Occurrence and dominance of six Pacific Northwest conifer species

Todd A. Schroeder, Andreas Hamann, Tongli Wang & Nicholas C. Coops

Abstract

Questions: Can probability of occurrence and dominance be accurately estimated for six important conifer species with varying range sizes? Does range size impact the accuracy of species probability of occurrence models? Is species predicted probability of occurrence significantly related to observed dominance?

Location: Pacific Northwest region, North America (60°–40°N, 140°–110°W).

Methods: This study develops near range-wide predictive distribution maps for six important conifer species (*Pseudotsuga menziesii*, *Tsuga heterophylla*, *Pinus contorta*, *Thuja plicata*, *Larix occidentalis*, and *Picea glauca*) using forest inventory data collected across the United States and Canada. Species model accuracies are compared with range size using a rank scoring system. A suite of climate and topographic predictor variables are used to investigate environmental constraints that limit species range and quantify relationships between species predicted probability of occurrence and dominance at both plot and landscape scales.

Results: Evaluation statistics revealed significant and accurate probability of occurrence models were developed for all six species. Based on ranked evaluation statistics, *Tsuga heterophylla* had highest overall model accuracy (statistic rank score = 5) and *Pinus contorta* the lowest (statistic rank score = 17). Across species, ranked evaluation statistics also revealed a pattern of decreasing model accuracy with increasing range size. At plot level, correlations between dominance and probability of occurrence were weakly positive for all species with only half of the species having statistically significant correlations. *Pseudotsuga menziesii* had the highest correlation ($r = 0.36$, $P < 0.001$) and *Thuja plicata* lowest ($r = 0.038$, $P = 0.799$). At the 50-km scale, correlations between dominance and probability of occurrence improved for all species except *Pinus contorta*. *Pseudotsuga menziesii* displayed the highest correlation ($r = 0.68$, $P < 0.001$) and *Thuja plicata* the lowest ($r = 0.07$, $P > 0.709$).

Conclusions: Species probability of occurrence model accuracy decreased with increasing range size. The strength and significance of correlations between probability of occurrence and dominance varied considerably by species and across spatial scales. Apart from *Pseudotsuga menziesii* and *L. occidentalis*, the results suggest that probability of occurrence is not a consistently reliable surrogate for species dominance in Pacific Northwest forests. We demonstrate how the degree of correlation between species occurrence and dominance can be used as an indicator of how well predictions of occurrence characterize the optimal niche of a species.

Keywords: Bioclimatic envelope modeling; Non-parametric multiplicative regression (NPMR); Probability of occurrence; Realized niche; Tree species dominance.

Nomenclature: Flora of North America Editorial Committee (1993)

Schroeder, T. A. (corresponding author, tschroeder@fs.fed.us) & **Coops, N. C.** (nicholas.coop@ubc.ca), Department of Forest Resources Management, University of British Columbia, 2424 Main Mall, Vancouver, BC, Canada.

Hamann, A. (andreas.hamann@ualberta.ca), Department of Renewable Resources, University of Alberta, Edmonton, 739 General Services Building, Edmonton, AB, Canada.

Wang, T. (tongli.wang@ubc.ca), Department of Forest Sciences, Centre for Forest Conservation Genetics, University of British Columbia, 2424 Main Mall, Vancouver, BC Canada.

Introduction

Ecological niche models have been used to effectively predict both the current and potential future distribution of tree species at various spatial scales (Iverson et al. 1999; Shafer et al. 2001; Hamann & Wang 2006; Rehfeldt et al. 2006). Although these habitat models can take a variety of statistical forms (e.g., generalized additive models, logistic regression, classification and regression tree), they all work under the same premise that a set of abiotic factors exerts dominant control over the natural distribution of plant species. Biotic interactions, such as competition or pests and diseases, are implicitly included in these models, because they model the realized niche as implied by species census data (Hutchinson 1957,

1965). High-quality soil and topographic data are often not available as predictor variables, and therefore models are often built exclusively on correlations between observed species distributions and climate variables, which is why they are commonly referred to as bioclimatic envelope models.

Models of the realized niche of a species can be problematic and difficult to interpret because current species distributions may not be in equilibrium with climate. Since the current realized niche of a species does not fully represent the extreme limits necessary for survival and growth (Pearson & Dawson 2003), it is likely that the future realized niche of a species will almost certainly be different under changing climatic conditions. As such, actual shifts in species ranges and plant communities may be quite different from those predicted by envelope models, even if given time to restore approximate equilibrium conditions (Iverson & Prasad 1998). However, rather than literally predicting future species distributions, bioclimatic envelope models can be useful to match management objectives to anticipated future climate, and they have been proposed as useful tools for guiding ecosystem restoration, plantation forestry, and genetic resource management (Hamann & Wang 2006).

In this paper, we emphasize another aspect of bioclimatic envelope models. While correlations do not imply causation, the relationship among probability of occurrence, abundance, and predictor variables can provide useful information about potential abiotic factors that govern range limits, dominance of species, and, potentially, forest productivity. Forest productivity models could benefit from this type of correlative information, as they require species-specific definitions of precipitation and temperature limits to accurately estimate photosynthetic capacity and net primary production (e.g., 3-PG model, Landsberg & Waring 1997). Additionally, spatial estimates of probability of occurrence could be used to complement incomplete forest inventory records or enhance remote-sensing data with information about tree species composition for regional and continental-scale models of forest growth. To be effectively used for this purpose, species probability of occurrence must in some way be related to its dominance on the landscape.

