

Extending Community Ecology to Landscapes

Dean Urban¹, Sarah Goslee², Ken Pierce, and Todd Lookingbill

Nicholas School of the Environment, Duke University, Durham NC 27708 USA

Citation:

Urban, D., S. Goslee, K. Pierce, and T. Lookingbill. 2002. Extending community ecology to landscapes. *Ecoscience* 9(2):200-202.

Disclaimer:

This PDF preprint is reasonably close to the final version as published. Slight editorial changes and formatting differences will surely arise on publication.

¹ Contact information:

Nicholas School of the Environment, Levine Science Research Center, Duke University, Durham, NC 27708 USA. phone 919-613-8076; e-mail deanu@duke.edu

² Current address:

USDA-ARS, Jornada Experimental Range, New Mexico State University, Las Cruces, NM 88003

Abstract

A goal of landscape ecology is to infer processes or constraints that generate spatial pattern in communities and ecosystems. The rich tradition of plant community ecology is now being extended to address spatial pattern in vegetation over large spatial extents. The challenge in this is that vegetation pattern on landscapes is fine-grained, which presents sampling problems for large study areas. Further, spatial autocorrelation in ecological data, coupled with strong patterns of correlation among environmental factors (such as the gradient complexes governed by elevation) make it difficult to make clear inferences about the agents patterning landscape-scale vegetation. Here we review the methods of plant community ecology as extended to landscapes, and illustrate the challenges with a case study from Sequoia-Kings Canyon National Park in California's southern Sierra Nevada. We outline an iterative approach to such studies, with three stages. The first stage is a pilot study to characterize the spatial scaling of environmental factors presumed to be important to vegetation; this stage can often be conducted virtually, using digital terrain data. The second stage is iterative, and consists of building a preliminary explanatory model using a combination of ordination, classification, and Mantel tests—all analyses based on the same ecological distance or dissimilarity matrices. This preliminary model is then attacked to find its uncertain or sensitive parts, and these parametric conditions are mapped into geographic space to identify candidate sites for follow-up field studies in the third stage. This approach ensures that the most uncertain aspects of the preliminary model are refined in an efficient manner. As the approach proceeds toward a richer understanding of species-environment relationships and vegetation pattern, a need emerges for new kinds of field studies and novel extensions to existing statistical analyses. We discuss possible extensions of these as a natural consequence of this iterative process of model construction and revision.

Keywords: classification and regression trees, cluster analysis, gradient analysis, landscape ecology, Mantel tests, nonmetric multidimensional scaling, ordination, vegetation pattern; Sierra Nevada, California

Introduction

There is a growing body of work in which the methods of plant community ecology are extended to studies of landscapes (see Jongman et al., 1995; Legendre & Legendre, 1998 for examples). The most direct route to landscape ecology from community ecology is quite simple: do the same sorts of studies but over a larger area. But this literal extension does not admit that there are aspects of landscapes that call for different approaches in data collection, analysis, and modeling.

Here we consider the extension of community-ecological methods to landscape ecology, by noting first that this extension has not been entirely successful. The key issues ultimately are logistical: to a plant ecologist, landscapes are fine-grained but of large extent. We present an overview of an iterative strategy for sampling and analysis which has emerged in our lab; analytic methods follow recommendations by Legendre and colleagues (Legendre & Legendre, 1998 and references therein), while the field components were developed partly by design and partly through trial and error. There are several components to this approach: (1) We are interested in developing explanatory models stated in terms of plant-relevant variables at the scales at which these are expressed. Much of this interest stems from applications concerned with forecasting the impacts of environmental change. For example, while vegetation patterns are often correlated with elevation, elevation itself is but a proxy for plant-relevant factors (temperature, soil moisture) that covary with it. Although elevation might serve adequately to explain vegetation under current conditions, under climate change scenarios temperature and precipitation might not covary in the same way (indeed climate models do not even agree on the direction of the possible changes in precipitation) and soil variables will not change over these decadal timescales. Thus, explanatory models for our purposes must decompose the elevation gradient complex into its component variables. (2) We have invested in multi-scale sampling designs that can capture fine-grained details, at preselected scales, over large spatial extent. (3) We have devised a general tactic for translating between the parameter space of statistics or models, and the geographic space of actual landscapes; this serves as an explicit guide to sampling schemes designed to test hypotheses. (4) We have resigned ourselves to an explicitly multi-stage approach of successive refinement of explanatory models, in which preliminary data are used to leverage follow-up field studies to refine our models as efficiently as possible. In this we rely on a small set of complementary statistical methods, which are reviewed as applied in community ecology and then extended to landscapes.

We illustrate this integrated approach with examples from an on-going research program in Sequoia-Kings Canyon National Park in the Sierra Nevada of California, USA. Our study is part of a larger research program concerned with global change (Stephenson & Parsons, 1993). In this context, one of our goals is to be able to make useful forecasts of the potential implications of anthropogenic climate change on these climate-sensitive forests. As we also are pursuing similar studies in other areas (Oregon Cascades, southern Appalachians) we are confident that the approach is quite general and robust.

Complementary Methods in Community Ecology

There is a tradition in community ecology of not only developing appropriate methods, but also testing them intensively and extensively using real and artificial data. The chief tools in this, by far, have been ordination and classification (Whittaker, 1967, 1978a,b; Gauch, 1982; Pielou, 1984; Jongman et al., 1995; Legendre & Legendre, 1998). But these ecologists have also shown a ready willingness to combine complementary analyses to improve the interpretability or robustness of the analyses. Although the methodology of community ecology is far from controversial, we will focus our discussion on a small set of complementary techniques that are emerging as favorites in this field. Primarily, these are techniques based on ecological distance or dissimilarity (Gower 1985).

Ecological Distance and Dissimilarity

Community ecologists are familiar with the notion of ecological distance as expressed by indices of pairwise dissimilarities (or reciprocally, similarities) between sample units (Gauch, 1973). Common examples include various metrics of species compositional dissimilarity, which capture patterns of presences or abundances shared between samples. There is a nearly overwhelming variety of such indices (Legendre & Legendre, 1998, review dozens), but for our purposes we will only consider a typical example, the familiar Bray-Curtis index (Bray & Curtis, 1957):

$$d_{ij} = \frac{\sum_{k=1}^s |y_{ik} - y_{jk}|}{\sum_{k=1}^s (y_{ik} + y_{jk})} = 1 - \frac{2W}{A + B} \quad [1]$$

where d_{ij} is the ecological distance between samples i and j , y_{ik} is the abundance of species k on sample i , and there are s species. The far right side of equation (1) shows a perhaps more familiar form of the same index, in which W is species abundances shared between two samples, and A and B are total abundances for the two samples, respectively. This index, called *percentage difference*, can be applied to quantitative abundances or to binary (presence/absence) data. Indices of species composition typically are computed to account for the troublesome ecological fact that the information imparted by shared absences of a species is not equivalent to that implied by shared presences. Of course, depending on how the abundance data are transformed (relativized by species maximum, sample sum, or both?) the same index might yield very different results.

More generally, distance or dissimilarity indices can be computed for any environmental variables measured over a set of sample units. In many cases, an appropriate distance index is simple Euclidean distance:

$$d_{ij} = \sqrt{\sum_{k=1}^s (y_{ik} - y_{jk})^2} \quad [2]$$

with variables as defined above. Of course, for this index to make sense all variables must be in the same units; this is often done by standardizing all variables to z -scores. A special case of equation (2) occurs when the measured variables are sample locations as X and Y coordinates (e.g., UTM coordinates), in which case the ecological distance is geographic. It is this focus on spatial relationships that distinguishes landscape ecology from the traditional perspective of community ecology.

Ordination: Nonmetric Multidimensional Scaling

The ordination technique based on ecological distance is nonmetric multidimensional scaling (NMS) (Kruskal, 1964; Minchin, 1987; Faith et al., 1987). The goal of NMS is to produce a low-dimensional ordination space in which sample separation in this space reflects sample separation in multi-dimensional “species” space. In the NMS algorithm, this is accomplished by finding the ordination scores that yield the highest possible rank correlation between ordination distance and ecological distance. The algorithm is iterative and for large datasets can be rather computer intensive, but with new computers this is rarely limiting. Relative to other popular ordination methods (especially correspondence analysis and its derivatives), NMS makes no assumptions about the nature of species response to underlying gradients; further, the algorithm seems remarkably robust to noise and nonlinearities in the data (Minchin, 1987). The program is not widely available in commercial statistical packages, but is available in specialized ecological software packages (e.g., PC-ORD: McCune & Mefford, 1999; DECODA: Australian National University 1998).

Further interpretative power is afforded the ordination by regressing sample ordination scores on measured environmental variables. Graphically this is done by plotting correlation vectors into the ordination space (e.g., Jongman et al., 1995; McCune & Mefford, 1999). The net result of an NMS ordination followed by regressions on the environmental variables is similar to the outcome of canonical correspondence analysis (CCA; ter Braak, 1986, 1987): ordination axes that display compositional gradients relative to environmental factors. The difference in the two approaches is that while CCA is constrained to find axes expressed in terms of the measured environmental variables, NMS is free to find whatever compositional trends exist—whether these are related to the measured variables or not.

Classification: Hierarchical Clustering

Hierarchical agglomerative classification (clustering) is a compelling complement for NMS because the clustering can proceed from the same distance matrix as the ordination. The choice of linkage method (e.g., nearest-neighbor, farthest-neighbor, centroid, average, and so on) can have a profound influence on the classification (Lance & Williams, 1967; Sneath & Sokal, 1973; Legendre & Legendre, 1998).

