Gradient analysis, the next generation: towards more plant-relevant explanatory variables

Todd R. Lookingbill and Dean L. Urban

Abstract: The long history of gradient analysis is anchored in the observation that species turnover can be described along elevation gradients. This model is unsatisfying in that elevation is not directly relevant to plants and the ubiquitous "elevation gradient" is composed of multiple intertwined environmental factors. We offer an approach to landscape-scale vegetation analysis that disentangles the elevation gradient into its constituent parts through focused field sampling and statistical analysis. We illustrate the approach for an old-growth watershed in the Oregon Western Cascades. Our initial model of this system supports the common observation that forest community types are highly associated with specific elevation bands. By replacing elevation and other crude environmental proxy variables with estimates of more direct and resource gradients (radiation, temperature, and soil moisture), we create a vegetative model with stronger explanatory power than the proxy model in both cross-validation analysis and validation using an independent data set. The resulting model is also more biologically interpretable, which provides more meaningful insight into potential forest response to environmental change (e.g., global climate change scenarios). Acquiring a better mechanistic understanding of the relationship between plant communities and environmental predictor variables presents the next great challenge to community ecologists conducting gradient studies at landscape scales.

Résumé : L'analyse de gradient s'appuie depuis toujours sur le fait que le renouvellement des espèces peut être décrit en suivant des gradients d'altitude. Ce modèle n'est pas satisfaisant puisque l'altitude n'est pas directement pertinente pour les plantes et que le terme omniprésent « gradient d'altitude » est composé de multiples facteurs environnementaux entremêlés. Les auteurs proposent une approche d'analyse de la végétation à l'échelle du paysage qui décompose le gradient d'altitude en ses parties constituantes en ciblant l'échantillonnage de terrain et l'analyse statistique. Ils illustrent cette approche à partir d'un bassin versant composé de vieux peuplements de l'ouest des Cascades, en Oregon. Leur modèle initial de ce système est en accord avec l'observation courante que les types forestiers sont étroitement associés aux bandes spécifiques d'altitude. En remplaçant l'altitude et d'autres variables environnementales indirectes par des estimations de gradients de ressources plus directs (radiation, température et humidité du sol), ils créent un modèle de végétation ayant un pouvoir explicatif plus grand que celui du modèle indirect, ce qui a été vérifié dans les cas d'une analyse de validation interne et d'une validation à l'aide d'un fichier de données indépendantes. Ce modèle facilite aussi l'interprétation biologique des résultats, ce qui permet d'approfondir nos connaissances sur la réaction potentielle de la forêt en fonction de changements environnementaux (p. ex., des scénarios de changement global du climat). L'acquisition d'une meilleure compréhension fonctionnelle des relations entre les communautés végétales et les variables environnementales de prédiction constitue le prochain grand défi pour les écologistes des populations qui s'intéressent aux études de gradient à l'échelle du paysage.

[Traduit par la Rédaction]

Introduction

The Humboldtian tradition of gradient analysis in forest ecology involves the association of communities of co-occurring species with various descriptors of the environmental template. The results of these studies are often projected into an aspatial parameter space derived from environmental "proxy" variables. For example, Whittaker (1956, 1960, 1967, 1978) and a host of others (reviewed by Stephenson 1990) have arrayed species or vegetation types as distinct domains within an environmental setting defined by the principal axes of elevation (an estimator of temperature) and a variety of estimators of plant-available moisture (usually related to exposure).

Many practical applications of these models require their rendering in a spatial context (Franklin 1995). An example of this approach and its implications is found in the global change literature, where several studies have predicted the spatial displacement of vegetation communities along eleva-

T.R. Lookingbill[1,2] and D.L. Urban. Nicholas School of the Environment and Earth Sciences, Duke University, Durham, NC 27708, USA.

[1]Present address: University of Maryland, Center for Environmental Science-Appalachian Laboratory, 301 Braddock Road, Frostburg, MD 21532, USA.
[2]Corresponding author (e-mail: tlookingbill@al.umces.edu).

tion gradients in response to a warming climate (Peters and Darling 1985; Fujiware and Box 1999). In light of these predictions, some have argued for specific management responses, such as the design of networks of forest reserves that incorporate elevation and latitudinal gradients to allow for future species migration (Noss 2001).

In evaluating model predictions of species range shifts, it is important to bear in mind that although elevation may be strongly correlated with forest pattern, elevation differences are not responsible for the distribution of forest species. Trees do not respond directly to elevation, but rather to variables such as temperature and precipitation that covary with elevation. Relying on correlations with indirect gradients such as elevation to predict future vegetation patterns becomes complicated if some of the factors underlying the gradient change while others do not. For example, temperature may increase in response to rising atmospheric CO_2 levels, but precipitation (which is also a function of elevation) is likely to have a much more spatially variable response (IPCC 2001), and other environmental factors (e.g., soil properties, hillslope topography) are unlikely to change at all. Observed correlations between species distributions and elevation, therefore, may be insufficient to predict forest response to climate change (Halpin 1997; Urban et al. 2000).

Austin and Smith (1989) have long argued that indirect environmental proxy variables should be replaced in community ecology studies with variables representing resource and environmental gradients that have a direct physiological impact on plant biology. Yet the literature is replete with references to species associations with indirect environmental gradients that are merely correlative (Austin 1987; Hobbs 1997). Only a small handful of gradient analysis studies (e.g., Gagnon and Bradfield 1987; Allen et al. 1991; Ohmann and Spies 1998; Bunn et al. 2005) have attempted to quantify and relate direct and resource gradients such as temperature and moisture to plant community pattern. As a consequence, ecological theory over the past few decades has stressed the importance of temperature and moisture as independent factors in explaining forest pattern on one hand, while empirical studies continue to apply elevation-based inference of pattern on the other.

