Climatic and biophysical controls on conifer species distributions in mountain forests of Washington State, USA

Donald McKenzie[1]*, David W. Peterson[2], David L. Peterson[1] and Peter E. Thornton[3] [1]*Seattle Forestry Sciences Lab, USDA Forest Service, Pacific Northwest Research Station, Seattle, WA, USA, [2]Wenatchee Forestry Sciences Lab, USDA Forest Service, Pacific Northwest Research Station, Wenatchee, WA, USA and [3]National Center for Atmospheric Research, Climate and Global Dynamics Division, Boulder, CO, USA*

Abstract

Aim The purpose of this study was to quantify relationships between conifer species distributions and climatic and biophysical variables, in order to provide better insight into the potential for redistribution of species on the landscape in response to climatic change.

Location Data are from 10,653 georeferenced sites in Washington State, USA, along a longitudinal gradient from west of the crest of the Cascade Range to the beginnings of the western slope of the Rocky Mountains, and across two physiographic provinces, the Northern Cascades, characterized by steep, rugged topography, and the Okanogan Highlands, presenting moderate slopes and broad rounded summits.

Methods Tree data were drawn from the USDA Forest Service Area Ecology Program database, collected in mature, undisturbed stands. We compared simple climatic variables (annual temperature, growing-degree days, annual and seasonal precipitation) to biophysical variables (soil, hydrologic, and solar radiation) derived from climatic variables. Climatic and biophysical variables were taken from the output of climatological and hydrological simulation models and estimated for each plot in the tree database. Generalized linear models were used, for each of fourteen tree species, at multiple spatial extents, to estimate the probability of occurrence of that species as a function of climatic and biophysical predictors. Models were validated by a combination of bootstrapping and estimating receiver operating characteristic (ROC) curves.

Results For the majority of species, we were able to fit variables representing both moisture and temperature gradients, and in all but a few cases these models identified a unimodal response of species occurrence to these gradients. In some cases the ecological/environmental niche of a species had been clearly captured by the model, whereas in others a longer gradient in the predictor variable(s) would be needed. Responses of most species were consistent across three spatial scales.

Main conclusions By identifying the ecological niches of multiple species, we can forecast their redistribution on the landscape in response to climatic change, evaluate the predictions of simulation models, and alert managers to particularly sensitive or vulnerable ecosystems and landscapes.

Keywords
Conifer species distributions, biophysical variables, DAYMET, VIC, scale, climatic change, generalized linear models, unimodal responses, ROC curves.

*Correspondence: Donald McKenzie, Seattle Forestry Sciences Lab, USDA Forest Service, Pacific Northwest Research Station, 400 North 34th Street, Suite 201 Seattle, WA 98103, USA. E-mail: donaldmckenzie@fs.fed.us

INTRODUCTION

Rapid climatic change is expected for the next century (Thompson *et al.*, 1998; Houghton *et al.*, 2001), and vegetation patterns are expected to change in complex ways, not only in direct response to climate (Neilson, 1995; Neilson & Drapek, 1998; Bachelet *et al.*, 2001; Walther *et al.*, 2002) but also as a function of patterns of land use and changing disturbance regimes (Baker, 1995; McKenzie *et al.*, 1996; Keane *et al.*, 1999; Dale *et al.*, 2001; He *et al.*, 2002). In anticipating redistributions of species on the landscape over time in response to climatic change, direct responses to climate and the biophysical environment are a key piece of the puzzle, onto which constraints imposed by land use, disturbance, and competitive influences of other species can be overlaid. The effects of climatic change may be particularly strong in mountains, because warmer temperatures affect the depth and duration of snowpacks (Cayan, 1996), which are key limiting factors for tree growth (Davis & Botkin, 1985; Peterson, 1998; Peterson & Peterson, 2001).

Climatic influences are difficult to assess in mountainous areas, because complex topography produces steep gradients in the biophysical environment, and climate-monitoring stations are sparsely distributed. In lieu of accurate climatic data at appropriately small spatial scales, geographic variables (latitude/longitude) and topographic variables (elevation, slope, aspect) are commonly used as surrogate predictors (Franklin, 1995; Ohmann & Spies, 1998; McKenzie & Halpern, 1999; Iverson & Prasad, 2001). However, with the increasing availability of climatic data at spatial resolution finer than 4-km (e.g., Daly *et al.*, 1994; Thornton *et al.*, 1997) and the development of topographically sensitive hydrological models (Nemani *et al.*, 1993; Liang *et al.*, 1994; Wigmosta *et al.*, 1994), it is becoming possible to develop quantitative models relating species distributions directly to climatic and biophysical variables (Franklin, 1998; Bolliger *et al.*, 2000b; Guisan & Zimmermann, 2000; Pearce & Ferrier, 2000b).

Describing species' environmental niches has long been a central task of ecology and biogeography, but has received renewed attention in attempts to understand the effects of climatic variability and change on species distributions and losses of biodiversity (Brockway, 1998; Austin, 1999; Bolliger *et al.*, 2000a; Dale *et al.*, 2001). Vegetation theory, along with a substantial volume of literature, suggests that the response of species to environmental gradients is unimodal, with presence/absence or abundance being greatest at the center of a species range along each gradient (Gauch *et al.*, 1974; Austin, 1987; ter Braak, 1987; ter Braak & Prentice, 1988; Austin & Gaywood, 1994). Although some disagreement exists over the exact shape of response curves (Austin *et al.*, 1994; Austin, 2002; Oksanen & Minchin, 2002), a multidimensional 'environmental niche space' of predictor variables can be estimated for a species if a nonlinear (unimodal) statistical model can be fit (Guisan *et al.*, 2000). If the predictor variables are spatially explicit climatic (precipitation, temperature) or biophysical variables, then changes in the *geographical* niches of species in response to

climatic change scenarios can be quantified from the models, i.e., 'predictive vegetation mapping', *sensu* Franklin (1995).

Empirical/statistical models of species distributions can be quite sensitive to the spatial extent and resolution of the data used for their estimation. For example, McKenzie & Halpern (1999) found that the estimated environmental niches of shrub species changed substantially with changes in spatial extent of the study area. If models are to be extrapolated to new climatic conditions, they need to be robust across scales (i.e. parameter estimates and the shapes of response curves should be similar). Additionally, parameter estimates should be different in separate geographical areas in which the mechanisms of climate–species interactions are different – for example, as a result of genetic variation in a wide-ranging species such as Douglas-fir (*Pseudotsuga menziesii* – Hermann & Lavender, 1990) or subalpine fir (*Abies lasiocarpa* – Alexander *et al.*, 1984). Thus, modelling at multiple spatial extents identifies both the robustness of parameter estimates and the geographic extent over which climate–species interactions are homogeneous.

In this paper, we quantify associations between climatic and biophysical variables and individual conifer species distributions in mountain forests across the state of Washington, USA, using generalized linear models (McCullagh & Nelder, 1989). Our objectives are to identify the key predictors of species presence/absence and how these predictors and the form of species response functions change across scales and along an east–west gradient. Predicting species abundance (cover, biomass, basal area) requires different modelling techniques, because the response variables are continuous rather than binary. One must also address the relative value of predicting mean vs. maximum responses (Scharf *et al.*, 1998; Cade *et al.*, 1999; McKenzie *et al.*, 2000a). Good predictors for abundance may be different from those for distribution (presence/absence); hence, we restrict this study to the latter.

Conifers are the dominant tree life form in the Pacific Northwest, and deciduous species are only sparsely represented in our database, thus we limited this study to conifers. We compare the predictive ability of simple climatic variables (annual temperature, growing-degree days, annual and seasonal precipitation) to that of variables (soil, hydrologic, and solar radiation) derived from climatic variables and output of spatially explicit simulations – hereafter referred to as 'biophysical' variables. Biophysical variables might have more predictive power than simple climate variables because they are more closely linked to resource needs/limiting factors. Conversely, climatic variables might have better predictive power than biophysical variables because climatic variables are effective surrogates for resource needs/limiting factors and because it is difficult to link biophysical variables to species at broad scales.

We also compare models from multiple spatial scales. Broad-scale (regional) models might have better predictive power than subregional models because species responses to climatic and biophysical environments are generally homogeneous and regional analyses incorporate broader ranges of predictor variables (or environmental gradients).

Conversely, subregional models might have better predictive power than regional models because of heterogeneity in limiting environmental factors within larger regions.

METHODS

Study area

Data are from 10,653 sites roughly along a longitudinal gradient from west of the crest of the Cascade Range to the beginnings of the western slope of the Rocky Mountains (Fig. 1), across two physiographic provinces (Franklin & Dyrness, 1988), the Northern Cascades and the Okanogan Highlands. In the Northern Cascades, topography is extremely rugged, with deep and steep-sided valleys with eastward- and westward-flowing streams (Franklin & Dyrness, 1988). Further east, the Okanogan Highlands present moderate slopes and broad rounded summits (Franklin & Dyrness, 1988). A variety of soil types appear in both provinces, reflecting the influence of Pleistocene glaciers, with glacial soils predominant on valley bottoms and residual soils on hillslopes and ridgetops.

Climate in our study area is intermediate between the maritime climate west of the Cascade Crest and the continental climate east of the Rocky Mountains, and is a result of the interaction of three air masses: (1) moist marine air from the west, (2) continental air from the east and south, and (3) dry arctic air from the north (Ferguson, 1997). Summer drought, caused by a seasonal northward shift in the jet stream in conjunction with high pressure over coastal Oregon and Washington, is common, even in areas with high annual precipitation. A small anomaly exists, however, over the central Okanogan Highlands, such that a greater percentage of total precipitation falls in the summer than in areas immediately east or west (Daly et al., 1994).