Previous studies exploring this relationship have produced mixed results. Studying both plant (bracken fern) and animal (moose) distributions, Nielson et al. (2005) cautioned against extrapolating abundance from probability of occurrence predictions, citing that environmental factors influencing abundance may differ from those limiting distribu-

tion. Others have suggested that, on average, there is a positive relationship between abundance and distribution simply as a result of chance agreement between two mutually occurring phenomena (Wright 1991; Gaston 1996). The strength of correlation observed between species probability of occurrence and dominance is also influenced by regionally specific factors such as forest type, management activities, disturbance regimes, and the quality and accuracy of the data used for comparisons. As the data used for comparison are spatially explicit, it is important that spatial autocorrelation be properly accounted for prior to interpreting the strength and significance of observed relationships (Fortin & Payette 2002; Kühn 2007). Methods such as geographically adjusting the effective number of degrees of freedom (e.g., Clifford et al. 1989; Dutilleul 1993) have been shown to be a simple yet effective solution for testing correlation coefficients in the presence of autocorrelation.

Here, we investigate the relationship between species occurrence and dominance of six coniferous tree species commonly found in Pacific Northwest forests using an integrated database of field plots ($n = 38\,598$) concatenated from tree measurements collected across both the United States and Canadian portions of the study area. Both ground data and spatial predictions of occurrence and dominance are used to explore two primary objectives. First, maps of probability of occurrence are evaluated using three separate accuracy metrics. As range size increases, it becomes more likely a species will inhabit a wider array of climatic gradients, making it more difficult to fit an accurate distribution model. Thus, we hypothesize that probability of occurrence model accuracy will decrease as species range size increases. Second, we examine the correlation between species occurrence and dominance at the plot and 50-km scales. In Pacific Northwest forests, species dominance is strongly governed by anthropogenic and natural disturbance regimes. Thus, we hypothesize that correlations between probability of occurrence and dominance will be weak (e.g., $r \leq 4.0$), especially after accounting for spatial autocorrelation.

Methods

Study area and tree species

Ranging from 60° to 40°N and 140° to 110°W, the study area contains approximately 1.3 million km^2 of forest land (based on 1-km resolution land
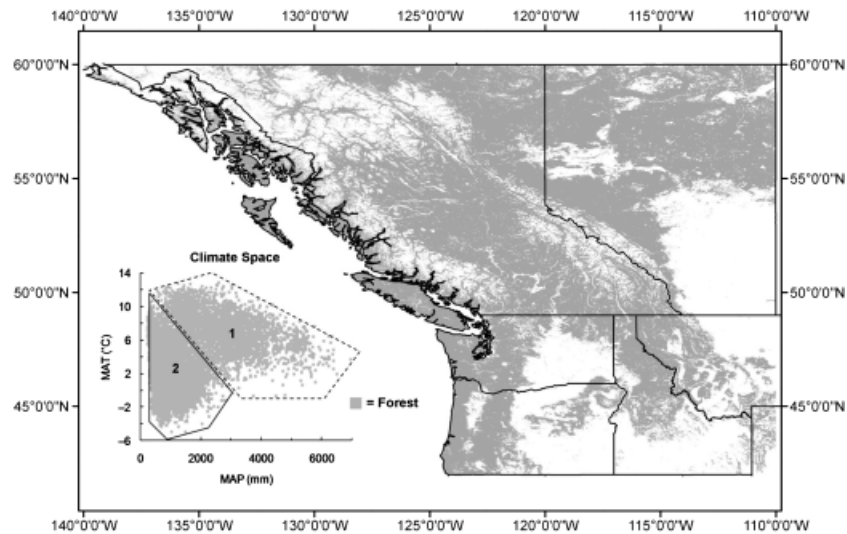
Fig. 1. The Pacific Northwest study area showing forest land based on 2003 Moderate Resolution Imaging Spectro-radiometer (MODIS) land-cover data. Inset shows the two distinct climate gradients, were MAT = mean annual temperature in °C and MAP = mean annual precipitation in mm.

cover from the 2003 Moderate Resolution Imaging Spectroradiometer MOD12Q1 data product) (Fig. 1). This large tract of forested land is made up of several mountain ranges, including the Coastal Mountains, Cascade Range, Columbia Mountains, Olympic Mountains, and the Canadian Rocky Mountains. Air masses originating over the Pacific Ocean typically gather large quantities of moisture. As these air masses make landfall, large deposits of precipitation are released as the warm, moist air rises over the windward side of the mountain ranges. Due to rain shadow effects, the leeward side of the mountain ranges receives substantially less rainfall and is characterized by broad semi-arid and arid plateaus (e.g., Interior Plateau in British Columbia and the Columbia Plateau in central Oregon and Washington). The combination of rugged topography and seasonal Pacific air masses creates two distinct climate regions (see Fig. 1 Climate Space inset): (1) the coast cordilleras (dominated mostly by moisture) and (2) the interior plateaus (dominated mainly by temperature).