This incongruity can be explained by the great difficulty in describing the variability in direct environmental gradients at the landscape scale, the operational unit of most gradient analysis studies. Intensive fine-grain studies are able to capture complex patterns in temperature and soil moisture explicitly (e.g., Yeakley et al. 1998), while much of the detail averages away at regional to global scales (e.g., Neilson 1991). Conventional empirical methods are poorly suited, however, to gathering fine-resolution information on the distribution of these variables at extents of kilometres to tens of kilometres (Chen et al. 1999). To meet the growing demands of environmental managers, vegetation models must be generated at these landscape scales (Christensen et al. 1996); therefore, modelers are forced to rely on the ubiquitous elevation gradient to parameterize their models.

We offer an approach to landscape-scale vegetation analysis that disentangles the indirect elevation gradient into its constituent parts through focused field sampling and statistical modeling. We illustrate the approach with an examination of the spatial transition in community dominance from the *Tsuga*

heterophylla (Raf.) Sarg. (western hemlock) to *Abies amabilis* (Dougl. ex Loud.) Dougl. ex Forbes (Pacific silver fir) vegetation zones in the Oregon Western Cascades. Historically, this transition has been described along an altitudinal band between 1000 and 1200 m (Fonda and Bliss 1969; Franklin and Dyrness 1988; Franklin and Halpern 2000), but efforts to identify the causes of the transition have been minimal. The relatively few studies that have focused on describing mechanisms for this transition have differed in their emphasis of underlying control factors. Krajina (1969), for example, suggested that low drought tolerance limited the distribution of *A. amabilis* and described the transition from *T. heterophylla* to *A. amabilis* along a gradient of increasing precipitation. Thornburgh (1969) argued that the influence of snowpack on *T. heterophylla* seedlings was critical to the transition. Zobel et al. (1976) proposed that the *T. heterophylla* and *A. amabilis* vegetation zones could be differentiated by an index of temperature, which was used by Packee (1990) to suggest that *T. heterophylla* was physiologically limited at low temperatures.

The competing postulates have different implications under greenhouse warming scenarios. For example, an ecotone formed primarily by differences in physiological response to temperature may respond linearly to changes in temperature, while an ecotone maintained primarily by drought tolerances may react less predictably to temperature shifts. An improved understanding of the direct and resource gradients underlying this community transition is needed to assess the potential impacts of changes in climate or management. In this study, we develop a predictive model of the *T. heterophylla – A. amabilis* ecotone based on modeled environmental variables (temperature, soil moisture, and radiation) that have a direct physiological influence on plant processes. We compare this model to one based entirely on indirect topographic variables (elevation, topographic convergence, and aspect) and discuss the differences with respect to global warming scenarios.

Materials and methods

Study area

The H.J. Andrews Experimental Forest (HJA) is located 80 km east of Eugene, Oregon, on the west slope of the Cascade Mountains. At the time of its designation as an experimental watershed in 1948, the HJA was an intact forest with about 65% of the land in old-growth stands (i.e., 400–500 years old). Since that time, old-growth forest has been reduced to approximately 40% of the total area, and it has been adopted into the Long Term Ecological Research (LTER) network (Cissel et al. 1999). The 6400-ha LTER site ranges in elevation from 410 to 1630 m (McKee 1998).

Ohmann and Spies (1998) argued that large-scale climatic factors are more important than soil variability in controlling vegetation in this region. *Pseudotsuga menziesii* (Mirb.) Franco (Douglas-fir), *T. heterophylla*, and *Thuja plicata* Donn ex D. Don (western redcedar) are the dominant species at lower elevations, while *A. amabilis*, *Abies procera* Rehd. (noble fir), and *Tsuga mertensiana* (Bong.) Carr. (mountain hemlock) dominate upper elevations (Franklin and Dyrness 1988). Climate is characteristic of the Pacific Northwest, with dry summers and wet winters. Annual precipitation ranges from 2200 mm at the base of the catchment to 3400 mm at upper

Fig. 1. Locations of 175 sample plots in the H.J. Andrews Experimental Forest. Sampling of vegetation was conducted in old-growth stands, stratified across elevation, slope and aspect, and vegetation type. Shading indicates decreasing elevation with increasing shading. Elevation ranges from 410 to 1630 m. Filled circles indicate 11 sample plots used for model validation.
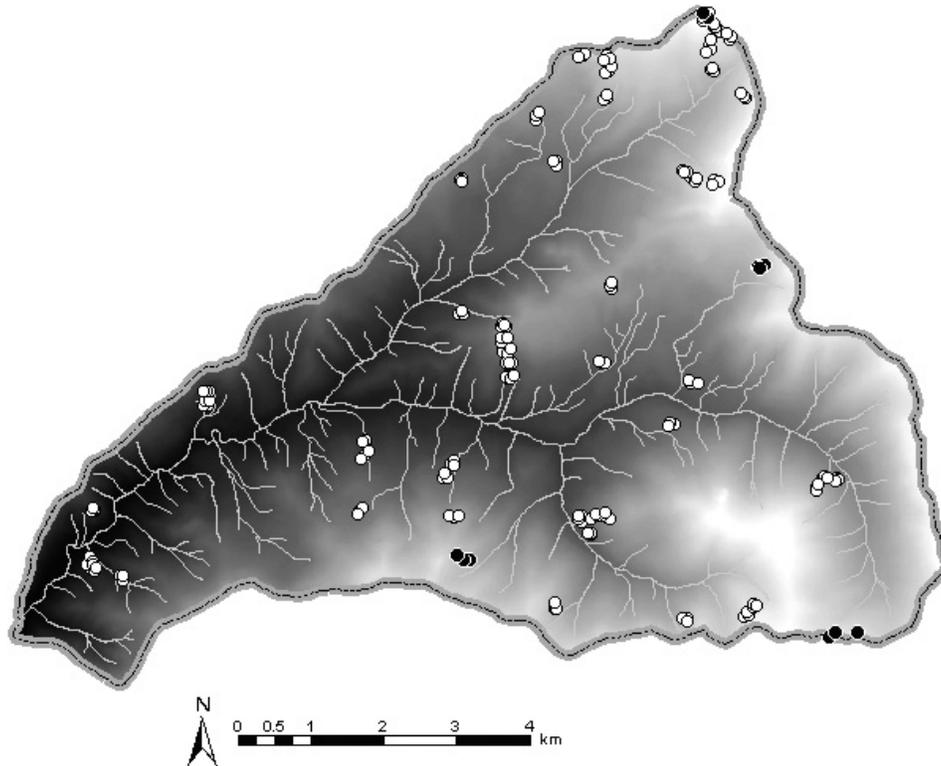
elevations, with less than one-tenth of total precipitation falling during the summer growing season (Grier and Logan 1977). Soils are mostly deep, well-drained Inceptisols derived primarily from colluvial and residual parent materials. Textures range from gravelly, silty clay loam to very gravelly clay loam. Rooting occurs almost entirely in the upper 200 cm of soil.

