Methods

Future Scenario Analysis

2020 Land Cover Change Projection

Population growth can result in conversion of land to residential or agricultural uses (Wheeler *et al.* 1998), but these two land uses result in different stressors. Thus, distributing these changes spatially was critical to projecting changes in stresses such as aquatic non-point source pollution (e.g., percent impervious surface or agriculture on steep slopes) and forest productivity. Land use changes can also directly alter estimates of resource condition or abundance (e.g., wildlife habitat; Browder *et al.* 1989).

The SLEUTH (Slope, Land use, Exclusion, Urban extent, Transportation, Hillshade) model (formerly known as the Urban Growth Model; Clarke *et al.* 1997) was used to determine the likelihood of urbanization. Briefly, the SLEUTH model applies growth rules to geographic data on a cell-by-cell basis. The model produces raster land-cover maps for the projected period, and cumulative probability maps that show the likelihood of urbanization for each 1-km square cell during the selected time period. For more details and case studies of SLEUTH, see *Project Gigalopolis: Urban and Land Cover Modeling* (http://www.ncgia.ucsb.edu/projects/gig/); *Methods and techniques for rigorous calibration of a cellular automaton model of urban growth* (Clarke *et al.* 1996); and *A Self-Modifying Cellular Automaton Model of Historical Urbanization in the San Francisco Bay Area* (Clarke *et al.* 1997).

To create the 2020 land cover map (Figure 3), areas of the 1992 NLCD coverage predicted by SLEUTH to have a 50% or higher probability of being developed were modified. Planned roads and road expansions and areas where new mining permits had been granted were also added. Projected future roads were obtained by state from Departments of Transportation (DOT). Mining permits were obtained from Pennsylvania (Anthracite coal only), West Virginia, and Virginia. All permitted areas were assumed to be mined by 2020. Locations where mines and urban were coincident were left as mines. Areas that didn't coincide with new urban, roads, or mining retained their 1992 land cover.

Projection of Migratory Bird Flights

Forest-dwelling Neotropical migratory birds require intact forested stopovers during migration. The greater the number of paths that pass through a given HUCs, the more important that HUC is in the migratory system. Modeled migratory flights were based on flight distance and direction to examine how nightly flights link stopovers into flyways (Tankersley 2004). Stopover habitats were quantified in 10-km radius hexagons for the entire study area, modeled on forest density, percent agriculture, and road density. All models were developed using the ReVA future land-use projection. Field observations made in 1999 were used to determine high-quality stopovers and as points for movement models created in Arc/Info using the Eucdirection function. Each scenario represents a unique combination of distance and compass direction, with a maximum of 32 scenarios supported by any one HUC. Importance of a HUC to the migratory systems was based on the number of scenarios supported by a HUC. The resulting output highlighted portions of the landscape that are important for the continued success of migratory birds. Areas where many different migration scenarios overlap are particularly important, as these areas will support a diverse collection of migratory strategies and populations.