The six conifer species selected for analysis in this study are among the most abundant and commercially important species in this region: *Pseudotsuga menziesii* (Douglas-fir), *Tsuga heterophylla* (western hemlock), *Pinus contorta* (lodgepole pine), *Thuja plicata* (western redcedar), *Picea glauca* (white spruce), and *Larix occidentalis* (western larch). The distribution, life history characteristics, including temperature and precipitation patterns of each species, are well described in Franklin & Dyrness (1988). In general, *Pseudotsuga menziesii, Tsuga*

heterophylla, Pinus contorta, and *Thuja plicata* are more adaptable species, as expressed by their occupancy of both the coast cordilleras and interior plateau environmental gradients. Of these species *Tsuga heterophylla* and *Thuja plicata* have distinct limits in minimum required precipitation. On the other hand, *Picea glauca* and *Larix occidentalis* occupy smaller, more narrowly defined regions of the temperature-dominated portion of climate space. *Picea glauca* has a distinct maximum temperature limit of ~7°C, whereas *Larix occidentalis* is not found when precipitation exceeds 2000 mm. Overall, the six conifer species selected for analysis represent a wide range of niche environments, making for an interesting subset to evaluate the relationship between species probability of occurrence and dominance.

Tree species data

For Canada, tree species data were assembled from a variety of sources. In Alberta (AB), data were taken from provincial permanent sample plots[1] (AB-PSP, $n = 393$) used to monitor timber growth and yield over time (Alberta Sustainable Resource Development 2005), and from ecological site information plots[2] (AB-ESIS, $n = 6951$), which are a standardized collection of plots from ecological land

[1]http://www.srd.gov.ab.ca/forests/pdf/PSP%20FIELD%20 PROCEDURE%20MANUAL%20Mar%2005.pdf
[2]http://srd.alberta.ca/lands/geographicinformation/resource dataproductcatalogue/esis.aspx

classification and stand management studies. In British Columbia (BC), data were taken from the centroids of stand-level polygons located in protected forested areas (BC-PA, $n = 11\,865$) and from vegetation resource inventory plots (BC-INV, $n = 15\,557$; B.C. Ministry of Forests 2001) collected across all forest lands using a three-phase, photo- and ground-based sampling design. Coordinates for BC-INV plots were not directly available to this study, but were used by a third party to sample the predictor grids for use in modeling and validation. For all Canadian plots, estimates of species percentage cover were used to obtain presence–absence (presence = % cover > 0), as well as to estimate dominance (D), which was calculated as percentage cover for an individual species divided by the total percentage cover of all species $\times 100$. Only trees greater than the DBH cut-off for seedlings were included. Since each data source used different criteria to define and sample seedlings, we opted not to include seedlings in our analysis. Locational accuracy of the Canadian plot coordinates was estimated at roughly ± 500 m.

For the United States, tree species data were taken from U.S. Forest Service, Forest Inventory and Analysis plots[3] (US-FIA, $n = 3482$). FIA data are recorded on a permanent sampling grid established across the conterminous United States at a density of approximately one plot per 2400 ha (Bechtold & Patterson 2005). To match as closely as possible the type of forest land sampled by the Canadian data, only single condition, forested FIA plots were used. Thus, only plots falling entirely within a particular forested land class were included in the sample. All other plots that contained only partial forest cover or non-forested land cover were excluded from the analysis. Only trees with DBH > 2.54 cm were considered in the sample (i.e., to match the Canadian data, no seedlings were included). Estimates of basal area were used to obtain presence–absence (presence = basal area > 0), as well as to estimate dominance (D), which was calculated as basal area for an individual species divided by total basal area of all species $\times 100$. Actual FIA plot locations are confidential, so we opted to use the publically available coordinates which have similar spatial accuracy to the Canadian data (i.e., ± 500 m).

To represent the forest conditions occurring between 2000 and 2005, tree data from both Canada and the United States were assembled into one database. To minimize uncertainty resulting from

Table 1. Tree species presence–absence data used to model probability of occurrence.

Species	Presence	Absence	Prevalence	Range (km²)
Pseudotsuga menziesii	8248	30 000	0.27	808 305
Tsuga heterophylla	4490	33 758	0.13	430 982
Pinus contorta	11 127	27 121	0.41	1 356 620
Picea glauca	6840	31 408	0.20	1 040 368
Larix occidentalis	879	37 369	0.20	162 346
Thuja plicata	4745	33 503	0.14	410 494

Table 2. The 12 climate and topographic predictor variables used to model species probability of occurrence.

Variable	Description, unit
MAT	Mean Annual Temperature, °C
MWMT	Mean Warmest Month Temperature, °C
MCMT	Mean Coldest Month Temperature, °C
EMT	Extreme Minimum Temperature, °C
DD<0	Estimated number of Growing Degree-Days < 0°C
DD>5	Estimated number of Growing Degree-Days > 5°C
TD	MWMT – MCMT
MAP	Mean Annual Precipitation, sum of monthly precipitation, mm
MSP	Mean Summer Precipitation, sum of summer precipitation (May–Sept), mm
AHMI	Annual Heat Moisture Index, (MAT+10)/(MAP×1000)
DEM	Elevation, m
PRR	Potential Relative Radiation, unitless

locational inaccuracies, we compared the elevation recorded in each plot's respective database to that taken from a 250-m digital elevation model (DEM). All plots found to have an elevation difference greater than ± 150 m were discarded from the analysis ($n = 6842$). The remaining plots ($n = 38\,428$, Table 1) were used to further evaluate the relationship between probability of occurrence and dominance.