Data

We collected georeferenced data on vegetation and site characteristics at a total of 175 plots (Fig. 1). The initial analyses were conducted on 164 of these plots, which were sampled from 1997 to 1999. The remaining 11 plots were sampled in 2002 in areas deemed most difficult to classify by the models and were used primarily for model validation. All plots were located exclusively in old-growth forest communities using a stratified-cluster sampling design, whereby clusters of 20 m × 20 m (0.04 ha) plots were distributed along transects across the landscape. Stratified clusters have been shown to be more efficient at capturing fine-scaled pattern over large extents than either random or stratified random sampling (Urban et al. 2002). Clusters consisted of three or four plots located at random distances (<100 m) and random azimuths from cluster center points along the transects. Cluster center points were separated by 200–400 m. Transects were separated by hundreds to thousands of metres.

Samples were intended to represent a homogeneous slope facet, so the plots were oriented on the slope with the baseline of the plot running across the slope and the left edge of the plot aligned in the direction of maximum slope. Each tree with a diameter at breast height (dbh, 1.37 m) greater

than 2.5 cm was tallied and its dbh and species identity recorded. Tree diameters were converted to basal area (cross-sectional area at breast height), and basal area was summed over each species at each plot.

Physical proxy variables: aspect, elevation, and topographic convergence

Aspect, elevation, and topographic convergence (common proxies for radiation, temperature, and soil moisture) were used to construct the first of two alternative models describing the distribution of forest tree community types. The lower-left corner of each plot was staked and its position recorded with a global positioning system unit. The locational data were used to extract the proxy measures from a 10-m resolution digital elevation model. Aspect was transformed to a more direct measure of radiation load on a scale of –1.0 (northeast-facing slopes) to 1.0 (southwest-facing slopes) (TAspect = –cos(45 – Aspect); after Beers et al. 1966). A topographic convergence index (TCI; Beven and Kirkby 1979) was also derived from the digital elevation model to describe relative wetness. We will refer to these terrain-based indices as "physical proxies" for environmental variables.

Plant-relevant variables: radiation, temperature, and soil moisture

The second model was constructed using more direct estimates of the environmental factors that potentially influence the distribution of forest communities: temperature, radiation, and soil moisture. We will refer to these measures as "plant-relevant variables". The specific models used to cre-

ate spatially explicit estimates of these factors are briefly summarized below and described in greater detail elsewhere (Lookingbill and Urban 2003; Lookingbill and Urban 2004; Pierce et al. 2005).

We developed and used a simple geographic information system (GIS) algorithm to describe the spatial distribution of summer radiation within the catchment. This algorithm captures both local hillshading effects and the temporal variability in sun angle over the course of the day and year. In model comparison testing (Pierce et al. 2005), 1000 randomly selected points from a data-intensive, computationally rigorous radiation model of the HJA (Smith 2002) were more highly correlated with estimates from our GIS-based model ($r^2 = 0.59$) than with the proxy variable, transformed aspect ($r^2 = 0.20$).

We also used GIS data to locate stratified field samples of temperature and soil moisture and to build local regression equations for estimating the spatial distribution of these variables for the HJA watershed. The temperature model extrapolates relative differences in monthly mean temperature measurements, as recorded in July 2000, using a sampling network of 45 portable dataloggers. The dataloggers were stratified across major watershed units, elevation, aspect, and distance from stream. Likelihood ratio tests of competing nested models identified the most parsimonious statistical model. This model reduced the prediction error of an independent validation data set recorded in July 1999 ($N = 33$, the majority of which were located on vegetation plots included in this study) by over 30% when compared to a lapse rate model using only elevation to predict temperature (Lookingbill and Urban 2003).

We estimated soil moisture using a statistical description of 88 synoptic volumetric moisture measurements collected during a critical period of summer drought (4 July 2001). The model incorporates both macroscale (climate) and mesoscale (drainage) influences on the water balance and recreates the characteristic spatial scaling of soil moisture in geostatistical variogram analysis (Lookingbill and Urban 2004). Importantly, it incorporates both the positive (climate) and negative (drainage) relationships between moisture and elevation. We collected 114 validation samples the following year, and these data were more strongly correlated with predictions from the statistical model ($r^2 = 0.33$) than with either TCI ($r^2 = 0.16$) or output from RHESSys (Band et al. 1991), a process simulation model parameterized for the HJA ($r^2 = 0.23$).

Each of the environmental models was designed to describe spatial patterns in plant-relevant environmental variables specifically for vegetation analysis; therefore, we did not expend the effort to account for the comprehensive suite of processes that influence radiation, temperature, and soil moisture variability. The models provide improved estimates of plant-relevant environmental variables given advancements in technology, data accessibility, and statistical theory, but are purposively kept simple enough for the methods to be easily translated to most montane systems. It is important to note, however, that the models are meant to provide relative differences only. The estimates are properly used to describe locations that are warmer–cooler, wetter–drier, and sunnier–shadier. The numerical estimates provided in this analysis should be interpreted in that context (e.g., plots with mean July 2000 temperatures greater than 14 °C = warmer sites).

