DOI: 10.1007/s00267-003-1084-0

Potential of Multivariate Quantitative Methods for Delineation and Visualization of Ecoregions

WILLIAM W. HARGROVE*
FORREST M. HOFFMAN
Environmental Sciences Division
Computer Science and Math Division
Oak Ridge National Laboratory
P. O. Box 2008, M.S. 6407
Oak Ridge, Tennessee 37831-6407

ABSTRACT / Multivariate clustering based on fine spatial resolution maps of elevation, temperature, precipitation, soil characteristics, and solar inputs has been used at several specified levels of division to produce a spectrum of quantitative ecoregion maps for the conterminous United States. The coarse ecoregion divisions accurately capture intuitively-understood regional environmental differences, whereas the finer divisions highlight local condition gradients, ecotones, and clines. Such statistically generated ecoregions can be produced based on user-selected continuous variables, allowing customized regions to be delineated for any specific problem. By creating an objective ecoregion classification, the ecoregion concept is removed from the limitations of human subjectivity, making possible a new array of ecologically useful derivative products. A red–green–blue visualization based on principal components analysis of ecoregion centroids indicates with color the relative combination of environmental conditions found within each ecoregion. Multiple geographic areas can be classified into a single common set of quantitative ecoregions to provide a basis for comparison, or maps of a single area through time can be classified to portray climatic or environmental changes geographically in terms of current conditions. Quantified representativeness can characterize borders between ecoregions as gradual, sharp, or of changing character along their length. Similarity of any ecoregion to all other ecoregions can be quantified and displayed as a ''representativeness'' map. The representativeness of an existing spatial array of sample locations or study sites can be mapped relative to a set of quantitative ecoregions, suggesting locations for additional samples or sites. In addition, the shape of Hutchinsonian niches in environment space can be defined if a multivariate range map of species occurrence is available.

Ecoregions are designed to help users visualize and understand similarities across complex multivariate environmental factors by grouping areas into like categories. The basis for such groupings has many contentious conceptual underpinnings. Debate abounds as to whether ecoregions should be specialized for a particular use or general purpose, spatially contiguous versus disjunct, nestable versus nonhierarchical, and whether ecoregions can be defensible as units of management, legislation, or even ecological triage (Omernik 1995, 2003; Overton and others 2002; Leathwick and others 2003). Supreme among these issues, however, is the question of whether ecoregions can (or should) be delineated using quantitative statistical methods or whether they can only be drawn using human expertise in a qualitative, weight-of-evidence approach (McMahon and others 2001).

It is not our intention to add to this debate. These conceptual issues are well represented in this special issue and elsewhere. Rather, we sidestep the question as to whether ecoregions are computable and demonstrate the ramifications of quantitatively deriving ecoregions. We argue that the spate of ancillary products resulting from the quantitative treatment of ecoregions enhances and expands the utility of the ecoregion concept and makes substantial new contributions to niche modeling, network and sample design, change detection, and conservation.

Regionalizations are models, whether quantitatively or qualitatively derived. Quantitative models are, however, more explicit, repeatable, transferable, and defensible than subjective models based on human expertise. This transferability and repeatability makes quantitative models more objective than their qualitative counterparts. Human experts might be able to rationally defend drawing a particular borderline between ecoregions, but they might be unable to elucidate the method used to place it at that precise location. Quantitative ecoregionalization techniques

KEY WORDS: Clustering; Climate change; Ecotone; Environmental envelope; Fences; Gradient; Network; Niche; Preserve design; Range; Representativeness; Similarity; Time series

Published online ■

*Author to whom correspondence should be addressed; *email:* hnw@fire.esd.ornl.gov

are not perfectly objective, because they still require subjective ecological expertise in the choice of data layers to include and in the interpretation of the resulting ecoregions. Nevertheless, the pursuit of a fully describable quantitative method for ecoregionalization is a desirable goal, whether as a replacement or an augmentation to qualitative, expert-opinion-based approaches.

The ability to explicitly select and control the input variables allows quantitative regionalizations to be customized for specific uses. Using quantitative methods, special-purpose ecoregions can be created specifically for particular uses. Such customized regionalizations can provide additional discrimination in places where popular generalized ecoregion schemes might miss particular features of importance.

In qualitative ecoregionalization, human experts might intentionally adjust the weighting of particular input layers in their mental model in particular spots across the map, whereas quantitative methods usually produce regionalizations in which all input variables receive equal weighting. Such evenly weighted ecoregion products provide, at the very least, an initial basis from which factor weights can be subsequently spatially modified by human expertise. Explicit spatial maps of altered weights, if known, could be considered directly as inputs into a fully quantitative model.

Sampling the Toolbox of Quantitative Methods

A full review of quantitative methods that have been used in regionalization is beyond the scope of this article. Although not exhaustive, brief treatment of major types of quantitative approach might provide context, and counterbalance more typical subjective approaches described elsewhere in this special issue. Most quantitative approaches rely on Gleasonian relationships between environmental patterns and species occurrence, but some have been tested directly using species distribution data, which include Clementsian effects of biotic interactions on geographic distributions. Strengths and weaknesses of each quantitative method are highlighted.

The Holdridge Life Zone model, usually shown as a set of hexagons arranged inside a triangular plot, was an early quantitative multivariate approach for defining ecoregions (Holdridge 1947). Plotting mean annual "biotemperature" against mean annual precipitation and potential evapotranspiration ratio places any location into one of a set of predetermined equal-area hexagons, which are assigned ecoregion names *a priori*. Homogeneity is not controlled, because several hexagons could represent a single ecoregion type. Lugo and others (1999) recently revisited the Life Zone approach for global vegetation, finding it adequate for forests, but of limited utility for grasslands, shrublands, and nonvegetated lands.

Kohonen's (1982, 1995) Self-Organizing Maps (SOMs) use a two-layer neural network to divide a multivariate map into ecoregions. During training, neurons are reinforced within a gradually shrinking network neighborhood around the best predictors (Hung 1993). Using the first five principal components of seasonal averages of climate data from 18 stations, Malmgren and Winter (1999) used a one-dimensional SOM to divide Puerto Rico into four natural climatic zones. It might be difficult, however, to transfer the "learning" from one neural network into a form that can be used by others.

Quantitative Ecoregions for Single Species

Mapping the geographic range of a single species of animal or plant is a special case of ecoregion delineation. Ecological niche theory is central to understanding how environmental change affects species abundance patterns (Jackson and Overpeck 2000). Hutchinson (1957) conceived of the niche as a multidimensional "hypervolume" with dimensions defined by the environmental factors that influence the fitness of individuals of that species.

Although Hutchinson's conception was static, environmental change involves temporal alteration of combinations of niche variables. Hutchinson's niche envelope assumes that a steady-state equilibrium with present environmental conditions has allowed adequate time for perfect adaptation and exhaustive migration to all parts of the potential range. Acclimation to changing conditions and historical factors limiting geographic dispersal are not considered by most quantitative range-prediction techniques [but migration can be simulated in a subsequent step (i.e., Peterson and others 2003)].

Generalized Additive Modeling (GAM) uses regression modeling to establish empirical relationships between a response variable (i.e., presence of a particular species at a location) and an individually smoothed set of spatial predictor variables (Hastie and Tibshirani 1990). GAM additively calculates the component response and can handle nonlinear and nonmonotonic relationships between the response and predictor variables. No parametric assumptions are necessary, but the probability distribution (e.g., binomial, Poisson, Gaussian) of the response variable must be specified. Generalized Linear Modeling (GLM) is a special case of GAMs in which predictors are parameterized instead of being smoothed (McCullagh and Nelder 1997).

Generalized additive modelings may be biased if they are fitted using presence-only datasets (e.g., museum collection localities). Because of their sequential nature, GAMs are poor at capturing interactions among predictor variables. A crucial step is the selection of an appropriate level of spatial smoothing for a predictor. Relationships between predictor variables and the response variable are empirical rather than mechanistic or process based. Therefore, GAMs fitted to data in a small region generally do not extrapolate well across space or time.

Generalized additive modeling was used by Austin and others (1990) to model the niches of five Eucalypt species. Overton and others (2001) used GAM to generate geographic frameworks in New Zealand for the purposes of ecosystem management and sustainability. Overton and others (2000), Leathwick (2001), and Lehmann and others (2002a) have used GAM repeatedly to predict occurrence of many species, building up estimates of community structure and biodiversity. They followed a Predict First, Classify Later (PFCL) paradigm. Brooker and others (2002) demonstrated the utility of ecoregions in epidemiology and human health by using logistic regression modeling to quantitatively delineate schistosomiasis regions in Africa. Lehmann and others (2002b) created Generalized Regression Analysis and Spatial Prediction (GRASP), which formalizes the regression modeling approach to species distribution modeling using GAM. See extensive reviews of GAM in Guisan and others (2002) and Guisan and Zimmerman (2000).

Classification and Regression Trees (CART) and Regression Tree Analysis (RTA) for categorical and continuous response variables, respectively, have been used for both regionalization (Stoms and Hargrove 2000) and geographic range prediction (Iverson and others 1999). A regression tree is a binary decision tree in which branching at each step is defined by test criteria involving a single best predictor variable, which can be continuous or categorical. CART and RTA overfit a tree on a training sample, which then has almost as many terminal nodes as there are training observations. Nodes are then pruned from the tree or shrunk (at the cost of decreased accuracy) to achieve generality. CART and RTA choose the locally best discriminatory feature at each stage in the divisive process rather than the globally best discriminator (Stockwell and Noble 1992), and they enforce a sequential univariate model rather than a true multivariate approach. White and others (1999) used RTA to predict bird occurrence data in Oregon in terms of 10 environmental variables, and Rathert and others (1999) found environmental correlates with richness of Oregon freshwater fishes. Iverson and Prasad (1998) used RTA to generate predicted ranges for 80 tree species in the eastern United States following climate change. Prince and Steininger (1999) used RTA with six forcing variables, including rainfall, temperature, and photosynthetically active radiation, to stratify sampling in the Large Scale Biosphere–Atmosphere Experiment in Amazonia.

Several quantitative environmental envelope-based methods have been developed to ascertain habitat suitability. Busby (1991) developed BIOCLIM, which uses a simple bounding hyperbox method to capture species occurrences in data space. Although widely used, BIOCLIM overpredicts when species distribution is influenced by a combination of environmental predictors rather than by each one individually (Carpenter and others 1993). Walker and Cocks (1991) produced HABITAT, which attempts to form a convex envelope more tightly containing all species occurrences than the simple rectilinear envelope of BIOCLIM. Carpenter and others (1993) found that BIOCLIM overpredicted and HABITAT underpredicted habitat, and they proposed a new method, DOMAIN, based on similarity with all occupied points using Gower's metric. This metric is the arithmetic mean of the differences between the two points in each dimension, after being standardized by the range to equalize the contribution from each predictor. Hirzel and Arlettaz (2003) point out that the Gower metric does not consider the density of occupied points and is, therefore, subject to influence by outliers.

Hirzel and others (2002) describe Ecological Niche Factor Analysis (ENFA), which does not require true absence data. ENFA discriminates environments occupied by species from all environmental combinations occurring within a larger study area. ENFA models the environmental niche relative to some set of environmental variance found within a larger study area. The range of conditions represented in the larger area chosen therefore constrains the definition of the environmental niche envelope based on its context.

Hirzel and others (2001) tested ENFA and GLM against three virtual species that were spreading, at equilibrium, and overabundant. GLM was badly affected for the spreading species, but produced slightly better results than ENFA when the species was overabundant. Both methods produced equivalent results when the virtual species was at equilibrium.

Zaniewski and others (2002) compared GAMs with presence-only data to GAMs using computer-generated "pseudoabsences" and ENFA models. By using the same presence data for all models, absence data were isolated as the varying factor, allowing different tech-

niques for modeling presence-only and pseudoabsence data to be compared. Although presence-only GAMs predicted individual species distributions more accurately than ENFA, they were less effective than ENFA in highlighting biodiversity "hot spots" from the summing of species predictions. Engler and others (2004) found that using ENFA-weighted pseudoabsences enhanced the quality of GLM-based potential distribution maps for rare and endangered species.

Stockwell and Noble (1992) described a Genetic Algorithm for Ruleset Prediction (GARP), which uses a genetic algorithm and four types of rule to inductively develop a rule set for predicting a geographic range from a set of presence-only data for a particular species. Starting with an initial set of approximate rules, a genetic algorithm modifies them in ways that might (or might not) lead to an increase in predictive power. Randomize, mutate, and concatenate operators alter individual rules in a rule set, and optimized sets of rules with the best predictive ability are retained.