Climate data

Climate surfaces were developed for the study area through downscaling of PRISM[4] (Parameter-elevation Regressions on Independent Slopes Model, Daly et al. 2000, 2002) climate data. The downscaling approach from 4-km resolution to 250-m resolution recovers small-scale temperature gradients in mountainous regions and is based on regional lapse rate adjustments for individual monthly temperature variables (Hamann & Wang 2005, Wang et al. 2006). The downscaling procedure was used to derive 48 base variables, including monthly maximum, minimum, and average temperature and average precipitation. Separate

[3]http://www.fia.fs.fed.us/

[4]http://www.prism.oregonstate.edu/

equations were used to transform the 48 base variables into fewer, biologically relevant climate predictors. The 10 derived climate variables used to describe the coldness, warmness, and wetness of the study area are shown in Table 2. For additional details regarding development of the derived climate variables see Wang et al. (2006).

Topographic data

To describe the physiography of the study area we used two topographic predictor variables (Table 2). First, a DEM was assembled from Shuttle Radar Topography Mission (SRTM) data downloaded from the Consortium for Spatial Information.[5] Individual tiles of hydrologically corrected DEM data were mosaiced in ArcInfo Grid, then resampled from 90-m resolution to 250-m using nearest neighbor resampling. The final DEM for the study area was then used to calculate potential relative radiation (PRR), a relative estimate of the effect of solar orientation caused by local topography (e.g., shadowing) (Pierce et al. 2005). To estimate the effects of monthly Earth–Sun movements, the day closest to the average solar period for each month was used, together with hourly specific solar azimuth and inclination angles to produce a series of hillshade grids in ArcInfo. These hillshade grids were summed to form "potential" relative radiation surfaces for each month. The monthly grids were summed to produce an "annual" relative radiation grid for the study area.

Species probability of occurrence mapping

Non-parametric multiplicative regression (NPMR) models were fitted and applied to GIS predictor variables using the Hyperniche software package[6] (McCune & Mefford 2004). A type of kernel smoother, NPMR is non-parametric; hence, it requires no advanced knowledge regarding the shape of species response to environmental gradients. In addition, predictor variables are considered multiplicatively, allowing the effect of one predictor to covary in complex ways with other predictors. NPMR is a local mean estimator, meaning that the proportion of a species occurrence in a locally defined environmental neighborhood is used to estimate the probability of the species occurring at a new target site. The environmental neighborhood consists of plots that lie close to the target site in multidimensional predictor space. The size of the

neighborhood is defined by a tolerance range around the target site. The shape of the neighborhood can have sharp edges if all observations falling within the neighborhood are given equal weight (e.g., uniform window where all sites inside the neighborhood = 1, all sites outside the neighborhood = 0); or can diminish gradually from the target point using weights based on a smoothing parameter (e.g., standard deviation of the Gaussian distribution).

The NPMR model is expressed in mathematical notation in McCune (2006) and a thorough explanation of the modeling process can also be found in Yost (2008). Using the default NPMR settings (i.e., improvement criterion = 0, step size = 5, maximum allowable missing estimates = 10%, data predictor ratio = 10, and minimum neighborhood size for acceptable model = n plots×0.05), we calculated probability of occurrence by taking the local mean based on a Gaussian weighting function. Models were fitted separately for each species through a leave-one-out, cross-validation process. A hold-out dataset consisting of 20% of presence and 20% of absence plots ($n = 7650$) was used to validate the accuracy of each species model. The validation data were taken at random for each species, so slightly different subsets of the full dataset were used to train and validate each of the species models.

Preferred models for each species were identified using three evaluation statistics. First, log likelihood ratios (log B), which are similar to Bayes factors, were used to evaluate the models. Expressed in powers of 10, this ratio is the likelihood of cross-validated estimates from the fitted model to estimates from the "naïve" model. The naïve model in this case is the species average frequency of occurrence (i.e., prevalence) in the dataset. Log B is an unbounded measure, thus it can become very large when strong relationships are modeled with large data sets. Its utility as a descriptive statistic lies in the fact that it increases as the weight of evidence for the model increases. Second, we used the improvement %, or the ratio of plots that receive probability estimates considered improvements over the naïve model (i.e., observed species prevalence). Therefore, improvements are considered as the percentage of presence plots that have a higher probability estimate than observed prevalence; or plots with species absence having a lower probability estimate than observed prevalence. Third, we used the area-under-the-curve (AUC) statistic, which provides a threshold independent measure of presence–absence model quality (Hanley & McNeil 1982). Taken as an integrated measure of area under the receiver oper-

ating curve, AUC represents the chance that a randomly selected presence plot will have a predicted probability higher than a randomly selected absence plot. Strong models have AUC near 1.0, while poor models have AUC near 0.5.

As model selection can be a major source of uncertainty in predicting species distributions (Dormann et al. 2008), we elected to use an ensemble approach (Araújo & New 2006), where the final probability of occurrence maps were derived by averaging the top three NPMR models for each species. Areas of non-forest land cover were removed from the final probability of occurrence maps using the MODIS 1-km forest/non-forest mask.