Analyses

Community ecologists have shown an affinity for combining complementary analytical methods to improve the interpretability or robustness of their analyses. The chief tools in this, by far, have been clustering and ordination (Whittaker 1967; Gauch 1982; Pielou 1984; Jongman et al. 1995). This combination of methods can be used to identify discrete community types, whose relationships with environmental gradients can be explored in ordination space. Classification and regression tree (CART) analysis offers a rather new and compelling alternative to ordination for exploring differences among groups (Breiman et al. 1984; Moore et al. 1991; Vassières et al. 2000). We rely on the complementary tools of clustering and classification tree analysis to define the forest communities of the HJA and to describe the association of these communities with the environmental template.

Community identities

Hierarchical clustering analysis was used to identify natural breaks indicative of major community types in the 164 plots sampled from 1997 to 1999 (Sneath and Sokal 1973). This analysis is highly dependent upon the choice of distance measure used to assess group similarity and the linkage criteria used to determine the distance between groups for joining purposes (Legendre and Legendre 1998). We used the Bray–Curtis index (Bray and Curtis 1957) as the distance measure, because this index has been shown to perform well with compositional data (Legendre and Legendre 1998; McCune and Grace 2002). We used unweighted pair-group method with arithmetic means (UPGMA; Sneath and Sokal 1973) as the joining criteria, because it minimized the amount of chaining in the analysis (5.6%). Chaining, the sequential addition of small groups to a few large groups, can cause considerable difficulties in clearly defining subgroups (McCune and Grace 2002).

Indicator Species Analysis (ISA) was used to assign species to the group for which they had the highest indicator value (Dufrene and Legendre 1997). Indicator values combine information on species relative frequency and relative abundance in different groups. Relative frequency was calculated as the percentage of plots in which a species was present for a given group. Relative abundance was calculated as the average basal area of a species in a given group of plots divided by the average basal area of that species in all plots. Statistical significance of group assignment was tested by Monte Carlo simulation with 1000 permutations.

In addition, ISA was used to determine the appropriate number of groups to be used in the final classification, as suggested by Dufrene and Legendre (1997). Because groups are defined at multiple levels in hierarchical clustering algorithms, choosing a meaningful level of clustering can present a challenge for which ISA can serve as a useful tool. Indicator values are low for poorly defined clusters and typically peak at an intermediate level of clustering. By plotting the sum of the indicator values for all the species versus the level of clustering, we determined the level of clustering that maximized the group differentiation. We used this level of classification for all subsequent analyses.

Community associations

To determine how the environmental variables influence the distribution of community types, we created classifica-

Table 1. Seven dominant species observed on 164 plots in the H.J. Andrews Experimental Forest used to construct the models.

Scientific name	Common name	Frequency (%)	Density (no./ha)	BA (m²/ha) Mean	BA (m²/ha) Max.	BA (m²/ha) SD
Abies amabilis	Pacific silver fir	42.7	173.6	5.2	57.9	10.7
Abies procera	Noble fir	24.4	47.6	6.0	96.4	16.0
Pseudotsuga menziesii	Douglas-fir	78.0	110.1	50.5	168.2	42.6
Taxus brevifolia	Pacific yew	40.9	58.4	0.9	13.3	2.2
Thuja plicata	Western redcedar	38.4	52.7	5.2	83.1	12.2
Tsuga heterophylla	Western hemlock	82.9	231.4	15.5	71.0	16.6
Tsuga mertensiana	Mountain hemlock	13.4	42.2	2.6	107.8	12.1

Note: BA, basal area; SD, standard deviation.

tion trees of the species groups using the physical proxy and plant-relevant environmental data as predictor variables. CART analysis is the recursive partitioning of a data set into increasingly homogenous subsets (Breiman et al. 1984; De'ath and Fabricius 2000; Urban 2002). Applied to a categorical response variable (here, community types), the goal of a classification tree is to partition the samples such that each node of a treelike diagram comprises only one group type. At each branch of the partitioning, the environmental variables are examined to find the best variable to separate groups.

The technique offers several attractive qualities. First, the interpretation of the tree is rather intuitive and can be easily implemented into a GIS for visualization of the results (Moore et al. 1991; Urban et al. 2002). Second, it is non-parametric and recursive and thus can handle a variety of ecological situations that are unwieldy using conventional linear classification models (see Vassières et al. (2000) for a comparison to other models). For example, because each branch of the tree is defined independently of other branches and the decision rules rely on no assumptions regarding the underlying model structure, a given group (e.g., *T. heterophylla*) may appear on multiple branches of the tree, each branch representing alternative (possibly compensatory) environmental conditions under which that group occurs.

Unfortunately, the high degree of precision with which CART analysis can classify a data set frequently results in an overfitting of the model to most ecological data (Legendre and Legendre 1998). The resulting model may explain the input data extremely well, but be too specific to those data to be generalized to broader patterns. To account for this tendency to overfit, we split the initial 164 plots into 10 roughly equal parts for cross-validation analysis (Venables and Ripley 2002). We generated cross-validated trees using combinations of nine of the parts and tested them with the excluded part. Classification errors were then determined from an average of the 10 cross-validated trees. The effect of this cross-validation was to penalize classification trees that overfit the data. We assessed the branching patterns on our final trees to identify splits that caused an increase in the cross-validation misclassification rate. In addition, we conducted a validation of the complete classification trees with the 11 independent, "difficult to classify" samples that were collected in 2002. Based on these validation criteria, we compared the CART model using the indirect physical proxy variables as explanatory factors with the model using radiation, temperature, and soil moisture. All models were generated using the RPART

Table 2. Group membership assignment for four-cluster solution.

Group	N	Species	Indicator value
TSHE	130	*Tsuga heterophylla*	80
		Pseudotsuga menziesii	57
		Taxus brevifolia	54
		Thuja plicata	50
ABPR	11	*Abies procera*	84
ABAM	18	*Abies amabilis*	64
TSME	5	*Tsuga mertensiana*	89
Mean			68

Note: Species were assigned to the group for which they had the largest indicator value.

library (Therneau and Atkinson 1997) in S-plus version 6.1 (Insightful, Seattle).