GARP predicts a slightly different geographic range using each optimized rule set. These alternative ranges are usually overlaid geographically and used as a single pseudosuitability map (Peterson and others 2003). Because GARP can suffer problems with repeatability, lack of absence data, and variable prior proportions, Stockwell and Peters (1999) developed a web-based system that eliminates these and other potential sources of error. Peterson and Cohoon (1999) found that GARP predictions were sensitive to the number of environmental variables that are included. Including five of eight environmental layers was necessary to avoid broad variance in predicted ranges for three species of birds. GARP can utilize museum collection data and has been used to estimate the ultimate potential distributions for invasive species (Peterson and Vieglais 2001). Martinez-Meyer and others (2004) used GARP to predict geographic distributions of 23 extant mammal species reciprocally between the Last Glacial Maximum and the present, suggesting that ecological niches are relatively constant over time. Such longitudinal evolutionary conservatism suggests that niche modeling can be used successfully to anticipate climate change effects on biodiversity.

Quantitative Clustering of Biotic Assemblages

Statistical clustering is the ordination and classification of multiple nonidentical objects into subgroups based on their similarity. Hierarchical clustering provides a series of divisions, based on some measure of similarity, into all possible numbers of groups, from one single group containing all objects, to potentially as many groups as there are objects. Hierarchical clustering is computationally intensive, so the assemblage to be classified must be limited to relatively few objects. Nonhierarchical clustering provides a single, user-specified level of division into groups; however, it can be used to classify a large number of objects because it does not divide exhaustively.

Cluster analysis can use occurrence data from a set of geographic sites to identify biotic communities that coexist in space. Overpeck and others (1985) clustered fossil pollens through time to identify modern analogs for ancient vegetation assemblages retrieved from cores. Campbell and McAndrews (1991) used a cluster analysis to group 33 lakes in southern Ontario. They described community changes in response to the Little Ice Age within the region by grouping similar pollen diagrams.

Multivariate Clustering to Form Homogeneous Regions

Several investigators have recognized the potential of geographic multivariate clustering for delineating homogeneous regions objectively within small maps (Belbin 1993; Omi and others 1979; Host and others 1996; Bunce and others 1996). Host and others (1996) used clustering to establish separate climatic and physiographic regions for northern Wisconsin, but then combined them using a simple GIS overlay. Environmental characteristics have been clustered to produce uniform regions of geology (Harff and Davis 1990), regions of uniform crop yield (Lark 1998), and regions of constant soil fertility (Carter 1997). Bernert and others (1997) used cluster analysis to further subdivide an existing expert-derived ecoregion, the Western Corn Belt Plains, and Krohn and others (1999) used geographic clustering to create hierarchical biophysical regions of Maine at 21 km resolution. Soriano and Paruelo (1992) used normalized multidimensional ordination of remotely sensed data to form homogeneous "biozones" for Patagonia in southern Argentina.

Hessburg and others (2000) used TWINSPAN, a divisive hierarchical method to create groups of subwatersheds of the Columbia River Basin. Repeated divisive classifications of overlapping regions allowed construction of pedigree trees showing similar analysis ancestries among subregions. Region separation was evaluated using discriminant analysis and cross-validation. Hessburg and others (2000) prestratified using an expert-derived regionalization and binned continuous ordinal variables before starting their quantitative process. Not all of their resultant regions nested within even the largest expert-derived domains with which they were initially constrained. Sequentially divisive methods are unlikely to result in equal-variance re-

gions, particularly if some regions are subsequently recombined.

Jensen and others (2001) used 19 indirect biophysical variables to hierarchically classify subwatersheds in the Columbia River Basin using agglomerative clustering with Ward's method. Subsequent analysis of variance (ANOVA) showed that the hydrologic subregions produced were highly significant in explaining Forest Service watershed and stream management hazard ratings, although no comparison was made with the regionalization of Hessburg and others (2000) or with expert-derived regions.

Zhou and others (2003) created objective ecoregions of Nebraska using an agglomerative hierarchical clustering procedure based on multitemporal satellite data, along with climate and soil information. Their aggregation procedure combined 2024 polygons to form 66 and 23 hierarchical regions. Because only spatially adjacent regions are merged, island and barrier features (i.e., water bodies and rivers) require human intervention during processing to prevent lingering arbitrary separations.

Leathwick and others (2003) created "environmental domains" at 1 km resolution for New Zealand using a 2-stage multivariate classification based on 10 climatic and landform variables affecting plant physiological processes. They first used nonhierarchical clustering to produce 350 geographic groups, then used sequential agglomerative clustering to obtain 20 final domains. In both stages, they used the Gower metric as a similarity measure, which is sensitive to outliers (Hirzel and Arlettaz 2003). They presented a tree showing the similarity of the 20 final domains, each of which were intuitively recognizable.

We developed a supercomputer-based multivariate statistical clustering algorithm to define ecoregions within extensive, high-resolution maps containing many multivariate descriptors (Hargrove and others 2001) (Figure 1). The user can specify the number of clustered ecoregions that result from the process, making it possible to divide the map into a few large, coarsely-defined ecoregions or a larger number of small, highly specified ones.

The nonhierarchical algorithm consists of a reversible transformation between two realms: one in two-dimensional geographic map space and one in multidimensional data space. Normalized variable values from each map raster cell are used as coordinates to plot each map cell in an environmental space with as many axes as there are multivariate environmental dimensions. Normalization gives environmental parameters measured in different units equal spacing by establishing a mean of zero and a unit standard deviation. Because the plotted location of map cells in data space pinpoints the combination of environmental variables within that map cell, two map cells that are plotted close to one another in data space will have similar mixtures of environmental conditions and are likely to be classified into the same ecoregion cluster. Thus, similarity is coded as separation distance in environmental data space.

We use the iterative *k*-means algorithm of Hartigan (1975), which begins with a user-specified number of ecoregion clusters, *k*, into which the map cells are to be grouped. All map cells are examined sequentially to find the most widely separated set of cells that will provide *k* initial "seed" centroids, one for each of the desired *k* cluster groups. In a single iteration, each map cell is assigned to the closest (i.e., environmentally most similar) centroid. At the end of the iteration, once all map cells are assigned to a centroid, the coordinates of all map cells within each group are averaged to produce a new, adjusted centroid for each cluster, and another iteration of assigning map cells to these new centroids begins. The iterative process of classifying map cells and adjusting centroid locations continues until fewer than a predetermined number of map cells change cluster assignments during an iteration. After the process has converged on a particular grouping scheme, the *k* ecoregions have been statistically defined.

Once cluster assignments have stabilized, map cells are reassembled in geographic space, retaining their ecoregion classifications. Although geographic coordinates are not used directly in the classification, ecoregions tend to be geographically cohesive because of the spatial autocorrelation that is usually present in the environmental data. Because of the Euclidean assignment method, the *k*-means algorithm tends to fit globular clusters of equal size in data space. Thus, all large ecoregions share a similar upper limit on within-group variance and have a similar maximum radius around each centroid. Although other clustering methods (i.e., Ward's, average, single, or complete linkage) are available, the uniform heterogeneity across ecoregions provided by *k*-means prevents the creation of side-by-side ecoregions that have vastly different within-region variance.

We call this empirical process Multivariate Geographic Clustering (MGC) and have implemented it in a parallel algorithm coded in C using the Message Passing Interface (MPI). Our code is dynamically load balancing and fault tolerant and performs both initial seed-finding and iterative cluster assignment in parallel (Hoffman and Hargrove 1999). Individual nodes independently classify subsets of cells, then combine

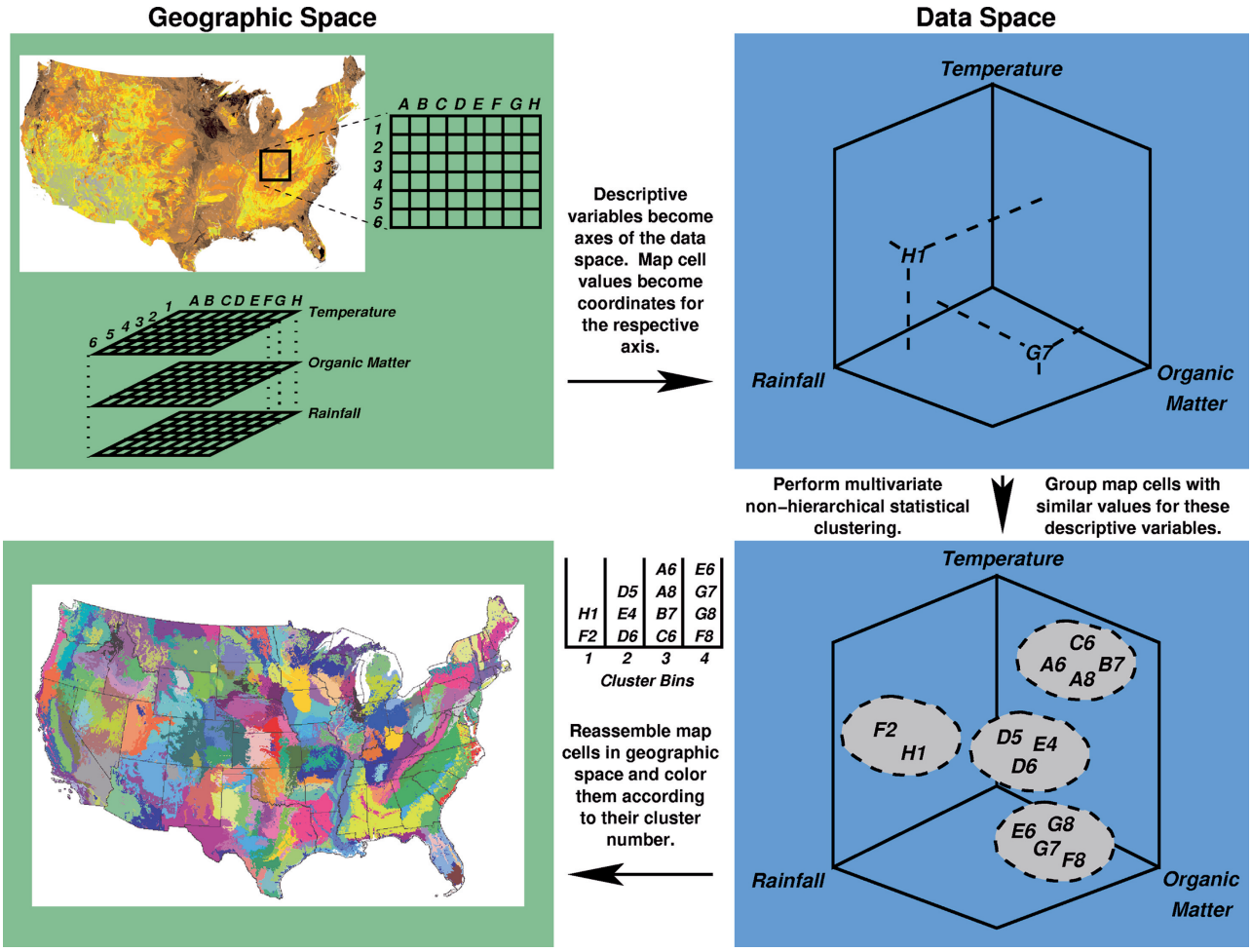

Figure 1. The Multivariate Spatio-Temporal Clustering procedure involves a transformation (moving clockwise from upper left) from geographic space (green) to data space (blue) and back. Normalized values of multivariate conditions in each map cell are used as coordinates to locate each map cell in an abstract data space. Although only three axes are shown, the process considers many multivariate characteristics. An iterative *k*-means clustering procedure assigns each cell to the closest of *k* centroids. At the end of each iteration, centroid positions are recomputed. After convergence, cells are reassembled in geographic space, colored by their final cluster assignments. Each resultant quantitative ecoregion contains roughly equal environmental heterogeneity.

results at the end of an iteration. As a result, the quantitative ecoregionalization process is not computationally limited.

Ecoregions determined using MGC are self-describing, in that the coordinates of the final centroids quantitatively define the synoptic conditions for each ecoregion. Nominative multivariate conditions within a particular ecoregion are described by the N coordinates of its centroid. The technique is parametric only in that means are used to calculate each new centroid. Rather than imposing a preconceived external grouping upon the map, the variance structure present in the environmental conditions is used, allowing a uniform classification structure to emerge from the data.