Field data collection

Vegetation data were obtained from the Area Ecology program, USDA Forest Service, for the Wenatchee, Okanogan (now combined into Okanogan-Wenatchee NF), and Colville National Forests in eastern Washington, USA, and for the Grizzly Bear Habitat Study (Gaines et al., 1990). The Grizzly Bear study sampled vegetation over a wide area encompassing parts of the Wenatchee and Okanogan National Forests, high-elevation sites in the Mt Baker/ Snoqualmie National Forest, and a few plots in North Cascades National Park. Our study sites occupy a longitudinal gradient from the crest of the Cascade Range to the western slope of the Rocky Mountains (Fig. 1).

The Area Ecology program has a wide range of objectives (see Williams & Lillybridge, 1983; Williams et al., 1990; Lillybridge et al., 1995); here we describe only aspects of data collection relevant to the objectives of this paper. Circular, 0.2-ha plots were established in stands encompassing a range of elevations, aspects, and slopes that met the following criteria: (1) more than 75 years old, (2) relatively undisturbed, and (3) relatively uniform in vegetation composition. Partly because of these criteria, and partly because of time and financial constraints, sampled stands in the Area Ecology Program were not spatially random. The clustered patterns in both these and plots from the Grizzly Bear Habitat Study reflect the severe physiographic constraints on locations of mature, undisturbed forests.
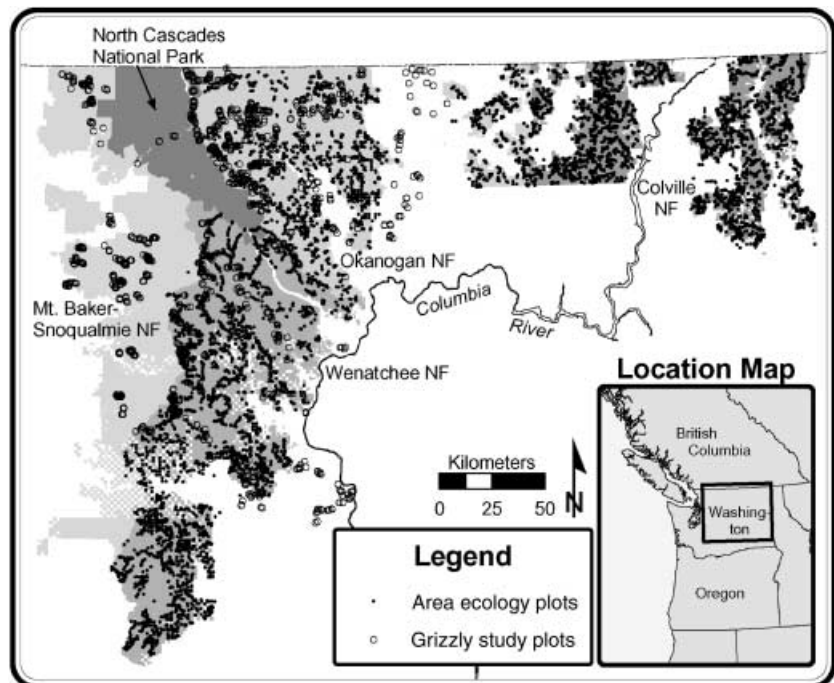

Figure 1 Vegetation plots from the Area Ecology Program and Grizzly Bear Habitat Study used in the analysis.

Elevation, aspect, and slope were recorded at each plot, and ocular estimates of cover were made for all species, including herbs, shrubs, regenerating trees (<3.5 m tall), and mature trees (>3.5 m tall). Field plots were geo-referenced to UTM coordinates. Using these, we created point coverages in ARC-INFO [Environmental Systems Research Institute (ESRI), 2000] for each of the four study areas (hereafter 'forests').

Generation of climatic and biophysical variables

Climatic and physical variables

The DAYMET model (Thornton *et al.*, 1997) was used to estimate daily meteorological variables (36 total) for the period 1980–97 over the coterminous United States. DAYMET combines interpolation algorithms that operate on raw data from weather stations with a set of adjustments for elevational lapse rates and topography (Hungerford *et al.*, 1989). The daily outputs, computed at 1-km resolution, were combined into monthly and annual summaries (mean and standard deviations) and transformed to raster coverages (Thornton *et al.*, 1997). For the current study we selected annual and monthly summary coverages of climatological variables and net solar radiation. Net solar radiation at the earth's surface was calculated by assuming attenuation from cloud cover as a function of diurnal temperature range (Bristow & Campbell, 1984). We imported the coverages into ARC-INFO (ESRI, 2000) as grids, and clipped the grids to the geographical range of the vegetation plots (Table 1). Vegetation plots were overlaid on the grids and values for each of the climatic and physical variables were extracted for each plot.

Soil and hydrological variables

Biophysical variables were generated using the Variable Infiltration Capacity (VIC) hydrologic model (Liang *et al.*, 1994) and the mountain climate simulation (MT-CLIM) model (Hungerford *et al.*, 1989). Climate forcing data for the models consisted of a gridded network (1/8th degree resolution) of 51-year time series of daily temperature (maximum and minimum), precipitation, and wind speed developed from historic climate station data for the Land Data Assimilation System (LDAS) – North America project (Maurer *et al.*, 2002). Vegetation sampling points were paired with the corresponding climate time series based on the recorded latitude and longitude at the sample point.

The MT-CLIM model was used to adjust climate time series for topographical effects on air temperature and to estimate daily solar radiation fluxes for use by the hydrology model. The LDAS climate forcing data were assumed to be for a flat grid cell with elevation equal to that of the mean elevation within the grid cell. MT-CLIM was used to calculate daily solar radiation fluxes based on diurnal temperature ranges (Bristow & Campbell, 1984), with adjustments made for slope angle and orientation and topographic shading. Maximum and minimum temperatures were adjusted to account for elevation differences between the grid cell and the sample site (lapse rate of 6 °C per 1000 m) and for topographical effects on solar radiation. The daily climate time series output from MT-CLIM, combined with the LDAS wind speed data, were used as climate forcings for the VIC hydrologic model.

Biophysical variables (see Table 1) for each vegetation sample site were created from VIC model output summarized over a 50-year period (water years of 1950–99). Climate data from water year 1949 were included in the model runs only to establish initial soil water conditions (output was discarded). At each site, we assumed a reference vegetation cover of evergreen coniferous forest with leaf area index of 3.0. Soils were assumed to be sandy loams with a maximum tree rooting depth of 1.5 m.

Data analysis

Abundance measures for overstory conifers in the vegetation database were transformed to presence/absence and compiled into a model database with the climatic, physical, and biophysical predictors. We used generalized linear models (GLMs) of the binomial family (McCullagh & Nelder, 1989) to estimate the probability of occurrence for each species at each plot as a function of the predictor variables. Analyses were conducted with Splus 2000 for Windows (Insightful, 2000).

For this study, we eschewed a variety of other approaches (Guisan & Zimmermann, 2000; Moisen & Frescino, 2002; Yee & MacKenzie, 2002) such as classification and regression trees (CART – Breiman *et al.*, 1984), generalized additive models (GAM – Hastie & Tibshirani, 1990; Yee & Mitchell, 1991), or artificial neural networks (ANN – Ripley, 1994). These latter can be more flexible than GLMs but they are more complex and more difficult to interpret with respect to new observations. For example, predicted responses for new observations in CART are limited by its hierarchical structure to those already estimated by the original model (Clark & Pregibon, 1992).

Climatic and biophysical predictor variables were classified into two categories, or *proxy sets* (Booth *et al.*, 1994): those representing moisture (precipitation, most hydrological variables, and soil water index) and temperature [temperature, growing degree days, frost days, soil degree days, and evapotranspiration, actual (AET) and potential (PET)]. AET and PET were allowed to 'float', however, because they can represent a combination of both temperature and moisture (Stephenson, 1990, 1998). As we expected, collinearity of predictors within each category was high, suggesting that only one variable from each category should be used in a given model. We therefore applied an exhaustive procedure in which, for each species, each predictor within the 'moisture' category was paired with each predictor from the 'temperature' category. Thus, when the two 'physical' variables, solar radiation and aspect, were included, the maximum number of distinct predictor variables for a model was four. We explored quadratic terms in the models for predictors in the moisture and temperature categories, under the assumption that species will show a unimodal response to environmental gradients (see above – ter Braak, 1987).

Table 1 Variables used to predict species distributions

Climate	Physical	Biophysical	
		Hydrology	Soil
Precipitation (cm)	Aspect (cosine transformed)	Baseflow [annual export of water via subsurface flow (mm)]	SoilDD (number of degree days over a threshold of 5 °C)
Annual (PPTANN) Summer (PPTSUM) Winter (PPTWIN)	SRAD [daily total short-wave radiation (Wm^{-2})]	Runoff [annual surface runoff (mm)]	
Annual mean daily temperature (C) (TDAYANN)		Outflow (annual total export = baseflow + runoff)	SoilW (number of days soil moisture is <5% available)*
Annual GDD above 0 °C Annual frost days		PET (mm) AET (mm) MaxSNOW (cm) SWE on April 15 (mm)	

*Soil water indices were computed at three depths in the soil layer: 0–10, 10–40 and 40–100 cm.
GDD, growing degree days; SRAD, solar radiation; PET, potential evaporative water loss; AET, actual evaporative water loss; MaxSNOW, maximum winter snowpack accumulation; SWE, spring snowpack water equivalent; SoilDD, soil degree days; SoilW, soil water index. Summer = June, July, August, September. Winter = December, January, February. 'Degree days' = the sum of degrees over a threshold for all days whose (mean) temperature exceeds that threshold.