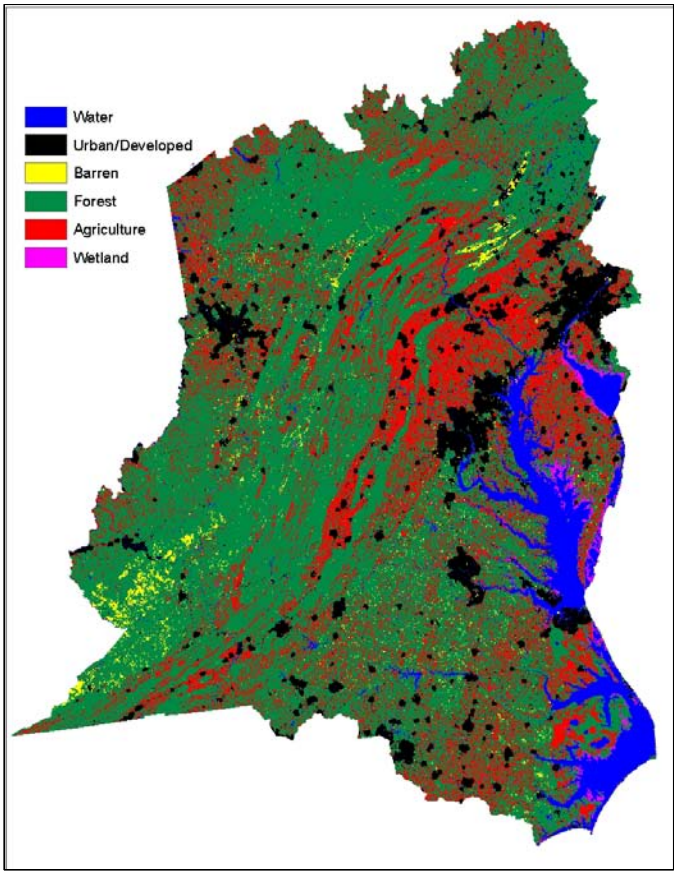

Figure 3. Map of predicted land cover for the Mid-Atlantic region in 2020.

Projection of Non-Indigenous Species

The Genetic Algorithm for Rule-set Prediction (GARP) model was used to create spatial maps of invasive species distributions. Briefly, GARP uses native species distributions to explore nonrandom relationships between point localities and the environmental conditions (niche modeling) surrounding the site (Grinnell 1917). Basic inputs include species records, temperature, precipitation, solar radiation, snow cover and frost-free days (Appendix 1). GARP uses multiple rule types including BIOCLIM, logistic regressions, and a genetic algorithm (artificial intelligence application) to generate a set of IF…THEN rule statements to describe the relationships between species and environmental conditions. The output from GARP can then be projected onto a landscape to visualize the species potential distribution. The distribution can also be projected onto areas of actual or potential invasion/introduction under different land cover and climatic conditions (Peterson *et al.* 2003). Current geographic distributions were used to project potential distributions onto a spatially explicit scenario for 2020. Detailed documentation can be found in Stockwell and Noble (1991) and Stockwell and Peters (1999).

Gypsy Moth was considered a special case of non-indigenous species because of the seriousness of the risk it poses and because of the availability of risk information available from the U.S. Forest Service (USFS). Forested areas with repeated annual defoliation by gypsy moth become more stressed and are at increased risk of permanent damage if further defoliation occurs in the future. A GIS was employed by the USFS to assemble, collate, and analyze locations of gypsy moth defoliation data (Eastman 1989).

When combined, data on suitable habitat, forest density, and geographic pattern of defoliation can be used to predict future potential for gypsy moth defoliation (Morin *et al.* 2005). The grid of future defoliation risk was provided by the United States Forest Service as a 2 x 2 km spatial map. The data were then summarized to add up all the grid cell values within each 8-digit HUC watershed.

Sources for aquatic species data include:

1. Nonindigenous Aquatic Species Web site (http://nas.er.usgs.gov/) (Hydrilla, Eurasian watermilfoil, giant salvinia)
2. Queensland Herbarium specimen data, Jardim Botanico do Rio de Janeiro specimen data (giant salvinia)
3. Auckland Museum specimen data, Te Papa Museum specimen data (New Zealand mud snail).

Sources for the terrestrial species included:

1. National Agricultural Pest Information System (NAPIS; http://www.ceris.purdue.edu/napis/) (hemlock woolly adelgid, greater pine shoot beetle)
2. Smithsonian Institution, National Museum of Natural History specimen data (greater pine shoot beetle)
3. Eduard Jendek, Institute of Zoology, Slovak Academy of Sciences (emerald ash borer)
4. OakMapper Web application (http://kellylab.berkeley.edu/SODmonitoring/OakMapper.htm) (Sudden oak death)
5. Literature (Peterson *et al*. 2003, garlic mustard; and Lingafelter and Hoebeke 2002, Asian long-horned beetle)

Predicted future distributions of aquatic and terrestrial species data were calculated as a weighted proportion of appropriate habitat overlapped by the potential distribution of a given species. These predictive models were built in the GARP (Stockwell and Noble 1991) and used a Hadley Centre climate model (CM2 GSdX20) (http://www.metoffice.com/research/hadleycentre/).