Strong correlations among input variables will affect clustering results. For example, elevation might be correlated with temperature and other climatic variables. Each input map should be selected or designed to contain unique information in order to preserve orthogonality in data space. Even strongly correlated inputs can be clustered by first performing a Principal Components Analysis (PCA) and then clustering in a data space formed by the PCA axes. Patterns resolved by MGC have proven robust to even strong correlations among a few input variables.

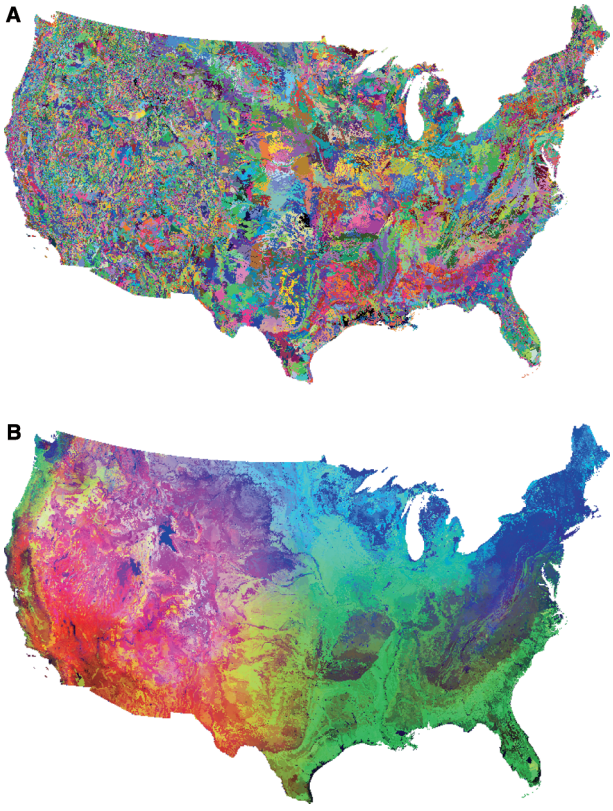

Figure 2. (**A**) The 3000 most different quantitative ecoregions in the United States based on nine variables, including precipitation, solar input, elevation, depth to water table, soil nitrogen and organic matter, soil water-holding capacity, heating degree-days during the growing season, and cooling degree-days during the nongrowing season, colored randomly. (**B**) When PCA is performed on the nine variables and the first three scores are assigned to red, green and blue, the color of each ecoregion indicates the relative mix of the nine environmental conditions inside each ecoregion. Red is "physiographic position" (i.e., low precipitation, high solar insolation, high elevation, and deep water table). Green is "plant nutrients" (i.e., high soil N, organic matter, and available water). Blue is "temperature" (i.e., few degree-days heat and many degree-days cool). Shown in these Similarity Colors, the borders between individual ecoregions disappear and the map now shows regional scale gradients in environmental conditions.

Multivariate Geographic Clustering Across Space

Initially, we performed a number of empirical regionalizations for the conterminous United States at 1 km resolution, dividing the United States into as many as 3000 distinct ecoregions (Figure 2A), (Hargrove and Luxmoore 1998). We included nine characteristics from three categories—elevation, edaphic factors, and climatic factors. The edaphic factors were (1) plant-available water capacity, (2) soil organic matter, (3) total Kjeldahl soil nitrogen, and (4) depth to a seasonally high water table. The climatic factors were (1) mean precipitation during the growing season, (2) mean solar insolation during the growing season, (3) degree-day heat sum during the growing season, and (4) degree-day cold sum during the nongrowing season. The growing season was defined by the frost-free period between mean day of first and last frost each year. A map for each of these characteristics was generated from best available data at a 1-km resolution for input into the clustering process. Each of the input maps contained more than 7.8 million cells. Such map, data, and ecoregion resolution surpasses that usually accomplished by ecoregion experts using qualitative methods. These maps appear to capture the ecological relationships among the nine input variables (Figure 2a). More recently, we have added new variables and divided the United States into as many as 5000 ecoregions. Twenty-five environmental factors, including elevation, mean and extremes of annual temperature, mean monthly precipitation, soil nitrogen, organic matter and water capacity, frost-free days, soil bulk density and depth, and solar aspect and insolation were included (Hargrove and Hoffman 2003).

Visualizing Ecoregion Similarity Using Similarity Colors

Randomly colored ecoregion maps emphasize the location of borders between ecoregions. However, ecologists might also wish for some indication of how different the mixture of environmental conditions is across the border between neighboring ecoregions. Because the final location of the cluster centroid is, by definition, the most centrally located point inside each cluster, the data space coordinates of the centroid provide a description of the average ecological conditions in this cluster ecoregion. Likewise, differences between the centroid coordinates from two ecoregions quantify the differences between the average environments found in each cluster ecoregion.

We devised a statistical coloring scheme to visualize similarities among environments within different ecoregions. If we use a PCA, either before or after MGC, to condense a larger number of "raw" environmental variables into orthogonal principal component axes, we can map the first, second, and third principal component scores to a red–green–blue (RGB) color triplet. In this way, the combination of values for the coordinates for each cluster centroid are used to

specify a unique color for that ecoregion that indicates the relative mixture of each environmental factor. Ecoregions containing similar environmental combinations are colored similarly.

The ecoregion map using Similarity Colors appears strikingly different from that using random colors. Individual ecoregion borders disappear and the Similarity Colors reveal smooth gradients that reflect the dominant suites of variables affecting environments in each region of the country (Figure 2B). The red Southwest is dominated by physiographic factors such as low precipitation, high solar radiation, and a deep water table. The blue Northeast is dominated by temperature factors, and the green Southeast is dominated by fertile soils. The upper Midwest is high in all factors, but is light blue because of the cold continental winter. The Pacific Northwest and the Central California valley are light green, indicating favorable conditions for plants. It is not possible to code maps by Similarity Colors when ecoregions are created qualitatively.

Characterizing Borders Between Ecoregions

Borders between ecoregions can be sharp, forming distinct ecotones. More commonly, however, gradual transitions cause edges to be indistinct, and it is difficult to locate a line of demarcation between distinct ecoregions (Bailey 1983). We have termed this type of gradual border an "ecopause" (Hargrove and Hoffman 1999). Indeed, a border can begin at one geographic location as an ecotone, and then transform slowly along its length into an ecopause. Approaches to distinguish these different types of border have included fuzzy set theory (Leung 1987; Lark 1998) and wavelet analysis (Csillag and others 2001; Csillag and Kabos 2002), but no single method has been widely adopted.

Fundamental to characterizing the sharpness of borders is quantifying how representative a particular location is of its parent ecoregion. In MGC, the Euclidean distance from each cell to its centroid measures its deviation from the cluster norm. Cells close to their centroids are more representative of their cluster ecoregions than cells far from their centroids in environmental space (Belbin 1993). If we depict these "representativeness" values as elevations, we can create a continuous surface whose height inversely corresponds to the representativeness of the cell at that geographic location (Hargrove and Hoffman 1999).

Because the edge properties are calculated for each pair of adjacent clusters, each side of every border has two distinct sharpness properties. We used contour lines of equal representativeness to visualize the sharpness characteristics of borders between ecoregions (Hargrove and Hoffman 1999). Closely spaced representativeness contours reflect steep edges and, therefore, a sharp ecotone. On the other hand, widely spaced representativeness contours indicate gradually sloping representativeness, characteristic of an indistinct ecopause. Representativeness contour lines have the flexibility to represent mixed gradual/sharp borders, as well as borders whose characteristics change along their length.

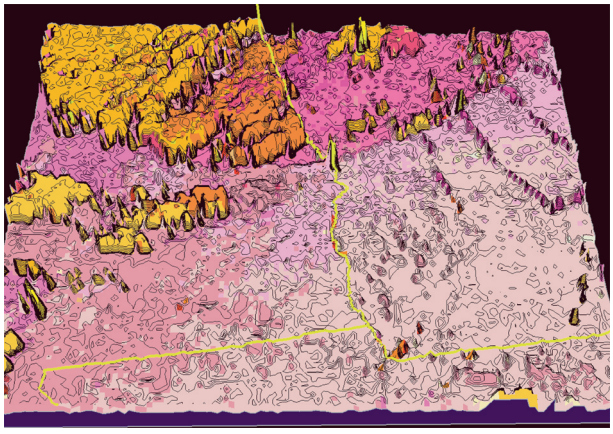

Figure 3. Borders between adjacent quantitative ecoregions can be characterized as sharp or gradual using representativeness contours. Georgia is at the right, Alabama is at the left, and the Florida panhandle is at the bottom. Ecoregions are shown in Similarity Colors, as in Figure 2B. The representativeness elevation of each cell is calculated by measuring the Euclidean distance in data space from the cell to the centroid of the quantitative ecoregion to which the cell was assigned. Representativeness contours are closely spaced and parallel adjacent to sharp borders, but meander near gradual borders.

Figure 3 shows the representativeness surface for southwest Georgia, Alabama, and northern Florida. The topography of each cell is obtained from the Euclidean distance from that cell's location in environmental data space to the centroid of its ecoregion. Cluster membership is shown in Figure 3 as the Similarity Colors of each cell. The random orientation and meandering character of the contours near the Coastal Plain and Piedmont ecoregions in southern Alabama clearly indicate that this border is an ecopause (i.e., environmental gradients are relatively unchanging). On the other hand, the closely spaced, parallel contour lines separating the Piedmont from the Ridge-and-Valley in northern Alabama reveal this border as a sharp ecotone.

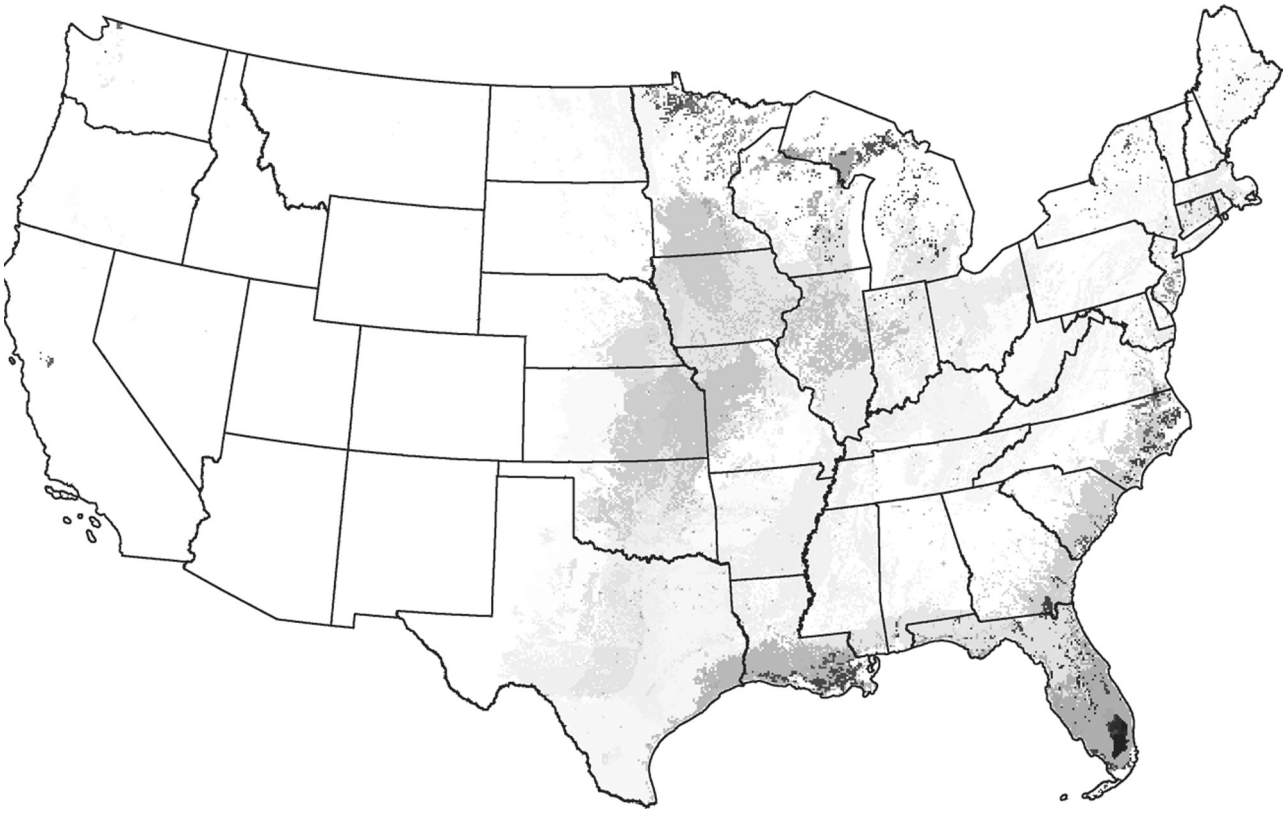

Figure 4. Maps of similarity to any selected quantitative ecoregion can be produced. The Euclidean distance in data space from the centroid of each ecoregion to the centroid of the chosen region is calculated. Ecoregions with closer centroids are more similar and are colored darker gray. The Everglades ecoregion was selected from this 1000-ecoregion map based on 25 primary environmental variables, so that the map shows the quantitative degree of ''Everglades-ness'' across the map.