Orthogonal polynomials were created to avoid collinearity in the predictors, and a quadratic term was accepted only if, in addition to the criteria applied to all predictors, it had a negative coefficient. (A positive quadratic term produces a concave-up response, which is not biologically meaningful except in rare cases where minima are at one extreme of the range of the predictor. We did not encounter any of these). When back-transformed via the logit link function, these quadratic terms produce a Gaussian (bell-shaped) response curve.

For each species, we applied four criteria to isolate an optimal model:

- Percentage deviance explained (PDE) for the model, penalized by the number of predictors, was the maximum of all models tested. PDE, roughly equivalent to R^2 in a regression model, is calculated as 1 − (residual deviance/ null deviance). Thus, we used the equivalent of adjusted R^2. A likelihood ratio test was also applied to quadratic terms for predictors at a significance level of 0.05.
- A strict criterion was applied to minimize collinearity in the predictors, because collinearity leads to large variances for parameter estimates. Therefore, no individual variance inflation factor (VIF) could be >3. In cases where VIF was >3, backward elimination (Neter *et al.*, 1990) was used to drop the predictor that reduced residual deviance the least.
- A model had to explain at least 10% of the deviance in the response variable to be retained; models that did not are not reported, although significance tests might have shown an association between response and predictors. In the context of this study, we considered such models to be too weak for meaningful interpretation.
- No models were attempted for species that were present on fewer than 5% of plots on a given forest. Because of the nature of GLMs, estimated probabilities of presence are

likely to be biased when a preponderance of observations of the response variable are either 0 or 1. We made one exception: whitebark pine (*Pinus albicaulis*), a rare sub-alpine species of considerable management interest and thought to be sensitive to climatic change (Tomback *et al.*, 2001).

A key objective of this study was to identify how robust the interpretations of models were to changes in spatial extent. Thus, for species that were present in multiple forests, we built models at multiple spatial extents: (1) individual forests; (2) either the two westernmost forests, Grizzly Bear and Wenatchee, or the three easternmost forests, Wenatchee, Okanogan, and Colville; and (3) all four forests.

Model evaluation
A variety of methods have been developed to evaluate predictive models with binary responses (Hosmer & Lemeshow, 1989). To properly evaluate such models of species distributions, one must quantify not only the success of predictions via some criterion for *goodness-of-fit*, but also their ability to discriminate between occupied and unoccupied sites (Guisan & Zimmermann, 2000). For the former we used the residual deviance (first criterion, above), and for the latter we used a combination of bootstrapping (Efron & Tibshirani, 1993) and receiver operating characteristic (ROC) curves (Fielding & Bell, 1997; Pearce & Ferrier, 2000a). ROC curves permit evaluation of a model's accuracy of discrimination between occupied and unoccupied sites for any threshold probability of occurrence (between 0 and 1), and are created in the following way.

On either side of any predicted threshold probability of occurrence, there will be observations that are predicted correctly and incorrectly. For example, for a threshold probability of 0.5, there will be sites with predicted values

above it where the species is present (true positives) and absent (false positives). Similarly, there will be sites with predicted values below 0.5 that are occupied (false negatives) and unoccupied (true negatives). For a given state (present or absent), the number of correct predictions plus the number of incorrect predictions equals the total number of observations in that state. For a specified threshold probability, the 'true positive fraction' is the proportion of correctly predicted presences, the 'false negative fraction' the proportion of incorrectly predicted absences, the 'true negative fraction' the proportion of correctly predicted absences, and 'false positive fraction' the proportion of incorrectly predicted presences. These fractions are in the range (0,1), true positive fraction + false negative fraction = 1, and true negative fraction + false positive fraction = 1 (Murphy & Winkler, 1987; Pearce & Ferrier, 2000a).

A model's ability to discriminate (i.e. identify correctly sites that are occupied) is expected to differ for different values of the 'decision threshold'. ROC curves allow multiple thresholds to be examined simultaneously. For a range of decision thresholds within the predicted values of a model, one can calculate any of the four fractions (above). Typically, the false positive and true positive values are used, and plotted on the X axis (range 0,1), and the Y axis (range 0,1), respectively (Pearce & Ferrier, 2000a). The smooth curve defined by these points is the ROC curve. The area under this curve corresponds to the ability of the model to discriminate between presences and absences. With no ability to discriminate, the expected true positive fraction would equal the expected false positive fraction, and the area under the ROC curve would be the area under the 45° line (= 0.5). With perfect discrimination, the area would be 1.0.

It can be shown (Bambar, 1975; Hanley & McNeil, 1982) that the area under an ROC curve is approximated by the Mann–Whitney statistic (from a non-parametric test of the difference between two samples), when this statistic is standardized by dividing by the product of the two samples sizes (Sokal & Rohlf, 1995). The result of the standardization lies between 0.5 (no discrimination) and 1.0 (perfect discrimination). Values below 0.7 show 'poor' discriminating ability, values between 0.7 and 0.9 show 'reasonable' ability, and values above 0.9 show 'excellent' ability (Swets, 1988; Pearce & Ferrier, 2000a).

Because of the large number of models and computer-intensive nature of bootstrapping with thousands of data points, we selected fourteen models at random (two from each forest and each combination of forests – Wenatchee/ Grizzly, Wenatchee/Okanogan/Colville, and all four forests), with the constraint that no more than two were from the same species. We calculated the area under ROC curves for these models via the standardized Mann–Whitney statistic, and then applied the models to obtain fitted values (probabilities) for 1000 bootstrap samples of the response variable. Standard bootstrap estimates of prediction error for linear models produce a statistic, 'error optimism', which reflects the percentage increase in residual deviance expected if the model were extrapolated to other data from a similar population (Efron & Tibshirani, 1993). We used the areas under

bootstrapped ROC curves (developing ROC curves from the bootstrap samples and predicted values), rather than residual deviance, to estimate error optimism as the mean percentage decrease in area under the ROC curves from the bootstrap predictions relative to predictions on the original data set:

$$E_{\text{opt}} = (A_{\text{data}} - A_{\mu\text{boot}})/A_{\text{data}}$$

where E_{opt} = error optimism, A_{data} = area under ROC curve computed from the model, and $A_{\mu\text{boot}}$ = the mean area under ROC curves computed from bootstrap resamples.

Using the models selected for evaluation, GAMs were estimated, to compare the quadratic term, corresponding to a Gaussian response, to other non-linear terms (e.g. higher-order polynomials). We replaced the quadratic terms with a stiff loess smoother (Hastie & Tibshirani, 1990) and examined the shape of response curves. Substantial departure from a bell-shaped curve in the non-parametric terms in GAMs, specifically skewness or bimodality, would suggest the need for more complex models to accurately estimate the relationship to explanatory variables (Austin & Gaywood, 1994). Finally, we plotted empirical variograms of model residuals, to check graphically for spatial auto-correlation in error terms. Monotone-increasing variograms would suggest that sites were more similar, on average, to their neighbours than to distant sites, and that incorporation of a spatial component would improve the models.

RESULTS

Climate and biophysical variables differ noticeably in their ranges among forests (Table 2), with wider ranges, suggesting steeper environmental gradients, generally found within the Grizzly Bear and Wenatchee forests. Conifer species vary in their constancy within forests and their distributions among forests. Eight species are found in all four forests (Table 3), although two of these, mountain hemlock and western hemlock, have very low constancy in one and two forests, respectively. None of the species is restricted to one forest, although whitebark pine has low constancy in the two forests in which it was found (Table 3). Douglas-fir is the most common species on all four forests, occurring on more than 50% of plots in the Wenatchee, Okanogan, and Colville forests.

Model results
Percentage deviance explained was >10% in fifty-three of sixty-one total attempts to fit models at the three spatial scales. All of the eight failures involved the same four species: Alaska yellow cedar, western larch, lodgepole pine and Engelmann spruce. PDE for the fifty-three models ranged from 10.7% to 51.4%, with a mean of 29.6%. PDE was highest, on average, for the models for the Grizzly bear and Wenatchee, and for these two forests combined; it was lowest on the Colville (the easternmost forest), and surprisingly, at larger spatial scales on the combinations of the three eastern forests and all four forests.