Plotting samples in ordination space, coded by cluster group, is a powerful method for interpreting vegetation pattern. Groups that are compositionally “real” appear as tight clusters in ordination space, and the correlation vectors that “point to” each group suggest the environmental factors associated with the group. Conversely, a cluster group that was diffused throughout the ordination space would be less satisfying, indicating that the group’s cohesion is expressed on a higher ordination axis or that it has been too finely divided (and is not “real”). Likewise, a cluster

group to which no environmental factor “points” would suggest that some other (unmeasured) factor or constraint governs the group.

Community ecologists have a long tradition of using ordination and classification as complementary methods for depicting vegetation pattern. Ordination provides a continuous model of gradient response, while classification highlights discrete community types or discontinuities in the data. This approach is especially powerful when the ordination and classification are based on the same basic information, ecological distance (Legendre & Legendre, 1998). To this tandem approach we can add a third method, the Mantel test, which is based on correlations between distance matrices.

Regressions on Distance Matrices: Mantel Tests

Mantel’s (1967) test is based on a simple cross-products term and is normalized so that it is equivalent to the familiar Pearson product-moment correlation (Legendre & Fortin, 1989; Legendre & Legendre, 1998):

$$r = \frac{1}{(n-1)} \sum_{i=1}^n \sum_{j=1}^n \frac{(x_{ij} - \bar{x})}{s_x} \cdot \frac{(y_{ij} - \bar{y})}{s_y} \quad [3]$$

where x and y are variables measured at locations i and j and n is the number of elements in the distance matrices ($= m(m-1)/2$ for m sample locations), and the s_x and s_y are standard deviations for variable x and y . In practice, computing this index entails “unfolding” the lower triangle of each distance matrix into a long column vector of elements in order $d_{21}, d_{31}, d_{32}, d_{41}, \dots, d_{m,m-1}$ (many distance matrices are symmetric; if the matrix is not symmetric, all the elements are needed). The standardization in equation (3) allows one to consider variables of different measurement units within the same framework, rescaling the statistic to the range of a conventional linear correlation coefficient bounded on $[-1,1]$. In practice, a negative Mantel correlation is rare and would be difficult to interpret ecologically. The magnitude of Mantel correlations is often comparatively small even when highly significant statistically (Dutilleul *et al.* 2000).

Because the elements of a distance matrix are not independent, a Mantel test of significance is evaluated via permutation. In this, the rows and columns of the distance matrices are randomly rearranged. Mantel statistics are recomputed for these permuted matrices, and the distribution of values for the statistic is generated via many iterations (~ 1000 for $\alpha=0.05$, ~ 5000 for $\alpha=0.01$, $\sim 10,000$ for greater precision; even more near critical P values: Manly 1991a, 1997; Jackson & Somers, 1989).

The simple Mantel correlation addresses the basic ecological question, *Are samples that are environmentally similar also similar in species composition?* which is computed as the correlation between compositional and environmental dissimilarity matrices. Likewise, one could correlate either compositional or environmental dissimilarity with geographic distance, and ask whether there is spatial structure (autocorrelation) in these data. But the test is quite versatile and a number of variations lend additional inferential leverage in special cases.

One extension of a Mantel test uses a model or design matrix as one of the distance

matrices (Legendre & Fortin, 1989). For example, one might construct a test in which the samples are each assigned to a clustered group (*e.g.*, community type) and the predictor variables are measured environmental variables. The question is, *Are samples in the same group (community type) also similar in terms of the environmental variables?* In this case two samples are similar (distance=0) if they are both assigned to the same group, otherwise they are dissimilar (distance=1). A simple Mantel's using this matrix tests group means by comparing among- to within-group dissimilarities—much like an *F*-ratio. This is analogous to the test developed by Clarke (1993) and provided in the program ANOSIM.

A special use of a design matrix is to partition or subset the analysis into a series of discrete distance classes. That is, a first distance matrix is evaluated for all pairs of points within the first distance class; then a second matrix is scored for all pairs of points within the second distance interval, and so on. The result of this analysis is a Mantel correlogram, completely analogous to an autocorrelation function but performed on a (possibly multivariate) distance matrix (Oden & Sokal, 1986).

Smouse et al. (1986) extended the simple Mantel test in a regression framework, to compute partial Mantel correlations (see Legendre, 2000). As a partial regression technique, the Mantel test provides not only an overall test for the relationships among distance matrices, but also can test the contribution of each predictor variable for its pure partial effect on the dependent variable. If geographic location is included as one of the predictor matrices, then the test returns the pure spatial residuals (the effect of “space itself”) as well as the partials for each of the predictor variables (Borcard et al., 1992; Leduc et al., 1992). In the examples above, the relevant partials address the questions *Are samples that are similar in environment also compositionally similar, once the autocorrelation in composition has been accounted?* and *Are samples that are close together also similar in species composition, once the influence of environmental variables has been removed?*

Relative to ordination and classification, the family of Mantel tests offers an inferential model for explaining compositional trends in terms of measured environmental variables, in the same conceptual framework, but with explicit attention to the spatial structure of ecological data. Programs for computing simple Mantel correlations are increasingly available (McCune & Mefford, 1999) but the larger family of partial tests and Mantel correlograms are still restricted to specialized software packages (R: Casgrain & Legendre, 2001; RT: Manly, 1991b). Using a regression framework, these analyses can be coded readily in commercial packages such as S-PLUS (Insightful Corporation, Seattle, WA) that provide for user-written routines (Goslee and Urban, *unpublished ms*; S-PLUS library available on request).

Extensions to Landscape Ecology: a Sierran Case Study

In the Sierra Nevada we began simply, by collecting data and analyzing the samples as community ecologists might do. This was more difficult than we expected, and three special concerns quickly emerged to focus our on-going efforts. First, the study area is huge and yet we are concerned with patterns expressed over distances on the order of tens to hundreds of meters. This forced a sampling design that could capture fine grain over large extent, a departure from the conventional community-scale studies. Second, it became obvious that it would require several iterations between field studies and analysis to achieve a satisfyingly robust and predictive model of tree species-environment relationships in this system. This, in turn, required a facile means of

translating ecological information readily between the parameter spaces of statistical analysis (e.g., ordination space, regression models) and geographic space. Finally, our goal of devising predictive models using plant-relevant explanatory variables forced us to modify existing statistical methods to address more explicit and narrowly focused questions about our data, and at the same time to pursue a new suite of field methods that can provide new data to resolve uncertainties raised by preliminary analyses. Here we revisit these issues as a general approach to conducting community ecological analyses at the landscape scale.

Stage 1: Sampling Design

The logistical challenge of capturing fine-scale pattern over large extent naturally arises when extending community ecology to landscapes. If sampling intensity is a function only of limited time and human resources, then there are two natural consequences for sampling: either the study area remains small for a reasonably fine-grained spacing of sample locations, or the extent increases and the between-sample spacing of sampling increases. Either result has serious implications for subsequent analyses. In the former case, the sampling frame captures a small range of environmental conditions on the landscape and thus misrepresents the system; in the latter case, the full range of environmental variability is captured but between-sample spacing is too large to capture a host of relevant fine-grained ecological processes and constraints.

This latter condition is easily envisioned by noting the characteristic spatial scaling of a variety of environmental variables or processes that community ecologists hold to be important to vegetation pattern. For example, soils and microtopography vary over distances of meters to tens of meters, while hillslope gradients are expressed over scales of tens to hundreds of meters (Urban et al., 2000). Likewise, the process of seed dispersal by forest trees is expressed over distances of tens of meters (Clark et al., 1999). Clearly, a sparse sampling of a landscape will miss many variables likely to be useful in explaining vegetation pattern.

One straightforward approach to this sampling issue is to use prior knowledge of the system to postulate likely explanatory variables, identify the characteristic spatial scaling of these variables, and then devise a sampling scheme that will capture information at these scales. For many variables that influence vegetation, proxies can be developed from digital elevation models (DEMs) to do this analysis (Moore et al., 1990). For example, Urban et al. (2000) used a DEM from Sequoia National Park to identify the characteristic scaling of several terrain variables likely to influence the soil water balance: (1) elevation serves as a proxy for variation in temperature (via lapse rates) and precipitation (via orographic lifting); (2) slope aspect, transformed relative to the maximum insolation of a southwestern orientation (Beers et al., 1966), proxies for evaporative demand; (3) a topographic convergence index (Wolock & McCabe, 1995) represents the effects of local drainage on soil moisture. Each of these variables can be computed from a DEM, and their scaling identified via geostatistical techniques such as variograms or correlograms (Legendre & Fortin, 1989; Urban et al., 2000). For other variables such as soil depth or seed dispersal distance, prior field studies must be used to estimate their characteristic scaling.

Given information on the spatial scaling of variables of interest, simulation exercises can be used to design an effective sampling scheme that will capture this scale of variation while also covering the study area. For example, as an exercise we used a topographic convergence index (TCI) computed from a 30-m resolution DEM of Sequoia National Park as a test case. We “sampled” the digital TCI grid with an inordinately large number of random points (clearly well

beyond any logistically feasible field sampling effort). We computed a correlogram from these data, and used this as a reference case or benchmark. We then explored a variety of sampling designs, with the goal of reproducing the reference case with a logistically feasible number of sample points. To compare two correlograms (trial and reference case), we computed a distance-weighted sum-of-squared-differences for values of the two correlograms:

$$c = \frac{1}{n} \sum_{i=1}^n \frac{(r_i - t_i)^2}{\ln(d_i)} \quad [4]$$

where r_i and t_i are the autocorrelation values at distance class i for the reference and trial correlogram, d_i is the actual distance for class i , and there are n distance classes. The inverse (natural log) weighting of distance reflects our concern that the close distances are much more important ecologically than far distances (at which the autocorrelation often tends toward 0 and nonsignificance).

