cells also can make it easier to notice patterns in indicator maps. Finally, grid cells can be scaled, meaning that the user may select any size for the cells. Unfortunately, grid cells do not match either decision-making boundaries or natural boundaries, which is a significant drawback. Further, grid placement is arbitrary, so shifting the grid may substantially change some indicator values in some cells.

### *Data Reapportionment*

Some data, such as socio-demographic or economic data from the USCB, are collected by specific areas; these data are generally enumerated by county. When the collecting or enumerating area and reporting unit boundaries do not match, data must be apportioned from one area to the other. The easiest way to do this is by area-weighting. For example, if 20% of a county is located in HUC 1 and 80% is located in HUC 2, then 20% of the population for the county would be assigned to HUC 1 and 80% would be assigned to HUC 2 (Figure 3). An area-weighting method involves the assumption that values (the number of people, in the current example) are evenly distributed across the county, which is obviously incorrect in many or all cases.

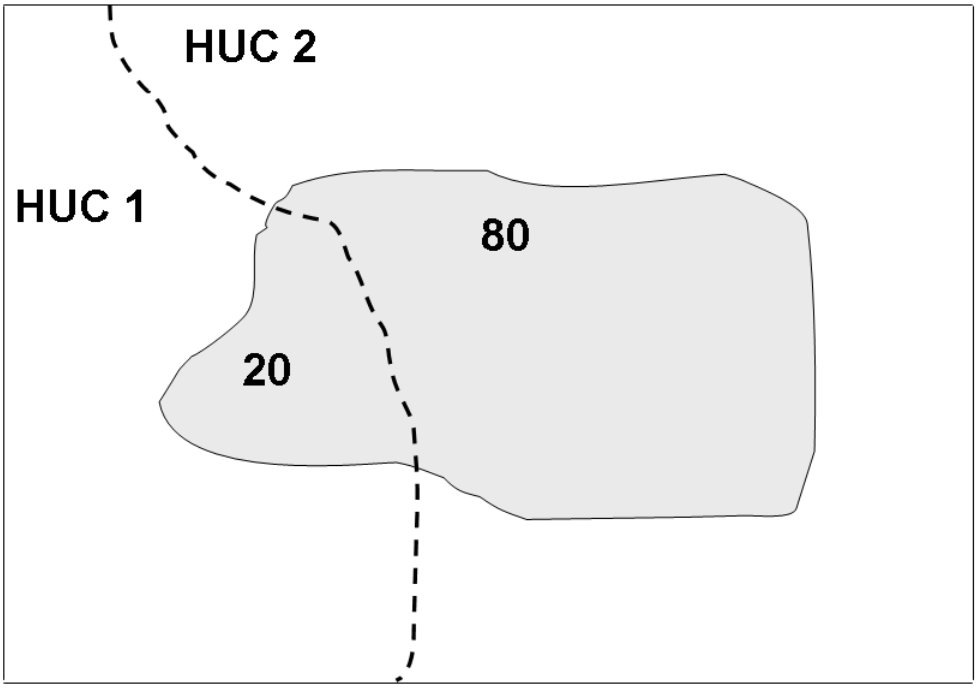


**Figure 3. Example of simple apportionment of data by area-weighting. Hydrologic Unit Code 1 (HUC 1) contains 20% of the population designated by the shaded area, while HUC 2 contains 80% of the population. The boundary between HUC 1 and HUC 2 is represented by the dashed line.**

For census data, a better approach for reapportionment makes use of the fact that population and housing units are enumerated by smaller block groups within counties. County-level variables, such as the number of children under five years of age, can be apportioned to block groups based on proportion of the county population contained in the block group. If, for example, a county has 1,000 children under five, and block group 1 contains 2% of the county's total population, then 20 children can be assigned to that block group. This apportionment method involves the assumption of an even distribution of demographic and economic conditions across the county – a more realistic possibility, compared to assuming an even spatial distribution of people. An example of population reapportionment by small reporting units located within two HUCs is given in Figure 4.

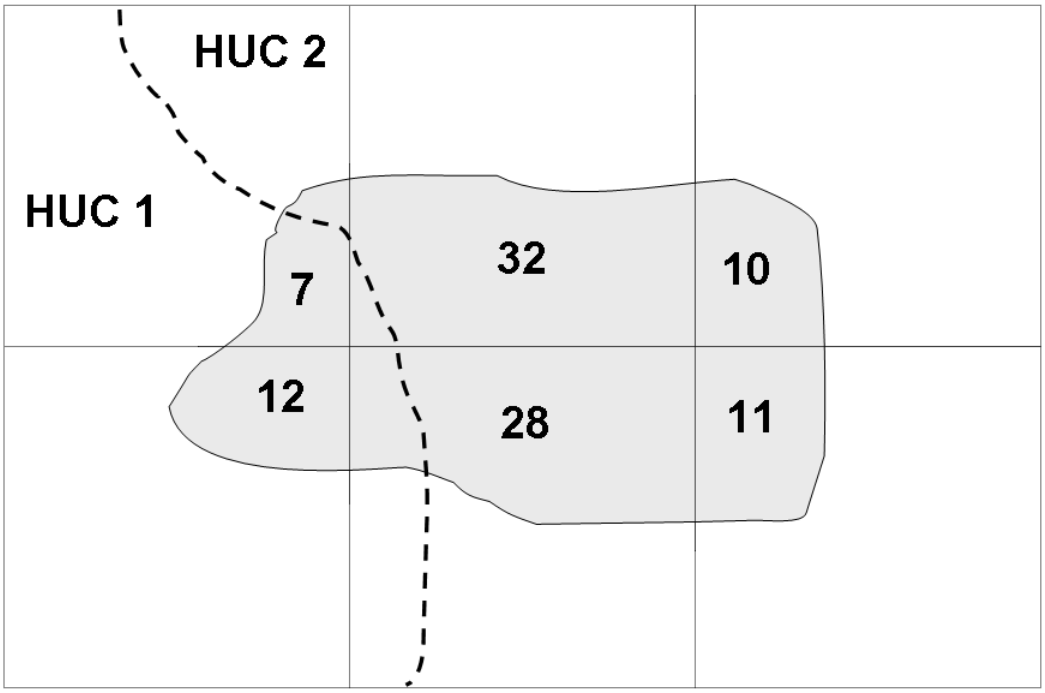


**Figure 4. Example of apportioning population data for small reporting units (represented by squares) for an urban area (shaded area) that occurs in two HUCs. Values shown in the counties represent percentage of the urban area's population; thus, they sum to 100.**

Continuing with the Mid-Atlantic ReVA example, block groups were then intersected with HUCs (although some other reporting unit could be used, as noted previously) and values were apportioned using the area-weighted method described above. In this process, using the much smaller block groups, rather than counties, was expected to mitigate much of the error introduced by the assumption of even spatial distribution. Values from each of the block groups in the HUC, partial or whole, were then summed to determine the overall estimated value of the 1990 population total for the HUC.

### *Missing Data*

Missing data can be estimated using various interpolation or extrapolation methods. These techniques are meant to be used only with continuous data, such as elevation; they are not appropriate for categorical data, such as land cover.

Extrapolation is the process of using known data to predict values for areas or times beyond the spatial or temporal extent of the known data (Figure 5). An important assumption of extrapolation is that observed patterns or trends are consistent in space and time. Therefore, extrapolation is usually more reliable over short distances or time

intervals. This method of estimating missing data becomes progressively more suspect when applied to larger distances or longer time intervals.

Interpolation is the process of estimating values between two or more known values (Figure 5). As with extrapolation, data can be interpolated over time and space. Linear interpolation is the most straightforward method of estimating values, but other functions can be used for interpolation. Common spatial interpolation methods include Inverse Distance Weighting (IDW), splining, trend surface analysis, and kriging.

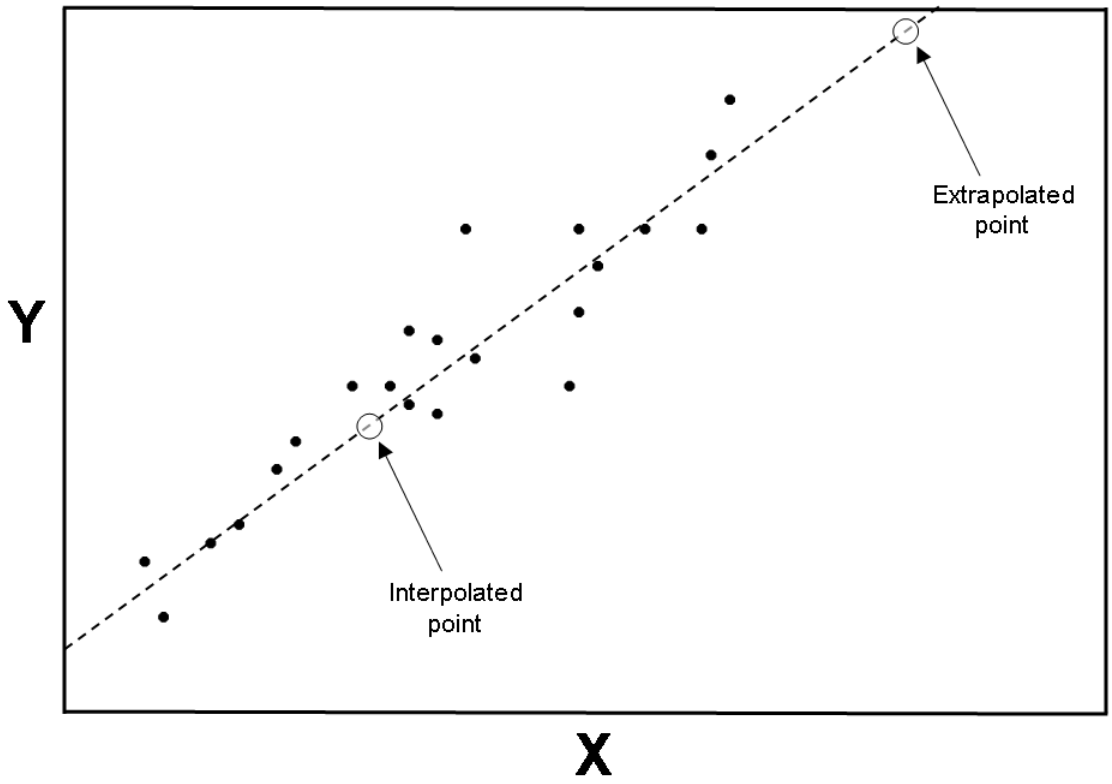


**Figure 5. Graph (hypothetical X and Y axes) showing measured data (solid circles), an interpolated point, and an extrapolated point.**

Kriging is a method of spatial interpolation that minimizes the variance of estimation error. It is a linear, unbiased, least-squares method that uses spatial covariance to help estimate values at locations that have not been sampled. Kriging is often used with point data, such as air quality samples, to create a surface map where every cell has a value. An excellent reference on kriging is Cressie (1993).

### *Metric Selection and Preparation for Integration*

After datasets for individual variables are assembled and documented, the variables must be examined carefully for relevance, consistency, and interdependence. Then comes the hard part: variables (or metrics), or combinations of variables, must be selected for use in integration. In an earlier document (Smith et al., 2003), we reviewed 11 methods for integrating metrics into indices for use in ReVA. That review provides details on how each of the 11 indices are calculated and it provides discussion on each method's advantages and disadvantages. Our objective in this section is more basic: we note that while some ReVA metrics are developed from models (described in more detail in Section 4 below), others are simpler and can be calculated without the use of models. Examples of metrics that do not depend on the use of models are percent of forest cover and road density within the reporting unit. Percent forest cover simply involves overlaying land cover on the reporting unit and dividing forest area by total area, then expressing the proportion as a percentage. Road density is a similarly simple metric and is estimated by overlaying roads on reporting units using standard GIS tools to determine the sum of road length within each unit. In the Mid-Atlantic study, many of the indicators related to land cover were generated using an ArcView extension called the Analytical Tools Interface for Landscape Assessment (ATtILA http://www.epa.gov/nerlesd1/land-sci/attila/index.htm).[3]

As a first step in preparing metrics for ReVA, it is important to examine the relevance of each variable to the assessment being performed (Smith et al., 2003). Expert judgment often is required in this process. One might decide, for example, to include both total fish biomass and biomass of a fish species known to be highly sensitive to water pollution. One might decide to include numbers of people employed in forestry as a resource variable vulnerable to urbanization, and yet exclude numbers employed in financial industries as marginally related to the current assessment. Including variables that are not of immediate relevance can bias the integrated estimates of vulnerability across the region.

The second step in index development is to examine the frequency distribution of each variable across the region. The examination can reveal outlier data that need to be explained. In the Mid-Atlantic dataset, for example, several watersheds had values for sedimentation that were nearly an order of magnitude greater than elsewhere. Close inspection revealed that the high sedimentation values had been derived from an independent study that had estimated sedimentation values using a linear regression between landscape variables and sedimentation. The watersheds that had unusually large sedimentation values had landscape values that were well outside the range used in the original regression model. No other method for modeling sedimentation was available so it was necessary to eliminate the sedimentation variable from the dataset. In other types of study, it might be acceptable to simply truncate the frequency distribution and eliminate the outliers. In the ReVA approach, eliminating the outlier values means

[3]ATtILA is a free software application developed by EPA's Landscape Sciences Program. It is used to calculate many of the landscape metrics used in ReVA-type assessments and can be applied to any type of land-use/land-cover data (i.e., any scale).

eliminating those watersheds from any further analysis, because there must be a value for every variable that is used, for every watershed.