Of the predictors, the climate variables were most frequently used, with at least one (and frequently two) appearing in 79% of the models (Table 4). The other classes

Table 2 Mean values and ranges of climatic and biophysical variables on the four forests. See Table 1 for units

Variable	Wenatchee	Okanogan	Colville	Grizzly Bear
Annual precipitation	133.78	83.96	91.32	161.48
	41.67–326.23	39.18–186.05	35.99–160.68	22.70–344.01
Summer precipitation	14.48	14.80	23.11	21.55
	5.46–41.30	6.60–23.98	11.09–37.84	3.67–51.06
Winter precipitation	57.66	32.37	26.77	64.83
	18.04–128.65	11.63–74.36	9.37–48.56	9.04–133.96
Mean annual temperature	4.72	3.21	4.56	3.52
	−1.66–8.94	−2.41–8.44	0.34–8.44	−2.14–11.08
Growing degree days	2237.14	2046.72	2326.22	2031.09
	1096.89–3496.51	1022.69–3436.81	1364.39–3403.80	980.81–4255.78
Frost days	200.77	231.51	199.12	216.05
	136.72–312.16	154.93–313.18	106.83–263.44	59.09–312.31
Baseflow	636.20	268.20	276.01	873.00
	23.00–1552.00	16.00–1221.00	15.00–697.00	12.00–2183.00
Runoff	221.49	129.99	87.07	440.16
	15.00–986.00	12.00–836.00	12.00–391.00	5.00–1573.00
Outflow	857.70	398.19	363.07	1313.16
	39.00–2328.00	28.00–1863.00	28.00–1021.00	17.00–3581.00
PET	585.14	615.95	641.71	525.40
	343.00–963.00	328.00–884.00	401.00–873.00	283.00–1186.00
AET	439.24	420.90	503.08	421.79
	296.00–575.00	295.00–560.00	350.00–630.00	189.00–670.00
Maximum winter snowpack	448.36	360.91	256.34	760.18
	50.00–1620.00	22.00–1431.00	23.00–763.00	12.00–5276.00
Snow-water equivalent	320.59	259.27	141.20	664.45
	0.00–1572.00	0.00–1397.00	0.00–696.00	0.00–5217.00
Soil degree days	1077.18	1046.48	1172.46	848.77
	213.00–2703.00	205.00–2188.00	332.00–2173.00	96.00–2884.00
Soil drought days (0–10 cm)	105.99	120.01	100.05	77.64
	17.00–202.00	26.00–186.00	46.00–179.00	4.00–244.00
Soil drought days (10–40 cm)	108.21	131.92	110.47	76.24
	7.00–216.00	18.00–215.00	41.00–210.00	1.00–270.00
Solar radiation	13.01	13.21	12.73	12.55
	10.97–14.40	10.83–14.22	6.31–13.85	6.41–14.30

Table 3 Percentage of plots present for each species on each forest. Numbers in parentheses are the total number of plots on the forest

Species	Grizzly Bear (1416)	Wenatchee NF (4105)	Okanogan NF (1975)	Colville NF (3157)
Douglas-fir (PSME) *Pseudotsuga menziesii* (Mirb.) Franco	24.2	65.2	66.4	69.5
Ponderosa pine (PIPO) *Pinus ponderosa* Dougl. ex Laws.	9.0	27.0	28.1	16.8
Subalpine fir (ABLA2) *Abies lasiocarpa*	21.5	28.6	30.6	30.9
Pacific silver fir (ABAM) *Abies amabilis* Dougl. ex Forbes	18.9	29.0	3.7	–
Western redcedar (THPL) *Thuja plicata* Donn ex D. Don	9.8	12.3	2.1	32.9
Western hemlock (TSHE) *Tsuga heterophylla* (Raf.) Sarg.	12.8	22.3	1.0	18.7
Mountain hemlock (TSME) *Tsuga mertensiana* (Bong.) Carriere	17.2	17.6	1.7	0.1
Western larch (LAOC) *Larix occidentalis* Nutt.	–	13.6	20.6	52.4
Lodgepole pine (PICO) *Pinus contorta* Dougl. Ex Loud.	8.1	16.1	32.0	35.5
Whitebark pine (PIAL) *Pinus albicaulis* Engelm.	4.1	5.5	–	–
Engelmann spruce (PIEN) *Picea engelmannii* var. glabra Goodman	10.7	17.1	37.0	26.7
Grand fir (ABGR) *Abies grandis* (Dougl. ex D. Don) Lindl	–	34.6	–	26.0
Western white pine (PIMO) *Pinus monticola* var. minima Lemmon	–	–	–	7.0
Alaska yellow cedar (CHNO) *Chamaecyparis nootkatensis* (D. Don) Spach	5.2	5.1	–	–

of predictors occurred with nearly equal frequency: physical variables (49%), soil variables (42%), and hydrological variables (40%). In two forests, Okanogan and Colville, climatic variables were included in models for every species and the physical variables for every species but one, whereas the soil and hydrological variables were very sparsely

Table 4 Generalized linear model results. Signs of coefficients are in parentheses. See Table 3 for species names

Forest	Species	Climate (DAYMET)	Physical	Soil (VIC)	Hydrology (VIC)	PDE	ROC
Grizzly bear	ABAM	GDD*	–	–	Baseflow*	26.5	0.835
Mean PDE = 33.8	ABLA2	GDD* PPTSUM*	–	–	–	19.2	0.786
Mean ROC = 0.877	PIAL	TDAYANN*	–	–	–	32.7	0.915
	PICO	PPTWIN*	–	–	AET*	28.4	0.847
	PIEN	PPTWIN* GDD*	–	–	–	22.6	0.818
	PIPO	–	–	SoilDD* SoilW*	–	47.2	0.941
	PSME	PPTSUM*	–	SoilDD*	–	34.9	0.877
	THPL	PPTWIN*	SRAD*	–	AET*	42.8	0.918
	TSHE	–	–	SoilDD*	PET-AET* Baseflow(+)	47.6	0.937
	TSME	PPTWIN*	SRAD(−)		AET*	35.9	0.897
Wenatchee	ABAM	–	–	SoilW*	SWE(−)	45.3	0.916
Mean PDE = 31.3	ABGR	Frost days(−)	–	SoilW*	–	28.9	0.841
	ABLA2	–	–	SoilW* SoilDD(−)	–	32.8	0.866
Mean ROC = 0.860	CHNO	Frost days*	–		Baseflow*	19.7	0.838
	LAOC	GDD*	Aspect(+)	–	Runoff*	23.8	0.834
	PIAL	TDAYANN*	–	–	MaxSNOW*	51.4	0.957
	PICO	PPTWIN* TDAYANN*	–		–	13.2	0.757
	PIEN	PPTANN* TDAYANN*	–	–	–	16.2	0.776
	PIPO	PPTSUM* TDAYANN*	SRAD(+)	–	–	39.9	0.898
	PSME	PPTSUM(−)	–	–	–	30.2	0.836
	THPL	–	SRAD(−)	SoilDD*	AET(+)	25.7	0.850
	TSHE	PPTANN*	–	SoilDD*	PET-AET*	37.0	0.890
	TSME	PPTWIN*	–	– SoilDD*	PET-AET*	43.2	0.917
Okanogan	ABLA2	PPTANN* TDAYANN*	Aspect(+)	–	–	22.2	0.820
Mean PDE = 28.2	LAOC	PPTWIN* TDAYANN*	Aspect(+)	–	–	44.3	0.896
Mean ROC = 0.832	PICO	TDAYANN* PPTWIN*	Aspect(+)	–	–	12.6	0.756
	PIEN	PPTSUM* GDD (−)	Aspect(+)	–	–	18.5	0.785
	PIPO	GDD* PPTSUM(−)	Aspect(−)	–	–	39.6	0.896
	PSME	GDD*	–	SoilW*	–	32.2	0.836
Colville	ABGR	PPTWIN* Frost days*	–	–	–	28.9	0.839
Mean PDE = 22.3 Mean ROC = 0.805	ABLA2	TDAYANN(−) PPTANN(+)	Aspect(+)	–	–	27.3	0.833
	PIMO	PPTWIN*	SRAD(+)	–	AET(+)	12.9	0.757
	PIPO	TDAYANN(+) PPTSUM(−)	SRAD(+) Aspect(−)			24.5	0.843
	PSME	TDAYANN* PPTWIN(−)	Aspect(−)		–	11.8	0.710
	THPL	PPTANN*	SRAD(−)	SoilDD*	AET*	24.4	0.818
	TSHE	PPTWIN* Frost days*	Aspect(+)	–	–	26.3	0.837
Grizzly & Wenatchee	ABAM	PPTWIN*	–	SoilDD*	–	41.2	0.903
	ABLA2	TDAYANN*	–	–	SWE*	26.2	0.838
Mean PDE = 33.8	PIAL	TDAYANN*	SRAD(+)	–	PET-AET*	36.1	0.908

Table 4 *continued*

Forest	Species	Climate (DAYMET)	Physical	Soil (VIC)	Hydrology (VIC)	PDE	ROC
	PICO	PPTWIN*	–	–	PET-AET*	15.9	0.777
Mean ROC = 0.870	PIPO	–	SRAD*	SoilW*	PET*	42.5	0.904
	PSME	GDD*	–	SoilW*	–	36.1	0.863
	THPL	–	SRAD*	SoilDD*	AET*	29.3	0.852
	TSHE	TDAYANN*	SRAD*	–	Baseflow* AET(+)	37.1	0.882
	TSME	PPTWIN*	Aspect(+)	SoilDD*	–	39.7	0.907
Wenatchee Okanogan Colville	ABLA2	–	SRAD*	SoilDD(–) SoilW*	–	26.3	0.833
Mean PDE = 24.9 Mean ROC = 0.815	LAOC	–	Aspect(+)	–	PET-AET* Runoff*	16.7	0.770
	PIPO	GDD*	SRAD(+) Aspect(–)	SoilW*	–	34.7	0.876
	PSME	GDD*	SRAD(+)	SoilW*	–	21.9	0.782
All 4 forests	ABLA2	–	–	SoilDD* SoilW*	–	22.2	0.815
Mean PDE = 24.15 Mean ROC = 0.810	PIEN	TDAYANN*	–	–	Baseflow*	10.7	0.723
	PIPO	GDD*	SRAD(+) Aspect(–)	SoilW*	–	36.3	0.888
	PSME	GDD*	SRAD(+)	SoilW*	–	27.4	0.812

*Negative quadratic term, suggesting unimodal response. See Table 1 for definitions of variables and Table 3 for species names and botanical authorities.

PDE, percentage deviance explained; ROC, area under ROC curve (see text for details).

Species with PDE <10% are not reported. Further details on the models (e.g. parameter estimates, partial R^2, summary statistics for fitted values) are available from the senior author.

represented (Table 4). PDEs for these forests were lower than for the Wenatchee and Grizzly Bear, in which there was more representation of the soil and hydrological variables, and less of the physical variables (Table 4).