Results

Dominant tree species are provided in Table 1, along with the basal area, density, and frequency of occurrence of each species observed in the 164 sample plots. None of the other species sampled contributed as much as 1% to the total basal area observed in the sampling, and these rarer species were excluded from the analyses. *Pseudotsuga menziesii* was by far the greatest contributor to total basal areas (nearly 60%), while *T. heterophylla* contributed the most to total density (32%).

The sum of the indicator values across all seven species was maximal for the four-cluster solution (Table 2). The four major communities can be described as follows:
(1) *Tsuga heterophylla – P. menziesii* group (TSHE): characterized by *T. heterophylla* and *P. menziesii*. *Thuja plicata* and *Taxus brevifolia* present about 50% of the time in this group and never found in the other three communities.
(2) *Abies procera* group (ABPR): indicated by *A. procera* with some *A. amabilis* and *P. menziesii*. Small number of *T. heterophylla* also commonly observed.
(3) *Abies amabilis* group (ABAM): dominated by *A. amabilis* with *T. mertensiana* and *A. procera* often represented.
(4) *Tsuga mertensiana* group (TSME): indicated by *T. mertensiana* with *A. amabilis* also present but in lower abundance. No other species common.

Fig. 2. Classification tree from classification and regression tree (CART) analysis using environmental proxy variables based on topography. Groups were defined by a clustering analysis and are described in the text and in Table 2. Length of branch corresponds to the amount of variance explained by that variable. Parentheses at terminal nodes provide the misclassification errors as number misclassified divided by total *N* in node.

Fig. 3. Classification tree from classification and regression tree (CART) analysis using modeled estimates of summer radiation, temperature, and soil moisture as potential explanatory variables. Groups were defined by a clustering analysis and are described in the text and in Table 2. Length of branch corresponds to the amount of variance explained by that variable. Parentheses at terminal nodes provide the misclassification errors as number misclassified divided by total *N* in node.
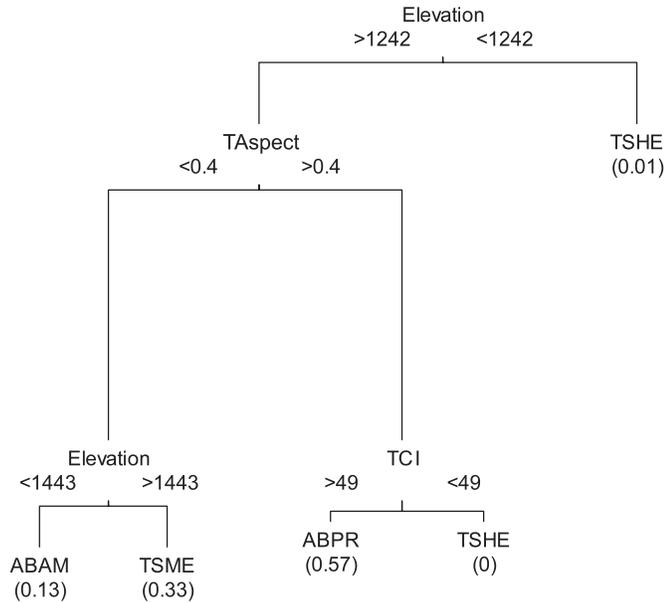
Table 3. Misclassification matrix for four community types based on environmental proxy model.

	Predicted group				Total	
Actual group	TSHE	ABPR	ABAM	TSME	Corr.	Misclass.
TSHE	121	8	1	—	121	9
ABPR	—	9	1	1	9	2
ABAM	1	3	13	1	13	5
TSME	—	1	—	4	4	1

Note: Rows are actual community types, columns are predicted community types. Row totals are listed as number correct (corr.) and number misclassified (misclass.). Total misclassification rate is 10.3%. See Table 2 for description of community types.

Table 4. Misclassification matrix for four community types based on plant-relevant explanatory model.

	Predicted group				Total	
Actual group	TSHE	ABPR	ABAM	TSME	Corr.	Misclass.
TSHE	123	7	—	—	123	7
ABPR	—	8	1	2	8	3
ABAM	6	1	10	1	10	8
TSME	—	—	—	5	5	0

Note: Rows are actual community types, columns are predicted community types. Row totals are listed as number correct (corr.) and number misclassified (misclass.). Total misclassification rate is 11.0%. See Table 2 for description of community types.

The first CART analysis, based on physical proxies, indicated that elevation and transformed aspect were the environmental variables that most strongly sort the four communities (Fig. 2). Elevation was the strongest predictor variable, separating the *T. heterophylla – P. menziesii* community from the high-elevation species at the first branch of the tree (though nine *T. heterophylla – P. menziesii* plots were found at elevations greater than 1242 m). Misclassification rates provide a measure of the ability of the model to discriminate among groups (Table 3). Total misclassification in the model was 10.4%, most of which was at the *A. procera* end node.

In the second CART model, we examined the ability of our modeled estimates of the plant-relevant variables temperature, soil moisture, and radiation to explain patterns in community composition (Fig. 3). Temperature was the strongest predictor variable in this classification tree, and three distinct zones were identified: (1) a high-temperature zone (mean July

2000 temperature greater than 14 °C) in which all but 1% of the plots were *T. heterophylla – P. menziesii*; (2) a low-temperature zone (mean July 2000 temperature less than 13.4 °C) in which none of the plots were *T. heterophylla – P. menziesii*; and (3) a zone in between these two temperature extremes in which *T. heterophylla – P. menziesii* plots were found mixed with *A. procera* (in areas of high radiation) and *A. amabilis* (in areas of low radiation). The total misclassification rate for this model was 11.0%, with better prediction of the actual *T. heterophylla – P. menziesii* plots and worse prediction of the *A. amabilis* plots than for the proxy variable model (Table 4).