Projection of Nitrogen and Phosphorus in Surface Water

Excess export of nitrogen and phosphorus to streams was calculated using a statistical model developed by Reckhow *et al.* (1980) which uses the amounts of land cover to estimate loadings. Furthermore, the loadings model was available in ATtILA (USEPA 2004). The projection of nitrogen and phosphorus loadings in surface water by the year of 2020 was obtained with the use of ATtILA on the 2020 land-cover projection.

Projection of Nitrate in Groundwater

Available water-quality well data obtained from U.S. Geological Survey National Water-Quality Assessment Program studies were used in association with geographic data to develop logistic-regression equations to predict the probability of nitrate exceeding a specified threshold (Greene *et al.* 2005). Independent variables include geographic data such as 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. To project groundwater vulnerability for the 2020 scenario, the projected land cover was used as input; all other variables were held at their current values.

Combining Various Projections in Future Scenario Analysis

Each projection was reflected via changes in one or several specific variables used in the analysis. For example, projection on land-cover change was accounted for via changes in AGSL, EDGE2, EDGE65, FORCOVDEFOL, FORPCT, INT2, INT65, PSOIL, RDDENS, RIPAG, STRFOR, STRD, UINDEX, and WETLNDSPCT. Demographic change was reflected in POPDENS, POPGROWTH, PRAGFM, and PRMINE. Projected changes in water quality due to land-cover change and other factors were estimated in DISSOLVED, NITRATEGW, and TOTALN. Projection of non-indigenous species was represented in AQUAEXOTIC and TERREXOTIC. Finally, projection of migratory bird flights was seen in MIGSCENARIO. The future scenario analysis for the Mid-Atlantic region was carried out via addressing a set of assessment questions with the use of multiple integration methods on the set of available variables for both current values and projected ones for the year 2020. The integration methods used in the analysis are presented in the next section.

Integration Methods Used in Scenario Analysis

The methods section below is intended to provide a brief description of the integration methods used in this reports; methods are fully explained and documented elsewhere. For a detailed discussion of methods, see EPA*/600/R-03/082 Regional Vulnerability Assessment for the Mid-Atlantic Region Evaluation of Integration Methods and Assessment* (Smith *et al.* 2003). Error analysis is presented in Smith *et al.* (2006). Additional information and list of related publications are available at http://www.epa.gov/reva/products.htm.

Simple Sum

Normalized values of environmental variables are summed into a single index to produce the Simple Sum where values can range from 0 (best) to 1 (worst). An advantage of the Simple Sum method is that it is easily understood and communicated. Furthermore, it is not sensitive to discontinuities or non-normal distributions and does not depend on meeting assumptions about the statistical distribution of the data. The Simple Sum method can, however, lead to occlusion and this method cannot account for covariance in the data set.

Principal Components Analysis

Principal Components Analysis (PCA) is a statistical technique used to reduce a set of complex multivariate data into a simpler set of uncorrelated variables, each of which is a linear combination of the original variables. Varimax rotation was used to minimize the number of variables having high loadings on each factor, simplifying the interpretation of the factors. The eigenvectors (loadings) derived from the PCA were then used to compute the principal component (PC)-based indices. The PC-based indices were weighted sums of the environmental indicators where the weights were the loadings' absolute values. The averages of the PC-based indices were then used as an integrated index for ranking and clustering. The primary advantage of the PCA method is the replacement of a set of multivariate data with a new set of uncorrelated variables. However, it is sometimes difficult to interpret environmental meanings of the new set of uncorrelated variables and PCA can be strongly influenced by data abnormalities (e.g., non-normal distribution, discontinuities).

State Space Analysis

State Space analysis was used to determine the distance (Mahalanobis 1936) of each watershed from the most vulnerable watershed in the region. Theoretically, the most vulnerable watershed in the region has relatively large amounts of intact valued resources, but these resources are comparatively more threatened by degradation from stress. The primary advantage of the State Space method is that it maintains the full dimensionality of the data set while modifying the calculation of distance to account for data covariance. The primary disadvantage of the State Space approach is not any mathematical property of its calculation, but that the method requires a choice be made to identify the “most vulnerable” watershed. At the moment, there is no entirely objective definition for identifying the most vulnerable watershed.