Because many of the integration methods in ReVA are statistically based, it is necessary to examine the frequency distributions for discrete variables that can violate the assumptions of the statistical methods. Discrete variables may enter environmental data sets because presence/absence data are common. The result of presence/absence data is a frequency distribution with peaks at 0 and 1, and no intermediate values. In the original Mid-Atlantic dataset, presence/absence data were available for many individual species. This problem was solved by aggregating the presence/absence data into continuous variables representing the number of terrestrial and aquatic species within a watershed.

Sometimes it is not feasible to aggregate variables to overcome the problem of presence/absence data. Then, the variable should be eliminated before using ReVA integration methods to prepare regional estimates of vulnerability. The variable can still be retained for some specific analyses, such as mapping regional patterns of presence and absence.

The third step in index development is to examine the candidate variables for mathematical dependence. Mathematical dependence means that some variable "X" is simply a mathematical combination of other variables. For example, one cannot include native forest acres, nonnative forest acres, and total forest acres in an index, because the third variable, total forest acres, is simply the sum of the other two. This type of problem is solved by eliminating any one of the three variables.

Many of the integration methods in the ReVA approach assume that the variables are statistically independent. Therefore, the fourth step of index development is to examine the dataset of variables for statistical interdependencies. The simplest way to do this is to search the variance-covariance matrix for all variables across all watersheds for unusually high correlations.

In a variance-covariance matrix, high covariance values (i.e., those near 1.0) may indicate that the two variables are essentially measurements of the same stressor or resource. For example, "number of families below the poverty line" and "low annual household income" are two very similar measures of a social group that might be vulnerable to environmental degradation. One of the variables should be dropped from the dataset, carefully choosing and retaining the variable that is more relevant to the current assessment objective. As a rule of thumb, variable pairs that have covariance values above ~0.95 may need to be considered closely for the possibility of inappropriate redundancy.

In other cases, high values of covariance may represent subtle mathematical dependencies. Two variables which logically appear to be independent may actually be mathematically related. This can occur, for example, with landscape cover metrics attempting to measure contiguous habitat on the watershed. High values of covariance between calculated values in this case may indicate that the different equations have

converted to the same measure of contagion, at least on the watersheds under consideration. If this is discovered to be the case, one of the variables should be dropped from the dataset.

In considering the covariance matrix, high values of covariance between two stressor variables may mean that the two are measures of the same underlying stress on the ecological system. In this case, one of the two variables should be eliminated, as noted previously. However, the high values for covariance between two stressor variables also may be due to other factors. Stressors such as air pollution and water pollution may co-occur in watersheds, even though the two stressors originate from different sources and have independent mechanisms. The co-occurrence of two independent stressors in this case means that the stress on the ecological system is significantly increased. Thus, both stressors are appropriate for assessing environmental vulnerability in this case and both variables should be retained.

Significant correlations also can occur between a resource variable (such as biodiversity) and a stressor variable (such as forest fragmentation). When these correlations result from an underlying cause-effect relationship, both variables should be retained in the dataset.

To facilitate combining the variables into integrated measures of condition and vulnerability, the data are normalized. Normalization is used to ensure that all variables have the same numerical range and can thus be compared. In the Mid-Atlantic dataset, for example, a range of 0 to 1.0 was chosen, where 0 represents the "best" value of a variable across the region and 1 represents the "worst" value. Having found the "worst" value for a variable in the region, the values of that variable in the remaining watersheds can be divided by the "worst" value to standardize the values between 0.0 and 1.0.

Variables also must be "directionalized" before integration. This ensures that all variables that represent a negative or positive condition are aligned such that high values mean the same thing, and that low values have the same meaning. For variables that are clearly resources (a positive attribute with a normalized value tending toward 0) and stressors (a negative attribute, with normalized values tending toward 1), this is simple and may require only a change in sign before normalization. However, for other variables, such as socioeconomic data or other descriptive data, it may not be as clear how to directionalize. In some cases, for example, a variable (e.g., population density within the watershed) is considered as a stressor on the ecological system and thus is normalized with a value that tends toward 1. In other cases, a variable (e.g., number of threatened and endangered species) is considered a resource, but one that renders the ecosystem more vulnerable to additional stress. In this case, the variable might be coded such that its normalized variable tends toward 1. In previous ReVA applications, we have generally evaluated if the value of the variable increases the overall sensitivity of the reporting unit to additional stresses. If so, the variable is considered to move condition and vulnerability in a negative direction, so directionalization should tend toward 1. In short, careful judgment must be exercised in such cases and the direction of

variable standardization may need to be adjusted depending upon the assessment question.

The highest value of a stressor such as human population growth is considered to be the "worst" value, and thus is assigned a value of 1.0. Conversely, the highest value of a resource (such as native aquatic fauna) is considered to be the "best" value, and thus is assigned a value of 0.0. This method of coding is advantageous in that it allows the assessor to quickly evaluate all variables for a watershed. A watershed that has many variables with scores near zero is considered to be in relatively good environmental condition, because the low scores mean that resources are high and stressors are low.

The method of variable normalization and directionalization used in the ReVA Mid-Atlantic dataset provides a relative estimate of "best" and "worst," because the limits are chosen as the extremes within the region. This coding strategy has the advantage of spreading the data across the extremes within the region, which simplifies the task of distinguishing between watersheds. But the approach has drawbacks, too: the coding strategy does not use objective criteria of "good" or "bad." The result is that watersheds might be considered to be in relatively good condition within the region, even though all of the watersheds within the region might be in poor condition if judged against an objective standard. Unfortunately, the present state of knowledge does not allow objective criteria to be developed for most variables, so the analyses are limited to evaluating relative vulnerability.

A better method of standardizing the variables would be based on thresholds established by statute or scientific study. Individual variables then could be coded by the extent to which variable values were above or below the specified threshold. Thresholds may be available as ecotoxicological ECx values (ECx refers to a concentration above which an associated adverse effect occurs, for "X" percent of the individuals in a population). Thresholds also may be based on an expert opinion or on a societal consensus as expressed in statutes that limit human activities. An extensive literature review was conducted in an attempt to find thresholds for the variables used in the Mid-Atlantic region. This review revealed that thresholds existed for only a small percentage of the variables and could not be used as the basis for standardization in this application.

A final factor that must be considered before integrating the variables into useable indicators is whether there is an imbalance in the dataset between different factors influencing vulnerability. For example, a dataset may have five measures of stressors on the aquatic system (e.g., riparian vegetation, agriculture on steep slopes, inputs of pesticides and herbicides, and roads crossing streams), but only one measure of the biotic community (e.g., the number of native aquatic species). In this case, if one were to calculate an integrated measure by summing the coded variables, one would be assigning five times more weight to the stressors than to the single resource. To avoid an imbalance between the stressors and the resource, one can average the five stressors, then use this average to represent a single composite stressor.

In general, the best approach to an imbalance between stressor and response variables is to first categorize the dataset into groups of discrete factors, such as terrestrial stressors and terrestrial resources. Then one can average within the groups before combining the data to assign equal weights to the different factors. The need to balance and the exact groupings needed to achieve balance is determined by the purpose(s) of the assessment. If, for example, one wishes to determine which of the individual aquatic stressors is most important, then averaging the several stressors would not be appropriate. Alternatively, if one wants to assess the relative condition of watersheds across the region, then balancing the dataset is generally appropriate. The easiest way to do this is to give equal weight to each of the factors determining condition. The choice is a matter of judgment and the answer may differ for different analyses done on the same regional dataset.

# Section 4

# Spatially Explicit Models

## *Why Models?*

Spatially explicit data are required to compare risk across a region (Hunsaker et al., 1992). Typically, data for regional assessments include infrastructure (e.g., roads), stressors (e.g., atmospheric deposition, chemical inputs), landscape features (e.g., geology, elevation, vegetative cover), sensitive resources (e.g., wetlands), and ecological endpoints (e.g., avian biodiversity). Unfortunately, in many cases these data are not in a format that can easily be incorporated into a regional analysis. For example, consistent monitoring data for surface water and ground water are usually only available at relatively fine scales and these data are unevenly distributed across a region. To use these types of data, models are needed to estimate values for points where data are not available. For this reason, models are an important part of the overall ReVA process. However, it is critical to keep in mind that models are only tools to guide the researcher to further inquiries about the nature of the system under evaluation. Models are an abstraction or simplification of a more complex system: they are not truth but "the lie that helps us see the truth" (Fagerstrøm, 1987).

## *Types of Models*

Mathematical models translate our understanding about relationships (e.g., cause-effect processes) into equations. Such models help reduce the vast quantities of available data and facilitate the generation of useful hypotheses. Further, data which appear to be "outliers" based on the model are more evident, which makes it easier to determine if the outliers are really outliers or if the model needs to be adjusted. The two classes of mathematical models that are most commonly used during any assessment are empirical models and process models.

### Empirical Models

Empirical models are used to examine the relationships between single and multiple variables without incorporating the underlying mechanisms responsible for the relationship. In ReVA, for example, statistical (empirical) models relate land cover to a dependent variable, such as nutrient load, pollutant deposition, or bird migration. The relationship between land cover and a dependent variable can be linear, exponential, bimodal, or any of a large number of other forms; the relationship only needs to be simple and statistically strong (assuming a large geographic area is used to capture a broad range of variability). The simplified structure of an empirical model is both its strength and its weakness. Empirical modeling allows an investigator to evaluate large

quantities of data, but it does not provide information on the fundamental cause of the observed relationships.

Process Models

Process models, as the name implies, include known processes or mechanisms in nature. In the case of ReVA, a model such as AQUATOX could be used to predict changes in biological and ecological endpoints such as the abundance of phytoplankton, the abundance of game and bottom fish, the concentrations of nutrients and dissolved oxygen, or even the percentage of organic matter in sediments in response to toxic organic chemicals. The predicted conditions could then be expressed spatially and thus be incorporated as changes in resources. The major drawback to process models is the extensive effort needed to "fit" the model with reasonable values for its constituent parameters (e.g., current populations, population growth rates, land use, pesticide application rates, lake dimensions, etc.) which are needed to operate the model.

## *Examples of Spatial Models Used in ReVA*

When doing an assessment at a broad regional or national scale, the lack of or uneven distribution of monitoring sites often requires the development of spatial models for filling in areas not covered. Several examples of the types of models and model output ReVA has used to meet assessment needs are given below as "thumbnail" examples. The models range from regression to Bayesian analyses to more complex combinations of statistical and mathematical algorithms.

Nitrate and Sulfate Deposition Modeling

Nitrate and sulfate deposition estimates used by ReVA came from an empirical model developed by Grimm and Lynch (2000). This model addressed the sparseness of the National Atmospheric Deposition Program monitoring sites by using a multiquadric equation developed by Hardy (1971) which provides the density needed for use in a spatial-weighted linear least-squares regression algorithm. This model yields deposition estimates as a function of latitude, longitude, elevation, slope, and topographic aspect. The elevation, slope, and aspect parameters all are derived from USGS DEM datasets.

Mercury Deposition Modeling

Bayesian statistical methods were used to develop models which derive interpolated maps of weekly mercury deposition. Data on monitored samples of mercury deposition were supplied by the National Atmospheric Deposition Network – Mercury Deposition Network (http://nadp.sws.uiuc.edu/mdn/). However, due to the small number of monitoring sites available, additional data were needed. By including nitrate, sulfate, and precipitation, all of which correlate with mercury, we were able to supplement the amount of spatial information. Using these related sources of information, we developed a space-time model that provided spatial predictions of nitrate, sulfate, mercury, and precipitation, as well as their associated uncertainties, including spatial and temporal

misalignment among the networks. Since the depositions of these constituents occur in response to precipitation, we also modeled a spatial field of the probability of precipitation for the area of interest. Depositions and probabilities of precipitation, in turn, are jointly modeled through time-varying linear models of co-regionalization (Banerjee et al., 2004). The mercury-deposition model therefore provides a constructive specification of the cross-covariance function allowing for non-stationarity and dependence among the fields.

Invasive Species Modeling

Genetic Algorithm for Rule-set Production (GARP) modeling was used to create spatial maps of the potential distributions of invasive species within the region of interest (see figure 6). GARP uses occurrence data for a species within its current range to predict the species' likely distribution within the area of interest. Input data include spatial data on species occurrence and environmental factors such as temperature, precipitation, solar radiation, snow cover, and frost-free days. GARP uses multiple rule types including BIOCLIM, logistic regressions, and a genetic algorithm (an artificial intelligence application) to generate a set of "IF…THEN" rule statements that describe the relationships between the species and the environmental conditions. The output from GARP can then be projected onto a "new" landscape to visualize the species' potential distribution. The distribution also can be projected onto areas of an actual or potential invasion/introduction under different land cover and climatic conditions (Peterson et al., 2003).