Because classification tree models are developed through recursive partitioning, error rates can be calculated for models of increasing complexity (i.e., more splits in the tree). The models using the two different sets of predictor variables had similar decreases in error rates with increasing

Fig. 4. Relative error versus the number of splits. The relative error is calculated as the total number of tree misclassifications divided by the number of misclassifications given zero splits in the tree. Proxy model uses TAspect, elevation, and topographic convergence index (TCI) as possible explanatory variables. Plant-relevant model uses the modeled estimates of radiation, temperature, and soil moisture.

Fig. 5. Relative error from cross-validation versus the number of splits. Values represent relative errors from an average of 10 cross-validated trees in which one-tenth of the entire data set was excluded from tree construction and used to test the model. Proxy model uses TAspect, elevation, and topographic convergence index (TCI) as possible explanatory variables. Plant-relevant model uses the modeled estimates of radiation, temperature, and soil moisture.
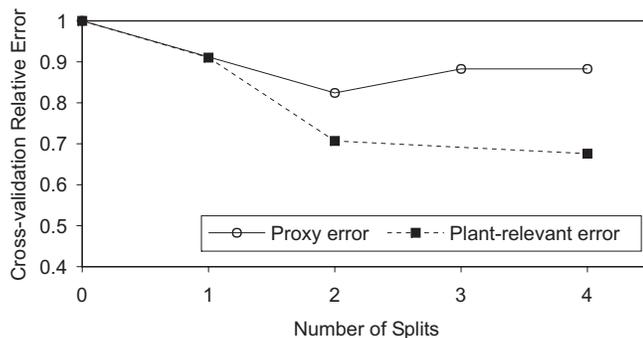
model complexity (Fig. 4). However, the cross-validation error was much higher for the model using the indirect proxy variables (Fig. 5), suggesting that this model overfit the data below the second split in the tree. In addition, nearly all the validation plots (9 of 11) were misclassified by the proxy model, while only 3 of 11 validation plots were misclassified by the plant-relevant model.

Discussion

An underlying tenet of many climate-change predictions is that elevation acts as a proxy for temperature in ecological systems (Peters and Darling 1985). For years, ecologists have called for the use of direct and resource gradients in describing forest vegetation patterns to test these types of assumptions (Austin and Smith 1989, and many others). However, efforts to infuse more physiologically meaningful data into gradient studies were hindered by the difficulty of collecting these data at fine resolutions over large spatial extents. As a result, theoretical and simulation modeling (e.g., Urban et al. 1993) have been the best and often only options to represent

the relationships between forest vegetation and factors such as temperature, soil moisture, and radiation. Recent advances in data collection technology, sampling methodology, and GIS-based modeling have greatly enhanced the ability to capture detailed environmental variability across large areal extents. Our study serves as an example of how these advances can provide valuable information for testing and refining gradient analysis models.

Model comparison of community associations

The communities characterized by the cluster analysis in our study agree with the published sorting of forest tree species within this watershed (Dyrness et al. 1974; Zobel et al. 1976). Classification tree analysis is specifically designed to distinguish among the different group types by addressing the companion questions: (1) Are plots that are compositionally dissimilar also different in terms of their environmental attributes? (2) If so, which environmental variables best differentiate the groups (Urban 2002)? The CART analysis with physical proxy variables supports the existing community model (e.g., Franklin and Dyrness 1988) that the *T. heterophylla – P. menziesii* community is found below approximately 1250 m, *T. mertensiana* is found above approximately 1450 m, and in between these elevations is found a mix of *A. amabilis*, *A. procera*, and *T. heterophylla – P. menziesii*. The results illustrate two of the benefits of the flexible recursive partitioning algorithm used in CART analysis: (1) the ability to use predictive variables multiple times (e.g., elevation) to represent situations in which interactions among variables are not simply multiplicative and (2) the ability to describe multiple environmental settings associated with the same group (e.g., *T. heterophylla – P. menziesii*). The model thus identifies not only the average case (*T. heterophylla – P. menziesii* below 1242 m) but also cases that emerge as exceptions to the rule (nine *T. heterophylla – P. menziesii* plots above 1242 m on south-facing aspects).

The model using the more traditional variables of gradient analysis (indirect elevation, topographic convergence, and aspect gradients) can be compared to the one created using our spatial estimates of more direct environmental and resource gradients. Because of the flexibility of the recursive partitioning algorithm used in CART analysis, the structure of the models (i.e., the branching patterns) differ slightly, and a direct comparison of the individual decision rules used in the classification trees is not appropriate. CART provides an overall assessment of how groups are different and which variables distinguish the groups. Interpreting the complete classification trees in this context, the results provide corroborating evidence that upper (1443 m) and lower (1242 m) elevation thresholds are associated with temperature (mean July 2000 temperatures of 13.4 and 14 °C in Fig. 4).

We did not find a 1:1 correspondence between elevation and temperature as explanatory variables, however. For example, the transition from *A. amabilis* to *T. mertensiana*, which occurs at the upper elevation threshold according to the proxy model, is more strongly correlated with soil moisture differences than temperature; thus, a warming climate would not be expected to stimulate an encroachment of *A. amabilis* on *T. mertensiana*. It may instead create a constriction of the *A. amabilis* range through drought effects, if this community's

upper boundary is physiologically defined and a warmer climate translates to a drier climate.

Both models indicate a similar habitat envelope for the *A. procera* community: high elevation and high exposure in the proxy model compared to low temperature and high radiation in the more plant-relevant model. These conditions are consistent with the literature (Franklin and Dyrness 1988). The suggestion in the proxy model that *A. procera* is found in areas of high convergence (TCI > 49) is difficult to interpret physiologically and is not supported by the model of more direct explanatory variables. As CART models can tend to be statistically overfitted to the data, it is appropriate to ask whether this branch of the classification tree is real or an artifact of the modeling process. The increase in error rate with increasing complexity in the cross-validation analysis (Fig. 5) and the failure of this model to predict the independent validation plots suggest that the TCI split is likely an artifact.