Optimal models were not substantially superior, for most species, to alternatives. In a number of cases, there were several other models that were close in explanatory power to the optimal model (PDE within 1% of the optimal model for the same number of predictors). These sub-optimal models always involved replacement of one predictor in the temperature or moisture category with another. This is not surprising because most variables within these categories were strongly correlated (Pearson's R-statistic $>\pm0.85$).

For the majority of species, we were able to fit variables from both the moisture and temperature categories of predictors, and in all but a few cases these models included negative quadratic terms, indicating a unimodal (concave down) response of species occurrence to a climatic or biophysical variable (Table 4). Models varied in how 'centered' this unimodal response was in the space of predictor variables. Some indicated that the 'environmental niche' of a species, with respect to one predictor, had been clearly captured by the model, in that the range of non-zero predicted probabilities is within the range of the predictor variable (Fig. 2). Others suggested that a longer gradient in the predictor variable(s) would be needed (Fig. 2).

Conifer species were fairly consistent in their responses across the four forests and at different spatial scales. For example, from models at three scales, Douglas-fir is predicted to be most likely to occur when growing degree days are

between 2500 and 3000, and soil drought days are between 100 and 150 (Fig. 3). Similarly, models at two spatial scales predict a sharp increase in the occurrence of mountain hemlock as winter precipitation approaches 80 cm (Fig. 4). The models for these two species had relatively high explanatory power (PDE range = 27.4 – 43.2%). For species whose models had low explanatory power, responses across scales were less consistent. For example, the peak response of lodgepole pine to winter precipitation is predicted to be at about 45 cm in the Wenatchee forest, but only about 30 cm when the Wenatchee and Grizzly bear forests are combined.

Model evaluation

Areas under ROC curves (ROCs) ranged from 0.710 for Douglas-fir on the Colville forest to 0.941 for ponderosa pine on the Grizzly Bear forest (Table 4). Spearman's rank correlation between PDEs and ROCs for the fifty-three models was 0.8; thus, discriminating ability was not quite a monotonic function of goodness-of-fit. Of the fifty-three ROCs, 79% (forty-two) were in the reasonable category of discriminating ability, and 21% (eleven) were in the excellent category. Error optimism as predicted by bootstrapped ROCs was essentially zero for most models, and never above 2.4% (Table 5).

Substitution of non-parametric terms for the polynomial terms in the models did not suggest substantial departures from unimodality, nor did they suggest substantial skewness when examined visually (not shown). Empirical variograms of model residuals (not shown) were not monotonic, that is, residuals nearer to each other in space were no more similar
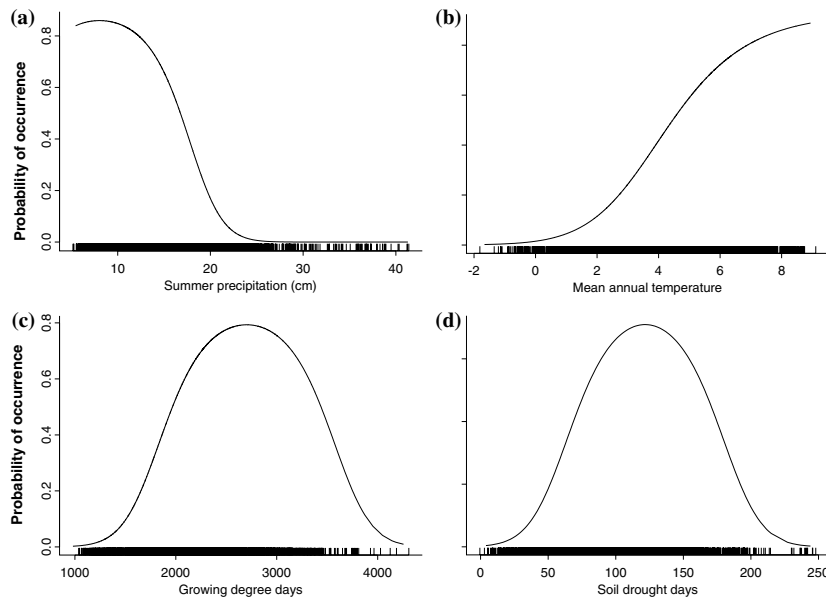
Figure 2 Unimodal/Gaussian responses are predicted by the models. (a) and (b) ponderosa pine on the Wenatchee NF. (c) and (d) Douglas-fir on the Wenatchee and Grizzly Bear forests. Density bands along the *X*-axis represent individual values of the predictor variables.

than widely separated residuals. Thus, we concluded that spatial autocorrelation was not a significant factor in the models and that observations could be considered independent.

DISCUSSION

We used statistical models to identify and quantify the environmental niche space occupied by fourteen conifer species in the Pacific Northwest. In several cases, one predictor variable proved to be optimal for a species at multiple scales (e.g. Fig. 5), and the similarities among response curves in different forests or at different scales suggest that the models are robust and accurately portray the response of that species across the region. In other cases, optimal predictors for a species came from different members of proxy sets, but the strong collinearity within sets and the similarity of response curves across scales and forests suggest that using a single predictor from the set would produce near-optimal models and an equally robust interpretation. For example, every model for ponderosa pine (present on all forests) includes a term from the temperature set; either growing degree-days, annual temperature, or soil degree-days.

We expected the strength and robustness of models to vary among species for two reasons: (1) adequacy of samples in representing populations, and (2) more importantly, the resolution of predictor variables at spatial scales that match species responses. For example, models for two species, lodgepole pine and Engelmann spruce, were consistently poor (Table 4), even in forests in which they were moderately abundant. Lodgepole pine is abundant on sites where conditions such as shallow, low-fertility soils and frequent summer frosts exclude the establishment of other conifers (Steele *et al.*, 1981). Neither of these limiting factors is represented by our climatic and biophysical variables; they

are both likely to be heterogeneous at finer scales than our data. The habitat of Engelmann spruce is difficult to characterize because, although it is generally associated with cold environments at high elevations (Alexander & Shepperd, 1990), it is found at middle elevations on cold-air drainages, alluvial terraces, and wet benches (Daubenmire, 1969; Pfister *et al.*, 1977). Factors associated with its distribution are, therefore, like those for lodgepole pine, likely to be heterogeneous at scales too fine for our models.

In contrast, models for two other species, mountain hemlock and ponderosa pine, were consistently strong (Table 4). The key predictor for mountain hemlock models was winter precipitation. A detailed study of the response of mountain hemlock to climatic variability identified snowpack in May as a key factor limiting growth (Peterson & Peterson, 2001). This is likely a function of total precipitation during winter months, so PPTWIN is an effective surrogate (our variable MaxSNOW, maximum snowpack, is less so). Similarly, the environmental niche space of ponderosa pine may be sharply defined on both ends of the moisture gradient (Kusnierczyk & Ettl, 2002), as suggested by key predictors PPTSUM and SoilW (Table 4). In the Pacific Northwest, ponderosa pine dominates lower treeline east of the Cascade Range, but at higher moister sites it is out-competed by more shade-tolerant, less drought-tolerant species (Franklin & Dyrness, 1988).

If we have identified variables that directly influence the mechanisms of a species' survival and growth, then we would expect that statistical associations with species occurrence would not change substantially with geographic distance, composition of co-occurring species, or spatial scale. The robustness of models among forests and across scales implies that the responses of these species to climate and biophysical factors are relatively homogeneous across two contrasting physiographic types (see Study Area) and a number of
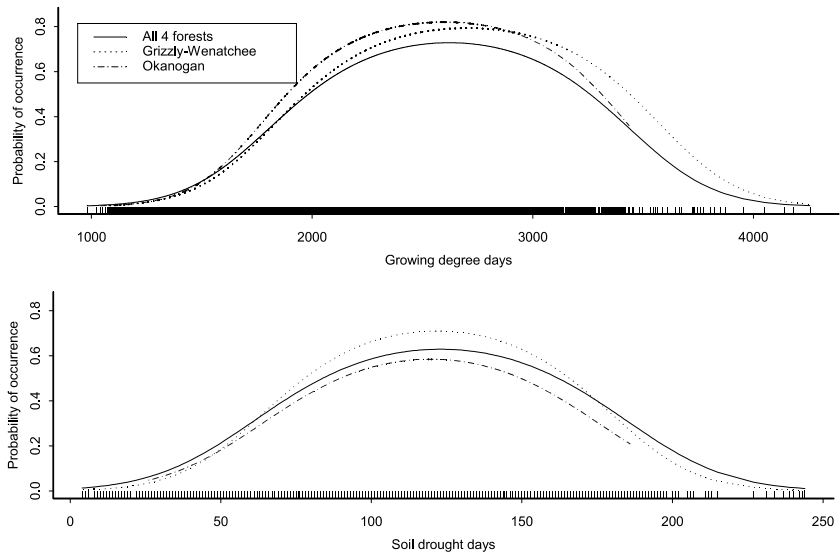
Figure 3 The response of Douglas-fir to two predictor variables is consistent across spatial scales. Density bands along the *X*-axis represent individual values of the predictor variables.
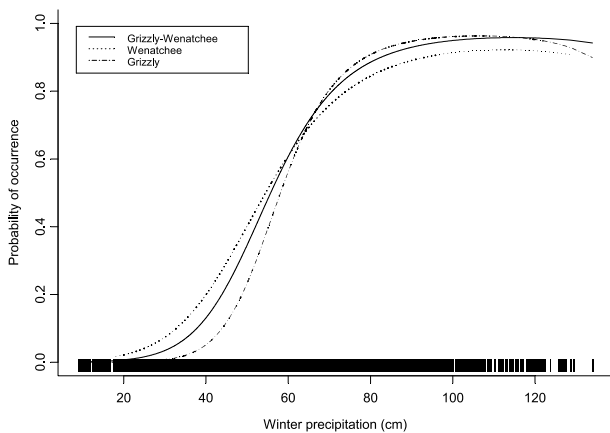

Figure 4 Response of mountain hemlock to winter precipitation is consistent at two scales along a geographic gradient. Density bands along the *X*-axis represent individual values of the predictor variables.

vegetation types, or plant associations (Williams *et al.*, 1990; Lillybridge *et al.*, 1995). We have not explained all the variance in species responses. The remainder of the variation is probably because of one or more of the following: competitive interactions among species that share environmental niches can cause one species to be eliminated from a site where a broad-scale model based on climate would predict it to be present. Local or small-scale effects, such as noted above for lodgepole pine and Engelmann spruce, are unlikely to be captured in models built at 1-km resolution. The inertia of successional processes may allow mature individuals of a residual species from a earlier climatic regime to remain on a site long after a model would predict their absence (e.g. Dunwiddie, 1986; Campbell & McAndrews, 1993). Finally, there is intrinsic variability in all ecological processes. Nevertheless, we believe that our predictive models can be applied, where appropriate, with a high level of confidence.