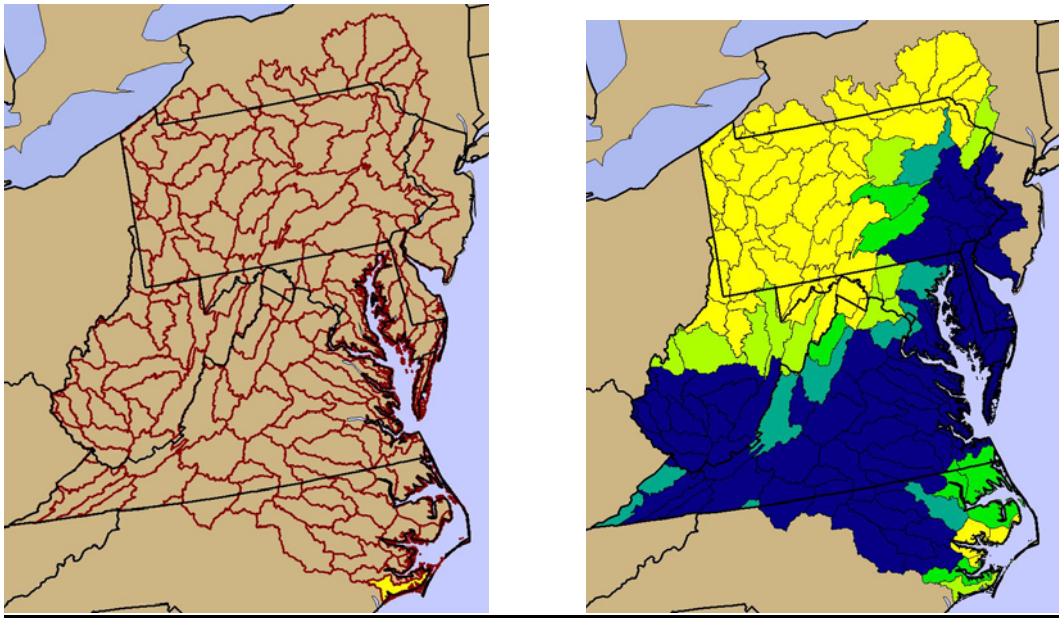

**Figure 6. Graphic depicting 2000 distribution of Giant Salvinia (*Salvinia molesta*) and estimated 2020 distribution using the Genetic Algorithm for Rule-set Prediction (GARP) model.**

Nitrate in Ground Water Modeling

Data on ground water quality obtained from USGS National Water-Quality Assessment Program (NAWQA) studies were used in association with geographic data to develop logistic-regression equations to predict the probability of nitrate exceeding a specified management concentration threshold (Greene et al., 2005). The geographic data included land cover, soil permeability, soil organic matter, depth of soil layer, depth to water table, clay content of the soil, silt content of the soil, and hydrologic groups within a specified area of influence. The relationship of these factors with nitrate concentrations above a threshold was determined using logistic regression. Since well data were not uniformly distributed across the study area, the coefficients calculated from the significant geographic features were used to create a surface map of the likelihood for exceeding acceptable levels of nitrate across the study area (figure 7).

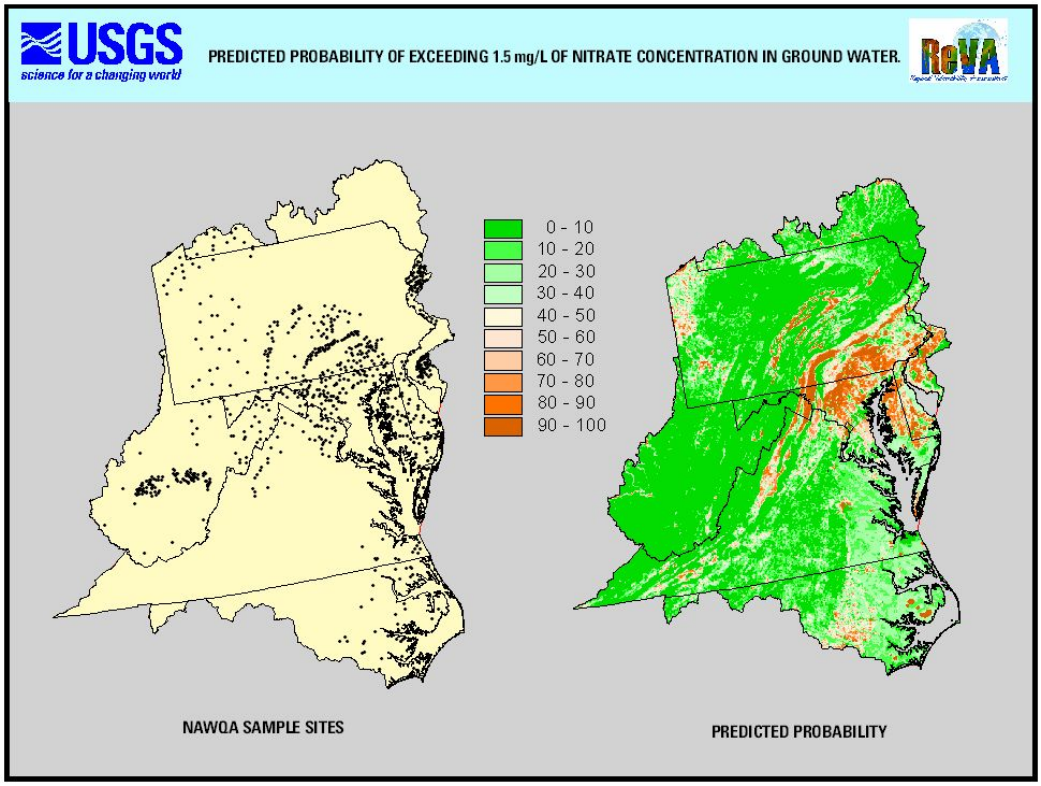


**Figure 7. Graphic showing NAWQA sample sites (map on left) and results of logistic regression model that estimates probability of exceeding a threshold of nitrate concentration in shallow ground water aquifers across the Mid-Atlantic region.**

## *Forecasting Drivers of Change*

ReVA relies on various models to evaluate current and future ecological conditions and vulnerabilities at a broad spatial scale. At the regional scale, a number of drivers of

ecological change operate to effect changes that are observable at this broad scale and over a temporal scale that may span decades. Changes may not be observable at a local scale until they reach some threshold, yet may have irreversible consequences if not anticipated and addressed strategically. Regional-scale drivers of change include the following:

- Land-use change
- Resource extraction (e.g., over-fishing, timber harvest, mining)
- Changes in pollutants (e.g., nonpoint source pollution, agricultural runoff) and pollution (e.g., changes in atmospheric deposition)
- Spread of invasive, non-indigenous species (e.g., pests and pathogens, introduced species)
- Climate change (e.g., changes in weather patterns)

Of these drivers of change, land use is probably the most important, because land cover and land-use pattern affect every other driver of change. Thus, land use is often one of the most important parameters in any model that estimates a future distribution of stressors related to resource extraction, changes in pollutants and pollution, the spread of invasive species, and even changes in local weather patterns.

Resource extraction often follows development, as roads are constructed to access remote areas where resources have not yet been exploited. Models of nonpoint source or agricultural runoff specifically include land use as input parameters; these can represent the amount of chemical applications for farmland and sediment loading in areas that lack riparian buffers. Models of air deposition include mobile-source estimates, as well; thus, the pattern of road networks has implications for regional air quality. Many invasive non-indigenous species are transported by people and the spread of such organisms is generally facilitated by transportation networks. Similarly, land cover provides habitat for invasive species, which relates to the range of their spread. And finally, regional-scale models of climate change can include land-use/land-cover information as inputs, because local weather patterns are influenced strongly by surface roughness and reflectance, in addition to shading afforded by vegetative cover.

### Land-Use Change Models as a Component of ReVA

Land-use change models are an important component of ReVA. A particular problem is posed by projecting land-use changes caused by population growth – that of apportionment. For example, population growth can result in conversion of land to residential and agricultural uses (Wheeler et al., 1998). Distributing these changes spatially is critical to projecting changes in stressors such as aquatic nonpoint source pollution (e.g., percent impervious surface or agriculture on steep slopes) and forest productivity. Land-use changes also can directly alter estimates of resources (e.g., wildlife habitat, wetlands, etc.). To identify the most appropriate model for forecasting land-use change in the Mid-Atlantic region, ReVA reviewed and evaluated several land-use change models (Wagner et al., 2006). These models ranged from simply documenting plans for highway construction and new employment centers to estimating

land demand from state census projections to customizing applications of a traditional resource economics model (Hardie and Parks, 1997) to a state-of-the-art cellular model of urban growth (Clarke et al., 1997).

In the Mid-Atlantic region, ReVA chose to use output from the Slope, Land use, Exclusion, Urban, Transportation, Hillshading (SLEUTH) model in combination with other sets of land-use data (figure 8). SLEUTH uses a cellular automata simulation approach (Clarke et al., 1997) to illustrate future urbanization based on historic patterns of land transition. We chose SLEUTH because it distributes change spatially, employs more complex rules than those of a typical cellular automata simulation method, and uses numerous data sources (including topography, road networks, and settlement distributions) to accumulate probabilistic estimates based on Monte Carlo methods (Jackson et al., 2004). It is, however, an *urban* growth model and thus may not effectively represent regional land-use change processes, such as changes in rural land use or the creation of new urban centers.

For forecasting, SLEUTH assigns each 1-km cell a probability of being developed in any given time frame. We chose 50% as the threshold and created a binary map of developed/not developed in 2020. SLEUTH does not address any other land-cover changes (e.g., conversion of forest to agriculture).

The following steps were used to create the final future land-use map for use in regional analysis.

1. Begin with NLCD 1992 (30-meter resolution).

2. Add new urban areas, based on outputs of the SLEUTH model. SLEUTH produced 1-km raster output of projected areas of urban growth. Areas predicted to have a 50% or greater probability of being developed were "burned in" as urban cover on the NLCD map.

3. Planned roads and road expansions were overlaid with NLCD and "burned in" as new developed cover in 2020.

4. Mining permits were obtained from Virginia, West Virginia, and Pennsylvania. All permitted areas were assumed to be mined in 2020; each of these areas was also "burned in" as mined area in 2020.

5. Areas where mines and urban were coincident were coded as "mines."

6. Areas that did not coincide with new urban area, roads, or mining retained their 1992 land cover status.

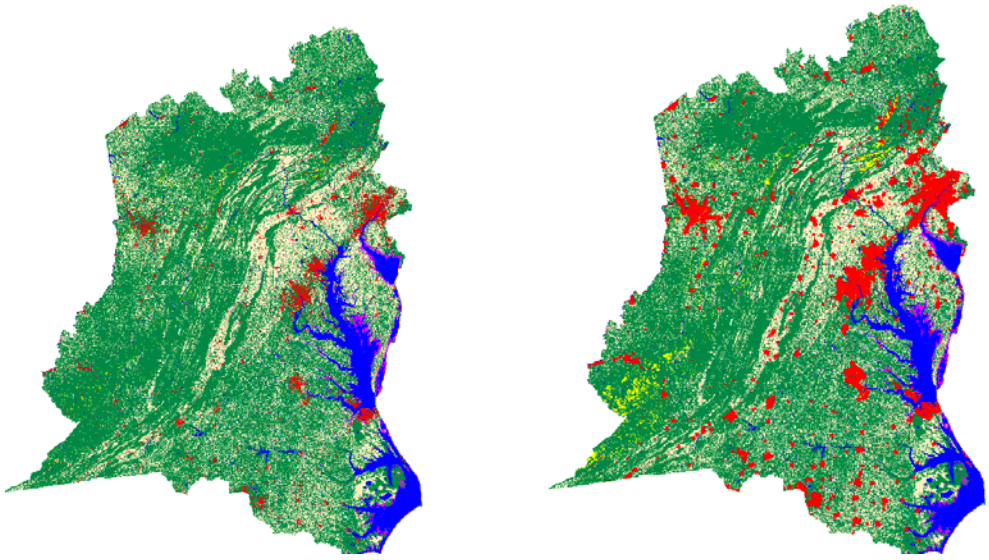

**Figure 8. Graphic depicting current land use/land cover based on the National Land Cover Database (NLCD) 1992 (left) and estimated 2020 land use/land cover using the Slope, Land use, Exclusion, Urban, Transportation, Hillshading model (SLEUTH) in combination with planned roads and permitted mines (right).**

## Models that Use Land Cover/Land Use as Input

As future scenarios are developed, the challenge is to translate the projected scenarios into spatial changes in stressors and resources. In most cases, the changes can be extrapolated using the same models that are used in assessing current conditions. Since population growth and urbanization results in changes to land-use proportions, it is simply a matter of applying the model to the new land cover.

## Resource Extraction

Many ecological resources are considered vulnerable, yet the use of these resources provides society with valued goods, services, and other benefits. Such benefits can involve resource extraction (e.g., forests and minerals), recreation (e.g., hiking and fishing), waste treatment, and nutrient recycling. Vulnerable ecological resources in this category are critical because damage to them can impact society immediately. In the East, which was the focus of our initial ReVA studies, forests are one of the largest resources of concern. Forests provide numerous goods and services, including recreation, economic timber harvest, and species habitat. Forests are vulnerable to urban growth, fragmentation caused by timber harvests and accompanying roads, and introduction of exotic pests and pathogens. Mineral extraction is a driver of ecological change largely due to the impacts to other resources (e.g., water quality, habitat, etc.).

The USDA Forest Service's Forest Inventory and Analysis data were used to estimate current and future forest conditions at the watershed scale (Schaberg and Abt, 2004). A timber economic forecasting model (Subregional Timber Supply Model; see Prestemon and Abt, 2002) was used to project trends in timber harvest and forest sustainability into

the future. This model included information on projected land-use change, because we expected that much of the timber resource extraction would follow new developments.

As mentioned in the section under land use, we used available state mining to predict where mining would likely occur in the future. This was reflected as a change in land use for our future land-use/land-cover map.

Pollutants/Changes in Water Quality

The susceptibility of a landscape to erosion is estimated by semi-empirical models such as the Universal Soil Loss Equation (USLE). The USLE has been widely used to estimate average annual soil loss (mass per unit area) according to known erosion mechanisms: rainfall, soil type, slope, vegetative cover, and agricultural management practices. By incorporating the new values developed for future land cover, it is possible to get an estimate of potential future erosion. Modifications of the USLE, such as the RUSLE (Revised USLE), make use of meteorological data to estimate soil erosion with temporal responses for specific time periods or rainfall events. Although many soil transport mechanisms have been characterized for small watersheds, existing models for estimating sediment delivery to surface waters require an extensive local calibration. As a result, these models lack utility at scales suitable for regional assessments.