Abrupt changes in Similarity Colors are accompanied by the numerous parallel contours of an ecotone, whereas subtle color changes are accompanied by the meandering contours of an ecopause. Close representativeness contours create thick black lines where ecoregion borders are sharp ecotones. Representativeness contours combine the traditional notion of exclusive membership of each location in a single ecoregion with a quantitative measure of goodness of fit or belonging.

Quantifying the Similarity of Ecoregions

When ecoregions are quantitatively derived, one can select a single ecoregion of interest and then produce a sorted list of the similarity of all other ecoregions to the one selected. The chosen ecoregion establishes an origin in data space and, using the Euclidean distance from this origin to the centroid of every other ecoregion, pairwise similarity measures can be calculated. Coding these pairwise similarity values as gray levels, the degree of similarity of all ecoregions to the selected ecoregion can be mapped.

Thus, maps can be drawn that show the degree of innate similarity between a particular selected ecoregion and the rest of the map. For example, starting with the 1000 most different ecoregions based on the 25 primary environmental factors described earlier, we produced a map of ''Everglades-ness'' that shows how similar other regions are to the Florida Everglades (Figure 4). Variables considered include elevation, temperature, precipitation, soil characteristics, and solar inputs (Hargrove and Hoffman 2003). Darker areas are most similar to the selected ecoregion. The Okefenokee Swamp in Georgia, the Great Dismal Swamp in Virginia, the Mississippi Delta, and the Wisconsin/Minnesota ''Land of a Thousand Lakes'' all rank high in their degree of ''Everglades-ness.'' These comparative representativeness maps quantify the ecoregion comparisons that ecologists have always wanted to make but, using traditional expertise-based ecoregions, could only subjectively estimate.

Quantitative Ecoregions as a Basis for Sampling Network Design

If we invert the quantitative comparison concept to consider nonrepresentativeness in the context of an existing network of sites or sample locations, we can quantify how well a particular established network is representative of a larger map that contains it. A network in this sense consists of a geographic constellation of sites or facilities or can simply represent locations where samples have been taken. Network analysis shows how well the sampled environments represent the rest of the map and identifies the best locations for new sites or installations. The best location for an additional site will be in places that are the least well represented by the network of existing sites.

Instead of a one-to-one centroid comparison like "Everglades-ness," network analysis entails a one-to-many centroid comparison in data space. To quantify network coverage, we determine how different each ecoregion is from the most similar network site or sample. For each ecoregion in the map, we find the Euclidean distance in data space to the single closest ecoregion that contains a site from the network. As earlier, this distance is coded to a gray level. Unlike the "Everglades-ness" maps, however, darker areas represent areas that are poorly represented by the existing network. Because this method quantifies coverage or presence of sites, sites will always sit within well-represented ecoregions, which will be colored white.

Maps showing the geographic areas represented by each individual site can be generated, and importance values can be calculated for each site based on the marginal representation it adds to the network. Quantifying the contribution of each site to network representation can minimize the impact of site elimination on representation. Finally, a network with a given number of sites can be designed that is theoretically optimal, having the highest possible representation of environmental conditions on a map (given that number of underlying ecoregions).

When submitted to a network analysis based on an ecoregionalization into the 2000 most different ecoregions based on 25 environmental factors, the National Science Foundation's Long-Term Ecological Research (LTER) study site network indicates that additional LTER sites in the Olympic peninsula and the Klamath, Sierra Nevada, and Northern California Coast mountains would increase the degree to which the LTER network represents environments in the United States (Figure 5). Gulf coastal environments are also poorly represented by the existing LTER network.

Networks of installations like LTER and AmeriFlux represent significant investments of research capital. Sites in most national-scale networks might not be located by design, but, instead, by opportunity or logistic convenience. Network analysis, as an outgrowth of the quantitative treatment of ecoregions, provides objective guidance about the design, performance, and modification of such networks and can improve on more idiosyncratic design approaches.

Statistical Modeling of Environmental Niche Envelopes

A variation of cluster-based ecoregionalization uses Hutchinson's theoretical underpinnings to forecast a species' geographic range in new locations or under altered environmental conditions (Hargrove and Hoffman 2000). First, each cell within the species' environmental range is located in a multidimensional environmental space. Rather than specifying the number of ecoregion groups desired, variance-based clustering is used to define as many fixed-radius, equal-variance clusters as necessary to ensure that each of the cells occupied by the species is contained within at least one cluster. All clusters, when considered together as multidimensional volumetric pixels, form a model of the location and shape of the niche within data space. Selection of the radius (variance) for the clusters controls the resolution with which the niche envelope is defined and serves to fill gaps within the Hutchinsonian hypervolume.

If the current geographic range data include some surrogate measures of fitness under each combination of environmental conditions, this information can be attached to the clusters defining the niche hypervolume. Hypervolume definitions are used to make geographic range predictions for new areas by projecting all map cells in the area into standardized environment space, then testing each cell to determine if it is within one of the clusters that define the niche hypervolume. The mean surrogate fitness associated with the closest cluster centroid containing a cell is used to predict the fitness of the species within that cell location in the new geographic range.

We defined niche hypervolume models for loblolly pine, *Pinus taeda* L., and sugar maple, *Acer saccharum* Marsh., in terms of the 25 environmental condition gradients of climatic, physiographic, and soil factors listed earlier. A within-cluster variance radius of 0.75 standardized units defined a niche model containing 49,324 clusters for *P. taeda* and 45,490 clusters for *A. saccharum*. We tested the niche hypervolume models

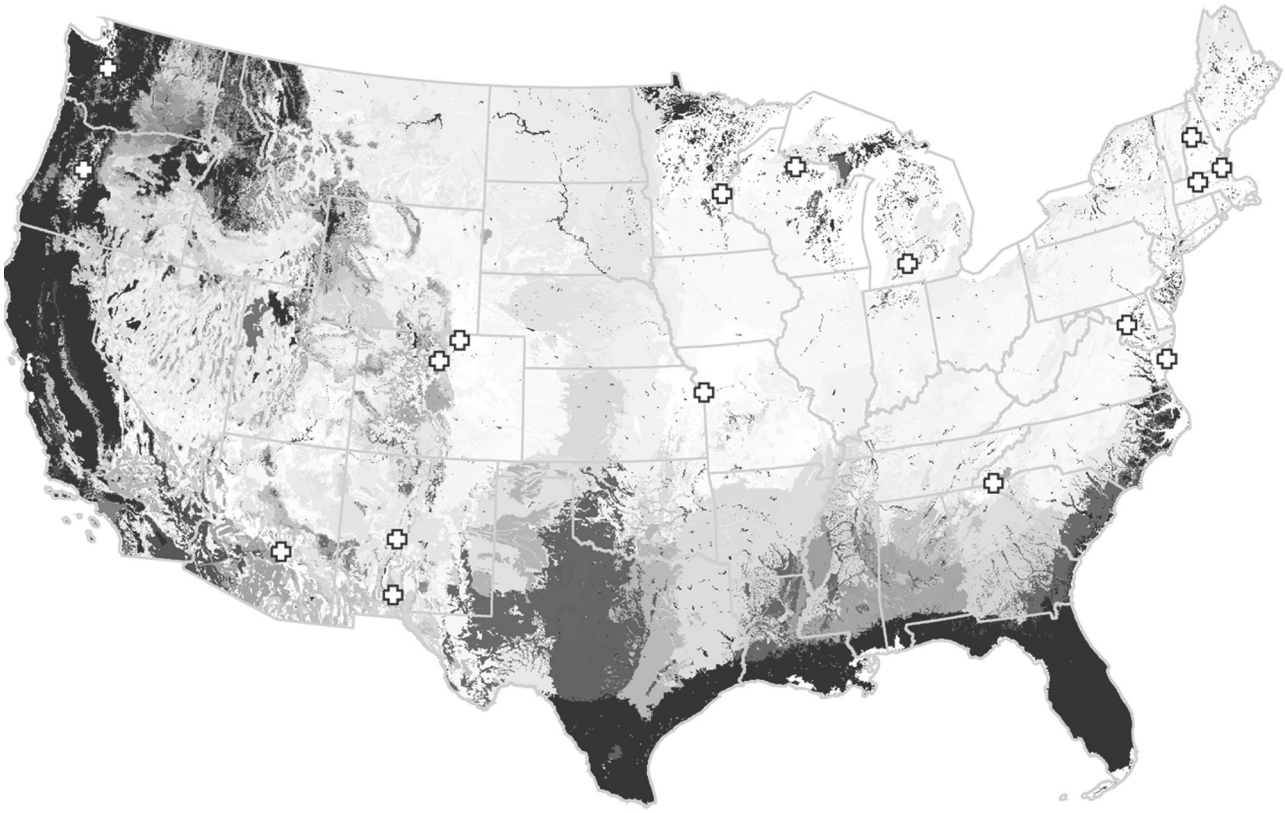

Figure 5. Quantitative ecoregions provide a basis for the analysis of sampling networks to show their coverage and representation. Distance in data space to the closest quantitative ecoregion containing a site quantifies the representation of a network to each ecoregion in the map. Environments in darker ecoregions are poorly represented by this network, the National Science Foundation's Long-Term Ecological Research (LTER) sites. Additional LTER sites in these areas would increase the representativeness of this sampling network.

by predicting the current ranges for each species within the continental United States. Figure 6 compares the current range prediction with the known current range for these two species.

When the current conditions within the United States are compared with the niche hypervolume definitions, the predicted distributions strongly resemble the known current distributions for both of these tree species (Figure 6). Areas outside the current geographic distributions of these species were not predicted by the niche hypervolume model (Figure 6). The current ranges were successfully predicted by a niche model even when half of the training data were randomly discarded (Hargrove and Hoffman 2000). Improvements resulting from the addition of biotic interaction effects would be limited to the outer margins of the geographic ranges for these two species and could be tested by including in the environmental data set information on the distribution of competitors, pests, or pathogens.

Unlike GAM or CART, cluster-based niche modeling first builds a model of the niche envelope itself and then uses this model to predict the species range. The niche model itself can be studied directly, including ranking niche breadth along each of the environmental dimensions. Overlap of multiple modeled species can be studied in environmental space as well as geographic space. Unlike ENFA, the environmental envelope is determined without reference to variance within some arbitrarily larger extent or context. Explicit absence data are not required, as they are with GAM. Rather than considering each environmental variable individually in sequence, cluster-based modeling considers the niche in a true simultaneous multivariate way, mirroring the way that organisms experience the environment.