To help uncover the relationship between range size and model accuracy, we opted to combine the evaluation statistics using a rank scoring system. For each of the three evaluation statistics (taken from the ensemble surfaces), we assigned the species with the most accurate model a rank score of 1, and the species with the least accurate model a rank score of 6. The rank scores were summed, yielding a relative measure of overall model accuracy. For example, a rank score of 3 would be given to a species that had the most accurate model across all three evaluation statistics, whereas a rank score of 18 would be given to a species that had the least accurate model across all three measures. For comparison, we similarly rank-scored the species according to range size (e.g., 1 = smallest range, 6 = largest range).

Species dominance mapping

Species dominance is highly variable at the plot level, and thus requires a smoothing approach for use at broader spatial scales. To map dominance, we used a four-step process of resampling and interpolation. First, a systematic 50-km grid was laid over the study area. Second, the plot data falling within each 50-km grid cell were used to calculate average dominance (see Tree species data section for more details regarding dominance calculation). Third, a smooth continuous surface was created using the centroid points of each 50-km grid cell and natural neighbor interpolation (Sibson 1981). All grid cells with at least one presence plot were used in the interpolation process, ensuring that the final dominance maps adhered closely to observed presence–absence. In the final step, the dominance maps for each species were masked using the 1-km forest/non-forest map from MODIS, and then clipped to adhere to the Little Jr. (1971) range

boundary. Since dominance was upscaled and interpolated directly from the plot data, and since dominance maps were derived solely for comparison with probability of occurrence, no accuracy assessment was performed.

Species probability of occurrence versus dominance

For each species, probability of occurrence and dominance are compared at both the plot and 50-km scale. At the plot scale, dominance is taken directly from the plot database and probability of occurrence is taken from the NPMR maps. At the 50-km scale, the comparison is made using only the 50-km grid cells that have at least five or more presence plots. Although arbitrary in nature, this cutoff value helped ensure a more accurate upscaling of dominance by focusing the analysis on the grid cell locations that contained a sufficient number of plots to suitably estimate dominance at the 50-km scale (e.g., approximately one plot every 10 km). Here, a corresponding probability of occurrence was derived by averaging the NPMR pixel values falling within each qualifying 50-km cell. Both the dominance and probability of occurrence maps were masked using the same MODIS forest/non-forest mask, thus the same number of pixels were used in calculating the averages of both metrics. The relationship between probability of occurrence and dominance was quantified using Pearson correlation coefficients (r) adjusted for spatial autocorrelation using Dutilleul's modified t-test (Dutilleul 1993).

Results

Species probability of occurrence mapping

Using between two and four predictor variables, the top three NPMR models successfully estimated probability of occurrence for each of the six conifer species (Appendix S1). Four of the six species models required the use of only two predictor variables. Except for *Picea glauca*, each of the two variable models utilized one precipitation- and one temperature-related predictor. Elevation was the only topographic predictor found useful in predicting species probability of occurrence. The most frequently used predictors were AHMI (an annual index of dryness) and DD < 0 (a measure of winter length and harshness). The range in probability of occurrence estimates varied considerably by species, as can be seen in the maps in Appendix S2.

The evaluation statistics for the ensemble surfaces (i.e., average of top three NPMR models per species) are presented in Table 3. As each of the species was modeled and validated with the same number of samples, the log B statistic can be used as a relative measure of model strength. Of the six species models, *Pseudotsuga menziesii* had the highest log B (417.01) and *L. occidentalis* the lowest (93.57). The percentage improvement was used to quantify the extent to which the species probability of occurrence estimates could be considered improvements over their naïve models (naïve model = observed species prevalence in Table 1). Of the six species, *Tsuga heterophylla* had the largest improvement (85.20%) over its naïve model, whereas *Pinus contorta* offered improved estimates of occurrence in barely half (55.33%) of its validation samples. The AUC statistic was used to evaluate the extent to which the species probability of occurrence estimates could be accurately separated into discrete presence–absence categories. The results showed that *L. occidentalis* had the highest AUC (0.91) and *Pinus contorta* the lowest (0.71).

The ranked evaluation statistics presented in Table 4 show that *Tsuga heterophylla* had the highest overall accuracy (statistic rank score = 5) and *Pinus contorta* the lowest (statistic rank score = 17). To help visualize the relationship between model accuracy and range size, we plotted the ranked statistic scores against the range rank scores and fitted a second-order polynomial equation ($y = 0.9286x^2 - 4.9x + 13.4$, $R^2 = 0.90$) to the data points (Fig. 2).