Nutrient loading models such as the LTHIA (Long-Term Hydrologic Impact Assessment; Harbor and Grove, 1997), Reckhow's (1980) model (which was incorporated into the ArcView extension, ATtILA), or land-cover-based regression models from pour-point samples can be modified to use forecasted land-use change data. With LTHIA, the land-cover grid or the percentages of the land-use types can be used, depending on the model (Pandey et al., 2000). For the Reckhow model and other statistically based models, the percentages of each land-use type are used in conjunction with a set of land-use coefficients to calculate the overall nutrient load for a watershed. However, the ATtILA extension will convert percentage values and allow the user to set coefficients based on regional knowledge (U.S. EPA, 2004).

Spread of Invasive Non-Indigenous Species

Future scenarios of invasive non-indigenous species also were created using the projected land-use/land-cover map as input. GARP modeling (see section on examples of spatial models) was used to estimate the future distributions of several species of concern. The projected distributions were based on habitat requirements, which includes land cover and land use.

Models that Do Not Include Land Use/Land Cover as Input

Air pollution modeling is complex and requires multiple layers of data, including estimates of emissions from stationary and mobile sources. These data are needed to predict pollutant loadings to the landscape from the atmosphere. Land use and land cover are not extremely important in this case and these conditions generally are included at the regional traffic-demand modeling phase for mobile source emissions. Urban growth also

may be considered when estimating increased numbers or demand from stationary sources (i.e., Energy Generating Units). However, other factors, such as topography and meteorology, are very important factors in predicting air pollution. Two regional pollutant datasets used by ReVA include the National Air Toxics Assessment (NATA; see http://www.epa.gov/ttn/atw/) and the Community Multiscale Air Quality model (CMAQ; see http://www.epa.gov/asmdnerl/CMAQ/).

# Section 5

# Creating Alternative Future Scenarios

Alternative scenarios are *not* intended to predict the future. Rather, they present a series of plausible future states that are likely to include: (1) a mixture of modeled projections of current trajectories and/or prospective forecasts, (2) alternative policy and/or management options that will likely affect ecological goods and services, and (3) various spatial and monitoring data that are used to describe both the baseline and the alternative scenarios in terms of landscape characteristics and associated ecosystem services.

General guidelines for creating scenarios are provided by Liu et al. (2007), Pandey et al. (2000), and Weingand (1995). These guidelines suggest the following: (1) scenarios that include "best-case," "worst-case," and a "most-likely" case are both useful and informative, (2) scenarios should be distinct, or at least different enough to discern changes over space and time, (3) scenarios should explore the bounds of what is feasible, and (4) scenario creation should have a clear focus, purpose, or direction, thereby ensuring that the number of scenarios created, analyzed, and assessed is kept to a minimum (fewer are better than many). ReVA follows these guidelines in developing scenarios and recommends ReVA users to do the same.

## *Building Alternative Scenarios for Analyzing Future Trends*

Alternative future scenarios can be prepared either by creating a set of static future scenarios for a specified time in the future or by projecting trends in a series of time steps until a specific time period has been accommodated. Either method must include some projections of past trends (e.g., population growth, corn prices, etc.) that the ReVA user may want to combine with conditions that differ significantly from those that have been observed in the past.

## *Scale Considerations in Projective and Prospective Modeling*

Regardless of the mix of *projective* (extension of past trends) and *prospective* (significantly different from past trends) modeling, the appropriate scale of the model and the geographic extent of the region being modeled must be considered. This is particularly true of land-use change models, which in the past have been developed to represent *urban* growth trends, but only rarely have captured *regional* growth processes which include the development of new urban centers and rural to exurban land conversions.

### Spatial Scale

A good reference for local-scale growth models is U.S. EPA (2000). Examples of regional land-use change models include the Resource Economics Model (Parks et al., 2000) (this model is not spatially explicit), the Spatially Explicit Regional Growth Model (SERGoM; see Theobald, 2005), and the Integrated Climate Land Use Scenarios (ICLUS) project (e-mail from Britta Bierwagen, U.S. EPA Global Climate Change Program, to Elizabeth Smith, U.S. EPA National Exposure Laboratory, dated May 2007), which is being developed by the Global Change Program.

Various spatial scale issues are associated with climate change. If climate change is included as part of the scenario creation, for example, then a prospective model of projected changes in weather patterns at a suitable scale (regional or subregional) will be more appropriate than one at a national or broader scale.

### Temporal Scale

Alternative future scenarios can be created in ReVA either by projecting conditions for multiple time steps that build on one another or by creating a future scenario independent of intermediate conditions that is based on some vision of the future. The environmental decision-maker should consider the type of end product or decision tool that is envisioned and how the future scenario will feed into that product or tool. For example, if the goal is to display changes over time in response to user input, then fine-scaled time steps for the forecast models may be needed to create dynamic responses. These short time steps might also be needed to represent processes important for assessing changes in ecosystem services. Alternatively, if the objective is to compare a suite of discrete scenarios that cannot be altered by the user, then coarser time steps (e.g., decadal, as in ICLUS) may suffice. Generally, for large geographic regions, fine-scale temporal detail is not feasible because their inclusion can greatly increase the need for computational resources. Temporal detail also may be less critical at broad spatial scales because changes in ecological services generally take time to become evident. In other words, resolution will be coarser when the goal is to represent the broad spatial scale.

### Anticipating Responses to Policies

Beyond projecting change by continuing a current trend, alternative scenarios are used to explore futures that involve trade-offs of ecosystem services through alternative decisions, policy levers, or incentives. Determining what these policies are likely to "look like" can be done by obtaining input from experts (i.e., from EPA Program Offices). Alternatively, a forward-looking estimate of future policies can be citizen-driven. That is, planners and developers can provide information on what commercial and residential densities are feasible for certain sections of an urban area, or a group of stakeholders could envision a future they would like to see.

Five types of issues are associated with incorporating input from experts or citizens/ stakeholders: (1) future scenarios with too many details (clients who want a "perfect" prospective future – really more of a *prediction*), (2) too many scenarios (trying to please

everyone), (3) bias (i.e., listening to only a few sources of input), (4) plausibility, and (5) trouble converting the input into a spatial model. The first four issues can be dealt with by managing expectations, working iteratively with stakeholders, and clearly conveying the capabilities of the regional approach. Approaches to the fifth issue are discussed in the next section.

Using Other Spatial or Monitoring Data to "Spatialize" Scenarios

Alternative future scenarios must be spatially explicit to effectively represent effects on ecosystem services and the trade-offs associated among the alternative scenarios. However, not all of the information used to create the alternative future scenarios will be in this format. Therefore, available spatial data and GIS decision rules are used to "spatialize" nonspatially explicit model results and other features of the scenarios, such as possible policy alternatives. Here is an example of why "spatialization" is needed, and how it can be accomplished. In the Future Midwestern Landscapes project, ReVA uses an economic projection of crop plantings (acreages) based on prices for corn and other crops (dollars per acre). This projection is used to determine how much land is planted as feedstock for ethanol (used as a biofuel). The results of this model must be expressed spatially, using information such as SSURGO soil data, National Agricultural Statistical Survey (NASS) crop data, and spatial representations of tillage practices, streams, roads, protected areas, etc. For policy alternatives, GIS decision rules are needed to develop the models to reflect these alternatives.

It is possible that additional point or monitoring data (e.g., air deposition data) may be used to create baseline or alternative landscapes. To do this, it is necessary to create a surface from these points, using some form of extrapolation, interpolation, or other type of model that predicts conditions at specific locations, such as spatial statistical approaches.

# Section 6

# Synthesis

Once an extensive dataset that covers many aspects of environmental quality and vulnerability is assembled, it is necessary to synthesize or integrate the information. If the information is not integrated, it is difficult to evaluate the overall effectiveness of environmental policy. For example, restoration of riparian vegetation may improve stream water quality, but if agriculture on steep slopes, roads crossing streams, wetland loss, and urbanization are extensive on the watershed, then planting trees along the stream bank may not improve in-stream water quality.

Combining individual variables into an integrated indicator is inevitably controversial (Andreasen et al., 2001). Researchers, for example, may have a sophisticated understanding of the interplay of environmental variables and frequently will disagree on the overall impact of these variables. As a result, the scientific community rarely is content with a single integrated value; they generally will want to examine the original data and debate the implications. Few decision-makers, on the other hand, possess the scientists' sophistication in data interpretation, but decisions must be made and limited resources must be allocated, even if scientists disagree. Similarly, stakeholders will want to know if actions taken in the past have actually made the environment "better," and they may not agree as to which aspects of the environment should be prioritized. So the question is not whether or not to synthesize and integrate the data – instead, the challenge is to develop and test innovative approaches to integration.

### *Available Information and Data Preparation*

An important limitation on the ReVA approach is the quality and extent of the available data. In the case studies that have been examined by ReVA to date, information has been limited to variables measured in other programs (Smith et al., 2004). Indeed, one of the original motivations for the ReVA program was to synthesize the multiple physical, chemical, and biological datasets being gathered by disparate programs within EPA. Resources do not exist within the ReVA program to perform field measurements. Thus, the analyses and the conclusions drawn from the analyses are limited by the available data. In the Mid-Atlantic study, for example, adequate information was available on remotely sensed land cover, but relatively little information was available on biodiversity across the region. Data on the numbers of native, non-indigenous, and threatened and endangered species were only available for relatively small number of taxa. Therefore, the study could not examine or represent important aspects of biodiversity.

In some cases, variables can be calculated using models developed by other researchers. Examples range from well-studied air quality models for nitrate and sulfate deposition to regression models relating watershed land cover to stream water quality (Jones et al.,

2001). To date, resources have not existed within ReVA to develop complex simulation models requiring testing and validation, such as exposure models or dose-response models. Instead, the program relies on testing and validation operations performed by the originators of the models.

To apply the ReVA methodology, it is important to understand the relationship between the data and the objectives of the analysis. In applications such as the Mid-Atlantic study, the objective was to assemble the extensive available data, place the data into a regional spatial framework, and explore the possibilities of integrating the data to locate potentially vulnerable watersheds. The study did not begin with a problem; it began with the objective of locating spatial patterns of environmental quality and identifying potential problem areas that might not be identified by other methodologies. Other applications may well begin with more specific objectives, such as relating spatial patterns of development to air and water quality.

When the study involves a specific goal, the objective will determine the data needed. In some cases, assessment questions may involve smaller scales such as a single 8-digit HUC. However, the more common case will be that the data to address these questions simply do not exist. If data needed to address the assessment questions do not exist, the ReVA methodology cannot be used to address the assessment questions.

In many assessment projects, available data and models have been supplemented by the use of expert opinion. Expert opinion is often qualitative or, at best, can be described by principles of Fuzzy Logic (Klir and Bo, 1995). This presents major but surmountable problems for integrating expert opinion with measured data or modeled variables. Often, expert opinion is the only available option for supplying information required for a specific assessment. ReVA will likely be using expert opinion in future projects and the use of expert opinion will require developing the appropriate analytical tools for its integration.

### *Methods for Integrating Variables*

#### Simple Sum and PCA Sum

The simplest method for integrating the variables is to sum their normalized values. This is referred to as the Simple Sum. Because the summation method contains no prior assumptions about relative importance of the variables, this approach is easily understood. The purpose is to provide an overview of the spatial pattern of environmental quality by combining stressors, resources, and socioeconomic factors.

Because the Simple Sum does not account for the correlation structure of the regional dataset, ReVA also developed an integrating method referred to as the Principle Components Analysis (PCA) Sum. The PCA Sum method accounts for correlations by weighting variables by principal components. Details of the method can be found in Smith et al. (2004).

After some experience with using the Simple Sum and PCA Sum methods, we recommend that these two methods always be used together. The PCA Sum method removes potential bias if many stressors co-occur in space. On the other hand, the Simple Sum method might be more useful if the co-occurring stressors act synergistically.

The two summation methods are visualization tools that allow one to see all of the spatial patterns of all of the variables in a holistic or synthetic manner. Because these methods provide such a generalized picture, they should not be used for providing answers to assessment questions. Rather, they should be used simply to visualize the spatial pattern of potential environmental problems across a region.

A potential problem that arises with the two summation methods is an imbalance. This can occur, for example, if one has many measures of water quality and only one measure of land-use change. The resulting sum is heavily weighted toward the aquatic. To avoid this problem, one can average within categories (i.e., aquatic or terrestrial) and sum the averages.

Best and Worst Quintiles

To calculate the Best and Worst Quintiles, variables are ranked and subdivided into quantiles with the same number of watersheds. Each watershed is then evaluated in terms of the number of its scores that fall in the best and worst quantiles. Watersheds are then depicted in quantiles again, based on these counts. This method must be used with caution as it is not a very sophisticated analytical technique. It can, however, be used to highlight where favorable and unfavorable conditions tend to cluster within the region.

A Monte Carlo uncertainty analysis was performed for this approach using the Region 3 dataset (Tran et al., 2007a). The results of this analysis showed that data errors had little impact on which watersheds appeared in the "Best Quintile" (i.e., the 20% of watersheds in the best ecological condition) or the "Worst Quintile" (i.e., the 20% of watersheds in the worst ecological condition). Watersheds in intermediate positions often shifted quintiles when error was randomly applied across the variables. We concluded that the Best and Worst Quintiles could be reliably estimated, but there was significant uncertainty about the positions of intermediate watersheds.

The explanation for this uncertainty pattern appears to reside in the regional dataset. Watersheds in the Best Quintile tended to be mountainous and inaccessible. In these watersheds, the resources were abundant and there were few human stressors, so all variables tended to have values near the "good" end of the spectrum. Random errors changed the value of the individual variables but did not change the sum, because all variables indicated good ecological condition. Conversely, the watersheds in the Worst Quintile tended to be urban; they had relatively few natural resources and numerous, co-occurring human stresses. Therefore, these watersheds tended to remain in the Worst Quintile even when random error was introduced.