Model validation

By nearly every measure, the classification tree using radiation, temperature, and soil moisture provides a better explanatory model than the tree using less direct proxy variables. Although the error rates of the unvalidated models are similar (Fig. 4), the cross-validation error rates are much higher for the proxy model below the first split of the trees (Fig. 5). In other words, the predictive power of the proxy variables diminishes drastically after the first use of elevation as a predictor variable.

The use of independent data is preferred for model validation when such a data set exists (Reynolds and Chung 1986), and the model that uses more plant-relevant explanatory variables does a much better job than the proxy model at predicting the 11 independent validation plots. The practice of translating between the parameter space of models and the geographic space of the study area is fundamental to the integration of models and data. Using a series of "IF… THEN…" conditional statements, the CART predictions were mapped onto the study site. Given these maps, it was rather easy to choose validation sampling locations that would test the predictions of the two CART models. These validation plots were selected specifically because they were in areas of high community turnover that should be difficult to classify (Congalton 1991; Urban et al. 2002); yet, the plant-relevant model still had a 73% success ratio in predicting community type compared to a 18% success rate for the proxy model. By replacing elevation and other indirect measures with explanatory factors that are more directly responsible for the distribution of plants, we have increased our ability to predict the arrangement of forest communities in these highly sensitive regions of the watershed.

Most importantly, the model using more direct estimates of temperature, soil moisture, and radiation is more biologically interpretable and provides more meaningful insight into potential forest response to environmental change (e.g., global climate change scenarios). Collectively, the results illustrate how vegetation models using indirect environmental proxy variables should be used not as a final product, but as a point of departure to try to identify more physiologically meaningful relationships. While associations between forest communities and elevation are often assumed to reflect gradients in temperature, we show that other resources can contribute to the ubiquitous "elevation gradient". Acquiring a better understanding of forest community patterns relative to more plant-relevant environmental variables will yield better predictive power and management information for responding to any potential shifts in environmental conditions. We envision the next generation of gradient analysis to be one in which relationships between elevation and basal area are replaced with more ecologically meaningful relationships between plant-relevant environmental variables and direct demographic mechanisms.

Acknowledgements

References

Allen, R.B., Peet, R.K., and Baker, W.L. 1991. Gradient analysis of latitudinal variation in Southern Rocky Mountain forests. J. Biogeogr. **18**: 193–201.

Austin, M.P. 1987. Models for the analysis of species' response to environmental gradients. Vegetatio, **69**: 35–45.

Austin, M.P, and Smith, T.M. 1989. A new model for the continuum concept. Vegetatio, **83**: 35–47.

Band, L.E., Peterson, D.L., Running, S.W., Coughlan, J., Lammers, R., Dungan, J., and Nemani, R. 1991. Forest ecosystem process at the watershed scale: basis for distributed simulation. Ecol. Model. **56**: 171–196.

Beers, T.W., Dress, P.E., and Wensel, L.C. 1966. Aspect transformation in site productivity research. J. For. **64**: 691–692.

Beven, K.J., and Kirkby, M.J. 1979. A physically based, variable contributing area model of basin hydrology. Hydrol. Sci. Bull. **24**: 43–69.

Bray, R.J., and Curtis, J.T. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecol. Monogr. **27**: 325–349.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. 1984. Classification and regression trees. Wadsworth and Brooks/Cole, Monterey, Calif.

Bunn, A.G., Waggoner, L.A., and Graumlich, L.J. 2005. Topographic mediation of growth in high elevation foxtail pine (*Pinus balfournia* Grev. et Balf.) forests in the Sierra Nevada, USA. Global Ecol. Biogeogr. **14**: 103–114.

Chen, J., Saunders, S.C., Crow, T.R., Naiman, R.J., Brosofske, K.D., Mroz, G.D., Brookshire, B.L., and Franklin, J.F. 1999. Microclimate in forest ecosystem and landscape ecology. Bioscience, **49**: 288–297.

Christensen, N.L., Bartuska, A.M., Brown, J.H., Carpenter, S., D'Antonio, C., Francis, R. et al. 1996. The report of the Ecological Society of America committee on the scientific basis for ecosystem management. Ecol. Appl. **6**: 665–691.

Cissel, J.H., Swanson, F.J., and Weisburg, P.J. 1999. Landscape management using historical fire regimes: Blue River, Oregon. Ecol. Appl. **9**: 1217–1231.

Congalton, R.G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. **37**: 35–46.

De'ath, G., and Fabricius, K.E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology, **81**: 3178–3192.

Dufrene, M., and Legendre, P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecol. Monogr. **67**: 345–366.

Dyrness, C.T., Franklin, J.F., and Moir, W.H. 1974. A preliminary classification of forest communities in the central portion of the Western Cascades in Oregon. Coniferous Forest Biome Bulletin No. 4. USDA Forest Service, Corvallis, Ore.

Fonda, R.W., and Bliss, L.C. 1969. Forest vegetation of the montane and subalpine zones, Olympic Mountains, Washington. Ecol. Monogr. **39**: 271–301.

Franklin, J. 1995. Predictive vegetation mapping: geographic modeling of biospatial patterns in relation to environmental gradients. Prog. Phys. Geogr. **19**: 474–499.

Franklin, J.F., and Dyrness, C.T. 1988. Natural vegetation of Oregon and Washington. Oregon State University Press, Corvallis, Ore.

Franklin, J.F., and Halpern, C.B. 2000. Pacific Northwest forests. *In* North American vegetation. *Edited by* M.G. Barbour and W.D. Billings. Cambridge University Press, New York. pp. 121–159.

Fujiware, K., and Box, E.O. 1999. Climate change and vegetation shift. *In* Perspectives in Ecology. *Edited by* A. Farina. Backhuys Publishers, Leiden, Netherlands. pp. 121–126.

Gagnon, D., and Bradfield, G.E. 1987. Gradient analysis of west central Vancouver Island. Can. J. Bot. **65**: 822–833.

Gauch, H.G. 1982. Multivariate analysis in community ecology. Cambridge University Press, Cambridge.