Species probability of occurrence versus species dominance

Although not quantitatively validated, the dominance maps adhere well to the presence–absence data and reveal interesting and physically plausible patterns of species dominance at the landscape scale (Appendix S3). The range in modeled dominance varied by species, although not as much as probability of occurrence. At the plot level, correlations between dominance and probability of occurrence were found to be weakly positive for all species (Table 5). Only half of the six species at the plot level had statistically significant correlations after accounting for spatial autocorrelation. *Pseudotsuga menziesii* displayed the highest correlation ($r = 0.36$, $P < 0.001$) and *Thuja plicata* the lowest ($r = 0.038$, $P = 0.799$). At the 50-km scale, correlations between dominance and probability of occurrence improved for all species except *Pinus contorta* (Table 5). *Tsuga heterophylla* switched from significant correlation at the plot level to insignificant correlation at 50 km, whereas *L. occidentalis* switched from insignificant to significant correlation. Again *Pseudotsuga menziesii* displayed the highest correlation ($r = 0.68$, $P < 0.001$) and
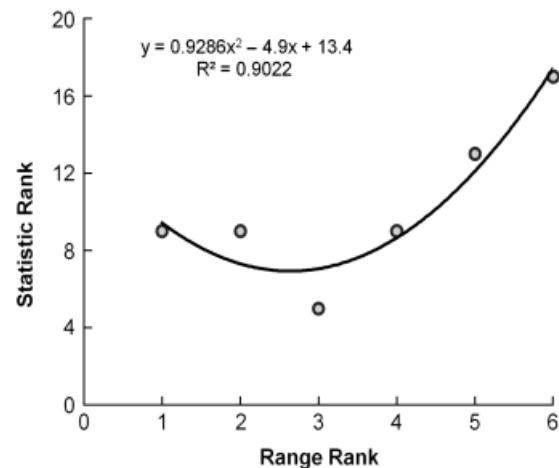

Fig. 2. Scatter plot of species range rank scores versus evaluation statistic rank scores. See Table 4 for individual species scores.

Table 3. Evaluation statistics for species probability of occurrence ensemble surfaces.

Species	log B	Improvement (%)	AUC
Pseudotsuga menziesii	417.01	73.47	0.83
Tsuga heterophylla	317.30	85.20	0.87
Pinus contorta	156.79	55.33	0.71
Picea glauca	271.51	61.73	0.80
Larix occidentalis	93.57	75.91	0.91
Thuja plicata	272.13	75.80	0.84

Table 4. Species evaluation statistic rank scores and range size rank scores.

Species	log B	Improvement (%)	AUC	Statistic rank	Range rank
Pseudotsuga menziesii	1	4	4	9	4
Tsuga heterophylla	2	1	2	5	3
Pinus contorta	5	6	6	17	6
Picea glauca	3	5	5	13	5
Larix occidentalis	6	2	1	9	1
Thuja plicata	4	2	3	9	2

Thuja plicata the lowest ($r = 0.07$, $P > 0.709$). *L. occidentalis* displayed the largest improvement in correlation from the plot to 50-km scale, followed by *Pseudotsuga menziesii*.

Discussion

Species probability of occurrence mapping

Overall, we found NPMR to be a useful method for predicting probability of occurrence for each of the six conifer specie. Although we do not explicitly compare the results of NPMR with results developed with other statistical approaches, we found that the reported NPMR accuracies are similar to other studies that have modeled species presence–absence over large areas (e.g., Moisen et al. 2006). The NPMR models were parsimonious in that many of the developed models required only two predictor variables. Aside from elevation, the best NPMR models almost exclusively utilized temperature- and moisture-related predictor variables.

Each of the three evaluation statistics used to validate the NPMR models focused on a slightly different aspect of presence–absence accuracy. While the accuracy measures were useful for selecting preferred statistical models, they did not individually uncover a salient relationship between model accuracy and range size. Combining the accuracy measures with a rank-scoring system helped lead to a clearer interpretation of the relationship between range size and model performance (Fig. 2). Although not conclusive, our results indicate that as range size increases, model performance decreases. This provides some evidence supporting our initial hypothesis that there might be a threshold of range size where species exceeding this threshold would yield lower model accuracies. Our initial hypothesis was based on the notion that species covering large areas would likely occupy

larger, more diverse regions of climate space, thus making it more difficult for NPMR to fit an accurate distribution model using primarily climate-based predictors. This notion is not entirely supported by our results as *Picea glauca* has the second-largest range size yet occupies a relatively small, well-defined region of climate space. Still, despite its small climate space, *Picea glauca* was found to have the second-highest statistic rank score, indicating the possibility that range size (or other unmeasured factors) might be contributing to low model accuracies.

Species probability of occurrence versus species dominance

Four of the six conifer species were found to have weakly positive correlations between dominance and probability of occurrence (Table 5). As the strength and significance of correlation varied considerably by species and across spatial scales, we conclude that, aside from *Pseudotsuga menziesii* and *L. occidentalis*, probability of occurrence is not a consistently reliable surrogate for species dominance. This supports our initial hypothesis that correlation between probability of occurrence and dominance would be weak in Pacific Northwest forests, likely the result of spatial variability in species dominance resulting from frequent disturbance events (e.g., fire and harvesting) and forest management activities. These factors, along with chance agreement, likely contributed to the observed correlations, but ascribing the degree to which these driving forces affected the strength of the relationships is beyond the scope of this study. However, we do believe one important contributing factor to the weak correlations is the difficulty of species models (such as NPMR) to assign high probability of occurrence predictions across both climatic gradients found in the study area.