State Space Method

The State Space Method measures the distance between two points (i.e., two watersheds) in multivariate space. The distance measure used by ReVA (Tran et al., 2006) avoids the potential bias in distance measures such as the Mahalanobis distance (see De Maesschalck et al., 2000).

The State Space Method is very versatile and can be used for various assessment applications. It can be used, for example, to measure the overall distance of each watershed in the region from a reference point. The reference point might be a nearly pristine area, such as a national park, in which case the distance is a measure of degradation from this pristine state. Alternatively, the reference point might be the most vulnerable watershed in the region, in which case the distance is a measure of resilience. In the current implementation, the user can choose the reference point and see how far the other watersheds deviate from this reference.

The State Space Method is particularly valuable in analyzing the results of scenario studies. In scenario studies, additional stressors, such as climate change or invading species, are imposed on the region. Conversely, the scenario may be designed to evaluate the regional improvement resulting from particular mitigation or restoration activities. The distance measure then indicates the degree of degradation or improvement on each watershed. This multivariate analysis is necessary because, for example, restoration activities alone may have little impact on overall quality in the region if all other stressors remain or worsen.

Criticality Analysis

Criticality Analysis is similar to the State Space method in that it measures distance from a reference state. But in this case, the reference state is a postulated prehuman or totally nondisturbed state. The logic is that this measures how far an ecological system has been disturbed away from the state under which it evolved. The greater this distance is, the greater the probability that the system will pass a critical stability point and change to a new state. The theoretical justification for this idea can be found in Smith et al. (2004).

Because Criticality Analysis measures a distance in multivariate space, any of several distance measures could be used. In the Region 3 study (Smith et al., 2004), ReVA used a fuzzy distance measure (Tran and Duckstein, 2002). This measure was chosen because the pre-disturbance state could not be defined with precision, so we estimated the pre-disturbance distributions of variables using fuzzy logic. While this choice seems reasonable, other measures of distance might be chosen in future applications.

Overlay Method

The Overlay Method attempts to identify watersheds where important resources still exist but the remaining resources are under significant stress. Such watersheds are vulnerable in the sense that further stress, e.g., from additional development, could result in the loss

of valued resources. Thus, the Overlay Method provides a direct measure of regional vulnerability.

The Overlay Method first divides a dataset into stressors and resources. In the current implementation, the coded variables are summed within stressor and resource classes. The method then classifies watersheds by comparing the number of resources with the number of stressors. When resources and stressors are both high, the watershed is likely to be highly vulnerable.

Stressor-Resource Matrix

In any regional analysis in which mitigation is a potential policy option, there is a need to identify the stressor(s) having the greatest impact on the valued resources and to identify the resources that are most intensively stressed. This has led the Society of Environmental Toxicology and Chemistry (SETAC) to develop a matrix methodology that uses expert opinion to identify the greatest stressor (Foran and Ferenc, 1999; Ferenc and Foran, 2000). The ReVA approach permits an explicit analysis based on regional data rather than on expert opinion.

The ReVA method constructs a matrix that blocks stressors and resources and connects the blocks with a vector. By raising the matrix to a large power, the influence of all stressors on all resources is captured in the vector. Mathematical details can be found in Tran et al. (2007). The largest vector element then indicates the most influential stressor. A similar matrix can be constructed to determine the resource receiving the greatest stress.

### *Moving to Smaller Scales*

While the Integration Methods are general and not limited to any specific scale, the ReVA methodology was designed for regional assessments and it is recommended that it be used only for that scale. The problem with applying the ReVA approach to smaller scales lies with the data. In general, the information available across the region cannot be directly applied to smaller scaled problems. For example, monitoring stations scattered across a region can be reasonably averaged upward to provide estimates at larger-scale watersheds. However, choosing smaller watersheds would result in missing data. The missing data can be supplied by spatial interpolation, but interpolation assumes that the monitoring stations adequately represented maxima, minima, and spatial trends – a condition that is rarely or ever the case. The result is that using interpolation to scale to a finer resolution typically introduces far greater error than averaging up to larger scales.

The greatest power of the ReVA methodology lies in assessing spatial patterns of vulnerability across large regions. Over large regions, remotely sensed land-use data, GIS technology, and advances in landscape ecology provide a powerful means for combining and analyzing spatial information. At larger scales (across topographic gradients, soil types, ecoregions, and human development patterns), the spatial pattern

and the differences among subregions can most clearly be shown and analyzed. At smaller scales (such as a single state or about twenty 8-digit HUCs), the patterns can be less interpretable and the statistical power of the integration methodology is diminished.

### *Uncertainty in ReVA Analyses*

All measurements have associated uncertainty, referred to as measurement error. At least two sources of measurement error are important to ReVA. First, the value assigned to the spatial unit has some associated uncertainty. Second, even if the metric value is known perfectly, there is uncertainty associated with the impact of the stressor on a response variable within that spatial unit. These two types of uncertainties are discussed in Wickham et al. (1997).

If the greatest strength of ReVA lies in integrating available data, its greatest danger of misapplication lies in assuming that the available data are sufficient. Implicit in the ReVA methodology is the possibility of false negatives. That is, based on available data, a given watershed may appear to be in reasonable ecological condition and not vulnerable to further stresses. In this assessment scenario, the watershed would not be given high priority for managerial action. However, the watershed may in fact be highly vulnerable due to stressors that are unknown at the time of analysis. For example, illegal dumping of toxic material or undetected leakage of raw sewage may be occurring. Such factors could make the watershed highly vulnerable to ecological damage, but may not be incorporated into the ReVA analysis.

Then again, it is unlikely that the ReVA methodology would produce many false positives. A watershed is identified as vulnerable in ReVA because it is known to contain important ecological resources and is known to be subject to multiple factors that stress the resources. While it is conceivable that the ecological system is uniquely resistant and resilient at this location, this possibility is unlikely. Therefore, when the methodology identifies a watershed as vulnerable, it is reasonable to assume that managerial action is called for – or, more conservatively, that the responsible officials need to examine these watersheds more closely.

# Section 7

# Results Communication

### *Audience and Assessment Needs*

How the results of any assessment are communicated depends largely on the intended audience and their specific assessment needs. ReVA's analyses are designed to be of value primarily for decision-makers, rather than stakeholders in general or for the general public, but the assessment information can be extended for environmental outreach. However, any kind of environmental outreach would require substantial work to interpret results in lay terms, and this goes well beyond the scope of the guidance offered here. Providing ReVA results as outreach information requires both a thorough understanding of the ecological data and results, and good communication skills to translate the scientific results into information that is readily understandable by nonscientists.

Within the broad category of decision-makers, different levels of detail are needed. These can vary depending upon the type of assessment question that is being asked and the expertise of the decision-maker that needs the information. Compare, for example, the needs of a U.S. EPA Deputy Regional Administrator, who must determine which division within the region should have the largest share of discretionary funds based on critical issues, versus a Water Division Director, who must determine if funding should go toward restoration efforts in one watershed or toward establishing partnerships with a local community to promote smart growth practices in another watershed. The higher-level decision-maker (in the current example, the Deputy Regional Administrator) may not need detailed information on individual endpoints such as water quality or future vulnerability of aquatic biota. Rather, his/her needs could include a review of all endpoints, using an index that represents the current conditions across the region. Similarly, the Water Division Director may have little interest in endpoints other than those specific to water. A specific example of how ReVA has addressed these differences in needs is provided in the Regional Growth Decision Tool (RGDT) that was created for the Sustainable Environment for Quality of Life (SEQL) project (Figure 9). This toolkit provides options for three levels of users, each of which have different assessment needs. The level of detail of information is reflected in the types of indices used to "roll up" information (e.g., across multiple endpoints for decision-makers at higher policy levels, versus individual endpoints for analysts who need information in its most detailed format).

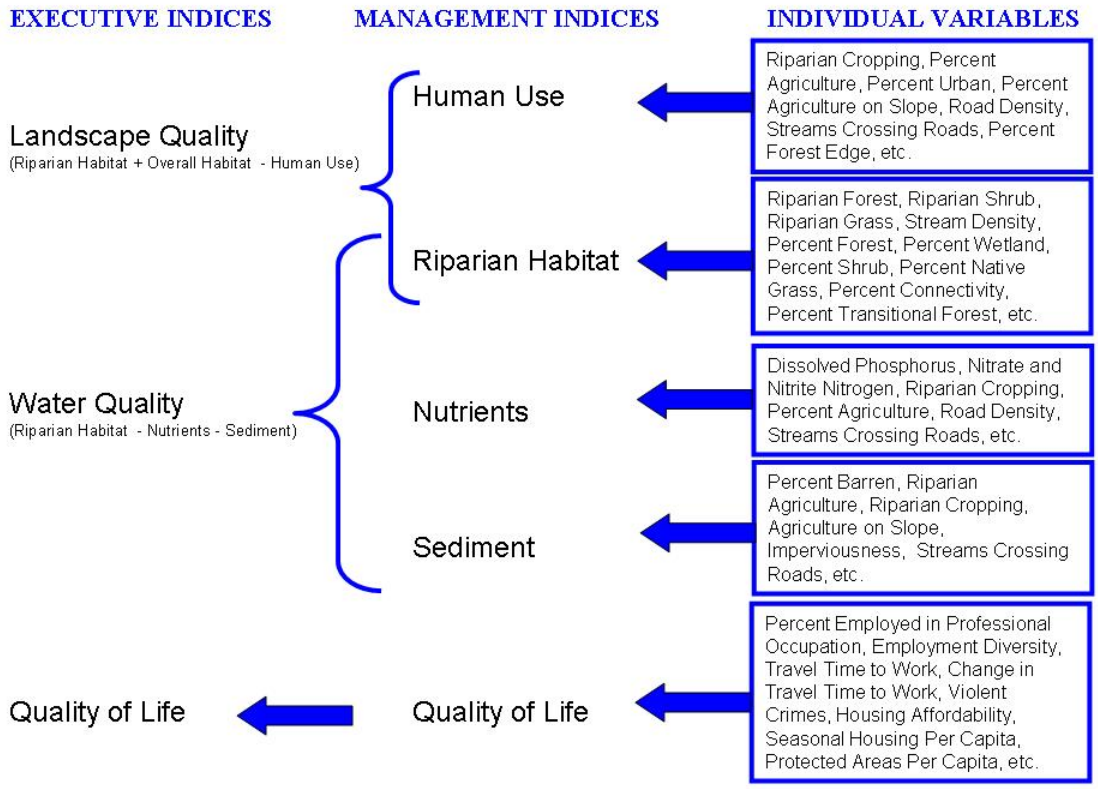


**Figure 9. Graphic depicting an example of variables and indices produced for different levels of users within the Sustainable Environment for Quality of Life (SEQL) project.**

### *Visualization*

Visualization of results is an effective way to communicate information. Since the approaches presented here are designed for spatially explicit analyses, mapped results are an assumed product. Careful attention to the details of how results are communicated is important even for an analytically-minded audience because differences in mapping can lead the user to very different conclusions. Examples of aspects that must be considered in mapping results include: (1) how relative differences in metrics or indices are "binned" across the region, (2) the choice of color codes for representing high/low or good/bad conditions, (3) the most appropriate representation of data distributions (normal versus skewed), and (4) how best to visualize metadata (i.e., the distribution of sample points or error/uncertainty maps).

The ReVA methodology generally encourages the use of overviews of individual metrics and indices for regional perspectives. However, ReVA users inevitably have the urge to drill down to individual reporting units and examine conditions at finer-than-regional spatial scales. Relationships between variables are best represented using standard statistical graphics such as scatter plots, bar charts, and box diagrams (Figures 10-12).

Quick visualization of integrated results for individual reporting units also can be accomplished by using graphics such as the "radar plot" (Figure 13).