In our tests, we had the best success with clusters of samples stratified regularly over the study area (figure 1). In this case, the between-sample distances within clusters capture the fine-scale variability in the data, while the between-cluster spacing captures larger scales of variability and provides for thorough coverage of the study area. The exact result will clearly vary for different variables, as indeed it also varies when the exercise is conducted over different spatial extents. But for any nontrivial spatial extent, it seems clear that the most effective sampling design almost certainly will be a multi-scale design such as stratified clusters or a nonaligned block design. In particular, random or stratified-random designs popularly used in ecological field studies are not as efficient because they require many more samples to capture fine-scale details. Regular (stratified) designs such as those laid out on a grid perform especially poorly in our trials using data with fine grain and large extent.

We should note that clustered sampling designs such as these have some implications for further statistical analysis. If the sampling indeed captures fine-grained pattern and spatial autocorrelation, the data cannot then be analyzed using conventional parametric statistics because the assumption of independence among samples will have been violated. As we recommend (below) nonparametric analyses based on randomization tests, this need not be a limitation. Alternative approaches for multivariate analysis of variance (MANOVA) and regression also are becoming more available (Anderson, 2001a, b).

Stage 2: Constructing and Dissecting a Preliminary Model

In the Sierra study, we used a sampling design of clusters of 3 or 4 plots within a 100-m radius, with the clusters themselves separated by hundreds of meters along transects aligned on elevation gradients. Because of the logistics of access from roads and trails, the actual arrangement of clusters appears rather haphazard (figure 2). Further, because of the huge area of the Park, we opted to focus on a smaller region within the 50,000-ha Kaweah Basin, one of three large basins comprising Sequoia-Kings Canyon National Park. This region includes a large infrastructure of previous or on-going studies, and our focus there thus piggy-backs on and complements these studies.

Sampling was conducted over the summers of 1997-1999, resulting in 99 sample quadrats.

Each sample is a 20x20-m plot in which all trees (> 2.5 cm diameter at breast height, dbh) were tallied by species and dbh. A total of 17 tree species were recorded; 11 common species were retained for analysis (table 1). A variety of environmental variables were recorded on each plot, including topographic indices and soils variables processed in the lab from three composite soil samples taken from the top 20 cm in the field (table 2).

Distance matrices were computed from these data by relativizing species data by species maximum, and then computing the Bray-Curtis dissimilarity coefficient (eq. 1). Environmental variables were standardized and summarized in terms of Euclidean distance. Simple Pearson correlation analysis led us to discard several redundant variables (retained variables are marked by asterisks in table 2). Geographic locations of plots were recorded in the field using global positioning satellites (GPS) and geographic separation was computed using Euclidean distance (eq. 2).

Step 2.1: Building a Preliminary Model

NMS ordinations reveal the striking elevation gradient in the forest (figure 3). A three-axis solution performed best, with most of the variation in species composition expressed on axes 3 and 1 respectively (numbering of NMS axes is arbitrary because all axes are computed simultaneously). Several cluster groups are readily interpretable in terms of this gradient, although three distinct cluster groups appear at lower elevations. A mixed-conifer cluster keyed by white fir dominates the lower and middle elevations, which is distinguishable from sites characterized by Jeffrey pine (on axis 1) and a third mid-elevation group keyed by Ponderosa pine (expressed on the second ordination axis, not shown). Elevation itself is strongly correlated with the third ordination axis, along with soil chemistry and depth. No environmental variables are strongly correlated with the other two ordination axes. Thus, elevation is the overriding correlate of vegetation pattern in this first approximation.

Mantel tests with these preliminary data echo the ordination results. There is a strong simple Mantel correlation between compositional dissimilarity and environmental dissimilarity, and also a strong spatial structure in composition as well as in the environmental variables (table 3). As partials, the correlation between species and space controlling for environment remains strong, while the partial on species and environment controlling for space is rather weak (all correlations are significant at $P < 0.001$, based on 10,000 permutations).

More resolution can be gained by separating the environmental dissimilarity matrix into its constituent variables, and computing dissimilarity matrices for each variable separately. This leads to a partial Mantel analysis using 12 dissimilarity matrices, each variable expressed in terms of absolute difference (the univariate form of Euclidean distance). At this resolution, elevation, transformed aspect, and soil depth emerge as important correlates of compositional dissimilarity; the partial (residual) on geographic distance is also significant (table 4).

These results are not particularly satisfying, for at least two reasons. First, we already knew that the elevation gradient was important (Vankat, 1982; Stephenson, 1990, 1998) and so this has been an extremely expensive confirmation of prior knowledge. Secondly, the variables that emerge as important are not very satisfying ecologically. Elevation itself means nothing to plants except that it proxies for important variables such as temperature, precipitation, and other variables. For a global change research program, a correlation with elevation is useless because

under greenhouse scenarios temperature and precipitation might change but elevation will not. Thus, we need to discover the actual ecological variables of plant response, rather than their correlation with elevation. Likewise, a correlation with geographic distance (space) is uninteresting except that it points to two other possibilities. Space can act as a proxy for a purely spatial process such as seed dispersal or a spatially contagious disturbance (fire, in this system). Alternatively, space might denote an ecologically important but unmeasured environmental variable that is autocorrelated. Thus, given our goals, the immediate tasks are to (1) replace elevation and space as explanatory variables, and (2) ensure that the correlations derived from the preliminary model are robust (i.e., that they are not mere artifacts of this particular sample).

Step 2.2: Targeting Areas for Model Refinement

Ordination and classification provide an easy framework within which to pursue a more refined understanding of species-environment relationships. The correlation vectors in ordination space suggest environmental variables that might be important vectors of species response. Partial Mantel tests confirm these correlations while isolating the best correlates of species response. This yields a parsimonious subset of environmental variables that we can expect to serve as useful predictors of plant response. At the same time, the classification analysis identifies discrete community types that lend themselves to a complementary predictive analysis, using classification and regression trees.

Classification and regression tree (CART) analysis is a recursive partitioning of a dataset with the goal of minimizing the local heterogeneity of the data (Breiman et al., 1984; Moore et al., 1991; Vassières et al., 2000). Applied to a categorical response variable (here, community types) the goal is to partition the samples into groups such that each type is reached as a node of a tree-like diagram. At each branch of the partitioning, the environmental variables are examined to find the best variable—and the best split of rank values on that variable—to separate the groups. A given group (e.g., white fir forest) may appear on multiple branches of the tree, and each branch thus indicates alternative (possibly compensatory) environmental conditions under which that group occurs. For our purposes, CART offers two important features. First, it is nonparametric and recursive and thus can handle a variety of ecological situations that are unwieldy using linear classification models (see Vassières et al. 2000 for a comparison to other models). Second, the CART solution can be implemented readily into geographic information systems (GIS) so that the model predictions can be displayed geographically (Moore et al., 1991, and see below).

There are two sorts of observations that are especially interesting from the perspective of trying to force an explanatory model. By “force” we mean to attack the model to expose its weakest points, so that we can improve it efficiently by focusing on these points. Species-environment correlations that appear as correlation vectors in ordination space, and especially as significant partials in Mantel tests, are prime candidates for further pursuit because we wish to verify that these correlations are robust and predictive. Likewise, the predictor variables selected by CART analyses to distinguish groups should also be verified with further testing. In particular, variables that separate pairs of groups in the CART tree identify potential ecotones, which should be especially informative about gradient response.

In the Sierran case study, CART analysis of seven compositional groups identified by clustering (color coded in figure 3) offers two sorts of insights into vegetation pattern (figure 4). The variables selected as predictors are the same as those obtained as significant (partial)

correlations using Mantel tests, which is comforting. More to the point, the CART tree identifies a few potential ecotones for further analysis. In particular, the analysis suggests that at high elevations, western white pine and lodgepole pine stands can be distinguished largely on the basis of slope aspect. As CART models can tend to be statistically over-fitted to the data (a result of the recursion), it is appropriate to ask whether this result is real or an artifact of this dataset.

Step 2.3: Translating Parametric Explanations to Geographic Predictions

The CART predictions about high-elevation forest types are mapped in figure 5. In this, all grid cells that satisfy the model conditions for western white pine are coded as that type, and sites predicted to support lodgepole are coded to contrast with these. This represents a translation of the CART model from parameter space into geographic space (Urban, 2000). Given this map, it is rather easy to choose new sampling locations that are logistically accessible and that will test the predictions of the CART model.

Likewise, the location and environmental settings of the red fir zone can be mapped in much the same way. In figure 6, the mapped colors denote sites predicted to be red fir sites and that are currently mapped that way by the Park's (preliminary and imperfect) vegetation map, sites mapped by the Park as red fir but predicted by CART to be something else, and sites predicted as red fir by CART but mapped as some other type by the Park. Again, it is straightforward to choose a set of new sample locations that will clarify these uncertainties.

Stage 3: Follow-up Sampling

The mapping exercises outlined above translate parametric domains into geographic space and thus delineate a set of candidate sites for follow-up field studies. These can be further stratified using a GIS to provide an efficient set of new locations to provide maximum leverage in improving the model. For example, one might use distance from roads or trails to further restrict the sampling domain to minimize travel costs and time (Urban, 2000). In practice, this second stage of sampling would entail a wider range of field conditions than those illustrated above, because there are several uncertainties in the model we have not discussed here. Also, to be conservative, follow-up sampling might be extended to cases not covered explicitly in the preliminary model, to minimize the likelihood of discovering still more artifactual results or uncertainties in the next model approximation.

New data collected in the second-stage sampling can then be incorporated into the statistical model, which should confirm some relationships while revising others that were less robust. At this point, the process can iterate to the second stage, of model dissection and refinement. This iteration continues until the model is confirmed and its uncertainties resolved.