Multivariate Geographic Clustering Through Time

Multivariate geographic clustering can be used through time to find any number of geographic areas with similar combinations of environmental conditions,

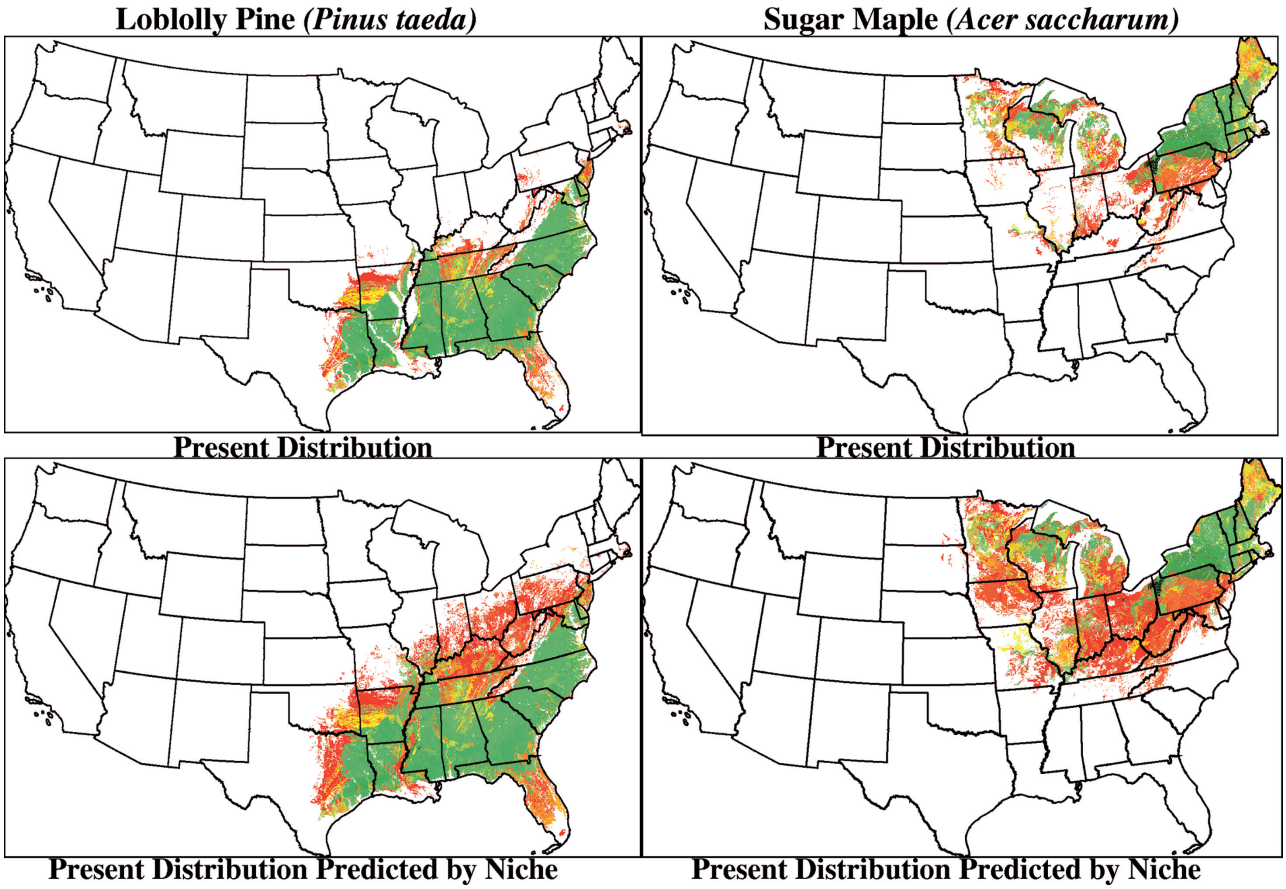

Figure 6. Variance-based clustering can be used to model the environmental niche hypervolume for a particular species. Niche models for loblolly pine and sugar maple were tested by predicting current ranges of these species within environments found in the conterminous United States. Green shows high-fitness habitat, red shows poor habitat, and yellow is intermediate. Predictions for each species agree with optimum habitats, but they overpredict the fringes of the ranges. Niche models can be used to predict species responses to climatic change.

wherever and whenever they occur. The Multivariate Clustering process outlined in Figure 1 can also be applied to a series of maps to visualize shifts in ecoregions under a set of dynamic conditions that are changing through time. Rather than disassembling individual map cells from a single map, cells from all maps in a chronological sequence are disassembled and plotted together in the same data space. In this way, groups of similar cells are objectively determined across space and through time. The number of maps in the chronosequence is unimportant; the Multivariate Spatio-Temporal Clustering (MSTC) process is extensible to any temporal resolution.

Comparing Past and Present

We used difference maps from a paleovegetation atlas (Frenzel and others 1992), along with present-day maps for temperature and precipitation, to produce spatial estimates of temperature and precipitation that occurred during the Last Glacial Maximum (LGM), 20,000 to 18,000 years before present. Then, we clustered the LGM and present-day growing condition maps together in a single pass using MSTC. The maps that resulted were both simultaneously divided by this empirical technique into clustered spatio-temporal combinations of conditions within which the growing conditions were similar. By examining the LGM and present-day maps that result, one can readily identify regions with similar growing conditions, both within and between the two points in time.

Nine factors deemed important for plant growth were included in the analysis: elevation, slope, bulk density of the soil, depth of mineral soil, soil depth to bedrock, mean annual temperature, mean annual precipitation, soil water-holding capacity, and mean annual solar insolation. Spatial patterns of all conditions except temperature and precipitation were assumed to be identical at the LGM to what they are

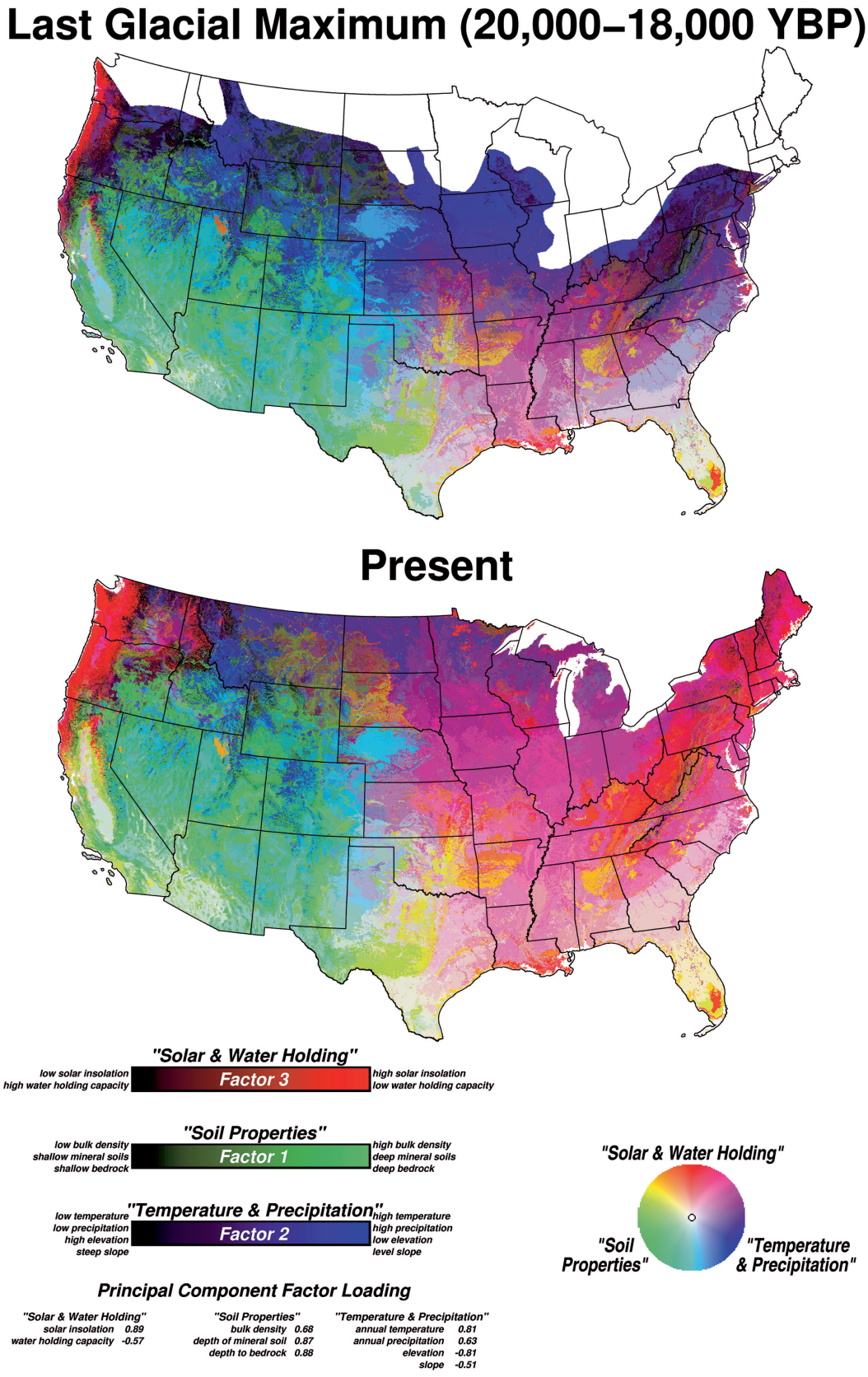

Figure 7. A common set of quantitative ecoregions can be found in a sequence of maps changing through time. Environments at Last Glacial Maximum (LGM) and the present were divided into a common set of 500 ecoregions in terms of nine environmental characteristics, shown here in Similarity Colors. The white region along the northern border of the LGM map represents the southernmost encroachment of glaciers.

today; topographic factors, like elevation and slope, and soil physical properties, like density, depth, and water-holding capacity, were assumed to have been constant, but were included to further differentiate growing conditions.

We used MSTC to simultaneously divide the LGM and present-day maps into 500 clustered spatio-temporal combinations on the basis of the nine input variables (Figure 7). Both maps are colored using the Similarity Colors encoding scheme, so that similar environmental combinations and regions with similar growing conditions are visible as similar colors, both within maps and across time.

In Figure 7, Factor 1, "soil properties," is green, Factor 2, "temp & precip," is blue, and Factor 3, "solar & water-holding," is red. Black results from small but balanced values of all three factors, and white results from large but equal values of all three factors. Thus, white areas in Florida, Texas, and California's Central Valley indicate high solar insolation, low water-holding, high bulk density, deep soils and bedrock, high temperature and precipitation, low elevation, and gentle slopes.

Figure 8 shows the maps clustered through time into 50 ecoregions shown with random colors. Because the same random color scheme is conserved across both maps, particular clusters can be tracked across time. For example, the red ecoregion in central Florida, the Panhandle, and southeastern Texas at the LGM can be seen to expand tremendously in the present-day map, growing to fill Florida, most of the coastal plain, and the eastern half of Texas. The purple ecoregion that had occupied the coastal plain during the LGM migrated to northern Texas.

Most of the 500 clusters are present at both times, but a given ecoregion might have decreased in area, remained about the same size, or increased in area across the pair of maps. Changes in the area of each ecoregion are easily quantified. Unique clustered combinations that show extreme behavior through time are of particular interest. For example, in the LGM map in Figure 9, black areas indicate ecoregions that went extinct (i.e., shrank to zero area in the present-day map). These locations during the LGM had unique combinations of growing conditions that no longer exist. Similarly, the black areas in the present-day map at the bottom indicate the locations of ecoregions with new combinations of growing conditions; there were no analogs for these environmental conditions during the LGM.

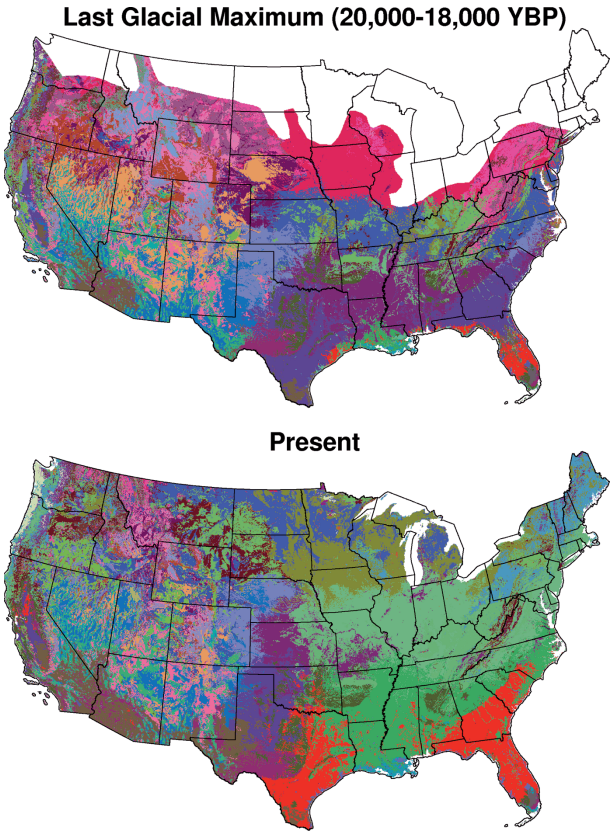

Figure 8. The same environments as shown in Figure 7, divided into 50 common ecoregions, and shown in random colors. Colors are conserved, such that the same color indicates the same ecoregion in each map.

Comparing the Present with Two Alternative Futures

We compared the present environment of the United States in terms of the 25 variables listed earlier with two alternative predictions for the future United States in the year 2099. We implemented forecasts from two global climate models by altering 16 of the 25 variables in spatially explicit ways and then used MSTC to find 100 common environmental combinations across this set of 3 maps. One prediction is from the Hadley United Kingdom Meteorological Office (UKMO) and the other is from the Canadian Climate Center (CCC). We used yearly predictions from 1994 to 2099 from these two models that were downscaled to 0.5-degree spatial resolution for the continental United States by the VEMAP program as part of the US National Assessment (Kittel and others 1997).