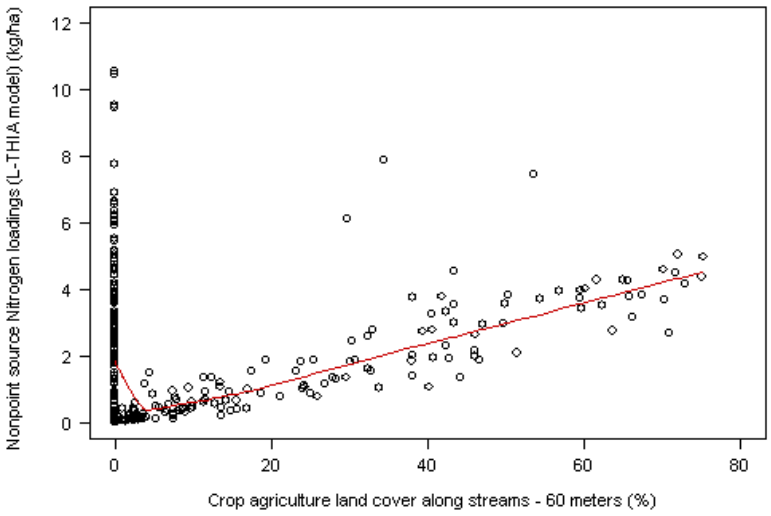


**Figure 10. Graphic depicting a scatter plot comparing two variables, nonpoint source nitrogen loadings as estimated by the model LTHIA and percent crop agriculture along streams within a 60-m buffer.**

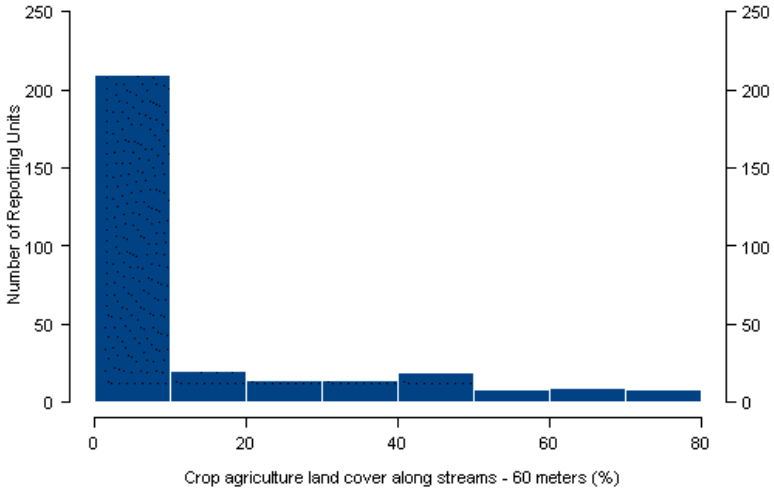


**Figure 11. Graphic depicting a histogram of number of watersheds and percent crop agriculture within a 60-m buffer along streams.**

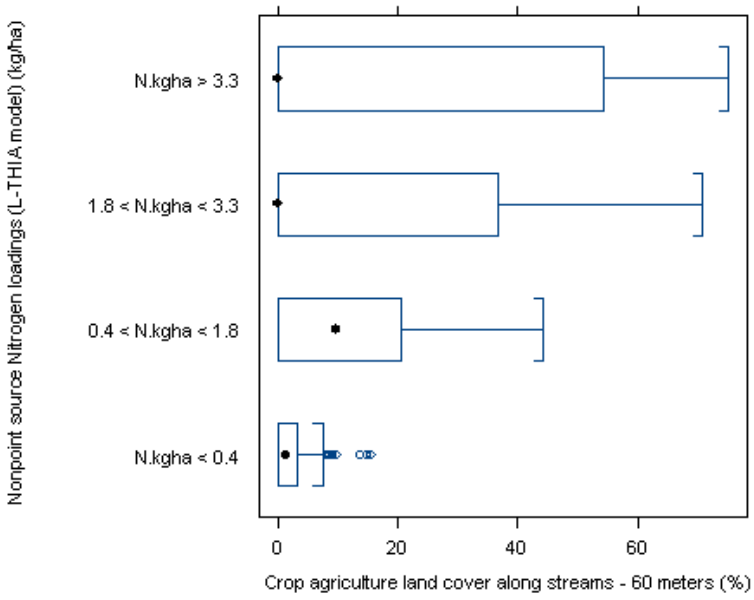


**Figure 12. Graphic depicting a box plot of nonpoint source nitrogen with percent crop agriculture within a 60-m buffer along streams.**

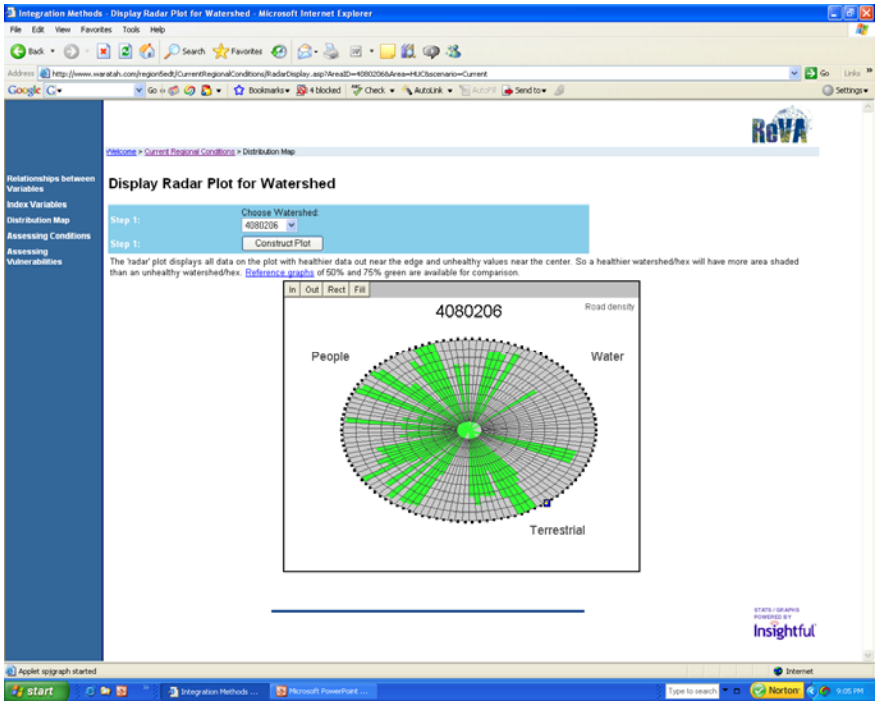


**Figure 13. Graphic depicting a screenshot from a ReVA Environmental Decision Toolkit (EDT) with a radar plot for a displayed 8-digit HUC. In a radar plot, each spoke of the wheel represents an individual variable and the amount of green represents the relative rank of that variable in relation to the same variable in all other 8-digit HUCs across the region (green represents good conditions, not green represents poor conditions).**

From the perspective of ReVA and many types of ecological analysis, it is ideal to have all data available as a surface map or data in a form that could be reformulated into a surface using some type of model. However, because data are not always available in this form, data are aggregated into reporting units and relative values for these reporting units are mapped across the region. Various options are available for dividing data into categories for comparison. These options include: (1) equal numbers of reporting units (e.g., watersheds or counties) in each bin or category (Figure 14), (2) equalized value ranges within each category (that is, if the range of values is 1-10 and the user wants five classes or bins, each bin would have a value range of 2) (Figure 15), (3) natural breaks in the data, where classes are based on natural groupings of data values (Figure 16), and (4) customized binning, in which classes are designed to highlight specific points in the data distribution, such as all reporting units exceeding a threshold and the spread of reporting units not exceeding this threshold.

The choice of color codes is important for several reasons. The first, and probably most important, reason is that choice of color can impart a subtle (or not so subtle) value judgment, such as occurs in the use of red-to-green colors selected to represent poor-to-good conditions across the map. This color-code choice may be the message a ReVA user wants to communicate. However, for some metrics, such as socioeconomic data, use of these colors may convey an unintended message. A good resource for selecting colors to represent relative differences in metric or index values is the ColorBrewer Web site, located at: http://www.personal.psu.edu/cab38/ColorBrewer/ColorBrewer_intro.html.

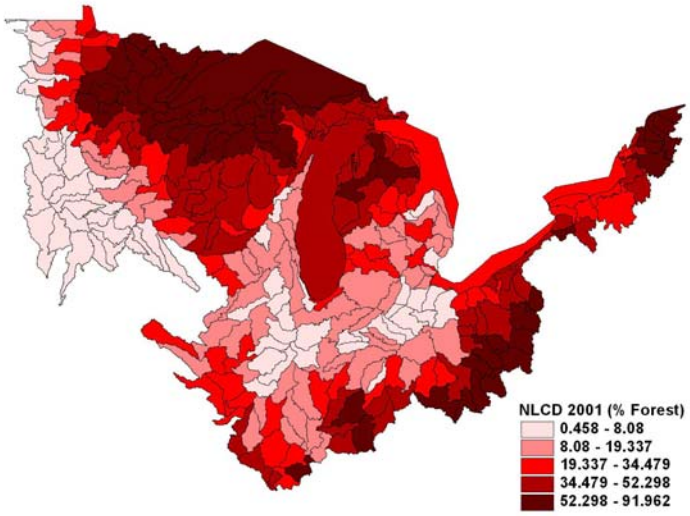


**Figure 14. Graphic showing the percent of forest cover for every 8-digit HUC across EPA Region 5 as displayed using quintiles as the binning method.**

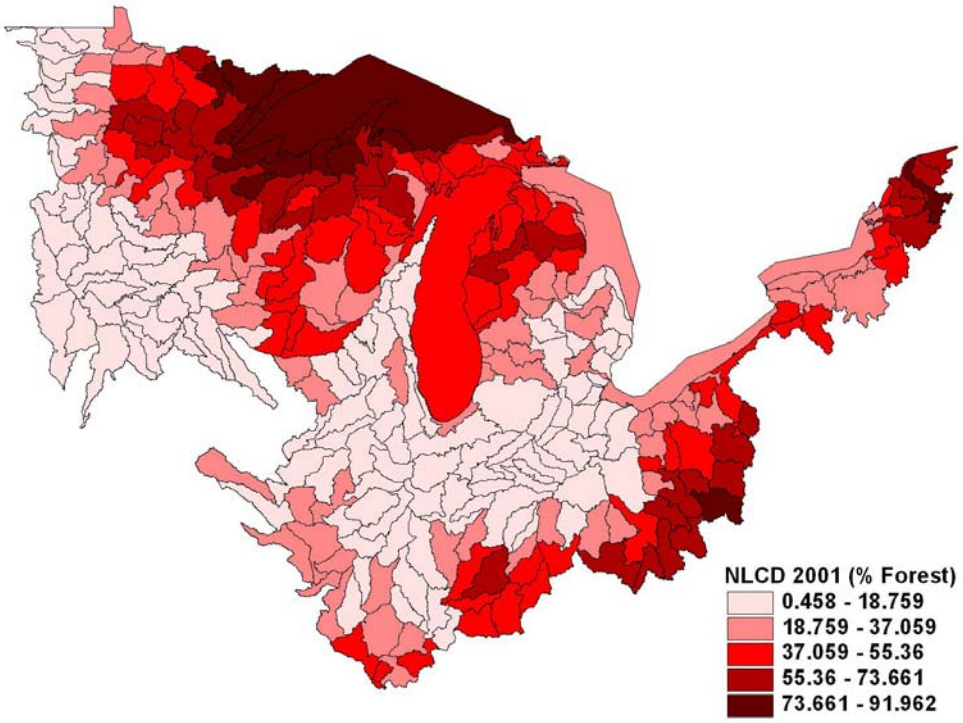


**Figure 15. Graphic showing the percent of forest cover for every 8-digit HUC across EPA Region 5 as displayed using equal intervals as the binning method.**

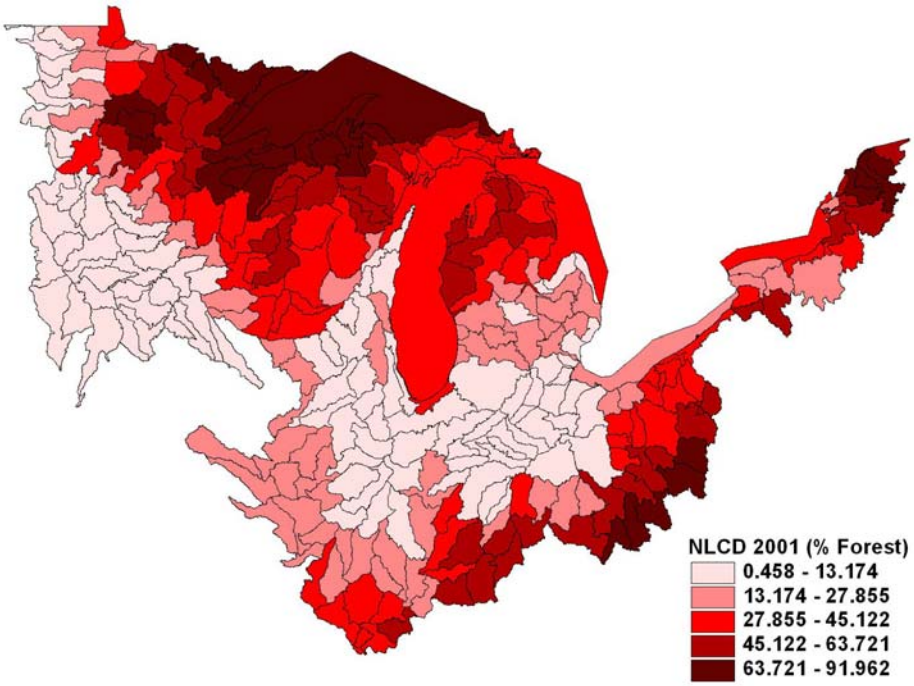


**Figure 16. Graphic showing the percent of forest cover for every 8-digit HUC across EPA Region 5 as displayed using natural breaks as the binning method.**

A clear understanding of data distributions is important when reviewing results, because differences in data distributions can inform the user as to how to interpret the mapped output. This understanding also is important because some of the data integration methods assume normal distributions. Bar charts, such as those that are provided by most statistical packages, are a good way to inspect data distributions; these charts allow a user to rapidly judge whether the different datasets are normally distributed, multimodal, or highly skewed. Evaluating data distributions also is useful in that it allows the user to better determine binning for reporting units when mapping results.

When feasible, metadata should be visualized, as this can be extremely valuable for users of individual data coverages. An example of this is the case in which a surface coverage has been developed using monitoring data. A map showing the number and distribution of monitoring points can be invaluable in communicating sampling density and areas of coverage. Similar benefits are evident for models that have error or uncertainty estimates that can be mapped; such maps can help communicate the validity of the model.

Other options for displaying results of analysis while maintaining the spatial context include using techniques such as *linked micromaps.* Key characteristics of the micromap template, which enhances graphical perception, are the ability to: (1) use position along a scale to represent estimates, (2) include multiple variables, (3) display confidence intervals for estimates, and (4) group large amounts of information into meaningful and manageable units for human interpretation. The micromap template consists of four elements: (1) parallel sequences of panels, (2) sorted study units, (3) partitioned study units, and (4) linked study units across corresponding panels (Carr et al., 1998; Carr et al., 2000; Carr et al., 2003). Figure 17 shows an example of linked micromaps used within a ReVA Environmental Decision Toolkit.

