Method selection for species distribution modelling: are temporally or spatially independent evaluations necessary?

David R. Roberts and Andreas Hamann

D. R. Roberts (drr3@ualberta.ca) and A. Hamann, Univ. of Alberta, Dept of Renewable Resources, 751 General Services Building, Edmonton, AB T6G 2H1, Canada.

To assess the realism of habitat projections in the context of climate change, we conduct independent evaluations of twelve species distribution models, including three novel ecosystem-based modelling techniques. Habitat hindcasts for 24 western North American tree species were validated against 931 palaeoecological records from 6000, 11 000, 14 000, 16 000 and 21 000 yr before present. In addition, we evaluate regional extrapolations based on geographic splits of > 55 000 sample plots. Receiver operating characteristic analyses indicated excellent predictive accuracy for cross-validations (median AUC of 0.90) and fair accuracy for independent regional and palaeoecological validations (0.78 and 0.75). Surprisingly, we found little evidence for over-parameterisation in any method. Also, given high correlations found between model accuracies in non-independent and independent evaluations, we conclude that non-independent evaluations are effective model selection tools. Ecosystem-based modelling approaches performed below average with respect to model sensitivity but excelled in specificity statistics and robustness against extrapolations far beyond training data, suggesting that they are well suited to reconstruct historical biogeographies and glacial refugia.

Species distribution models, also referred to as ecological niche models or bioclimate envelope models, are an important group of modelling techniques to predict habitat suitability, originally developed in the context of conservation biology (Araújo and Williams 2000). Over the last decade, such models have been extensively applied to project future species habitat under anticipated climate change (reviewed by Elith and Leathwick 2009). A search with ISI Web of Knowledge for (species distribution models) AND (climate change) reveals that, between 1990 and 2010, the number of publications increased in exponential fashion, with 538 papers published in 2010. While the amount of research addressing potential climate change issues is encouraging, it has also been noted that there is a lack of thorough and independent validation of the predictive accuracy for these models, and that some standard evaluation statistics may be flawed (Botkin et al. 2007, Elith and Graham 2009).

While it is not possible to validate future projections directly, models can be evaluated by other means, with statistical accuracies inferred by some form of cross-validation where a subset of the data is used for model training and the remainder used for model validation. However, these cross-validation methods are problematic: ecological data is often highly auto-correlated and random data-splitting methods do not result in truly independent validation datasets, which leads to overly optimistic assessments of model accuracy (Araújo et al. 2005a, Segurado et al. 2006). Instead, independent model validations should be performed with data sets sourced externally from the training data, which often

include data from new geographic regions or new time periods (Araújo et al. 2005a).

Several examples of independent model evaluations can be found in the recent literature. Projections into different geographic regions have been employed by Randin et al. (2006), Fløjgaard et al. (2009), and Morueta-Holme et al. (2010) to better assess the accuracy of their projections into the future or past. Palaeoecological data has also been employed, either by training models with fossil and pollen records to predict current species distributions (Martinez-Meyer and Peterson 2006), or by using modern census data for training and palaeoecological data for validation (Martinez-Meyer et al. 2004, Giesecke et al. 2007, Pearman et al. 2008, Rodríguez-Sánchez and Arroyo 2008, Roberts and Hamann 2011). All of these studies, however, assess only one or two model techniques, and a comprehensive assessment of species distribution model accuracy against independent data is lacking.

In this study, we contribute independent regional and palaeoecological validations for 12 modelling techniques that represent four major approaches to modelling species distributions (climate envelopes, machine learning techniques, regression-based techniques, and a novel ecosystem-based approach). We applied these techniques to predict habitat of 24 major tree species of western North America based on > 55 000 present-day sample plots and > 700 palaeoecological study sites. Our primary goal is to assess the predictive accuracy of a variety of species distribution models and to assess their robustness when extrapolating spatially or

temporally beyond data coverage. We further aim to identify model techniques that are prone to over-parameterisation by comparing independent and non-independent validations. We also assess, for the first time, a new approach to habitat modelling that relies on projections of ecosystem classes, and that predicts the distribution of all species in a single model run, and we conclude by discussing the suitability of available techniques for different objectives in ecological habitat modelling.