Discussion

We have outlined an iterative approach to community ecology conducted at the landscape scale, in which a final explanatory model for species-environment relationships and vegetation pattern is obtained by successive approximation. The three stages of this approach involve a preliminary sampling design phase, which to some extent can be conducted in the lab from terrain-based environmental proxies. Data collected with this design are then used to construct a

tentative, preliminary model of species-environment relationships. This model is then dissected to discover its weakest points, to isolate uncertain parametric conditions for tactical field testing. A key part of this dissection is the translation of model predictions from parameter space to geographic space. A follow-up stage of new field studies at these key locations provide an efficient new set of data to help refine the model. This process can then iterate until further dissection and testing confirm that the model predictions are robust.

Two key issues remain. The first of these concerns field studies and the nature of observations needed to refine an explanatory model. The second issue concerns statistical methods and their power to reveal the ecological relationships that underpin correlations in multivariate, spatially structured ecological data.

Explanatory Field Studies

Correlations between species abundance and environmental variables are at best indicative of species response, rather than causal. Thus, new kinds of field data are likely to be required to fully refine an explanatory model based initially on measured species abundances. Further, in montane systems correlations between elevation and most environmental variables make it difficult to isolate the effective, plant-relevant factors that govern species response. While somewhat beyond the scope of this paper, it is worth digressing somewhat to address these empirical issues because some of the techniques outlined above can also be used to help establish new kinds of field studies.

Gradient Focus Plots. A predictive model of species response to an environmental gradient would require some understanding of the actual demographic processes involved (establishment, growth, mortality) in the context of competition with other species along the gradient (Smith & Huston, 1989, Austin & Smith, 1989; ter Braak & Prentice, 1988). At the landscape scale, this implies a daunting field effort because there may be several variables associated with the gradient complex, arrayed over a large spatial extent. One way to constrain this logistical challenge is to use the parametric-geographic mapping approach outlined above to locate short and well-defined gradients for focused study.

In our preliminary efforts in this direction, we have established short but steep gradients spanning significant environmental change over distances of ~100 m (actual distances vary according to field settings). Examples include topographic contrasts along hillslopes or contrasting slope aspects, or transects that span locations predicted to cross species-compositional boundaries (ecotones). Active gradient response would require trends in seedling establishment success, growth, or survivorship along the gradient focus plot—patterns that can be measured directly in the field. While data collected at this scale are noisy and may be somewhat limited by small sample sizes, they offer the advantage that the environmental variables and demographic processes can be measured directly and thoroughly rather than inferred from larger-scale correlations. This approach is similar in design to conventional stem maps used in demographic studies, but with two important distinctions. First, we establish the focus plots in areas of obvious environmental heterogeneity (i.e., gradients), while stem maps are typically established in more homogeneous settings. Secondly, our sampling is destructive (we harvest seedlings and saplings to age them and estimate growth rates), an approach incompatible with long-term monitoring studies often conducted on stem maps.

Environmental Settings. A second empirical issue is due to the overwhelming statistical influence of elevation in montane gradients. For example, temperature decreases while precipitation increases with increasing elevation, so it is difficult to know whether temperature or soil moisture is the more crucial constraint on plant species response. Beyond this, many soil variables also covary with elevation (pH and texture, in our system), further complicating the correlations. We have initiated a new field exercise to resolve these uncertainties. In this, we establish sampling locations that provide for a strong and local contrast of topographic settings, such as contrasting aspects of hills or valleys. At these locations, we deploy remote temperature dataloggers or use portable soil moisture meters (reflectometers). The temperature loggers can record hourly temperatures for 2-4 years on a single battery, and their low cost makes it feasible to deploy dozens of these in clustered settings across a landscape. Similarly, the moisture probes can provide synoptic measurements of soil moisture across gradients, as long as the sampling locations are readily accessible in nearly adjacent but edaphically contrasting sites. In each case, the parametric mapping exercise outlined above can be used to locate candidate sites for these samples.

Preliminary results from montane landscapes suggest that temperature can be predicted at much higher precision and finer spatial resolution using this approach, as compared to predictions regressed from elevation using lapse rates (T. Lookingbill & K. Pierce, unpublished data). We anticipate similarly useful results using potential radiation proxies and topographic convergence indices to interpolate soil moisture over complex terrain. Thus, using these targeted field measurements we can separate temperature and moisture from elevation, and focus more usefully on variables important to plant demographic response.

Refined Explanatory Models

One goal of the iterative model approximation is to replace elevation and “space” as explanatory variables, in favor of variables more closely coupled to plant demography. The field studies outlined above offer some promise to provide better resolution of the environmental correlates of elevation. The role of “space” is somewhat more vexing. As we noted, a significant partial (residual) correlation between species distribution and space implies either that a purely spatial process is operating (seed dispersal, fire) or that we neglected to measure some important environmental variable(s). In both cases, the statistical leverage we would like to apply is not provided by available methods. While a possible solution is similar in both cases, the solutions we seek arise in somewhat different ways, and it will be helpful to explore these separately.

Consider the goal of demonstrating that there is an important effect of local seed dispersal governing forest pattern. By this we mean that local seed sources influence species composition beyond the local influences of the physical environment. In a Mantel framework, this would require a significant partial residual on “space” after controlling for environmental variables. But we know that seed dispersal occurs over distances of tens of meters in this system (Clark et al., 1999), and so this partial Mantel result cannot distinguish seed dispersal from some other spatial residual (say, a shared fire history that spans hundreds of meters distance). The analysis that comes to mind is a partial Mantel test restricted to specified (short) distance classes, in effect a partial Mantel correlogram.

In the second case of some unmeasured environmental variable(s), the spatial residual averaged over all distances is again not very informative. Intuition would lead us to expect that an

environmental variable would show its influence at its own characteristic spatial scale; for example, topographic constraint should be reflected in vegetation pattern at similar topographic scales. Again, this leads us to a test in which we seek a significant partial Mantel correlation, controlling for all other measured variables—but we would like to know the specific distances over which this residual is expressed. Given this, we could make some inference (or guess) about the environmental variable underpinning the residual, or at least we could focus further field studies at the appropriate scales. Again, the implied test is a partial Mantel correlogram, with the caveat that we do not know in advance which scale(s) might be important. To our knowledge, this test—a scale-specific partial Mantel correlation—has not yet been devised. While we are currently exploring variations on this theme, we confess some uncertainty about how to construct the matrix permutation tests required for a significance test (see Legendre, 2000, for an in-depth discussion of the complexities of matrix permutation tests).

In the Sierran case study, we regressed the third ordination axis on elevation and found a strong relationship ($r^2=0.72$, $P<0.0001$). The residuals from this regression are autocorrelated at three distinct spatial lags (<150 m, ~500 m, and ~1250 m). These residuals in turn are correlated with transformed aspect ($r=-0.24$) and various soil chemistry variables (pH, base saturation, $r\sim -0.2$ in both cases). But we have no further inferential leverage to determine how these variables might contribute to these spatial residuals.

Alternative Statistical Models

It is worth digressing at this point to distinguish the distance-based approach illustrated here from other popular approaches, especially canonical correspondence analysis (CCA; ter Braak, 1987; available as CANOCO; ter Braak, 1988; Ter Braak & Smilauer, 1998) or redundancy analysis (RDA; Borcard et al., 1992; Borcard & Legendre, 1994). See Legendre & Legendre (1998) for a in-depth review of these and related techniques. Both CCA and RDA amount to regressing a multivariate species dataset on a multivariate environmental dataset, and in each case spatial locations (e.g., UTM coordinates) can be included as explanatory variables. The two techniques differ in that CCA assumes a nonlinear, unimodal model of species response to environmental gradients, while RDA assumes a linear response model. A newer analysis, distance-based RDA (db-RDA; Legendre & Anderson, 1999; McArdle & Anderson, 2001) first computes principal coordinates on the species data (a distance-based ordination) and then performs RDA on the ordination scores—an approach similar conceptually (but not equivalent computationally) to regressing NMS ordination scores on environmental variables. While these methods are statistically compelling, the perplexing issue remains that the incorporation of “space” into these models is difficult to interpret. In a Mantel test, “space” is equivalent to “distance apart” and this is consistent with the interpretation of local spatial structure due to autocorrelated environmental variable(s) or a contagious spatial process (including disturbance history). By contrast, in CCA and RDA “space” is equivalent to “location” and this is a different concept ecologically. As an example, consider a case where vegetation pattern is controlled by a strong directional gradient (say, elevation) in addition to a local spatial process (say, seed dispersal). As an explanatory variable, “distance apart” would capture the dispersal process, while “location” might not. In fact, for the residuals from elevation (i.e., variation in species composition *not* predicted by elevation) to be correlated with “location”, the residuals would have to either increase or decrease monotonically with location (say, longitude or UTM easting). But we would expect seed dispersal to act similarly over the elevation gradient, an expectation

consistent with “distance apart” as an explanatory variable. Similar complications arise in the case of an unmeasured environmental variable (consider a strong slope/aspect effect in the example above): it is the distance range of autocorrelation in the residuals that would suggest the scale of the unmeasured variable, and this pattern is more readily interpreted in terms of “distance apart” than in terms of location.

This distinction about “space” can be viewed in terms of the conventional equation for a regionalized variable z measured at location i :

$$z(i) = f(i) + s(i) + \varepsilon \quad [5]$$

where f is a forcing or trend, s is local spatial structure, and ε is error. We would argue that “location” tends to capture trend, while “distance apart” might better capture local structure. Trends typically are fitted using polynomial trend surface regressions (Legendre & Fortin, 1989; Legendre & Legendre 1998) or smoothing operators such as splines or lowess (Cleveland 1979; Chambers et al., 1983). Local structure is fitted using autocorrelation or variogram analysis (Journel & Huijbreghts, 1978; Isaaks & Srivastava 1989; Cressie 1993; Legendre & Legendre, 1998). Only universal kriging attempts to model both components simultaneously, via a combination of trend surface regression and variogram fitting (ver Hoef, 1993).