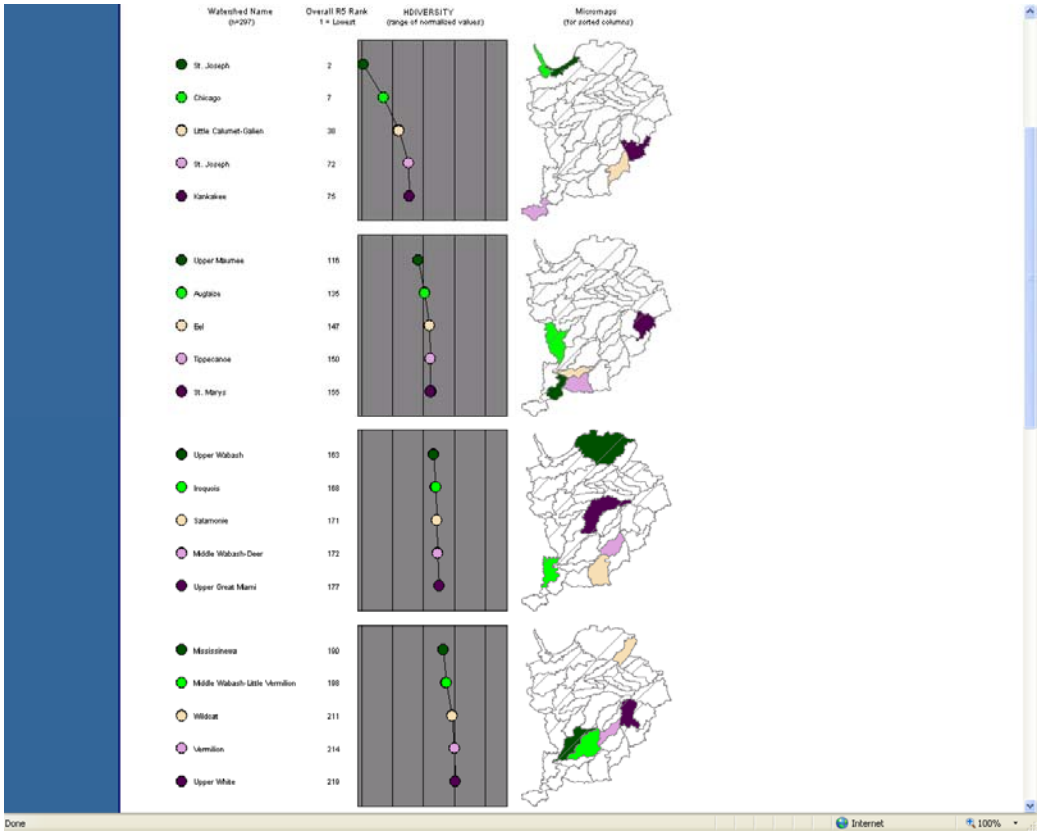


**Figure 17. Graphic depicting linked micromaps.**

Another consideration for visualizations of mapped results is that of adding locational information to help orient users of the information. This type of "orientational" information may include things such as state boundaries, major cities, county lines, etc.

Finally, especially when comparing alternative scenarios, difference maps are particularly effective. Difference maps highlight the differences between the current state and each of the alternatives. These maps allow users to see the trade-offs for the entire region and trade-offs among individual reporting units (Figure 18). If individual variables/metrics are also mapped with difference maps, the trade-offs can also be tracked among various endpoints. One watershed, for example, might gain in economic development under one scenario, but suffer declines in water quality.

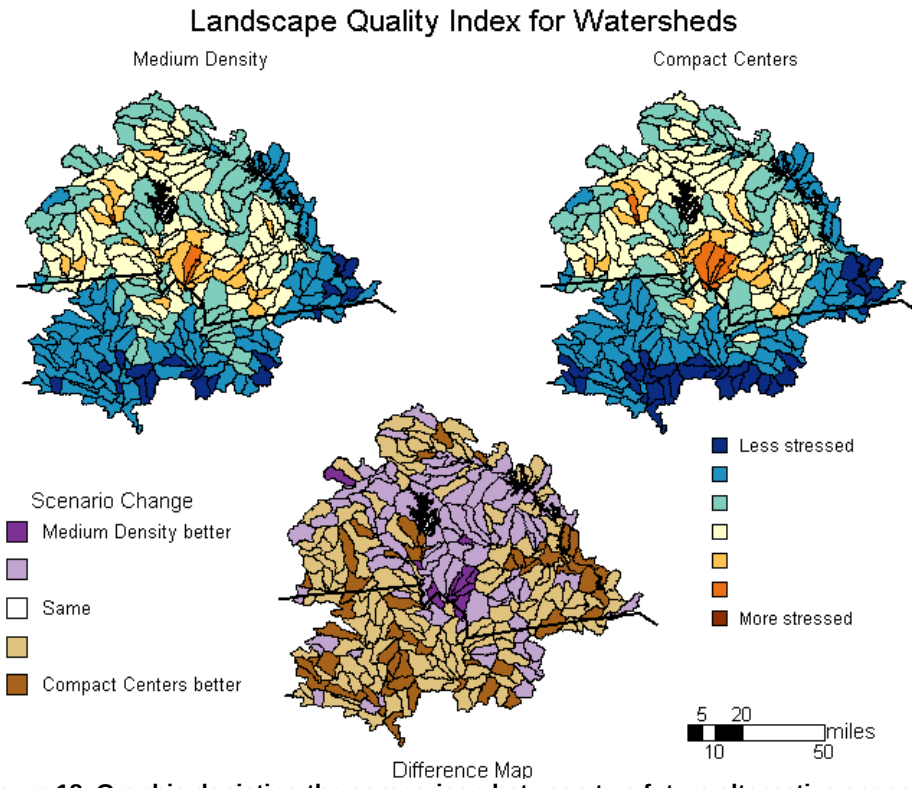


**Figure 18. Graphic depicting the comparison between two future alternative scenarios (upper maps) with a difference map highlighting both individual watershed differences as well as overall regional differences.**

# Glossary

**Area-weighting (areal interpolation):** A method of apportioning data from one geographic boundary to another when the boundaries do not match. For example, if 20% of a county is located in HUC 1 and 80% is located in HUC 2, then 20% of the population for the county would be assigned to HUC 1 and 80% would be assigned to HUC 2. An area-weighting method involves the assumption that values (the number of people, in this example) are evenly distributed across space (in this case the county).

**Bayesian statistical methods:** Statistical methods characterized by the updating of prior knowledge and estimation of conditional probabilities using Bayes' theorem and by the treatment of probabilities as subjective degrees of belief.

**Block groups:** As defined by the U.S. Census Bureau, a block group is a cluster of census blocks having the same first digit of their four-digit identifying numbers within a census tract. Block groups generally contain between 600 and 3,000 people, with an optimum size of 1,500 people.

**Continuous variables:** A quantitative variable that can take on any value over its range, including fractional values. Examples are measures of time, temperature, and chemical concentrations.

**Criticality analysis:** An integration method similar to the State Space method in that it measures distance from a reference state. But in this case, the reference state is a postulated prehuman or totally non-disturbed state**.**

**Difference map:** A GIS analytical technique where map algebra is used to subtract the values from one map from another.

**Directionalization:** In order to combine multiple variables into aggregate indices, some variables may have their values reversed (*directionalized*) in order to maintain a consistent definition for improvement or deterioration with a change in a variable score, such as "higher is better." Scores that formerly ranged from 0-100 might be reversed so that 0 becomes 100 and 100 becomes 0, with all other values inverted proportionally.

**Discrete (integer/categorical) variables:** Variables for which the values are not observed on a continuous scale because of the existence of gaps between possible values. Examples include integer values (number of people) or qualitative values (fair, good, moderate) that may be represented as numeric values.

**Ecoregion:** A large area whose boundaries are fixed by geography, topography, climate, vegetation, and other easily recognized natural features of landscape. Ecoregions contain many landscapes with different spatial patterns of ecosystems.

**Ecosystem:** The sum of the biotic and abiotic environment within which most or all nutrients are recycled.

**Ecosystem services:** The goods and services that people value that have natural functions or features as inputs. These goods and services cover a broad range, from food products to spiritual and cultural benefits. Ecosystem services can be divided into use and nonuse services, where use services are distinguished primarily by their requirement that users have direct access or proximity to sites generating goods and services, whereas nonuse services can accrue to those who are not close to the site and may never intend to visit the site.

**Ecotoxicological ECx values:** A concentration above which an associated adverse effect occurs, for "X" percent of the individuals in a population.

**Empirical model:** A mathematical model that is derived by fitting a function to data using statistical techniques or judgment.

**Endpoints:** A technical term used to describe the environmental value that is to be protected. An environmental value is an ecological unit and its characteristics. For example, salmon are valued ecological units; reproduction and age class structure are some of their important characteristics. Together "salmon reproduction and age class structure" form an endpoint.

**Euclidean distance:** The straight-line distance between two points on a plane. Euclidean distance, or distance "as the crow flies," can be calculated using the Pythagorean Theorem.

**Extrapolation:** The use of related data to estimate an unobserved or unmeasured value.

**Fuzzy logic:** A form of logic in which variables can have degrees of truth or falsehood.

**Geographic Information System (GIS):** A GIS is a system of hardware and software used for storing, retrieving, mapping, and analyzing geographic data. It is a computer technology that brings together all types of information based on geographic location for the purpose of query, analysis, and generation of maps and reports. GIS is both a database designed to handle geographic data and a set of computer operations ("tools") that can be used to analyze the data. In a sense, GIS can be thought of as a higher-order map.

**Geospatial data:** Information about the locations and shapes of geographic features and the relationships between them, usually stored as coordinates and topology; any data that can be mapped.

**Hydrologic Unit Code (HUC)**: A hierarchical, numeric code that uniquely identifies hydrologic units. The first two digits identify the region, the first four digits identify subregions, the first six digits identify accounting units, and the full eight digits identify subbasins. From the above example (definition of a hydrologic unit), the hydrologic unit codes are:

02 – the region (Mid-Atlantic)

0206 – the subregion (Upper Chesapeake)

020600 – the accounting unit (Upper Chesapeake. Delaware, Maryland, Virginia, and Pennsylvania)

02060002 – the subbasin (Chester-Sassafras. Delaware, Maryland, Pennsylvania)

Zeroes in the two-digit accounting unit field indicate that the accounting unit and the subregion are the same. Zeroes in the two-digit subbasin field indicate that the subbasin and the accounting unit are the same.

**Index:** A combination of multiple indicators.

**Indicator:** A concise measure of cumulative effects and ecosystem vulnerability.

**Interpolation:** A method of constructing new data points within the range of a discrete set of known data points. Interpolation can be performed on spatial or nonspatial datasets.

**Inverse distance weighting (IDW):** An interpolation technique that estimates values in a raster from a set of sample points that have been weighted so that the farther a sampled point is from the cell being evaluated, the less weight it has in the calculation of the cell's value.

**Kriging:** An interpolation technique in which the surrounding measured values are weighted to derive a predicted value for an unmeasured location. Weights are based on the distance between the measured points, the prediction locations, and the overall spatial arrangement (or autocorrelation) among measured points. The resultant interpolated points do not necessarily have to pass exactly through the input points.

**Land cover:** Anything that is visible from above the Earth's surface. Examples include vegetation, exposed or barren land, water, snow, and ice.

**Land use:** The way land is developed and used with respect to the kinds of anthropogenic (human-induced) activities that occur (e.g., agriculture, residential uses, industrial uses).

**Mahalanobis distance:** A multivariate distance measure that is based on correlations between several variables. It is a useful way of determining similarity of an unknown sample set to a known one. It differs from Euclidean distance in that it takes into account the internal correlations of the dataset and is scale-invariant, i.e., not dependent on the scale of measurements.

**Metadata:** Data that describe the content, lineage, quality, condition, and other characteristics of data. They are "data about data."

**Model:** A mathematical, physical, or conceptual representation of a system.

**Monte Carlo uncertainty analysis:** A computational method that involves repeated random sampling from the original (i.e., full) dataset in order to calculate results. Typically Monte Carlo simulations are performed when the underlying parameter cannot be estimated using deterministic methods.

**Non-Indigenous Species (NIS):** Nonnative plant, animal, or microbe species introduced into a region. Often, NIS can have significant impacts such as: overwhelming, crowding out, or disrupting relationships among native species, degrading habitats, and contaminating the gene pools of indigenous species. Examples include the wooly adelgid (an insect damaging hemlock trees in the Smoky Mountains), kudzu, and fire ants in southern U.S. and more than 160 known aquatic species in the Great Lakes.

**Nonpoint Source Pollution:** Pollution with a nonspecific location (i.e., those that are not discharged from a pipe outfall). The sources of the pollutant(s) are dispersed, not well defined, and typically not constant. Rainstorms and snowmelt often transport pollutants, increasing impacts. Examples include sediments from construction sites and chemical-bearing runoff from road surfaces and agricultural fields.

**Normalization:** A statistical technique that divides multiple sets of data by a common variable in order to negate that variable's effect on the data, thus allowing underlying characteristics of the different variables in a dataset to be compared. One common normalization technique subtracts the mean from each value and divides it by the standard deviation. This particular normalization will result in all the variables having a mean of 0 and a standard deviation of 1.

**Ordination:** A general class of multivariate statistical procedures used to create categories (or groups) of similar values. PCA is one of several different ordination techniques.

**Overlay method:** As applied by ReVA, an integration method that attempts to identify watersheds where important resources still exist but the remaining resources are under significant stress. Such watersheds are vulnerable in the sense that further stress, e.g., from additional development, could result in the loss of valued resources.