We averaged the VEMAP simulated forecasts for monthly minimum and maximum temperature, monthly solar irradiance, and monthly precipitation over a 5-year interval beginning with 1994, and another ending with 2099, and took the difference between the means as the predicted change. Difference layers were applied to our fine spatial resolution maps of current conditions within the United States in order to obtain

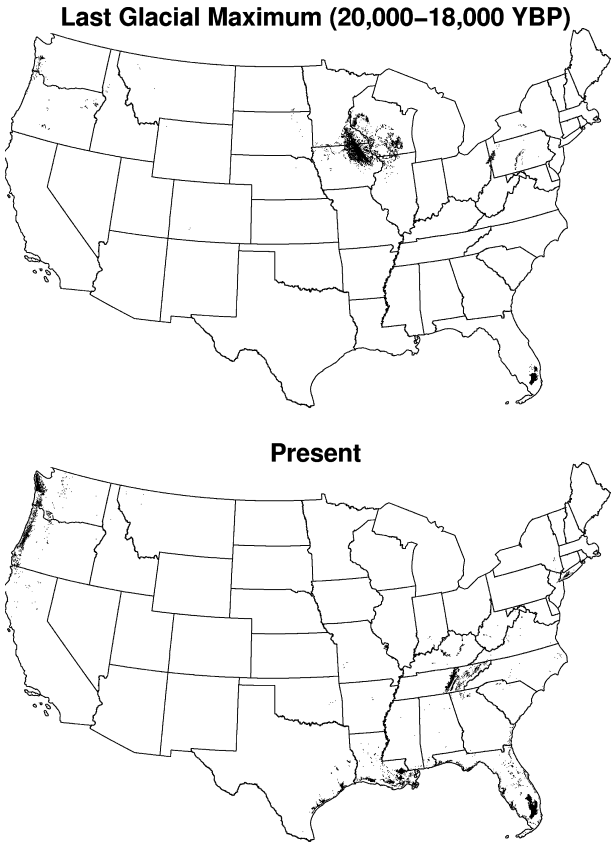

Figure 9. Changes in the area of a particular quantitative ecoregion can be tracked through time. Particular ecoregions can disappear entirely, and ecoregions with new environmental conditions can arise. The top map shows ecoregions that were present at the LGM but have no analogs today. Conversely, the bottom map shows new, novel environments that are present now but had no counterparts at the time of the LGM.

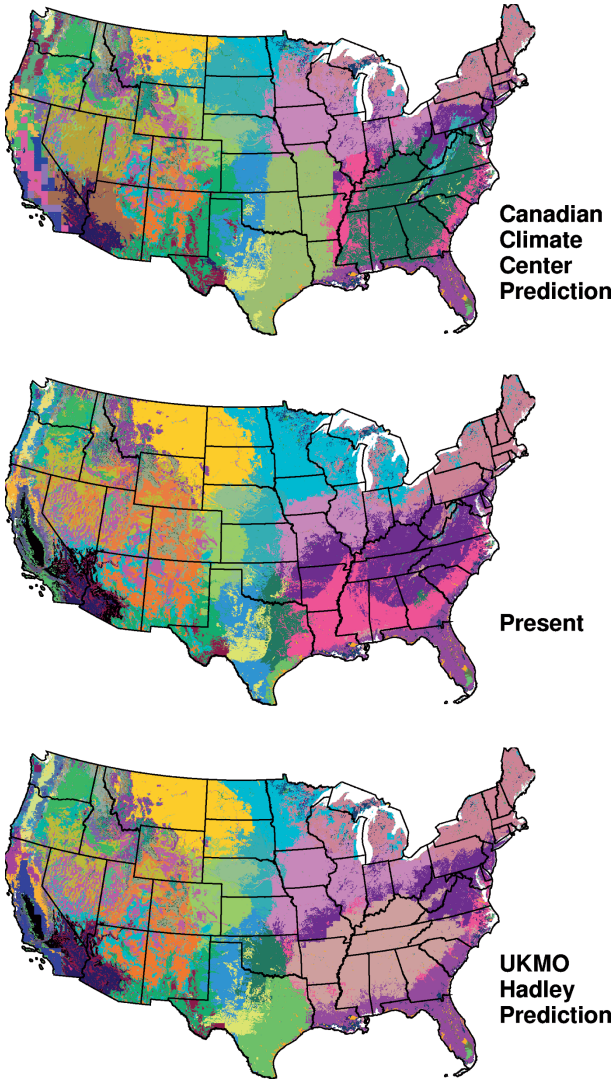

Figure 10. Quantitative ecoregions can show future climate changes in terms of present conditions. One hundred common ecoregions were found within a triad of maps: one representing present conditions and the other two representing predictions for the year 2099 from the Hadley and Canadian Climate Center (CCC) global climate simulations. Common quantitative ecoregions allow direct comparison of the two forecasts relative to present conditions.

predicted conditions. Because of the inherent spatial coarseness of the modeled conditions relative to the high resolution of the present-day conditions, some "slivers" near the coastline of the United States extended beyond the edge of the predictions. Conditions inside these small "slivers" remained unchanged.

One hundred common ecoregions delineated in the triad of maps are compared in Figure 10. Because the same random color table is used for all maps, changes in the location of areas affected by different environmental conditions can be traced between maps. According to the CCC future scenario, the northeastern United States experiences little change, except in Pennsylvania. The red cluster of the present coastal plain divides into a Mississippi valley component and an Atlantic seaboard component. A green coastal Texas cluster shrinks to a tiny area near Galveston Island. Severe changes are predicted for California by the CCC model, making the coarse resolution of the prediction visible there. In the Hadley UKMO model, the red coastal plain dissolves into a number of small remnant clusters around the periphery of a new brown ecoregion combination according to this forecast. The medium green coastal Texas ecoregion grows to become the dominant cluster in eastern Texas. Dividing present and future maps into a common set of ecoregions allows for direct comparison, letting the viewer

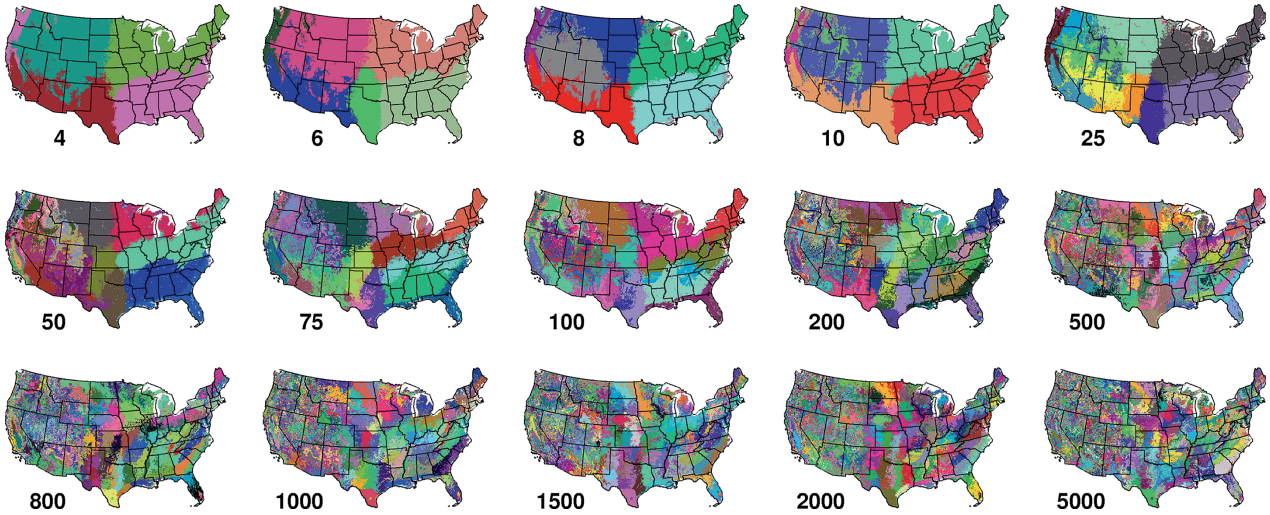

Figure 11. Multivariate Geographic Clustering can be repeated to create quantitative ecoregions at many levels of division. Dividing the United States based on 25 climatic, physiographic, and edaphic factors coarsely into a few ecoregions reproduces intuitively understood regional differences. Dividing finely into many ecoregions quickly surpasses the resolving ability of experts. All ecoregions from any single map have roughly equal environmental heterogeneity.

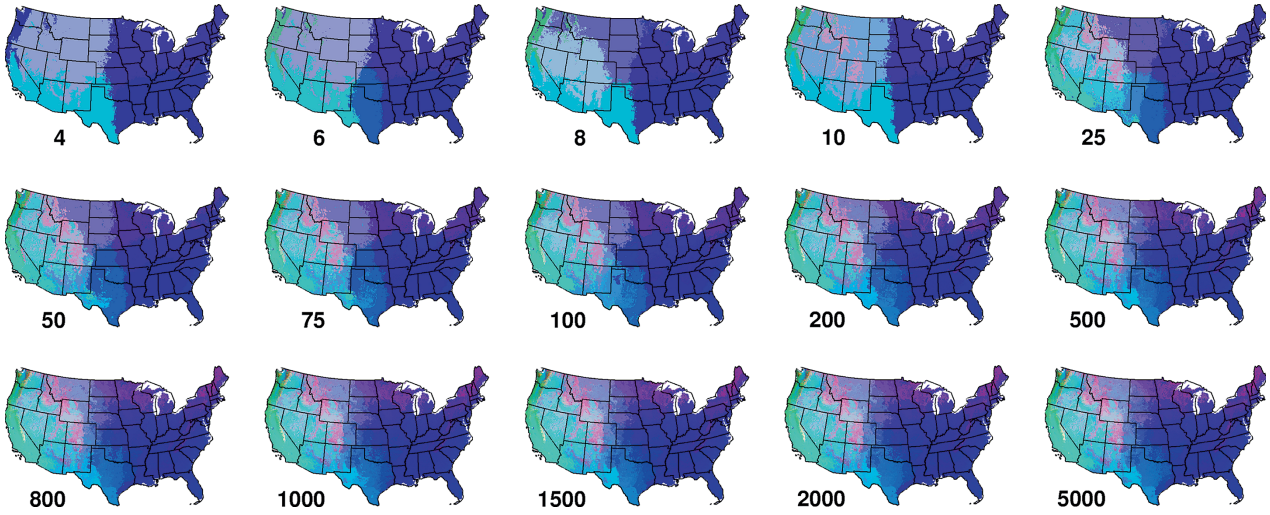

Figure 12. Quantitative ecoregions from Figure 11, colored using Similarity Colors. Beyond a certain level of division, all ecoregion maps converge and cannot be distinguished, even though the polygons underlying these maps are totally different.

see that under either scenario, the future ecoregion containing Pittsburgh becomes like the current ecoregion containing Atlanta, Minneapolis becomes like St. Louis, and Cleveland becomes like Kansas City.

Convergence Within a Spectrum of Ecoregions

Coarse thematic division of the United States on the basis of the 25 environmental variables into a few ecoregions accurately identified the commonly recognized subregions of the country (Figure 11). Increasingly finer divisions into more tightly defined ecoregions soon surpass the classification capabilities of any human expert.