For example, broadly tolerant species such as *Tsuga heterophylla*, *Thuja plicata*, and *Pinus con-*

Table 5. Correlation statistics (adjusted for spatial autocorrelation using Dutilleul's modified *t*-test) comparing species probability of occurrence and dominance at the plot and 50-km scales. Bold indicates model is not significant at the 0.05 level.

| Species	Plot scale					50-km scale				
	n	*r*	*t*	*df*	*P*-value	*n*	*r*	*t*	*df*	*P*-value
Pseudotsuga menziesii	8248	0.36	8.27	461.44	<0.001	353	0.68	4.87	26.97	<0.001
Tsuga heterophylla	4490	0.2	2.5	128.06	0.012	**208**	**0.3**	**1.2**	**12.15**	**0.254**
Pinus contorta	11 127	0.27	5.37	380.47	<0.001	598	0.23	2.96	156.38	0.04
Picea glauca	**6840**	**0.15**	**1.3**	**56.94**	**0.265**	**246**	**0.23**	**0.8**	**13.77**	**0.396**
Larix occidentalis	**879**	**0.13**	**1.79**	**196.66**	**0.075**	31	0.59	2.37	10.72	0.38
Thuja plicata	**4745**	**0.4**	**0.26**	**46.07**	**0.799**	**367**	**0.7**	**0.38**	**26.58**	**0.709**
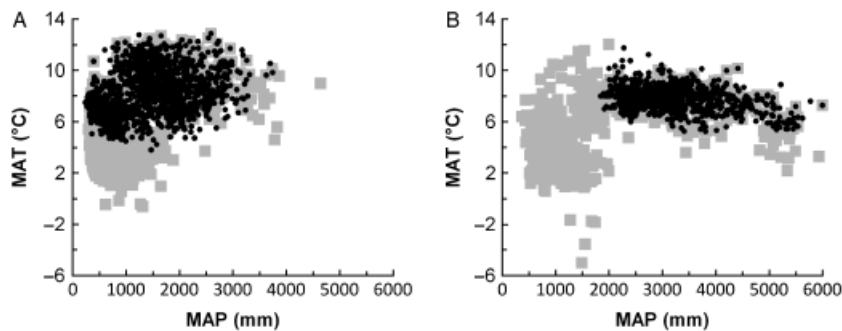
Fig. 3. Optimal qualitative niche of (**a**) *Pseudotsuga menziesii* and (**b**) *Thuja plicata* defined by upper quartile of dominance (■) and upper quartile probability of occurrence (●).

torta exhibited high dominance across portions of both climate gradients yet tended to have probability of occurrence predictions that were higher in one of the gradients than in the other. Although this geographic imbalance of predicted occurrences should not affect the models ability to correctly resolve presence–absence categories, it does contribute to the weak correlations observed between probability of occurrence and dominance. Although NPMR is fully capable of capturing a variety of simple and complex species response shapes (including bimodal), it is possible that the Gaussian weighting function smoothed over important secondary peaks in climate space where species occurrence was high. Future work will explore the use of the uniform weighting function in NPMR (previously described under Species probability of occurrence mapping) as this may help produce models that yield higher probability of occurrence predictions across both climate gradients. Another option for potentially improving correlations with dominance would be to model species probability of occurrence separately for each of the distinct climate gradients found in the study area.

Both the dominance and probability of occurrence maps presented here are useful for conservation and planning, as well as for quantifying the realized niche of a species. If correlations are reasonably strong between the two metrics, as is the case with *Pseudotsuga menziesii*, the qualitative optimal niche (Austin et al. 1990) defined by the upper quartile of occurrence and dominance will yield similar temperature and precipitation limits (Fig. 3a). If little or no correlation exists, as is the case with *Thuja plicata* (Fig. 3b), probability of occurrence predictions will not sufficiently characterize the full niche space of a species. Reasonably strong levels of correlation would also permit occurrence predictions to be combined to estimate species composition or habitat types (Yost 2008). While

quantifying reasonable correlation is difficult, we believe the results presented for *Pseudotsuga menziesii* represent a practical baseline of minimum correlation required to effectively use probability of occurrence as a surrogate measure of species dominance. If defining optimal growth limits for use in forest productivity models, dominance may be the preferred measure as it better identifies the areas on the landscape where a species has gained a natural or managed advantage over other species. If the goal is to estimate the geographical range of a species (i.e., presence/absence), occurrence may be the preferred measure of choice.

References

Anon. (Alberta Sustainable Resource Development) 2005. *Permanent sample plot (PSP) field procedures manual.* Public Lands and Forests Division, Forest Management Branch, Edmonton, AB, CA.

Anon. (B.C. Ministry of Forests) 2001. *Mensuration data from the provincial ecology program.* B.C. Ministry of Forests Workshop paper 62.

Araújo, M.B. & New, M. 2006. Ensemble forecasting of species distributions. *Trends in Ecology and Evolution* 22: 42–47.

Austin, M.P., Nicholls, A.O. & Margules, C.R. 1990. Measurement of the realized qualitative niche: environmental niches of five *Eucalyptus* species. *Ecological Monographs* 60: 161–177.

Bechtold, W.A. & Patterson, P.L. 2005. *The enhanced forest inventory and analysis program – national sampling design and estimation procedures.* General Technical Report SRS-80, U.S. Department of

Agriculture, Forest Service, Southern Research Station, Ashville, NC, US

Clifford, P., Richardson, S. & Hemon, D. 1989. Assessing the significance of the correlation between 2 spatial processes. *Biometrics* 45: 123–134.