In truth, none of the techniques currently available to ecologists does exactly what we would like, and none offers an unambiguous rendering of the pure effects and scaling of environmental factors as these govern species distributional patterns (Griffith, 1992). As in similar instances in community ecology, a combination of techniques is likely to prove more informative than any single technique—witness the accepted practice of combining ordination and classification in community ecology. In any case, a useful step in this analysis would be to map the residuals into geographic space, in a manner similar to the examples presented here. For example, we might predict MNS ordination scores on elevation, and map the residuals into geographic space. This would result to a recoloring of the points in figure 2, such that each sample plot was colored to indicate the deviation between its ordination score (on axes 3 from figure 3, the “elevation” axis) and this score regressed on elevation. Spatial trends or autocorrelation in these residuals would then suggest new factors to pursue in follow-up field studies.

Conclusions and Prospectus

Our goal in the Sierran study has been to arrive, by successive approximation, at a statistical model framed in terms of explanatory variables that are relevant to plant demographic processes, and which is sufficiently robust that it can be used to explore climate-change scenarios. This requires that we move beyond correlations with elevation, and that we capture relevant spatial processes. Logistical constraints and uncertainties mandated an iterative approach to this goal, a luxury not affordable for many short-term studies but crucial to this application.

Statistical approaches other than those we employ might be devised, perhaps by taking advantage of techniques used in other disciplines. The important point is that new techniques will not be devised until ecologists demand them, and we will discover or invent new techniques as we scrutinize our data in ways that ask more tactically focused questions of our field data. An

important component of this process is the willingness to push our empirical models until they fail, so that we can refine them in an efficient and effective manner.

In a sense, the approach we illustrate is not different than the approach long recommended for simulation models (Mankin et al., 1975; Gardner & Urban, *in press*). A difference is that the static, descriptive models typical of community ecology are rarely exercised in a more predictive manner; clearly this type of application warrants a mode of testing more like that used for simulators. Indeed, in our Sierra study the field data and empirical models are being used to develop and parameterize simulators (Urban et al., 2000; Urban, 2000). In the case of either type of model, the practice of translating strategically between the parameter space of models and the geographic space of the study area is fundamental to the integration of models and data.

Acknowledgements

This research was supported in part by the US Geological Survey (Biological Resources Division Contract CA8800-1-9004) and by the National Science Foundation (grant IBN 96-52656). We appreciate helpful comments by Bruce McCune and M-J. Fortin.

Literature Cited

- Anderson, M.J., 2001a. A new method of for non-parametric multivariate analysis of variance. *Australian Ecology*, 26:32-46.
- Anderson, M.J., 2001b. Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian J. of Fisheries and Aquatic Sciences*, 58:626-639.
- Austin, M.P., and T M. Smith, 1989. A new model for the continuum concept. *Vegetatio*, 83:35-47.
- Australian National University (ANU), 1998. DECODA: Database for ecological community data, version 2.05. ANUTech Pty. Ltd., Canberra.
- Beers, T. W., P. E. Dress, and L. C. Wensel, 1966. Aspect transformation in site productivity research. *J. Forestry*, 64:691-692.
- Borcard, D., and P. Legendre, 1994. Environmental control and spatial structure in ecological communities: an example using oribatid mites (Acari, Oribatei). *Environ. Ecol. Stats.*, 1:37-53.
- Borcard, D., P. Legendre, and P. Drapeau, 1992. Partialling out the spatial component of ecological variation. *Ecology*, 73:1045-1055.
- Bray, J.R., and J.T. Curtis, 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, 27:325-349.
- Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone, 1984. Classification and regression trees. Wadsworth and Brooks/Cole, Monterey, CA.
- Casgrain, P., and P. Legendre, 2001. The R package for multivariate and spatial analysis, version 4.0 d4 - User's manual. Département de sciences biologiques, Université de Montréal.
- Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey, 1983. Graphical methods for data analysis. Wadsworth, Pacific Grove, California.
- Clark, J., M. Silman, R. Kern, E. Macklin, and J. HilleRisLambers, 1999. Seed dispersal near and far: patterns across temperate and tropical forests. *Ecology*, 80:1475-1494.
- Clarke, K.R. 1993, Non-parametric multivariate analyses of changes in community structure. *Australian J. Ecol.*, 18:117-143.
- Cleveland, W.S. 1979, Robust locally-weighted regression and smoothing scatterplots. *J. Am. Statist. Assoc.*, 74:829-836.
- Cressie, N.A.C. 1983. Statistics for spatial data. Wiley and Sons, New York.

- Dutilleul, P., D. Sockwell, D. Frigon, and P. Legendre, 2000. The Mantel-Pearson paradox: statistical considerations and ecological implications. *J. Agricultural, Biological and Environmental Statistics*, 5:131-150.
- Faith, D.P., P.R. Minchin, and L. Belbin, 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio*, 69:57-68.
- Fortin, M.-J., P. Drapeau, and P. Legendre, 1989. Spatial autocorrelation and sampling design in plant ecology. *Vegetatio*, 83:209-222.
- Gardner, R.H., and D.L. Urban. Model validation and testing: past lessons, present concerns, future prospects. Chapter in proceedings of the 9th Cary Conference (*to appear*).
- Gauch, H.G., 1982. Multivariate analysis in community ecology. Cambridge University Press, Cambridge.
- Gauch, H.G., 1973. The relationship between sample similarity and ecological distance. *Ecology*, 54:618-622.
- Gower, J.C., 1985. Measures of similarity, dissimilarity, and distance. Pages 397-405 in S. Kotz and N.L. Johnson (eds.), Encyclopedia of statistical sciences vol. 5. Wiley, New York.
- Griffith, D.A., 1992. What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. *L'Espace géographique*, 3:265-280.
- Isaaks, E.H., and R.M. Srivastava, 1989. An introduction to applied geostatistics. Oxford Univ. Press, New York.
- Jackson, D.A., and K.M. Somers, 1989. Are probability estimates from the permutation model of Mantel's test stable? *Can. J. Zool.*, 67:766-769.
- Jongman, R.H.G., C.J.F. ter Braak, and O.F.R. van Tongeren, 1995. Data analysis in community and landscape ecology. Cambridge University Press, Cambridge.
- Journel, A.G., and C.J. Huijbregts, 1978. Mining geostatistics. Academic Press, London.
- Kruskal, J.B., 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29:1-27.
- Lance, G.N., and W.T. Williams, 1967. A general theory of classificatory sorting strategies. I. Hierarchical systems. *Comput. J.*, 9:373-380.
- Leduc, A., P. Drapeau, Y. Bergeron, and P. Legendre, 1992. Study of spatial components of forest cover using partial Mantel tests and path analysis. *J. Veg. Sci.*, 3:69-78.

- Legendre, P., 2000. Comparison of permutation methods for the partial correlation and partial Mantel tests. *J. Statist. Comput. Simul.*, 67:37-73.
- Legendre, P., and M.J. Anderson, 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.*, 69:1-24.
- Legendre, P., and M.-J. Fortin, 1989. Spatial pattern and ecological analysis. *Vegetatio*, 80:107-138.
- Legendre, P., and L. Legendre, 1998. Numerical ecology (2nd English edition). Elsevier, Amsterdam, The Netherlands.
- Mankin, J.B., R.V. O'Neill, H.H. Shugart, and B.W. Rust, 1975. The importance of validation in ecosystem analysis. Pages 63-71 in G.S. Innis (ed.), New directions in the analysis of ecological systems, part 1. Simulation Councils Proceedings Series, Simulation Councils, LaJolla, California.
- Manly, B.F.J., 1991a. Randomization and Monte Carlo methods in biology. Chapman and Hall, London.
- Manly, B.F.J., 1991b. RT: a program for randomization testing (version 1.02). Western Ecosystems Technology, Inc., Cheyenne, Wyoming.
- Manly, B.F.J., 1997. Randomization, bootstrap and Monte Carlo methods in biology (2nd edition). Chapman and Hall, London.
- Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27:209-220.
- McArdle, B.H., and M.J. Anderson, 2001. Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology*, 82:290-297.
- McCune, B., and M.J. Mefford, 1999. PC-ORD: Multivariate analysis of ecological data (version 4). MjM Software Design, Gleneden Beach, Oregon.
- Minchin, P.R., 1987. An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio*, 69:89-107.
- Moore, D.M., B.G. Lee, and S.M. Davey, 1991. A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. *Environ. Manage.*, 15:59-71.
- Moore, I.D., R.B. Gryson, and A.R. Ladson, 1990. Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrol. Processes*, 5:3-30.

- Oden, N.L., and R.R. Sokal, 1986. Directional autocorrelation: an extension of spatial correlograms to two dimensions. *Syst. Zool.*, 35:608-617.
- Pielou, E.C., 1984. The interpretation of ecological data: a primer on classification and ordination. Wiley & Sons, New York.
- Smith, T.M., and M.L. Huston, 1989. A theory of the spatial and temporal dynamics of plant communities. *Vegetatio*, 83:49-69.
- Smouse, P.E., J.C. Long, and R.R. Sokal, 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.*, 35:627-632.
- Sneath, P.H.A., and R.R. Sokal, 1973. Numerical taxonomy: the principles and practice of numerical classification. W.H. Freeman, San Francisco.
- Stephenson, N.L., 1990. Climatic controls on vegetation distribution: the role of the water balance. *Am. Nat.*, 135:649-670.
- Stephenson, N.L., 1998. Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. *J. Biogeography*, 25:855-870.
- Stephenson, N.L. and D.J. Parsons, 1993. A research program for predicting the effects of climatic change on the Sierra Nevada. Pages 93-109 in S.D. Veirs, Jr., T.J. Stohlgren and C. Schonewald-Cox (eds.), Proceedings of the Fourth Conference on Research in California's National Parks. U.S. Department of the Interior, National Park Service Transactions and Proceedings Series 9.
- ter Braak, C.J.F., 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67:1167-1179.
- ter Braak, C.J.F., 1987. The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio*, 69:69-77.
- ter Braak, C.J.F., 1988. CANOCO—a FORTRAN program for canonical community ordination by [partial] [detrended] [canonical] correspondence analysis, principal component analysis, and redundancy analysis (version 2.1). Agricultural Mathematics Group, Ministry of Agriculture and Fisheries, Wageningen.
- ter Braak, C.J.F., and I.C. Prentice, 1988. A theory of gradient analysis. *Adv. Ecol. Res.*, 18:271-313.
- ter Braak, C.J.F., and P. Smilauer, 1998. CANOCO release 4 reference manual and user's guide to Canoco for Windows - Software for canonical community ordination. Microcomputer Power, Ithaca, New York.