Although such a spectrum of divisions forms one sort of hierarchy, some desire a hierarchical framework of nestable ecoregions (e.g., Bailey 1983) or ecoregions that can be combined to form some larger biotic, political, or accounting polygons. Because such arbitrary preexisting polygons are unlikely to be equal in environmental heterogeneity, specifying a preset number of subregions per parent polygon is not

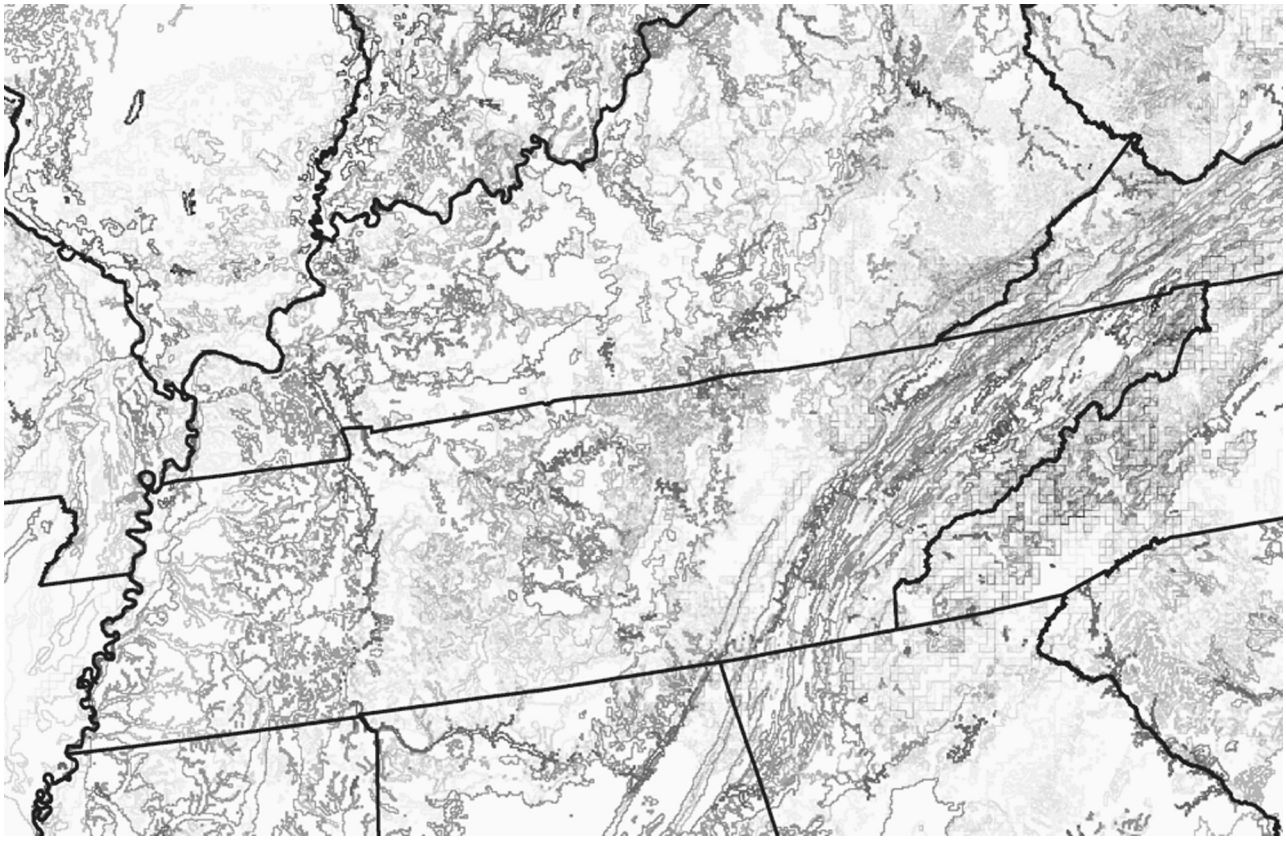

Figure 13. Borders that reoccur between ecoregions at many levels of division can be added together across maps in Figure 11 to create ''fences'' of varying heights. Fences depict sharp instantaneous linear differences in environment and might not form closed polygons. Darker gray fences are higher (more different). Portraying the environment as an open gradient punctuated by fences of linear discontinuities escapes the concept of closed homogeneous ecoregions.

advisable. However, variance-based MSTC, applied recursively, can be used to generate as many equal-variance subregions as are appropriate within each preexisting polygon. Any two subregions have comparable variability, but subregions will reconstitute the specified parent polygons.

As the number of specified ecoregions increases, ecoregion maps using Similarity Colors converge rapidly to show the same large regional trends in ecological relationships (Figure 12). If two ecoregionalizations based on the same environmental conditions are produced, but one is divided finely into many ecoregions and the other is divided coarsely into relatively few, the Similarity Colors versions of the two very different maps will be indistinguishable from each other at a large scale. This convergence occurs despite the fact that the polygons underlying each map are completely different—only the color-encoding technique is the same. The choice of the number of ecoregion divisions is relatively insensitive; beyond some minimum number of ecoregions, the same regional ecological patterns are revealed. Ecologists need only inspect such ecoregion visualizations to gain insight about regional environmental relationships.

Spatial ''Fences'' of Instantaneous Environmental Differentials

Figure 11 shows that some borders consistently reappear in the same locations across a series of ecoregion divisions. These common or persistent borders represent geographic locations where ecoregions are consistently divided in each map at several levels of overall division. The locations of borders can be examined across many levels of division to provide a measure of their robustness.

We call these recurring borders ''fences,'' and their height describes the magnitude of the environmental differential across them. We isolated the borders from the 15 ecoregion maps in Figure 11 and then added them together across maps to create ''fences'' of varying heights. Higher fences represent frequently recur-

ring borders and greater putative environmental differences. Showing fences taller than a particular height is equivalent to selecting a particular level of instantaneous environmental difference. Fences might not create closed regions at a particular height, but might appear as isolated line segments. If we walked around such a freestanding fence from a point on one side to the other, we would never experience as dramatic a shift in the mix of environmental conditions as we would if we climbed over the fence at that point. A gradual ecopause might not show a fence at all.

Figure 13 shows persistent ecoregion ''fences'' in Tennessee. Darker gray fences are taller (more different). The Ridge-and-Valley province, the Cumberland Plateau, and the Nashville Basin can be seen in terms of the darkest gray fences. The Appalachian Mountains show fences that form a regular grid, because the environmental conditions there are changing rapidly across space and are poorly represented at this resolution.

Like representativeness contours, fences represent a way to locate sharp ecotones in the environment. Fences portray the environment as a continuous gradient with sharp spatial discontinuities or steep differentials exceeding a selected level of heterogeneity. Classical ecoregions hide within-ecoregion variability, portraying all locations within a particular polygon as homogeneous (less than some accepted level of difference). Fences are more accurate, showing the environment as a gradient, but spatially locating ecotones steeper or sharper than a particular level of difference. Ecoregions no longer need to be shown as closed, encompassed polygons, but they can be viewed as open gradients punctuated with fences of varying heights representing instantaneous linear discontinuities at selected levels of difference.

Significance Testing of Ecoregionalizations

Although often demanded, there are no generally accepted quantitative methods for comparing alternative ecoregion schemes. Because they simply represent alternative grouped arrangements, no set of ecoregions is ''true'' or ''wrong'' in a statistical sense. Just as the correct time cannot be established from a consensus among many watches, no single set of ecoregions can be judged objectively as superior. Nevertheless, qualitative expert-based ecoregions have become the accepted standard against which alternative ecoregion schemes are compared.

A method for testing significant differences among alternative ecoregion maps would be valuable, yet statistics appropriate for the pairwise comparison of alternative categorical maps are lacking, particularly when those maps could contain different numbers of categories. Nor does having equal numbers of ecoregions necessarily improve a one-to-one comparison. Significance testing is exacerbated by the absence of appropriate null models for testing spatial pattern (but see Hargrove and others 2002). More fundamentally, a search for consensus among inherently different regionalizations created for unique purposes might be neither appropriate nor justified. A theoretical and technical framework for evaluating differences among ecoregions remains an open research question [but see Wolock (2004) and Thompson and others (2004) in this volume].

Conclusions

Statistical clustering objectively divides a complex continuous multivariate population into similar groupings for easier interpretation. This is precisely the object of ecoregionalization. We have outlined an empirical technique that can unambiguously locate, characterize, and visualize ecoregions and the borders that separate them, based on a chosen suite of environmental characteristics. Ecoregion maps coded with Similarity Colors and augmented with sharpness contours represent integrated portrayals of complex environmental datasets that are visually rich in ecological information.

Ecoregions should be created such that all regions in a single map have similar environmental heterogeneity. Some experts use high variability as an explicit ''characteristic'' to create single ecoregions, producing, for example, a single region for the Central Appalachian Ridge and Valley ecoregion (Omernik 1987). We suggest that ecoregions be divided on the basis of other properties, such that the heterogeneity within the regions that are produced is held constant. Our method would break such chimeric regions into their more homogeneous constituents (i.e., ridge tops and valley bottom). It is better to set a level of heterogeneity and then form ecoregions so that they all abide by this setting. The statistical clustering process ensures that all large ecoregions have the same upper limit on heterogeneity. Any two such ecoregions can be equitably compared, because they were created at the same level of a variance hierarchy. Equal-variance ecoregions are meaningful and interpretable, serving the purpose of ecoregions better than ecoregions produced *ad hoc*.

Ecoregions defined based on subjective opinions of an expert are limited by his/her geographic experience and knowledge. Subjective and qualitative techniques have restricted the application of the classical ecoregion concept to the realms of human experi-

ence. This requirement for direct expertise and personal familiarity with the ''data'' has been an underappreciated but severe constraint for the ecoregion concept.

Quantitative approaches liberate the ecoregion concept and allow ecoregions to be expanded into new realms of multitemporal comparisons, fine spatial resolutions, global extents, and numerous regional divisions. We have described machine-generated ecoregions created through time into the geologic past and the future—places where no human expert has been. When the delineation process does not depend on human expertise, ecoregions can be extended, even to temporally dynamic fluid environments like water and air. We are starting to experiment with oceanic and atmospheric ''ecoregions,'' and early results suggest that the same advantages that ecoregions hold for ecology might be applicable in these other realms (Hoffman and others 2004).

Indeed, a new set of terms might be needed to describe this extended ''ecoregion'' concept. The statistical creation of a set of ecoregions through time should perhaps be thought of as defining a set of frequently revisited environmental ''states'' or ''regimes'' that represent the actually realized combinations of factors seen within that geographic area during that interval (Hoffman and others 2004). Quantitative ecoregions offer an accounting procedure that can track changes as a geographic location shifts from one ecoregional state to another through time. This dynamic tracking aspect is new to the ecoregion concept and is of great potential utility.

This new-found extensibility of the ecoregion concept is the direct result of the quantitative approach. A few basic quantitative principles provide the basis for all of these products. The ability to calculate ecoregion centroids representing average or synoptic conditions allows centroids to be used as holotypes for newly defined ecoregions. Euclidean distance from cells to their centroids is a quantitative measure of representativeness within an ecoregion, and the distance between centroids is a useful measure of representativeness across ecoregions. The ability to quantify the Euclidean similarity of one ecoregion centroid to any other point or region provides a quantitative foundation for a wide diversity of formerly unavailable ecoregion-based analyses.

Ecoregionalization remains a rich and complex process — one with many subtle nuances that requires much expertise. The process may be sufficiently complex and irreducible that it is destined to remain more art than science. Whether bold or naive, we believe that efforts to capture and to quantify the techniques of such artists are a worthwhile scientific pursuit.

Acknowledgments

This article was improved by comments from Rebecca Efroymson, Holly Gibbs, Geoff Henebry, Paul Hessburg, Yetta Jager, and three reviewers, Ferko Csillag, Jake Overton, and Anthony Lehmann. Tom Loveland and Gerry McMahon paid for Hargrove's travel to EROS for the Ecoregionalization conference. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725.

References

Austin, M. P., A. O. Nichols, and C. R. Margules. 1990. Measurement of the realised quantitative niche: Environmental niche of five Eucalyptus species. *Ecological Monographs* 60:161–177.

Bailey, R. G. 1983. Delineation of ecosystem regions. *Environmental Management* 7:365–373.

Belbin, L. 1993. Environmental representativeness: regional partitioning and reserve selection. *Biological Conservation* 66:223–230.

Bernert, J. A., J. M. Eilers, T. J. Sullivan, K. E. Freemark, and C. Ribic. 1997. A quantitative method for delineating regions: an example for the Western Corn Belt Plains ecoregion of the USA. *Environmental Management* 21:405–420.

Brooker, S., S. I. Hay, and D. A. P. Bundy. 2002. Tools from ecology: Useful for evaluating infection risk models? *Trends in Parasitology* 18:70–74.

Bunce, R. G. H., C. J. Barr, R. T. Clarke, D. C. Howard, and A. M. J. Lane. 1996. Land classification for strategic ecological survey. *Journal of Environmental Management* 47:37–60.

Busby, J. R. 1991. BIOCLIM—A bioclimate analysis and prediction system. Pages 64–68 *in* C. R. Margules, and M. P. Austin. (eds.), Nature conservation: Cost effective biological surveys and data analysis. CSIRO, Melbourne.

Carpenter, G., A. N. Gillson, and J. Winter. 1993. DOMAIN: A flexible modeling procedure for mapping potential distributions of plants and animals. *Biodiversity and Conservation* 2:667–680.