Matrix Method

The matrix method was used to identify the most important stressors and resources in the Mid-Atlantic. This method is not novel and is used elsewhere to characterize risks from multiple stressors (Foran and Ferenc 1999, Ferenc and Foran 2000). The matrix method was originally proposed by Leopold *et al.* (1971) and a number of variations were reviewed by Canter (1977). In more recent applications, the emphasis has been on identifying important stressors (Cormier *et al.* 2000). The matrix represents stressors as rows and resources as columns and this has been used to organize complex assessment information for several decades (Phillips *et al.* 1978; Lumb 1982a, 1982b; Witmer *et al.* 1985; Clark 1986; Emery 1986; Risser 1988). In most applications, quantitative information is not available and a panel of experts is typically asked to assess the individual impacts and supply a qualitative value (e.g., 1 for a minor impact, 2 for a moderate impact, and 3 for a major impact). This value is then inserted into the appropriate cell of the matrix and when the matrix is complete; the values in each row were summed and taken as the total effect of each stressor across all resources. The row sums can then be ranked to indicate which stressors represent the greatest threat and, in a decision-making context, were in greatest need of control.

For ReVA’s analyses, the data were available for stressors and resources presenting a unique opportunity to apply the matrix approach quantitatively by using the correlation matrix (which measures the relationship between each stressor and resource pairing) (Smith *et al.* 2003). The correlation data were reduced to a matrix with the stressors as rows and resources as columns. The row sums represent the relationship between each stressor and all of the resources. The stressors with the largest row sums were then taken as the most important stressors as they were expected to have the greatest effect across all resources. A column sum of coefficients was done for each resource. The largest column sums were associated with resources that were the most stressed because they show the closest relationship with the stressors (considered across all stressors). Earlier testing of this approach (Smith *et al.* 2003) indicated that the row sums were reliable and stable, particularly when only the largest 2 or 3 row and column sums were considered. Therefore, while the range of row and column sums was divided into seven equal intervals our attention was restricted to only the top and bottom three septiles. To determine which watersheds may shift out of the top three or into the bottom three septiles, the septile ranges from the current condition were used to assess the 2020 data. It is possible that statistically insignificant coefficients represent spurious relationships and results were presented for both the total row sum and also the sum of the statistically significant correlation coefficients in each row.

Criticality Analysis

Criticality analysis calculates the distance between the vector of variable values representing current conditions on each watershed and a vector representing a hypothetical "natural" state. The distance is calculated using a fuzzy distance measure and attempts to reconstruct the set of conditions under which the components of the ecological system evolved. This approach assumes that systems in the natural state retain the feedback networks that permitted stable response to disturbances over the long period of evolutionary history. As humans add stressors (e.g., chemical pollutants), extract resources (e.g., lumber), and change landscape patterns (e.g., fragmentation), the natural feedbacks are disrupted and the system becomes more vulnerable to radical and potentially irreversible change. To deal with the uncertainties involved in defining the natural state, the Criticality analysis is based on "fuzzy" values.

The greatest strength of the Criticality approach is that it provides a unique perspective on regional watersheds. The phenomenon of catastrophic change in complex adaptive systems is a potentially important concept in large-scale assessment. Another strength of the Criticality approach is its relative insensitivity to the assumptions involved in defining a "natural" state. The greatest weakness of the Criticality approach is our inability to predict where the critical threshold lies. Although it appears reasonable to estimate relative vulnerability, it is not possible to pinpoint exactly which watersheds will undergo radical change given a natural disturbance or further development. For a detailed discussion of this method, see Tran and Duckstein (2002).

Stressor-Resource Overlay

The Stressor-Resource Overlay method was used to try to locate watersheds in which high amounts of valued resources occur together with high levels of stressors. For this analysis, stressor and resource variables were divided into quintiles and watersheds were scored on the number of stressor variables that fell into the worst two quintiles and also on the number of resource variables that fell into the best two quintiles. The advantage of the Stressor-Resource Overlay is in its ease of interpretation and it is the only method that directly addresses the geographic distribution of vulnerability. The primary disadvantage of the Stressor-Resource Overlay method is that it does not account for correlation between variables. However, the Stressor-Resource Overlay is not influenced by the correlation structure of the data as long as each resource is valued and stressors can interact to cause synergistic effects.