Accurate forecasting of the response of vegetation to climatic change depends upon the robustness of models to departures from the biophysical environment in which they were developed. One approach to this is to build 'process-based' models, under the assumption that the simulated mechanisms are so basic as to be invariant to those departures associated with climatic change (Shugart & Prentice, 1992; Running & Hunt, 1993; Urban *et al.*, 1993; Neilson, 1995; Acevedo *et al.*, 1996; Keane *et al.*, 1996a). A key advantage of process-based simulations is that they incorporate transient dynamics, which are clearly important for forecasting the responses of long-lived species. In contrast, empirical/statistical models like ours and many others (e.g. Franklin, 1995; Bolliger *et al.*, 2000b; Pearce & Ferrier, 2000b; Iverson & Prasad, 2001) take a static, or equilibrium approach. Associations are modeled at one point in time; their genesis must be inferred indirectly.

A disadvantage of process-based models is that they have multiple parameterizations, often based on statistical associations though grounded in biological theory (Korzukhin *et al.*, 1996). Models then need to be calibrated to produce 'reasonable' output. In contrast, empirical models are data-driven and self-calibrating, though still needing careful validation. Our models are robust across scales and geographical gradients, and the bootstrap validations suggest no loss of predictive power when applied to similar populations. Although there is substantial intrinsic uncertainty in empirical models (Elith *et al.*, 2002), they can be of considerable value, along with process-based models, in forecasting changes in species distributions in response to climatic change. For example, once a realistic climate-change scenario has been identified (e.g. Ferguson, 1997), the DAYMET and VIC GIS coverages could be adjusted stochastically and the predictive models applied to new values of the climate and biophysical variables. Sensitivity analysis would identify which species are sensitive to

Forest	Species	PDE	ROC	Bootstrapped ROC	Error optimism (%)
Grizzly Bear	ABLA	0.192	0.786	0.791	0.0
	TSME	0.359	0.897	0.880	2.4
Wenatchee	ABAM	0.453	0.916	0.917	0.0
	CHNO	0.197	0.838	0.834	0.5
Okanogan	LAOC	0.443	0.896	0.911	0.0
	PICO	0.126	0.756	0.753	0.4
Colville	ABGR	0.289	0.839	0.832	1.0
	PSME	0.118	0.710	0.694	2.3
Grizzly and Wenatchee	PIPO	0.425	0.904	0.909	0.0
	PICO	0.159	0.777	0.774	0.4
Wenatchee/Okanogan/Colville	PIPO	0.347	0.876	0.855	2.4
	LAOC	0.167	0.770	0.767	0.4
All 4 forests	PSME	0.274	0.812	0.805	0.9
	PIEN	0.107	0.723	0.728	0.0

Table 5 Bootstrap validation statistics for selected models. Two species were selected from each forest with contrasting PDEs (high and low) and to maximize the diversity of species
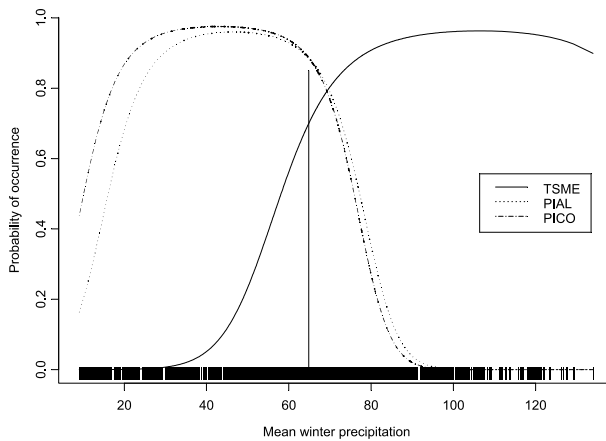

Figure 5 Responses of three species to winter precipitation on the Grizzly Bear forest. The vertical line is the current mean. Density bands along the *X*-axis represent individual values of the predictor variables.

changes in which variables, and in which geographical portion of their ranges they are most sensitive. Because our models were built at multiple scales, responses to climatic change could be forecast at multiple scales.

Recent climate is not a perfect surrogate for the climatic conditions under which now-mature trees established; this disparity between temporal scales (17–50 years of high-quality gridded climatic data vs. multi-century life spans of many conifer species) will always lend uncertainty to statistical models of the distributions of long-lived species. It is likely that environmental niches for regeneration are narrower than those for survival of mature trees. However, models of presence/absence are more likely to approximate regeneration niches than model of abundance, because 'presence' is directly linked to establishment success. It is reasonable to assume, therefore, that core areas of environmental space (high probability of presence) predicted by our

models are relatively robust to life-history stages of the organisms being considered.

The application of climate-change scenarios help to identify where species ranges may expand, shrink, or move. For example, suppose winter precipitation were to increase by 20 cm year^{-1}, on average, in the Grizzly Bear forest (straddling the North Cascades, at high elevation). Models predict that the occurrence of whitebark pine and lodgepole pine would be greatly reduced, whereas mountain hemlock occurrence would increase (Fig. 5). This, in conjunction with an introduced pathogen, white pine blister rust *Cronartium ribicola*) and fire exclusion, poses a significant threat to the persistence of whitebark pine (Tomback *et al.*, 2001). The identification of vulnerable species and development of possible mitigations are key aspects of forest planning (Flather *et al.*, 1998; Hann *et al.*, 1998; Hessburg *et al.*, 1999; Bolliger *et al.*, 2000b; Aber *et al.*, 2001; Hansen *et al.*, 2001).

A more comprehensive picture of potential response of conifer species to climatic change involves considering limits to regeneration, limits to abundance (as opposed to presence/absence), competitive interactions, and the effects of disturbance, chiefly fire. Environmental niche space is likely to change with different stages of life history (seedlings vs. mature trees). Methodology exists for modelling the effects of predictor variables on species composition, as a multivariate response, directly (Billheimer *et al.*, 2001; Cumming, 2001), thus implicitly incorporating competitive interactions, and for estimating the limits of species abundance or growth in response to multiple predictors (Scharf *et al.*, 1998; Cade *et al.*, 1999; McKenzie *et al.*, 2000a; Peterson & Peterson, 2001). Characteristics of fire regimes have been described and quantified at broad scales in the Pacific Northwest (Morgan *et al.*, 1996; McKenzie *et al.*, 2000b), and rule-based models, incorporating climate, have been developed to estimate the effects of fire regimes on physiognomic vegetation types (Lenihan *et al.*, 1998) and successional pathways (Keane *et al.*, 1996b). Process-based models simulate fire, succession, and forest composition at small

scales (e.g. Keane *et al.*, 1996a). A fair amount is known about the linkages between climate and fire regimes (Swetnam & Betancourt, 1990; Swetnam, 1993; Veblen *et al.*, 2000; Heyerdahl *et al.*, 2001). However, to date no quantitative synthesis exists of the linkages among species composition and distribution, climate and biophysical factors, and effects of disturbance. The models presented in this paper provide a key component of that synthesis.

ACKNOWLEDGMENTS

REFERENCES

Aber, J., Neilson, R.P., McNulty, S., Lenihan, J.M., Bachelet, D. & Drapek, R.J. (2001) Forest processes and global environmental change: predicting the effects of individual and multiple stressors. *BioScience*, **51**, 735–751.

Acevedo, M.F., Urban, D.L. & Shugart, H.H. (1996) Models of forest dynamics based on roles of tree species. *Ecological Modelling*, **87**, 267–284.

Alexander, R.R. & Shepperd, W.D. (1990) *Picea engelmannii* Parry ex Engelm. Engelmann spruce. *Silvics of North America: Volume 1. Conifers* (tech. coord. Burns, R.M. and B.H. Honkala), pp. 187–203. Agricultural Handbook 654. USDA Forest Service, Washington, DC, USA.

Alexander, R.R., Shearer, R.C. & Shepperd, W.D. (1984) *Silvical characteristics of subalpine fir*. USDA Forest Service General Technical Report RM–115, p. 29. Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO, USA.

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

Austin, M.P. (1999) The potential contribution of vegetation ecology to biodiversity research. *Ecography*, **22**, 465–484.

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157**, 101–118.

Austin, M.P. & Gaywood, M.J. (1994) Current problems of environmental gradients and species response curves in relation to continuum theory. *Journal of Vegetation Science*, **5**, 473–482.