- Urban, D.L., 2000. Using model analysis to design monitoring programs for landscape management and impact assessment. *Ecol. Applic.*, 10:1820-1832.
- Urban, D.L., C. Miller, N.L. Stephenson, and P.N. Halpin, 2000. Forest pattern in Sierran landscapes: the physical template. *Landscape Ecol.*, 15:603-620.
- Vankat, J.L., 1982. A gradient perspective on the vegetation of Sequoia National Park. *Madroño*, 29:200-214.
- Vassières, M.P., R.E. Plant, and B.H. Allen-Diaz, 2000. Classification trees: an alternative nonparametric approach for predicting species distributions. *J. Veg. Sci.*, 11:679-694.
- ver Hoef, J., 1993. Universal kriging for ecological data. Pages 447-453 in M.F. Goodchild, B.O. Parks, and L.T. Steyaert (eds.), Environmental modeling with GIS. Oxford Univ. Press, New York.
- Whittaker, R.H., 1967. Gradient analysis of vegetation. *Biol. Rev.*, 42:207-264.
- Whittaker, R.H., 1978a. Ordination of plant communities. Junk, The Hague.
- Whittaker, R.H., 1978b. Classification of plant communities. Junk, The Hague.
- Wolock, D.M., and G.J. McCabe, 1995. Comparison of single and multiple flow direction algorithms for computing topographic parameters in TOPMODEL. *Water Resources Res.*, 31:1315-1324.

Tables

Table 1. Species names and mnemonic codes for tree species recorded in samples in Sequoia National Park. Species denoted with an asterisk were retained for analysis, based on frequency of occurrence and average abundance (as basal area).

| Code | Scientific Name | Common Name |
|---------------|---------------------------------|----------------------|
| ABCO * | <i>Abies concolor</i> | White fir |
| ABMA * | <i>Abies magnifica</i> | Red fir |
| ACMA | <i>Acer macrophyllum</i> | Bigleaf maple |
| ARVI | <i>Arctostaphylos viscida</i> | Whiteleaf manzanita |
| CADE * | <i>Calocedrus decurrens</i> | Incense cedar |
| CEIN | <i>Ceanothus integerrimus</i> | Deer brush |
| CONU * | <i>Cornus nuttallii</i> | Pacific dogwood |
| PICO * | <i>Pinus contorta</i> | Lodgepole pine |
| PIJE * | <i>Pinus jeffreyii</i> | Jeffrey pine |
| PILA * | <i>Pinus lambertiana</i> | Sugar pine |
| PIMO * | <i>Pinus monticola</i> | Western white pine |
| PIPO * | <i>Pinus ponderosa</i> | Ponderosa pine |
| QUCH | <i>Quercus chrysolepis</i> | Canyon live oak |
| QUKE * | <i>Quercus kelloggii</i> | California black oak |
| SEGI * | <i>Sequoiadendron giganteum</i> | Giant sequoia |
| TOCA | <i>Torreya californica</i> | California torreyia |
| UMCA | <i>Umbellularia californica</i> | California bay |

Table 2. Names and definitions of environmental variables measured on sample plots in Sequoia National Park. Variables marked with an asterisk were retained for analysis (discarded variables were highly correlated with other variables).

| Variable | Description | Note |
|--------------------|------------------------------------|---|
| Elevation * | Elevation (m) | from GPS |
| Slope * | Maximum slope (°) | field measurement |
| TAspect * | Transformed Aspect ¹ | -1 x COS(45-Aspect) |
| TSI * | Terrain Shape Index ² | <1: dome; >1: cove |
| xDepth * | Mean soil depth (cm) | to maximum of 100 cm; from 30 measurements |
| sDepth * | standard deviation of depth | |
| pH * | pH | |
| C * | Soil Carbon (%) | total carbon |
| N | Nitrogen (%) | total nitrogen |
| C:N | Carbon:Nitrogen ratio | |
| P* | Phosphorus (mg/g) | total exchangeable |
| Ca * | Calcium (cmol+)/kg) | |
| Mg * | Magnesium (cmol+)/kg) | |
| K * | Potassium (cmol+)/kg) | |
| Ac | Acidity (cmol+)/kg) | total exchangeable |
| ECEC * | Effective Cation Exchange Capacity | S(Ca,Mg,K,Ac) |
| BS | Base Saturation (%) | S(Ca,Mg,K)/ECEC |
| xLitter * | Mean Litter Depth (cm) | depth to mineral soil |
| Clay * | Clay (%) | particles < 2 mm |
| Silt | Silt (%) ³ | particles 2-50 mm |
| Sand | Sand (%) | particles >50 mm |

¹ Aspect transformation based on Beers (1966), modified so that SW-facing slopes have value 1.0 and NE slopes have value -1.0. SW is direction of maximum radiation load.

² TSI is the mean slope (°) from measurements taken in 4 cardinal directions from plot center (after McNab 1989).

³ Percent silt is computed as 100-(Clay+Sand).

Table 3. Simple and partial Mantel correlations between species compositional dissimilarity, environmental dissimilarity, and geographic distance. Elements in the upper triangle of the matrix are simple correlations, while the lower triangle holds partial correlations.

| | Species | Environment | Space |
|--------------------|----------------|--------------------|--------------|
| Species | | 0.29 | 0.51 |
| Environment | 0.12 | | 0.38 |
| Space | 0.45 | na | |

Table 4. Simple and partial Mantel correlations between species compositional dissimilarity, dissimilarity on individual environmental variables, and geographic distance. Mantel coefficients are simple correlations with species composition, correlations with space (autocorrelation), partials with species controlling for space, and partials with species controlling for all other variables (*P* values in parentheses).

| | Y * Species | Y * Space | Y * Spp all |
|-------------------|--------------------|------------------|----------------------|
| Elevation | 0.54 (0.001) | 0.90 (0.001) | 0.20 (0.001) |
| TAspect | ns | ns | 0.04 (0.020) |
| Soil Depth | 0.29 (0.001) | 0.32 (0.001) | 0.15 (0.001) |
| Space | 0.51 (0.001) | na | 0.06 (0.001) |