Carter, S. E. 1997. Spatial stratification of western Kenya as a basis for research on soil fertility management. *Agricultural Systems* 55:45–70.

Csillag, F., B. Boots, Marie-Josee Fortin, K. Lowell, and F. Potvin. 2001. Multiscale characterization of boundaries and landscape ecological patterns. *Geomatica* 55:509–522.

Csillag, F., and S. Kabos. 2002. Wavelets, boundaries, and the spatial analysis of landscape pattern. *Ecoscience* 9:177–190.

Engler, R., A. Guisan, and L. Rechsteiner. 2004. An improved approach for predicting the distribution of rare and

endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology* 41:263–274.

Frenzel, B., M. Pecsi, and A. A. Velichko. 1992. Atlas of paleoclimates and paleoenvironments of the northern hemisphere, Late Pleistocene to Holocene. Geographical Research Institute, Hungarian Academy of Sciences, Budapest, Gustav Fischer Verlag, Budapest.

Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147–186.

Guisan, A., T. C. Edwards Jr., and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* 157:89–100.

Harff, J., and J. C. Davis. 1990. Regionalization in geology by multivariate classification. *Mathematical Geology* 22:573–588.

Hargrove, W. W., and F. M. Hoffman. 1999. Using multivariate clustering to characterize ecoregion borders. *Computers in Science & Engineering* 1:18–25.

Hargrove, W. W., and F. M. Hoffman. 2000. An analytical assessment tool for predicting changes in a species distribution map following changes in environmental conditions. Proceedings, GIS/EM4 Conference. Available at http://www.colorado.edu/research/cires/banff/pubpapers/104/.

Hargrove, W. W., and F. M. Hoffman. 2003. New analysis reveals representativeness of the AmeriFlux network. *EOS, Transactions, American Geophysical Union* 84:529–535.

Hargrove, W. W., and R. J. Luxmoore. 1998. A clustering technique for the generation of customizable ecoregions. Proceedings, ESRI Arc/INFO Users Conference. Available at http://research.esd.ornl.gov/~hnw/esri98/.

Hargrove W. W., F. M. Hoffman, and P. M. Schwartz. 2002. A fractal landscape realizer for generating synthetic maps. Conservation Ecology 6:2. Available at http://www.consecol.org/vol6/iss1/art2.

Hargrove, W. W., F. M. Hoffman, and T. Sterling. 2001. The do-it-yourself supercomputer. *Scientific American* 265(2):72–79.

Hartigan, J. A. 1975. Clustering algorithms. John Wiley & Sons, New York.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized additive models. Chapman & Hall, London.

Hessburg, P. M., R. B. Salter, M. B. Richmond, and B. G. Smith. 2000. Ecological subregions of the Interior Columbia Basin, USA. *Applied Vegetation Science* 3:163–180.

Hirzel, A. H., and R. Arlettaz. 2003. Modelling habitat suitability for complex species distributions by the environmental–distance geometric mean. *Environmental Management* 32:614–623.

Hirzel, A. H., V. Helfer, and F. Metral. 2001. Assessing habitat-suitability models with a virtual species. *Ecological Modelling* 145:111–121.

Hirzel, A. H., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological niche factor analysis: How to compute habitat suitability maps without absence data? *Ecology* 83:2027–2036.

Hoffman, F. M., and W. W. Hargrove. 1999. Multivariate geographic clustering using a Beowulf-style parallel computer. Pages 1292–1298 *in* H. R. Arabnia. (ed.), Proceedings of the international conference on parallel and distributed processing techniques and applications (PDPTA '99), Volume III. CSREA Press, Irvine, CA.

Hoffman, F. M., W. W. Hargrove, D. J. Erickson III, and R. Oglesby. 2004. Using clustered climate regimes to analyze and compare predictions from fully coupled general circulation models. *Earth Interactions* (in press).

Holdridge, L. R. 1947. Determination of world plant formations from simple climatic data. *Science* 105:367–368.

Host, G. B., P. L. Polzer, D. J. Mladenoff, M. W. White, and T. R. Crow. 1996. A quantitative approach to developing regional ecosystem classifications. *Ecological Applications* 6:608–618.

Hung, C. 1993. Competitive learning networks for unsupervised training. *International Journal of Remote Sensing* 14:2411–2415.

Hutchinson G. E. 1957. Concluding remarks—Cold Spring Harbor Symposia on Quantitative Biology 22:415–427. Reprinted in 1991: Classics in Theoretical Biology. *Bulletin of Mathematical Biology* 53:193–213.

Iverson, L. R., and A. M. Prasad. 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs* 68:465–485.

Iverson, L. R., A. M. Prasad, and M. W. Schwartz. 1999. Modeling potential future individual tree-species distributions in the Eastern United States under a climate change scenario: a case study with *Pinus virginiana*. *Ecological Modelling* 115:77–93.

Jackson, S. T., and J. T. Overpeck. 2000. Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology* 26(Suppl):194–220.

Jensen, M. E., I. A. Goodman, P. S. Bourgeron, N. L. Poff, and C. K. Brewer. 2001. Effectiveness of biophysical criteria in the hierarchical classification of drainage basins. *Journal of the American Water Resources Association* 37:1155–1167.

Kittel, T. G. F., J. A. Royle, C. Daly, N. A. Rosenbloom, W. P. Gibson, H. H. Fisher, D. S. Schimel, L. M. Berliner, and VEMAP2 Participants. 1997. A gridded historical (1895–1993) bioclimate dataset for the conterminous United States. pages 219–222 *in* Reno, N. V. (ed.), Proceedings of the 10th conference on applied climatology, 20–24 October 1997 American Meteorological Society, Boston.

Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. *Biological Cybernetics* 43:59–69.

Kohonen, T. 1995. Self-organizing maps. Springer Series in Information Sciences Vol. 30. Springer-Verlag, Berlin.

Krohn, W. B., R. B. Boone, and S. L. Painton. 1999. Quantitative delineation and characterization of hierarchical biophysical regions of Maine. *Northeastern Naturalist* 6:139–164.

Lark, R. M. 1998. Forming spatially coherent regions by classification of multi-variate data: An example from the analysis of maps of crop yield. *International Journal of Geographic Information Science* 12:83–98.

Leathwick, J. R. 2001. New Zealand's potential forest pattern as predicted from current species–environment relationships. *New Zealand Journal of Botany* 39:447–464.

Leathwick, J. R., J. McC. Overton, and M. McLeod. 2003. An environmental domain classification of New Zealand and its use as a tool for biodiversity management. *Conservation Biology* 17:1612–1623.

Lehmann, A., J. R. Leathwick, and J. McC. Overton. 2002a. Assessing New Zealand fern biodiversity from spatial predictions of species assemblages. *Biodiversity and Conservation* 11:2217–2238.

Lehmann, A., J. Mc C. Overton, and J. R. Leathwick. 2002b. GRASP: Generalized regression analysis and spatial predictions. *Ecological Modelling,* 157:189–207.

Leung, Y. 1987. On the imprecision of boundaries. *Geographical Analysis* 19:125–151.

Lugo, A. E., S. L. Brown, R. Dodson, T. M. Smith, and H. H. Shugart. 1999. The Holdridge Life Zones of the conterminous United States in relation to ecosystem mapping. *Journal of Biogeography* 26:1025–1038.

Malmgren, B. A., and A. Winter. 1999. Climate zonation in Puerto Rico based on principal components analysis and an artificial neural net. *Journal of Climate* 12:977–985.

Martinez-Meyer, E., A. T. Peterson, and W. W. Hargrove. 2004. Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. *Global Ecology and Biogeography* 13:305–314.

McCullagh, P., and J. A. Nelder. 1997. Generalized linear models. Chapman & Hall, London.

McMahon, G., S. M. Gregonis, S. W. Waltman, J. M. Omernik, T. D. Thorson, J. A. Freeouf, A. H. Rorick, and J. E. Keys. 2001. Developing a spatial framework of common ecological regions for the conterminous United States. *Environmental Management* 28:293–316.

Omernik, J. M. 1987. Ecoregions of the conterminous United States. *Annals of the Association of American Geographers* 77:118–125.

Omernik, J. M. 1995. Ecoregions: A spatial framework for environmental management. Pages 49–62 *in* W. S. Davis. and T. P. Simon. (eds.), Biological assessment and criteria: Tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

Omernik, J. M. 2003. The misuse of hydrologic unit maps for extrapolation, reporting, and ecosystem management. *Journal of the American Water Resources Association* 39:563–573.

Omi, P. N., L. C. Wensel, and J. L. Murphy. 1979. An application of multivariate statistics to land-use planning: classifying land units into homogeneous zones. *Forest Science* 25:399–414.

Overpeck, J. T., T. Webb III, and I. C. Prentice. 1985. Quantitative interpretation of fossil pollen spectra: dissimilarity coefficients and the method of modern analogs. *Quaternary Research* 23:87–108.

Overton, J. Mc C., J. R. Leathwick, and A. Lehmann. 2000. Predict first, classify later—A new paradigm of spatial classification for environmental management: a revolution in the mapping of vegetation, soil, land cover, and other environmental information. In 4th international conference on integrating GIS and environmental modeling (GIS/EM4).

Overton, J. Mc C., J. R. Leathwick, G. Barker, and A. Lehmann. 2001. Advances in ecosystem depiction: Frameworks for sustainable ecosystem management. *IALE Bulletin* 19:1–3.

Overton, J. Mc C., J. R. Leathwick, R. T. T. Stephens, and A. Lehmann. 2002. Information pyramids for informed ecosystem management. *Biodiversity and Conservation* 11:2093–2265.

Peterson, A. T., and K. P. Cohoon. 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. *Ecological Modelling* 117:159–164.

Peterson, A. T., and D. A. Vieglais. 2001. Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem. *BioScience* 51:363–371.

Peterson, A. T., R. Scachetti-Pereira, and W. W. Hargrove. 2003. Potential geographic distribution of *Anoplophora glabripennis* (Coleoptera: Cerambycidae) in North America. *American Midland Naturalist* 151:170–178.

Prince, S. D., and M. K. Steininger. 1999. Biophysical stratification of the Amazon Basin. *Global Change Biology* 5:1–22.

Rathert, D., D. White, J. Sifneos, and R. M. Hughes. 1999. Environmental correlates of species richness in Oregon freshwater fishes. *Journal of Biogeography* 26:257–273.

Soriano, A., and J. M. Paruelo. 1992. Biozones: Vegetation units defined by functional characters identifiable with the aid of satellite sensor images. *Global Ecology and Biogeography Letters* 2:82–89.

Stockwell, D. R. B., and I. R. Noble. 1992. Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation* 33:385–390.

Stockwell, D. R. B., and D. Peters. 1999. The GARP Modeling System: Problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science* 13(2):143–158.

Stoms, D. M., and W. W. Hargrove. 2000. Modeling potential NDVI to monitor environmental stress. *International Journal of Remote Sensing* 21:401–407.

Thompson, R. S., S. L. Shefer, K. H. Anderson, L. E. Strickland, R. T. Pelltier, and M. W. Kerwin. 2004. Topographic, Bioclimatic, and vegetation characteristics of three ecoregion systems in North America: Comparisons along continent-wide transects. *Environmental Management* 34(Suppl): (this issue).

Walker, P. A., and K. D. Cocks. 1991. HABITAT: A procedure for modelling a disjoint environmental envelope for a plant or animal species. *Global Ecology and Biogeography Letters* 1:108–118.

White, D., K. Freemark, E. Preston, and R. Kiester. 1999. A hierarchical framework for conserving biodiversity. Pages 127–153 *in* J. M. Klopatek, and R. H. Gardner. (eds.), Landscape ecological analysis: Issues and applications. Springer-Verlag, Berlin.

Wolock, D. M., T. C. Winter, and G. McMahon. 2004. Delineation and Evaluation of Hydrologic-Landscape regions in the United States and using geographic information system tools and multivariate statistcal analyses. *Environmental Management* 34(Suppl 1):(this issue).

Zaniewski, A. E., A. Lehmann, and J. Mc C. Overton. 2002. Predicting species spatial distributions using presence-only data: A case study of New Zealand ferns. *Ecological Modelling* 157:261–280.

Zhou, Y., S. Narumalani, W. J. Waltman, S. W. Waltman, and M. A. Palecki. 2003. A GIS-based spatial pattern analysis model for ecoregion mapping and characterization. *International Journal of Geographic Information Science* 17: 445–462.