Austin, M.P., Nicholls, A.O., Doherty, M.D. & Meyers, J.A. (1994) Determining species response functions to an environmental gradient by means of a ß-function. *Journal of Vegetation Science*, **5**, 214–228.

Bachelet, D., Neilson, R.P., Lenihan, J.M. & Drapek, R.J. (2001) Climate change effects on vegetation distribution and carbon budget in the United States. *Ecosystems*, **4**, 164–185.

Baker, W.L. (1995) Longterm response of disturbance landscapes to human intervention and global change. *Landscape Ecology*, **10**, 143–159.

Bambar, D. (1975) The area above the ordinal dominance graph and the area below the receiver operating graph. *Journal of Mathematical Psychology*, **12**, 387–415.

Billheimer, D., Guttorp, P. & Fagan, W.F. (2001) Statistical interpretation of species composition. *Journal of the American Statistical Association*, **96**, 1205–1214.

Bolliger, J., Kienast, F. & Bugmann, H. (2000a) Comparing models for tree distributions: concept, structures, and behavior. *Ecological Modelling*, **134**, 89–102.

Bolliger, J., Kienast, F. & Bugmann, H. (2000b) Risks of global warming on montane and subalpine forests in Switzerland – a modeling study. *Register of Environmental Change*, **1**, 99–111.

Booth, G.D., Niccolucci, M.J. & Schuster, E.G. (1994) *Identifying proxy sets in multiple linear regression: an aid to better coefficient interpretation*. USDA Forest Service Research Paper INT470, p. 12. Intermountain Research Station, Ogden, UT, USA.

ter Braak, C.J.F. (1987) *Unimodal models to relate species to environment*. Agricultural Mathematics Group, Wageningen, the Netherlands, p. 152.

ter Braak, C.J.F. & Prentice, I.C. (1988) A theory of gradient analysis. *Advances in Ecological Research*, **18**, 271–313.

Breiman, L., Friedman, J.H., Olshen, R. & Stone, C.J. (1984) *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA, USA, p. 358.

Bristow, K.L. & Campbell, G.S. (1984) On the relationship between incoming solar radiation and daily maximum and minimum temperature. *Agricultural and Forest Meteorology*, **31**, 159–166.

Brockway, D.G. (1998) Forest plant diversity at local and landscape scales in the Cascade Mountains of southwestern Washington. *Forest Ecology and Management*, **109**, 323–341.

Cade, B.S., Terrell, J.W. & Schroeder, R.L. (1999) Estimating effects of limiting factors with regression quantiles. *Ecology*, **80**, 311–323.

Campbell, I.D. & McAndrews, J.H. (1993) Forest disequilibrium caused by rapid Little Ice Age cooling. *Nature*, **366**, 336–338.

Cayan, D. (1996) Interannual climate variability and snowpack in the western United States. *Journal of Climate*, **9**, 928–948.

Clark, L.A. & Pregibon, D. (1992) Tree-based models. *Statistical Models in S* (ed. J.M. Chambers and T.J. Hastie), pp. 377–419. Wadsworth and Brooks, Pacific Grove, CA, USA.

Cumming, S.G. (2001) Forest type and wildfire in the Alberta boreal mixedwood: what do fires burn? *Ecological Applications*, **11**, 97–110.

Dale, V.H., Joyce, L.A., McNulty, S., Neilson, R.P., Ayres, M.P., Flannigan, M.D., Hanson, P.J., Irland, L.C., Lugo, A.E., Peterson, C.J., Simberloff, D., Swanson, F.J., Stocks, B.J. & Wotton, B.M. (2001) Climate change and forest disturbances. *BioScience*, **51**, 723–734.

Daly, C., Neilson, R.P. & Phillips, D.L. (1994) A statistical-topographical model for estimating climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, **33**, 140–158.

Daubenmire, R. (1969) Ecologic plant geography of the Pacific Northwest. *Madrono*, **20**, 111–128.

Davis, M.B. & Botkin, D.B. (1985) Sensitivity of cool-temperature forests and their fossil pollen record to rapid temperature change. *Quaternary Research*, **23**, 327–340.

Dunwiddie, P.W. (1986) A 6000-year record of forest history on Mount Rainier, Washington. *Ecology*, **67**, 58–68.

Efron, B. & Tibshirani, R.J. (1993) *An Introduction to the Bootstrap.* Chapman and Hall, New York, NY, p. 436.

Elith, J., Burgman, M.A. & Regan, H.M. (2002) Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecological Modelling*, **157**, 313–329.

Environmental Systems Research Institute (ESRI), Inc. (2000) *ARC-INFO version 8.0.2.* Redlands, CA, USA.

Ferguson, S.A. (1997) A climate-change scenario for the Columbia River Basin. USDA Forest Service Research Paper PNW-499. Pacific Northwest Research Station, Portland, OR, USA.

Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

Flather, C.H., Knowles, M.S. & Kendall, I.A. (1998) Threatened and endangered species geography. *BioScience*, **48**, 365–376.

Franklin, J. (1995) Predictive vegetation mapping: geographical modeling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, **19**, 474–499.

Franklin, J. (1998) Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *Journal of Vegetation Science*, **9**, 733–748.

Franklin, J.F. & Dyrness, C.T. (1988) *Natural vegetation of Oregon and Washington.* Oregon State University Press, Corvallis, OR. p. 452.

Gaines, W.L., Naney, R.H., Morrison, P.H., Ebey, J.R., Wooten, G.F. & Almack, J.A. (1990) Use of Landsat multispectral scanner imagery and geographic information systems to map vegetation in the North Cascades grizzly bear ecosystem. *International Conference on Bear Research and Management*, **9**, 533–547.

Gauch, H.G.J., Chase, G.B. & Whittaker, R.H. (1974) Ordination of vegetation samples by Gaussian species distributions. *Ecology*, **55**, 1382–1390.

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

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36.

Hann, W.J., Jones, J.L., Keane, R.E., Hessburg, P.F. & Gravenmier, R.A. (1998) Landscape dynamics. *Journal of Forestry*, **96**, 10–15.

Hansen, A., Neilson, R.P., Dale, V.H., Flather, C.H., Iverson, L.R., Currie, D.J., Shafer, S., Cook, R. & Bartlein, P.J. (2001) Global change in forests: responses of species, communities, and biomes. *BioScience*, **51**, 765–779.

Hastie, T. & Tibshirani, R. (1990) *Generalized additive models.* Chapman and Hall, London, p. 335.

He, H.S., Mladenoff, D.J. & Gustafson, E.J. (2002) Study of landscape change under forest harvesting and climate warming-induced fire disturbance. *Forest Ecology and Management*, **155**, 257–270.

Hermann, R.K. & Lavender, V. (1990) *Pseudotsuga menziesii* (Mirb.) Franco Douglas-fir. *Silvics of North America. Volume 1: Conifers* (ed. R.M. Burns and B.H. Honkala), pp. 527–540. USDA Forest Service Agricultural Handbook 654. Washington, DC, USA.

Hessburg, P.F., Smith, B.G. & Kreiter, S.G. (1999) *Historical and current forest and range landscapes in the Interior Columbia River basin including portions of the Klamath and Great Basins. Part I: Linking vegetation patterns and landscape vulnerability to potential insect and pathogen disturbances.* USDA Forest Service General Technical Report PNW-GTR-458. Pacific Northwest Research Station, Portland, OR, USA.

Heyerdahl, E.K., Brubaker, L.B. & Agee, J.K. (2001) Spatial controls of historical fire regimes: a multiscale example from the Interior West, USA. *Ecology*, **82**, 660–678.

Hosmer, D.W. & Lemeshow, S. (1989) *Applied logistic regression.* Wiley and Sons, New York, NY, USA, p. 307.

Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J. & Xiaosu, D. (2001) *Climate change 2001: the scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC).* Cambridge University Press, UK, p. 944.

Hungerford, R.D., Nemani, R.R., Running, S.W. & Coughlan, J.C. (1989) *MTCLIM: a mountain microclimate simulation model.* USDA Forest Service Res. Pap. INT-414. Rocky Mountain Research Station, Fort Collins, CO, USA.

Insightful, Inc. (2000) *Splus 2000 for Windows.* Seattle, WA, USA.

Iverson, L.R. & Prasad, A.M. (2001) Potential changes in tree species richness and forest community types following climate change. *Ecosystems*, **4**, 186–199.

Keane, R.E., Ryan, K. & Running, S.W. (1996a) *FIRE-BGC – a mechanistic ecological process model for simulating fire succession on coniferous forest landscapes of the Northern Rocky Mountains.* USDA Forest Service Research Paper INT-484. Rocky Mountain Research Station, Fort Collins, CO, USA.

Keane, R.E., Long, D.G., Menakis, J.P., Hann, W.J. & Bevins, C.D. (1996b) *Simulating coarse-scale vegetation dynamics using the Columbia River Basin Succession Model CRBSUM.* USDA Forest Service General Technical Report INT-GTR-340. Rocky Mountain Research Station, Fort Collins, CO, USA.

Keane, R.E., Morgan, P. & White, J.D. (1999) Temporal patterns of ecosystem processes on simulated landscapes in Glacier National Park, Montana, USA. *Landscape Ecology*, **14**, 311–329.

Korzukhin, M.D., ter-Mikaelian, M.T. & Wagner, R.G. (1996) Process vs. empirical models: which approach for forest ecosystem management? *Canadian Journal of Forest Research*, **26**, 879–887.

Kusnierczyk, E.R. & Ettl, G.J. (2002) Growth response of ponderosa pine (*Pinus ponderosa*) to climate in the eastern Cascade Mountains, Washington, U.S.A.: implications for climatic change. *Ecoscience*, **9**, 544–551.