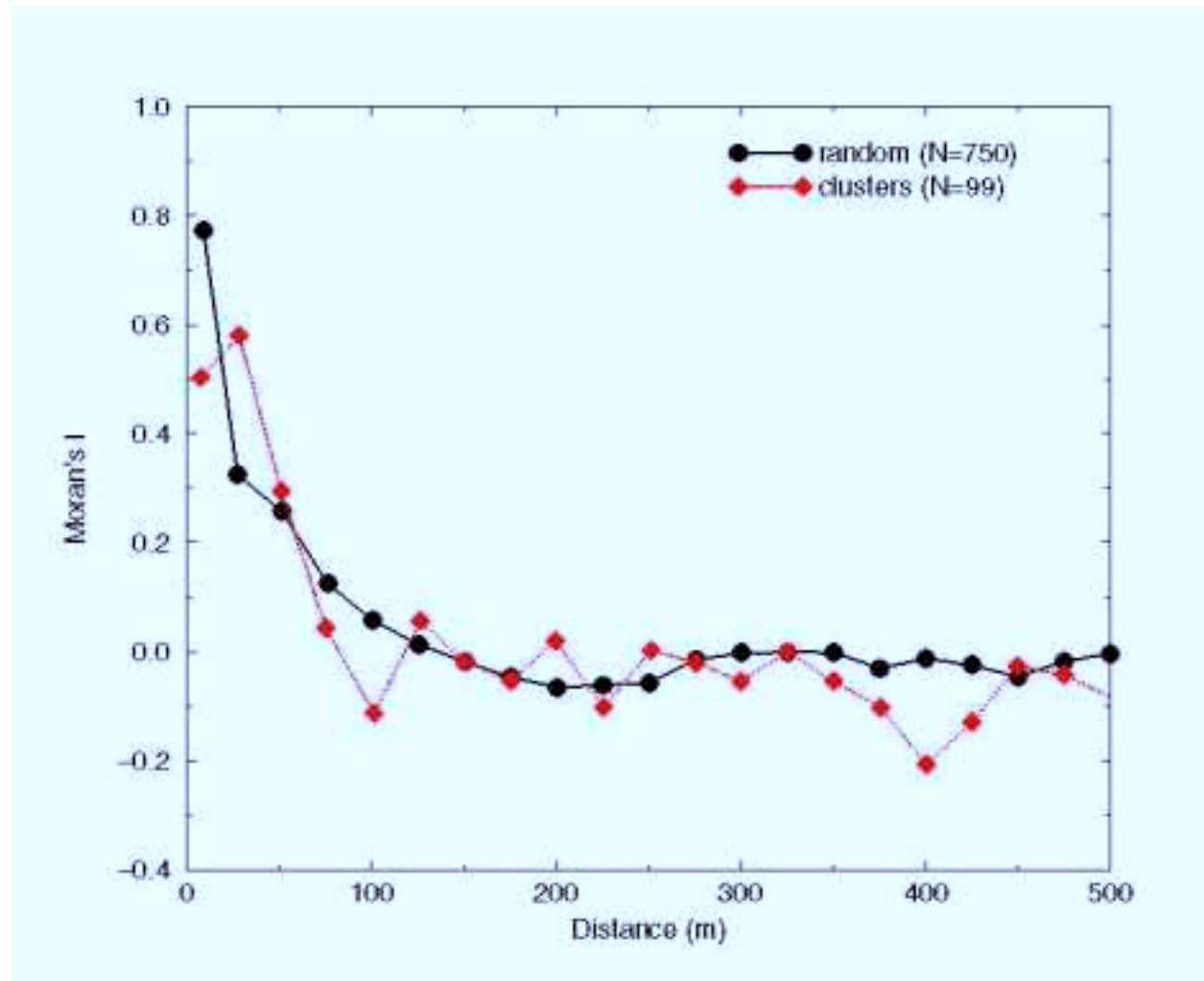


Figure 1. An example of a virtual sampling experiment in which a reference case based on a large number of samples is targeted by a sampling design constrained to comparatively few samples. The test is to reproduce the reference correlogram with as few samples as possible, by choosing an appropriate arrangement and intensity of sample points. This example is for a topographic convergence index, computed from a DEM of a small watershed within the study area in Sequoia National Park.

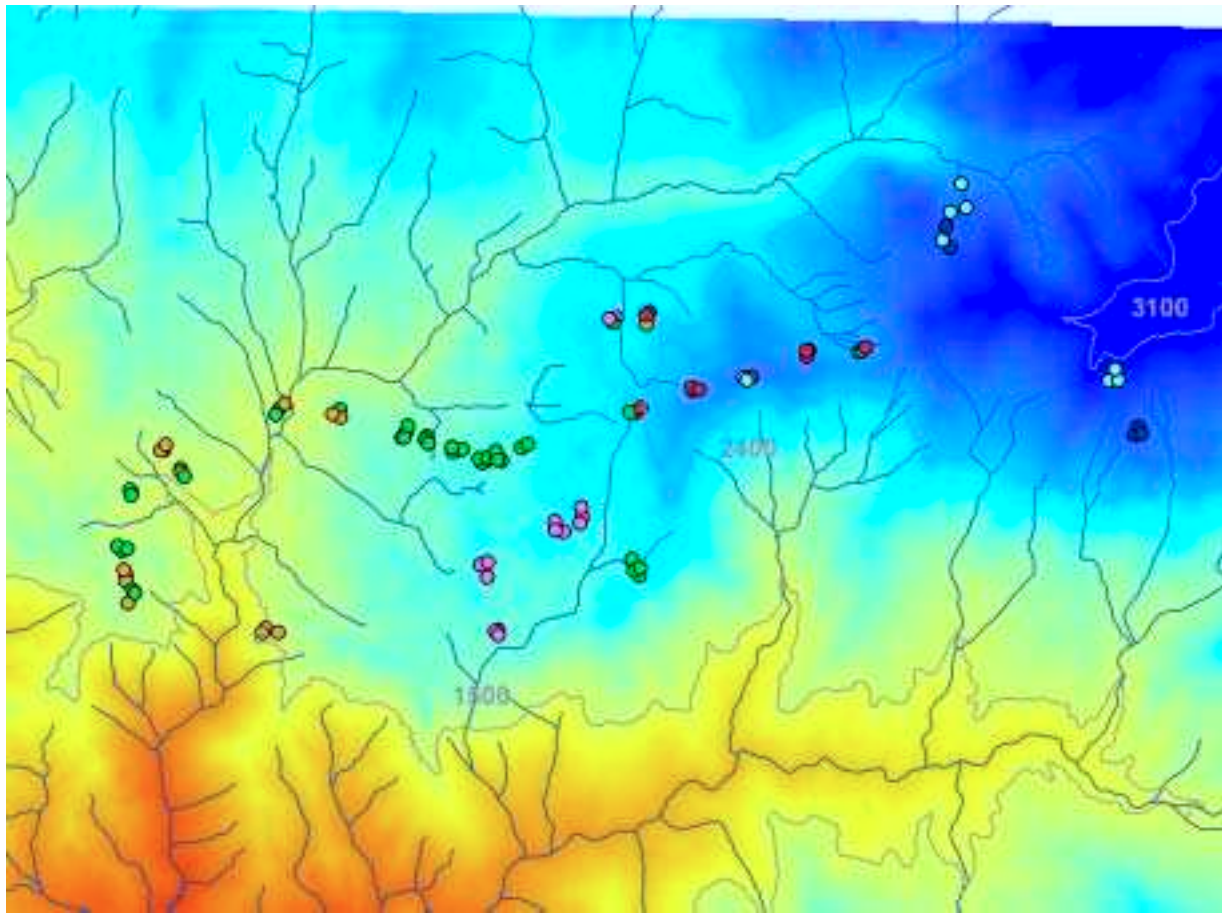


Figure 2. Arrangement of samples (point colors correspond to forest types in figure 3) within the study area, overlaid on digital elevation data (colored from red at lower elevations to blue at higher elevations). The study area is approximately 10 x 14.3 km.

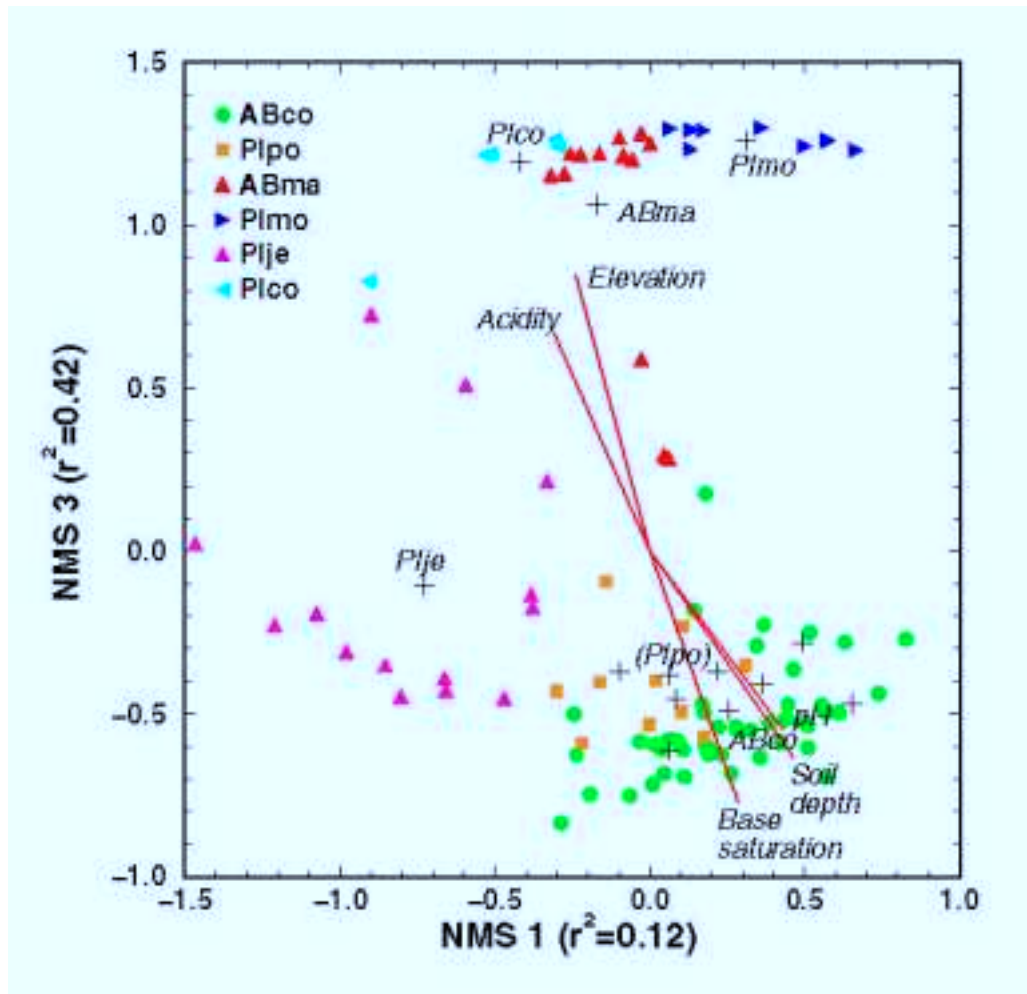


Figure 3. Ordination of the samples using nonmetric multidimensional scaling (NMS). Samples are color-code according to group membership derived by hierarchical clustering. Environmental variables are overlaid as correlation vectors.