The stressors and the sensitive receptors in a location are summed separately so they can be compared. The two sets of values are overlaid to create a 2-dimensional scoring system that includes the potential for four end-members: 1) low-stress, low-resource; 2) high-stress, low-resource; 3) low-stress, high-resource; and 4) high-stress, high-resource. The latter is the most vulnerable situation.

**Principle Components Analysis (PCA):** A widely used multivariate statistical technique which can be used to reduce the number variables analyzed. PCA orthogonally transforms the original variables into a new set of uncorrelated variables based on the covariance (or correlation) matrix.

**Projective modeling:** A model used to predict future conditions based on an extension of past trends.

**Prospective modeling:** A model used to predict future conditions based on a change in existing trends (e.g., change in management practices or land use).

**Quantile:** Points taken at regular intervals from the distribution of a variable, dividing ordered data into equal-sized data subsets. Quantile classification is well-suited to linearly distributed data and histograms are a common graphic representation of quantiles. When quantiles are used to display spatial data, results must be interpreted carefully because similar features may be separated into adjacent classes, or features with widely different values can be lumped into the same class. This distortion can be minimized by increasing the number of classes.

**Quintile:** A special name when data are split into 5-quantiles.

**Raster:** A spatial data model that defines space as an array of equally-sized cells arranged in rows and columns, and composed of single or multiple bands. Each cell contains an attribute value and location coordinates. Unlike a vector structure, which stores coordinates explicitly, raster coordinates are contained in the ordering of the matrix. Groups of cells that share the same value represent the same type of geographic feature.

**Reporting unit:** Any defined area (e.g., an 8-digit USGS hydrologic unit code "HUC," county) for which a landscape metric (e.g., percent urban) is calculated.

**Resource:** Any feature, good, or quality that can serve as an input into production of a desired outcome. A resource can be a natural endowment such as fresh water, a built product such as a road, or a social institution such as the people associated with a particular school.

**Revised Universal Soil Loss Equation (RUSLE):** A soil erosion model developed by the USDA Agricultural Research Service in 1993. It contains the same general formula as Universal Soil Loss Equation (USLE), but has several improvements in determining factors. These include some new and revised isoerodent maps, a time-

varying approach for a soil erodibility factor, a subfactor approach for evaluating the cover-management factor, a new equation to reflect slope length and steepness, and new conservation-practice values.

**Scale:** The spatial or temporal dimension over which an object or process exists, as in, for example, a landscape, or a forest ecosystem or community.

**Shapefile:** A vector data storage format specific to ESRI (Environmental Software Research Institute) for storing the location, shape, and attributes of geographic features.

**Spatially explicit:** An indication that geo-referenced data are used or created and that a relatively fine scale of spatial disaggregation is used in evaluation.

**Splining:** An interpolation method in which values are estimated using a mathematical function that minimizes overall surface curvature, resulting in a smooth surface that passes exactly through the input points. Splines can be mathematically adjusted by increasing or decreasing the tension in between points

**State space analysis:** An integration method that measures the distance between two points (i.e., two watersheds) in multivariate space**.**

**Stressor:** A physical, chemical, or biological factor that can disrupt, change, or otherwise alter ecosystem health and/or human health in a negative way. For example, pesticides used in agriculture can be stressors to both ecosystem health and human health.

**Trend surface analysis:** A surface interpolation method that fits a polynomial surface by least-squares regression through the sample data points. This method results in a surface that minimizes the variance of the surface in relation to the input values. The resulting surface rarely goes through the sample data points. This is the simplest method for describing large variations, but the trend surface is susceptible to outliers in the data. Trend surface analysis is used to find general tendencies of the sample data, rather than to model a surface precisely.

**Universal Soil Loss Equation (USLE):** A widely used soil erosion model first developed by the U.S. Department of Agriculture in the early 1960s. USLE predicts the long-term average annual rate of erosion on a field slope based on rainfall pattern, soil type, topography, crop system, and management practices.

**Variable:** A quantity that can take on discrete or continuous values to represent condition (e.g., of an ecosystem). Often used interchangeably with an indicator.

**Vector (vector element):** A coordinate-based data model that represents geographic features as points, lines, and polygons. Each point feature is represented as a single coordinate pair, while line and polygon features are represented as ordered lists of

vertices. Attributes are associated with each vector feature, as opposed to a raster data model, which associates attributes with grid cells.

**Watershed:** A watershed is an area of land that is drained by a single stream, river, lake, or other body of water. Ridges form the dividing lines between watersheds. Water on one side of the ridge flows into one stream and water on the other side of the ridge flows into a different stream. Thus, a watershed is a natural unit defined by the landscape.

# References


Andreasen, J.K., R.V. O'Neill, R. Noss, and N.C. Slosser. 2001. Considerations for the development of a terrestrial index of ecological integrity. *Ecological Indicators* 1:21-35.

Banerjee, S., B.P. Carlin, and A.E. Gelfand. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, New York, New York, USA. 472 pp.

Carr, D.B., A.R. Olsen, J.P. Courbois, S.M. Pierson, and D.A. Carr. 1998. Linked micromap plots: named and described. *Statistical Computing & Graphics Newsletter* 9(1):24-32.

Carr, D.B., A.R. Olsen, S.M. Pierson, and J.P. Courbois. 2000. Using linked micromap plots to characterize Omernik ecoregions. *Data Mining and Knowledge Discovery* 4:43-67.

Carr, D.B., S. Bell, L. Pickle, Y. Zhang, and Y. Li. 2003. The state cancer profiles web site and extensions of linked micromap plots and conditioned choropleth map plots. *Proceedings of the Fourth National Conference on Digital Government Research*.

Chambers, J.Q., G.P. Asner, D.C. Morton, L.O. Anderson, S.S. Saatchi, F.D.B. Espírito-Santo, M. Palace, and C. Souza Jr. 2007. Regional ecosystem structure and function: ecological insights from remote sensing of tropical forests. *Trends in Ecology and Evolution* 22(8):414-423.

Clarke, K.C., L. Gaydos, and S. Hoppen. 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design* 24:247-261.

Cressie, N.A.C. 1993. *Statistics for Spatial Data*. Wiley, New York, Revised edition. Wiley, New York. 900 pp.

De Maesschalck, R., D. Jouan-Rimbaud, and D.L. Massart. 2000. The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems* 50:1–18.

Fagerstrøm, T. 1987. On theory, data, and mathematics in ecology. *Oikos* 50:258-261.

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.

Foran, J.A., and S.A. Ferenc (eds.) 1999. *Multiple stressors in ecological risk and impact assessment.* SETAC Press, Pensacola, FL.

Greene, E.A., A.E. LaMotte, and K.A. Cullinan. 2005. *Ground-Water Vulnerability to Nitrate Contamination at Multiple Thresholds in the Mid-Atlantic Region Using Spatial Probability Models*. USGS Scientific Investigations Report 2004-5118.

Grimm, J.W., and J.A. Lynch. 2000. *Enhanced wet deposition estimates for the Chesapeake Bay watershed using modeled precipitation inputs*. Rep. CBWP-MANTA-AD-99-2. Maryland Department of Natural Resources, Annapolis.

Harbor, J., and M. Grove. 1997. *L-THIA: Long-Term Hydrological Impact Assessment - A Practical Approach*. Ohio Environmental Education Fund/Purdue University, Manual.

Hardie, I.W., and P.J. Parks. 1997. Land use in a region with heterogeneous land quality: An application of an area base model. *American Journal of Agricultural Economics* 79:299-310.

Hardy, R.L. 1971. Multiquadric equations of topography and other irregular surfaces. *Journal of Geophysical Research* 76:1905–1915.

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

Jackson, L.E., S.L. Bird, R.W. Matheny, R.V. O'Neill, D. White, K.C. Boesch, and J.L. Koviach. 2004. A Regional approach to projecting land-use change and resulting ecological vulnerability. *Environmental Monitoring and Assessment* 94:231–248.

Jones, K.B., A.C. Neale, M.S. Nash, R.D. Van Remortel, J.D. Wickham, K.H. Riitters, and R.V. O'Neill. 2001. Predicting nutrient and sediment loadings to streams from landscape metrics: a multiple watershed study from the United States Mid-Atlantic Region. *Landscape Ecology* 16(4):301-312.

Klir, G.J., and Y. Bo. 1995. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, Upper Saddle River, NJ. 592 pp.

Liotta, P.H. 2005. Through the looking glass: creeping vulnerabilities and the reordering of security. *Security Dialogue* 36(1):49-70.

Liotta, P.H., and J.F. Miskel. 2004. Redrawing the map of the future. *World Policy Journal* 21(1):15-21.

Liu, Y., H. Guo, Z. Zhang, L. Wang, Y. Dai, and Y. Fan. 2007. An optimization method based on scenario analysis for watershed management under uncertainty. *Environmental Management* 39:678–690.

Pandey, S., J. Harbor, and B. Engel. 2000. Internet based geographic information systems and decision support tools. *Urban and Regional Information Systems*. Park Ridge, IL. 36 pp.

Parks, J.P., I.W. Hardie, C.A. Tedder, and D.N. Wear. 2000. Using resource economics to anticipate forest land use change in the U.S. Mid-Atlantic region. *Environmental Monitoring and Assessment* 63:175-185.

Peterson, A.T., M. Papes, and D.A. Kluza. 2003. Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science* 51:863-868.

Prestemon, J.P., and R.C. Abt. 2002. The Southern timber market to 2040. *Journal of Forestry* 100(7):16-22.

Reckhow, K.H., M.N. Beaulac, and J.T. Simpson. 1980. *Modeling Phosphorus Loading and Lake Response Under Uncertainty: A Manual and Compilation of Export Coefficients*. USEPA 440/5-80-011.

Schaberg, R.H., and R.C. Abt. 2004. Vulnerability of Mid-Atlantic forested watersheds to timber harvest disturbance. *Environmental Monitoring and Assessment* 94: 101-113.

Smith, E.R., L.T. Tran, and R.V. O'Neill. 2003. *Regional Vulnerability Assessment for the Mid-Atlantic Region: Evaluation of Integration Methods and Assessments Results*. EPA/600/R-03/082.

Smith, E.R., L.T. Tran, R.V. O'Neill, and N.W. Locantore. 2004. *Regional Vulnerability Assessment of the Mid-Atlantic Region: Evaluation of Integration Methods and Assessments Results*. EPA/600/R-03/082 (NTIS PB2004-104952). U.S. Environmental Protection Agency, Washington, DC.

Theobald, D.M. 2005. Landscape patterns of exurban growth in the USA from 1980 to 2020. *Ecology and Society* 10(1):32

Tran, L.T., and L. Duckstein. 2002. Comparison of fuzzy numbers using a fuzzy distance measure. *Fuzzy Sets and Systems* 130(3):331-341.

Tran, L.T., R.V. O'Neill, and E.R. Smith. 2006. A generalized distance measure for environmental integrated assessment. *Landscape Ecology* 21:469–476.

Tran, L.T., R.V. O'Neill, E.R. Smith, and C.G. Knight. 2007. Sensitivity analysis of aggregated environmental indices with a case-study of the Mid-Atlantic Region. *Environmental Management* 39:506-514.

Tran, L.T., R.V. O'Neill, and E.R. Smith. 2007. Determining the most dominant stressors and most impacted resources for integrated environmental assessment. (In review).

U.S. EPA. 2000. *Projecting Land-Use Change: A Summary of Models for Assessing the Effects of Community Growth and Change on Land-Use Patterns*. EPA/600/R-00/098. U.S. Environmental Protection Agency, Washington, DC. 260 pp.

U.S. EPA. 2003. *Generic Ecological Risk Assessment Endpoints (GEAEs) for Ecological Risk Assessment*. EPA/630/P-02/004F. U.S. Environmental Protection Agency, Washington, DC. 67 pp.

U.S. EPA. 2004. *Analytical Tools Interface for Landscape Assessment (ATtILA).* EPA/600/R-04/083. U.S. Environmental Protection Agency, Las Vegas, NV. 39 pp.

U.S. EPA SAB (Science Advisory Board). 2002. *A Framework for Assessing and Reporting on Ecological Condition: An SAB Report*. EPA-SAB-EPEC-02-009.

U.S. EPA SAB (Science Advisory Board). 2005. *Advisory on ORD's Regional Vulnerability Assessment Program.* EPA-SAB-ADV-06-001. 41 pp.

Wagner, P.F., R.V. O'Neill, L.T. Tran, M. Mehaffey, T. Wade, and E.R. Smith. 2006. *Regional Vulnerability Assessment for the Mid-Atlantic Region: Forecasts to 2020 and Changes in Relative Condition and Vulnerability*. EPA/600/R-06/088. 37 pp. U.S. Environmental Protection Agency, Las Vegas, NV.

Weingand, D.E. 1995. Preparing for the new Millennium: the case for using marketing strategies. *Library Trends* 43(3):295-317.

Wheeler, J.O., P.O. Muller, G.I. Thrall, and T.J. Fik. 1998. *Economic Geography*. John Wiley and Sons, New York, NY. 416 pp.

Wickham, J.D., R.V. O'Neill, K.H. Riitters, T.G. Wade, and K.B. Jones. 1997. Sensitivity of selected landscape pattern metrics to land-cover misclassification and differences in land-cover composition. *Photogrammetric Engineering and Remote Sensing* 63:397-402.

United States
Environmental Protection
Agency

Office of Research
and Development (8101R)
Washington, DC 20460

Please make all necessary changes on the below label, detach or copy, and return to the address in the upper left-hand corner.

If you do not wish to receive these reports CHECK HERE ☐; detach, or copy this cover, and return to the address in the upper left-hand corner.

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT No. G-35

Official Business
Penalty for Private Use
$300

EPA/600/R-08/xxx
April 2008
www.epa.gov