Lenihan, J., Daly, C., Bachelet, D. & Neilson, R.P. (1998) Simulating broad-scale fire severity in a dynamic global

vegetation model. *Northwest Science (special issue)*, **72**, 91–103.

Liang, X., Lettenmaier, D.P., Wood, E.F. & Burges, S.J. (1994) A simple hydrologically based model of land surface water and energy fluxes for GSMs. *Journal of Geophysical Research*, **99**, 14,415–14,428.

Lillybridge, T.R., Kovalchik, B.L., Williams, C.K. & Smith, B.G. (1995) *Field guide to forested plant associations of the Wenatchee National Forest*. USDA Forest Service General Technical Report PNW-GTR-359. Pacific Northwest Research Station, Portland, OR, USA, p. 336.

McCullagh, P. & Nelder, J.A. (1989) *Generalized linear models*. Chapman and Hall, London, UK, p. 511.

McKenzie, D. & Halpern, C.B. (1999) Modeling the distributions of shrub species in Pacific Northwest forests. *Forest Ecology and Management*, **114**, 293–307.

McKenzie, D., Peterson, D.L. & Alvarado, E. (1996) *Predicting the effect of fire on large-scale vegetation patterns in North America*. USDA Forest Service Research Paper PNW-489. Pacific Northwest Research Station, Portland, OR, USA.

McKenzie, D., Halpern, C.B. & Nelson, C.R. (2000a) Overstory influences on herb and shrub communities in mature forests of western Washington, USA. *Canadian Journal of Forest Research*, **30**, 1655–1666.

McKenzie, D., Peterson, D.L. & Agee, J.K. (2000b) Fire frequency in the Columbia River Basin: building regional models from fire history data. *Ecological Applications*, **10**, 1497–1516.

Maurer, E.P., Wood, A.W., Adam, J.C., Lettenmaier, D.P. & Nijssen, B. (2002) A long-term hydrologically-based data set of land surface fluxes and states for the conterminous United States. *Journal of Climate*, **15**, 3237–3251.

Moisen, G.G. & Frescino, T.S. (2002) Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling*, **157**, 209–225.

Morgan, P., Bunting, S.C., Black, A.E., Merrill, T. & Barrett, S. (1996) *Fire regimes in the interior Columbia River basin: past and present*. Report on file at USDA Forest Service Intermountain Fire Sciences Laboratory, Rocky Mountain Research Station, Missoula, MT, USA.

Murphy, A.H. & Winkler, R.L. (1987) A general framework for forecast verification. *Monthly Weather Review*, **115**, 1330–1338.

Neilson R.P. (1995) A model for predicting continental-scale vegetation distribution and water balance. *Ecological Applications*, **5**, 362–385.

Neilson, R.P. & Drapek, R.J. (1998) Potentially complex biosphere responses to transient global warming. *Global Change Biology*, **4**, 505–521.

Nemani, R.R., Running, S.W., Band, L.E. & Peterson, D.L. (1993) Regional hydroecological simulation system: an illustration of the integration of ecosystem models in a GIS. *Environmental modeling with GIS* (eds M.F. Goodchild, B.O. Parks and L.T. Steyaert), pp. 296–304. Oxford University Press, Oxford.

Neter, J., Wasserman, W. & Kutner, M.H. (1990) *Applied linear statistical models*. Irwin Press, Homewood, IL, USA, p. 1181.

Ohmann, J.L. & Spies, T.A. (1998) Regional gradient analysis and spatial pattern of woody plant communities of Oregon forests. *Ecological Monographs*, **68**, 151–182.

Oksanen, J. & Minchin, P.R. (2002) Continuum theory revisited: what shape are species responses along ecological gradients. *Ecological Modelling*, **157**, 119–129.

Pearce, J. & Ferrier, S. (2000a) An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, **128**, 127–147.

Pearce, J. & Ferrier, S. (2000b) Assessing the predictive performance of wildlife-habitat models developed using logistic regression. *Ecological Modelling*, **133**, 225–245.

Peterson, D.L. (1998) Climate, limiting factors, and environmental change in high-altitude forests of Western North America. *The impacts of climate variability on forests* (ed. M. Beniston and J.L. Innes), pp. 191–208. Springer, New York, NY, USA.

Peterson, D.W. & Peterson, D.L. (2001) Mountain hemlock growth responds to climatic variability at annual and decadal time scales. *Ecology*, **82**, 3330–3345.

Pfister, R.D., Kovalchik, B.L., Arno, S.F. & Presby, R.C. (1977) *Forest habitat types of Montana*. USDA Forest Service General Technical Report INT-34. Intermountain Forest and Range Experiment Station, Ogden, UT, USA, p. 174.

Ripley, B. (1994) Neural networks and flexible regression and discrimination. *Advances in Applied Statistics*, **2**, 39–57.

Running, S.W. & Hunt, E.R. Jr (1993) Generalization of a forest ecosystem process model for other biomes, BIOME-BGC, and an application for global-scale models. *Scaling processes between leaf and landscape levels* (ed. J.R. Ehleringer and C. Field), pp. 141–158. Academic Press, New York, NY, USA.

Scharf, F.S., Juanes, F. & Sutherland, M. (1998) Inferring ecological relationships from the edges of scatter diagrams: comparison of regression techniques. *Ecology*, **79**, 448–460.

Shugart, H.H. & Prentice, I.C. (1992) Individual tree based models of forest dynamics and their application in global change research. *A systems analysis of the global boreal forest* (ed. H.H. Shugart, R. Leemans and G.B. Bonan), pp. 313–333. Cambridge University Press, New York, NY, USA.

Sokal, R.R. & Rohlf, F.J. (1995) *Biometry: the principles and practice of statistics in biological research*. W. H. Freeman and Co., New York, NY, USA, p. 887.

Steele, R., Pfister, R.D., Ryker, R.A. & Kittams, J.A. (1981) *Forest habitat types of central Idaho*. USDA Forest Service General Technical Report INT-114. Intermountain Forest and Range Experiment Station, Ogden, UT.

Stephenson, N.L. (1990) Climatic control of vegetation distribution: the role of the water balance. *American Naturalist*, **135**, 649–670.

Stephenson, N.L. (1998) Actual evapotranspiration and deficit: biologically meaningful correlates of vegetation distribution across spatial scales. *Journal of Biogeography*, **25**, 855–870.

Swetnam, T.W. (1993) Fire history and climate change in giant sequoia groves. *Science*, **262**, 885–889.

Swetnam, T.W. & Betancourt, J.L. (1990) Fire-Southern Oscillation relations in the southwestern United States. *Science*, **249**, 1017–1020.

Swets, J.A. (1988) Measuring the accuracy of diagnostic systems. *Science*, **240**, 1285–1293.

Thompson, R.S., Hostetler, S.W., Bartlein, P.J. & Anderson, K.H. (1998) A strategy for assessing potential future changes

in climate, hydrology, and vegetation in the western United States. U.S. Geological Survey Circular 1153.

Thornton, P.E., Running, S.W. & White, M.A. (1997) Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of Hydrology*, **190**, 214–251.

Tomback, D.F., Arno, S.F. & Keane, R.E. (2001) *Whitebark Pine communities: ecology and restoration*. Island Press, Washington, DC, USA, p. 328.

Urban, D.L., Harmon, M.E. & Halpern, C.B. (1993) Potential response of Pacific Northwestern forests to climatic change: effects of stand age and initial composition. *Climatic Change*, **23**, 247–266.

Veblen, T.T., Kitzberger, T. & Donnegan, J. (2000) Climatic and human influences on fire regimes in ponderosa pine forests in the Colorado Front Range. *Ecological Applications*, **10**, 1178–1195.

Walther, G., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J.C., Fromentin, J., Hoegh-Goldberg, O. & Bairlein, F. (2002) Ecological responses to recent climate change. *Nature*, **416**, 389–395.

Wigmosta, M., Vail, L. & Lettenmaier, D.P. (1994) A distributed hydrology-vegetation model for complex terrain. *Water Resources Research*, **30**, 1665–1679.

Williams, C.K. & Lillybridge, T.R. (1983) *Forested plant associations of the Okanogan National Forest*. USDA Forest Service Technical Paper R6-ECOL-132b-1983. Pacific Northwest Research Station, Portland, OR, USA, p. 140.

Williams, C.K., Lillybridge, T.R. & Smith, B.G. (1990) *Forested plant associations of the Colville National Forest, field guide*. USDA Forest Service, Pacific Northwest Research Station, Portland, OR, USA, p. 133.

Yee, T.W. & MacKenzie, M. (2002) Vector generalized additive models in plant ecology. *Ecological Modelling*, **157**, 141–156.

Yee, T.W. & Mitchell, N.D. (1991) Generalized Additive Models in Plant Ecology. *Journal of Vegetation Science*, **2**, 587–602.

BIOSKETCHES

Donald McKenzie is a research ecologist focusing on forest biogeography, the effects of climatic change on forests, fire and landscape ecology, climate–fire interactions, and spatial analysis of fire regimes.

David W. Peterson is a research forest ecologist whose interests include fire ecology and restoration of forest and savanna ecosystems, forest responses to climatic variability, and forest hydrology.

David L. Peterson is a research biologist focusing on fire ecology, synthesis and integration in fire science, and the effects of climatic variability and change on mountain ecosystems.

Peter E. Thornton studies terrestrial biogeochemical cycles and their interactions with the atmosphere, using a combination of observations and numerical modeling to integrate across spatial and temporal scales.