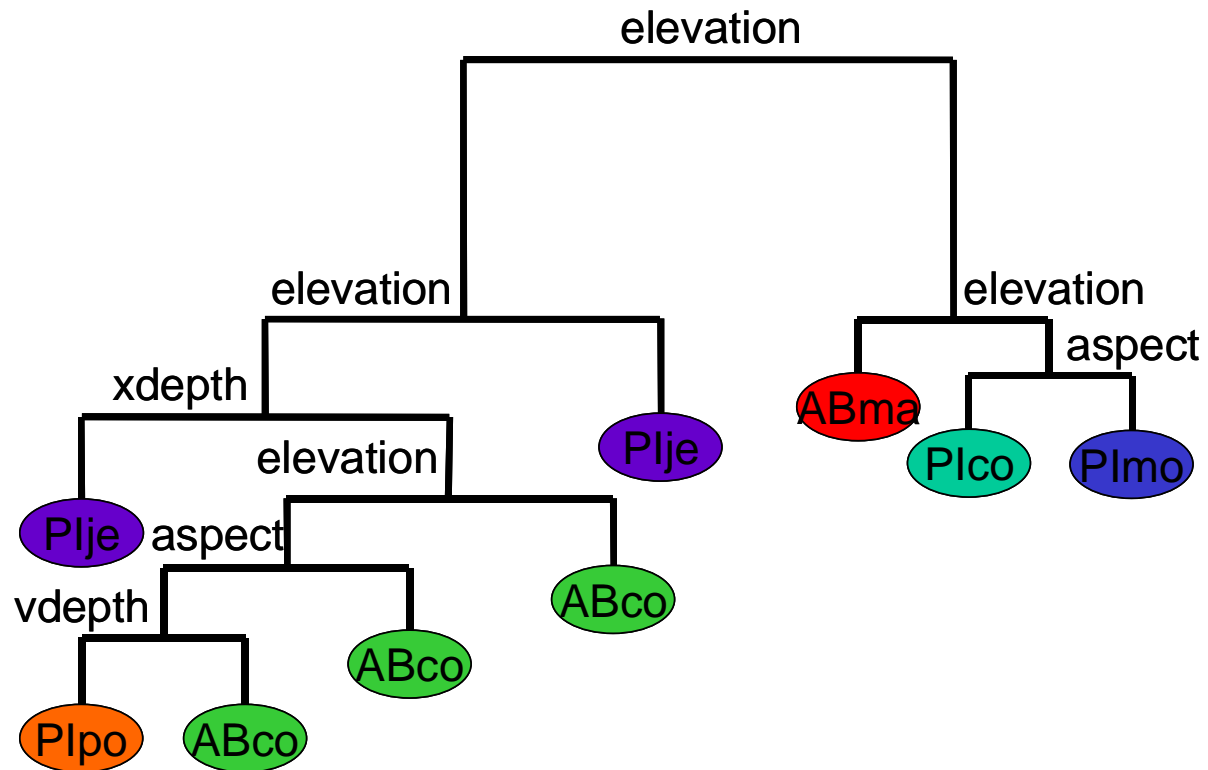


Figure 4. Partial classification tree for seven compositional forest types identified by cluster analysis, and predicted by environmental variables. The branch on the right, separating western white pine and lodgepole pine, is mapped in figure 5.

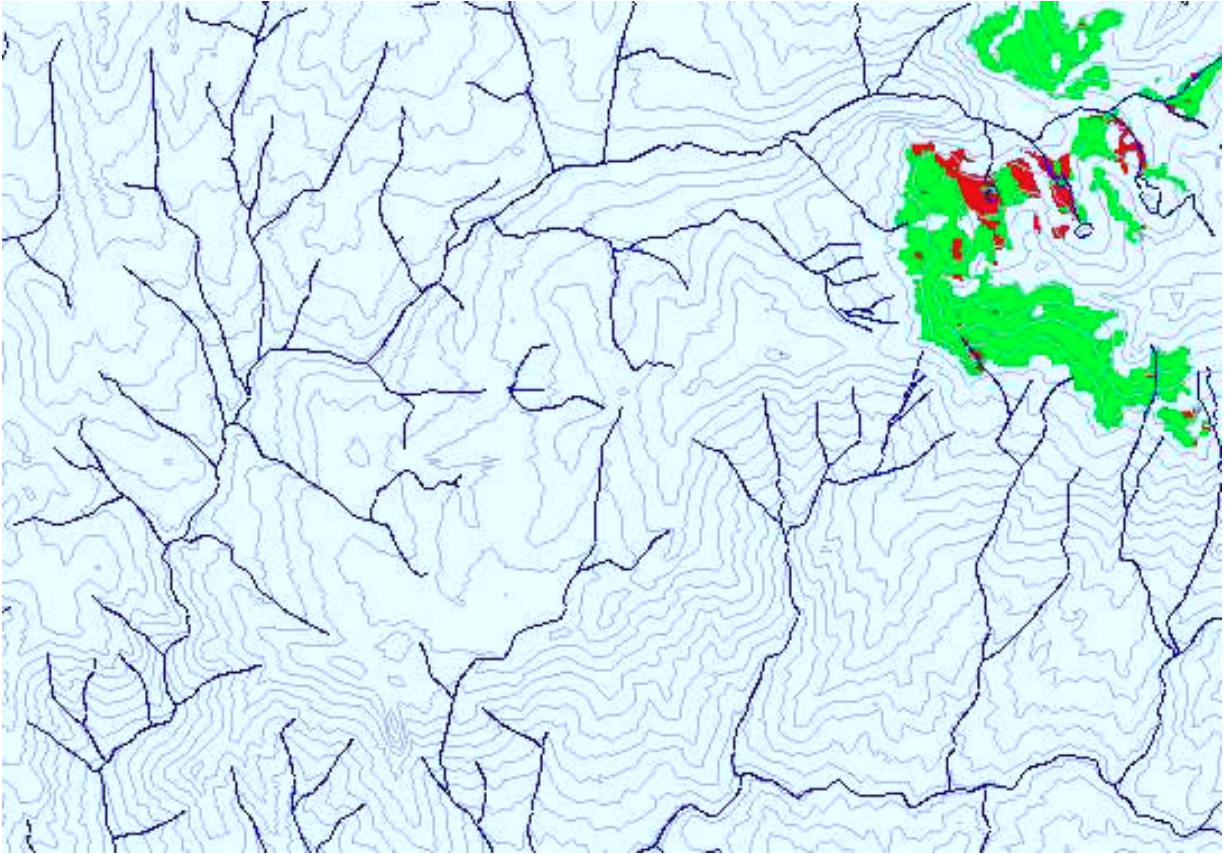


Figure 5. CART predictions of western white pine (green) and lodgepole pine (red) forest types, mapped into geographic space. Gray lines are 100-m contour intervals; streams are in blue. Geographically adjacent sites predicted to support different forest types are easy targets as sample locations to test the model.

figure 6

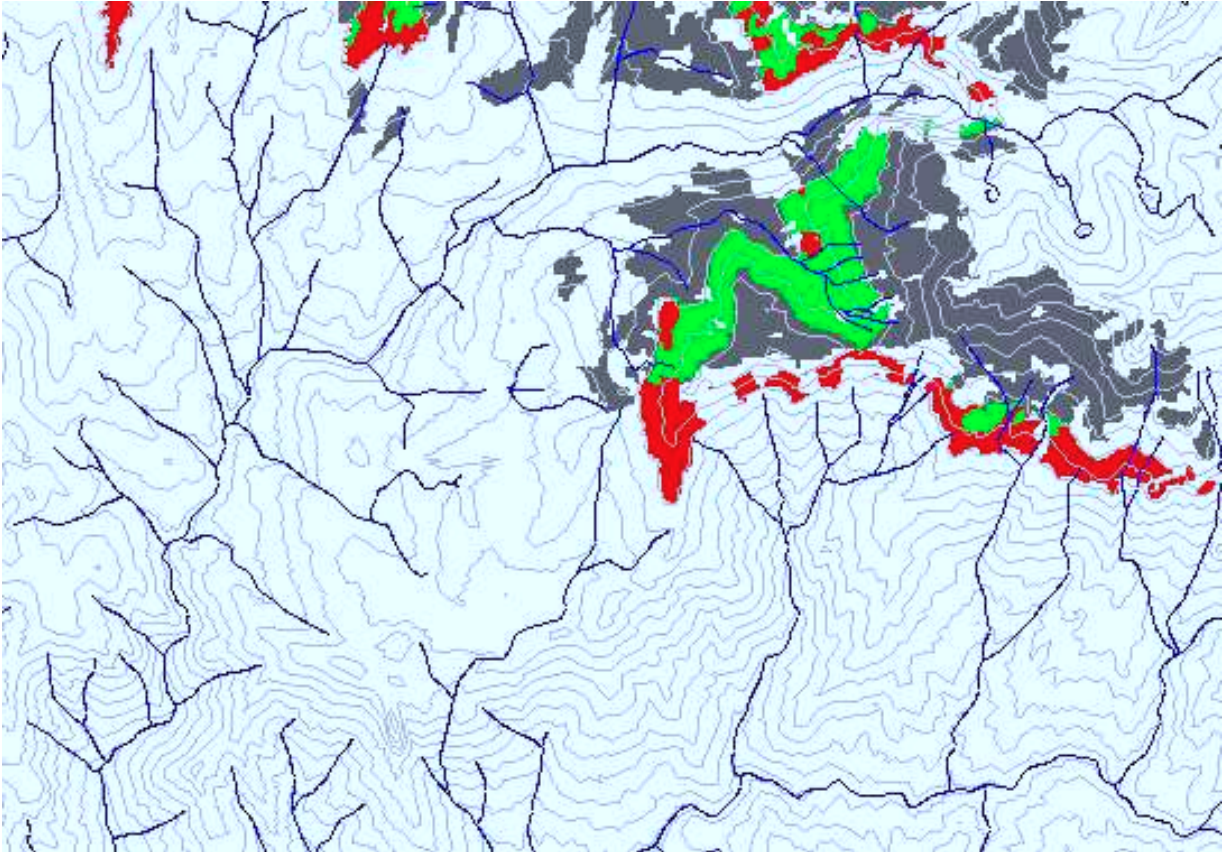


Figure 6. CART predictions of red fir forest type, compared to vegetation types in a preliminary map of the Park classified from satellite imagery and air photos. Green sites are predicted to be red fir by CART and currently classified that way in the Park map. Red sites are predicted by CART to be red fir, but mapped by the Park as some other vegetation type. Gray sites are mapped as red fir by the Park but not classified as such by CART. Gray lines are 100-m contour intervals; streams are in blue.