Daly, C., Taylor, G.H., Gibson, W.P., Parzybok, T.W., Johnson, G.L. & Pasteris, P.A. 2000. High-quality spatial climate data sets for the United States and beyond. *Transactions of the ASAE* 43: 1957–1962.

Daly, C., Gibson, W.P., Taylor, G.H., Johnson, G.L. & Pasteris, P. 2002. A knowledge-based approach to the statistical mapping of climate. *Climate Research* 22: 99–113.

Dormann, C.F., Purschke, O., Garćia Márquez, J.R., Lautenbach, S. & Schröder, B. 2008. Components of uncertainty in species distributions analysis: a case study of the great grey shrike. *Ecology* 89: 3371–3386.

Dutilleul, P. 1993. Modifying the *t*-test for assessing the correlation between 2 spatial processes. *Biometrics* 49: 305–314.

Flora of North America Editorial Comittee (eds). 1993. *Flora of North America North of Mexico*. 15 vols., New York and Oxford.

Fortin, M-J. & Payette, S. 2002. How to test the significance of the relation between spatially autocorrelated data at the landscape scale: a case study using fire and forest maps. *Ecoscience* 9: 213–218.

Franklin, J. & Dyrness, C. 1988. *Natural Vegetation of Oregon and Washington*. Oregon State University Press, Corvallis, OR, US.

Gaston, K.J. 1996. The multiple forms of the interspecific abundance–distribution relationship. *Oikos* 76: 211–220.

Hamann, A. & Wang, T.L. 2005. Models of climatic normals for genecology and climate change studies in British Columbia. *Agricultural and Forest Meteorology* 128: 211–221.

Hamann, A. & Wang, T. 2006. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. *Ecology* 87: 2773–2786.

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

Hutchinson, G.E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22: 415–427.

Hutchinson, G.E. 1965. The niche: an abstractly inhabited hypervolume. In: *The Ecological Theatre and the Evolutionary Play*. pp. 26–78. Yale University Press, New Haven, CT, US.

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

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

Kühn, I. 2007. Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions* 13: 66–69.

Landsberg, J.J. & Waring, R.H. 1997. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance, and partitioning. *Forest Ecology and Management* 95: 209–228.

Little, E.L. Jr. 1971. Atlas of United States trees. *Conifers and Important Hardwoods*. Vol. 1, USDA Miscellaneous Publication 1146, Washington, DC, US.

McCune, B. & Mefford, M.J. 2004. *HyperNiche. Non-parametric multiplicative habitat modeling. Version 1.38*. MjM Software, Gleneden Beach, OR, US.

McCune, B. 2006. Non-parametric habitat models with automatic interactions. *Journal of Vegetation Science* 17: 819–830.

Moisen, G.G., Freeman, E.A., Blackard, J.A. Frescino, T.S., Zimmerman, N.E. & Edwards, T.C. 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling* 199: 176–187.

Nielson, S.E., Johnson, C.J., Heard, D.C. & Boyce, M.S. 2005. Can models of presence–absence be used to scale abundance? Two case studies considering extremes in life history. *Ecography* 28: 197–208.

Pearson, R.G. & Dawson, T.P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimatic envelope models useful? *Global Ecology and Biogeography* 12: 361–371.

Pierce, K.B., Lookingbill, T. & Urban, D. 2005. A simple method for estimating potential relative radiation (PRR) for landscape-scale vegetation analysis. *Landscape Ecology* 20: 137–147.

Rehfeldt, G.E., Crookson, N.L., Warwell, M.V. & Evans, J.S. 2006. Empirical analysis of plant–climate relationships for the western United States. *International Journal of Plant Science* 167: 1123–1150.

Shafer, S.L., Bartlein, P.J. & Thompson, R.S. 2001. Potential changes in the distributions of western North America tree and shrub taxa under future climate scenarios. *Ecosystems* 4: 200–215.

Sibson, R. 1981. A brief description of natural neighbour interpolation. In: V. Barnett (ed.) *Interpreting multivariate data*. pp. 21–36. John Wiley & Sons, New York, NY, US.

Wang, T., Hamann, A., Spittlehouse, D.L. & Aitken, S.N. 2006. Development of scale-free climate data for western Canada for use in resource management. *International Journal of Climatology* 26: 383–397.

Wright, D.H. 1991. Correlations between incidence and abundance are expected by chance. *Journal of Biogeography* 18: 463–466.

Yost, A.C. 2008. Probabilistic modeling and mapping of plant indicator species in a Northeast Oregon industrial forest, USA. *Ecological Indicators* 8: 46–56.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix S1. Species probability of occurrence maps for a.) *Pseudotsuga menziesii*, b.) *Tsuga heterophylla*, c.) *Pinus contorta*, d.) *Thuja plicata*, e.) *Larix occidentalis*, and f.) *Picea glauca*.

Appendix S2. Predictor variables and evaluation statistics for the top three NPMR models per species.

Appendix S3. Dominance maps for a.) *Pseudotsuga menziesii*, b.) *T.suga heterophylla*, c.) *Pinus contorta*, d.) *Thuja plicata*, e.) *Larix occidentalis*, and f.) *Picea glauca*. Bold lines indicate Little Jr. (1971) range boundary.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.