Grier, C.C., and Logan, R.S. 1977. Old-growth *Pseudotsuga menziesii* communities of a western Oregon watershed: biomass distribution and production budgets. Ecol. Monogr. **47**: 373–400.

Halpin, P. 1997. Global climate change and natural-area protection: management responses and research directions. Ecol. Appl. **7**: 828–843.

Hobbs, R. 1997. Future landscapes and the future of landscape ecology. Landsc. Urban Plann. **37**: 1–9.

IPCC. 2001. Climate change 2000. The science of climate change. Contribution of Working Group I to the third assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

Jongman, R.H.G., ter Braak, C.J.F., and van Tongeren, O.F.R. 1995. Data analysis in community and landscape ecology. Cambridge University Press, Cambridge.

Krajina, V.J. 1969. Ecology of forest trees in British Columbia. Ecol. West. N. Am. **2**: 1–146.

Legendre, P., and Legendre, L. 1998. Numerical ecology. Elsevier, New York.

Lookingbill, T., and Urban, D. 2003. Spatial estimation of air temperature differences for landscape-scale studies in montane environments. Agric. For. Meteorol. **114**: 141–151.

Lookingbill, T., and Urban, D. 2004. An empirical approach towards improved spatial estimates of soil moisture for vegetation analysis. Landsc. Ecol. **19**: 417–433.

McCune, B., and Grace, J.B. 2002. Analysis of ecological communities. MjM Software Design, Gleneden Beach, Ore.

McKee, A. 1998. Focus on field stations: H.J. Andrews Experimental Forest. Bull. Ecol. Soc. Am. **79**: 241–246.

Moore, D.M., Lees, B.G., and Davey, S.M. 1991. A new method for predicting vegetation distributions using decision tree analysis and geographic information system. Environ. Manage. **15**: 59–71.

Neilson, R.P. 1991. Climatic constraints and issues of scale controlling regional biomes. *In* ecotones: the role of landscape boundaries in the management and restoration of changing environments. *Edited by* M.M. Holland, P.G. Risser, and R.J. Naiman. Chapman & Hall, New York. pp. 31–51.

Noss, R.F. 2001. Beyond Kyoto: forest management in a time of rapid climate change. Conserv. Biol. **15**: 578–590.

Ohmann, J.L., and Spies, T.A. 1998. Regional gradient analysis and spatial pattern of woody plant communities of Oregon forests. Ecol. Monogr. **68**: 151–182.

Packee, E.C. 1990. *Tsuga heterophylla* (Raf.) Sarg. Western hemlock. *In* Silvics of North America. Vol. 1. Conifers. *Technical coordinators* R.M. Burns and B.H. Honkala. USDA Agric. Handb. 654. pp. 613–622.

Peters, R.L., and Darling, J.D. 1985. The greenhouse effect and nature reserve design. Bioscience, **35**: 707–717.

Pielou, E.C. 1984. The Interpretation of Ecological Data: A Primer on Classification and Ordination. Wiley & Sons, New York.

Pierce, K., Lookingbill, T., and Urban, D. 2005. A simple method for estimating potential relative radiation (PPR) for landscape-scale vegetation analysis. Landsc. Ecol. **20**(2): 137–147.

Reynolds, J.M.R., and Chung, J. 1986. Regression methodology of estimating prediction error. Can. J. For. Res. **16**: 931–938.

Smith, J. 2002. Mapping the thermal climate of the H.J. Andrews Experimental Forest, Oregon. Oregon State University, Corvallis, Ore.

Sneath, P.H.A., and Sokal, R.R. 1973. Numerical taxonomy. 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.

Therneau, T.M., and Atkinson, E.J. 1997. An introduction to recursive partitioning using RPART routines. Mayo Foundation, Rochester, Minn.

Thornburgh, D.A. 1969. Dynamics of the True Fir-Hemlock Forests of the West Slope of the Washington Cascade Range. University of Washington, Seattle, Wash.

Urban, D.L. 2002. Classification and regression trees. *In* Analysis of ecological communities. *Edited by* B. McCune and J.B. Grace. MjM Software Design, Gleneden Beach, Ore. pp. 222–232.

Urban, D.L., Harmon, M.E., and Halpern, C.B. 1993. Potential response in Pacific Northwestern forests to climate changes: effects of stand age and initial composition. Clim. Change, **23**: 247–266.

Urban, D.L., Miller, C., Halpin, P.N., and Stephenson, N.L. 2000. Forest gradient response in Sierran landscapes: the physical template. Landsc. Ecol. **15**: 603–620.

Urban, D., Goslee, S., Pierce, K., and Lookingbill, T. 2002. Extending community ecology to landscapes. Ecoscience, **9**: 200–212.

Vassières, M.P., Plant, R.E., and Allen-Diaz, B.H. 2000. Classification trees: an alternative nonparametric approach for predicting species distributions. J. Veg. Sci. **11**: 679–694.

Venables, W.N., and Ripley, B.D. 2002. Modern applied statistics with S. Springer-Verlag, New York.

Whittaker, R.H. 1956. Vegetation of the Great Smoky Mountains. Ecol. Monogr. **26**: 1–80.

Whittaker, R.H. 1960. Vegetation of the Siskiyou Mountains, Oregon and California. Ecol. Monogr. **30**: 279–338.

Whittaker, R.H. 1967. Gradient analysis of vegetation. Biol. Rev. **42**: 207–264.

Whittaker, R.H. 1978. Ordination of plant communities. Junk, The Hague, Netherlands.

Yeakley, J.A., Swank, W.T., Swift, L.W., Hornberger, G.M., and Shugart, H.H. 1998. Soil moisture gradients and controls on a southern Appalachian hillslope from drought through recharge. Hydrol. Earth Sys. Sci. **2**: 41–49.

Zobel, D.B., McKee, A., Hawk, G.M., and Dyrness, C.T. 1976. Relationships of environment to composition, structure, and diversity of forest communities of the central western Cascades of Oregon. Ecol. Monogr. **46**: 135–156.