Methods

Species distribution models

Census data from 55 743 forest inventory plots, or a subset thereof, was used to build predictive species distribution models for 24 western North American tree species. Two approaches to model validation were used. First, we hindcasted suitable habitat based on climate parameters fo five periods of the Holocene and Late Pleistocene (6000, 11 000, 14 000, 16 000, 21 000 yr before present) and validated these projections against 931 fossil and pollen records from 737 unique sites, where records from different time periods from the same location were considered separate samples. Second,

we split the dataset geographically at 49° latitude and used plot data from Canada and Alaska to train the species distribution model. Projections were subsequently made for the continental United States and Mexico, geographically extrapolating into areas with generally warmer climate (as a proxy to projections under anticipated future climate) and evaluated the results against forest plot data from south of 49° latitude.

We employed a representative set of species distribution modelling techniques, including two climate envelope approaches, four machine learning methods, three modern regression techniques, and three ecosystem-based modelling techniques, one of them a novel method (described in Table 1). All techniques, except the ecosystem-based approaches and discriminant analysis were implemented using the BIOMOD package (Thuiller 2003) for the R programming environment (R Development Core Team). BIOMOD is a computational framework for multi-method modelling that generates probability of presence (PoP) outputs for multiple species for multiple methodologies while allowing the user control over each of the individual methods. It can generate a variety of ensemble projections, based on the outputs of the individual models (Thuiller et al. 2009). The BIOMOD package has several options for addressing pseudo-absence records, of which we selected

Table 1. Descriptions of each of the species distribution modelling techniques evaluated in this study, grouped by similar methodological approaches (based on groups described by Elith and Leathwick 2009, Franklin 2009).

Method	Description
Ecosystem-based methods	
Discriminant analysis	A standard multivariate approach to classification, using an ecosystem class as dependent class variable and climate data as predictor variables (Hamann and Wang 2006).
Minimum distance	A new classification approach based on the closest multivariate distance in climate variables to a ecosystem climate mean, using Euclidean distance of principal components that explain most of the variance in the climate dataset (essentially a modified Mahalanobis distance).
Random forest ensemble classifier	A classification tree implementation with an ecosystem class as dependent variable. Multiple classification trees are built based on random subset of variables and the final class prediction is obtained by majority vote from multiple classification trees (Mbogga et al. 2010).
Envelope techniques	
Discriminant analysis	A standard multivariate approach to classification, using a binary response variable (presence or absence).
Surface range envelopes	Data within the 5th and 95th percentile of the maximum and minimum range for each predictor variable is considered within the envelope and variable interactions are not considered (Beaumont and Hughes 2002).
Machine learning techniques	
Artificial neural networks	Networks are built of weighted hidden units (much like decision tree nodes) based largely on pattern recognition and are capable of incorporating feedback loops between the units (Segurado and Araújo 2004).
Classification tree analysis	Recursive data-splitting technique, iteratively creating homogenous subgroups (with the goal of minimising variance within each group). Cross-validation is used to prune the decision tree by balancing the number of terminal nodes and the explained variance (De'ath and Fabricius 2000).
Generalised boosting model	Iterative regression trees, where misclassified data from one classification tree is weighted heavier in subsequent classifiers, so each iteration places more emphasis on misclassified data (Leathwick et al. 2006).
Random forest ensemble classifier	Multiple classification trees are built based on random subset of predictor variables and the final predictions are derived by averaging probabilities over multiple classification trees (Prasad et al. 2006).
Regression-based techniques	
Generalised additive model	Generalised linear models for individual predictor variables are combined additively, using smoothing equations to generalise the data and fit to local data subsets (Guisan and Zimmermann 2000).
Generalised linear model	An extension of the general linear model for binomial data capable of capturing non-linear relationships (Guisan and Zimmermann 2000).
Adaptive regression splines	Fits splines to distinct but unequal intervals of the predictors before pruning excess spline connections through a stepwise analysis (Prasad et al. 2006).

the 'environmental envelope' option, an approach based on ranges of environmental predictors, shown to be effective by Zarnetske et al. (2007). We also included a standard discriminant analysis, implemented with PROC DISCRIM of the SAS statistical software package (SAS Inst.), where the dependent class variable is species presence or absence and the TESTOUT option provided a probability of presence value for habitat predictions.

We included three methods based on a species distribution modelling approach developed by Hamann and Wang (2006) and Roberts and Hamann (2011). The approach characterises the climate space of delineated ecosystem polygons, which represent a mapped area with relatively homogeneous species communities, topoedaphic, and climatic characteristics. We used 768 mapped ecosystem classes compiled from various public sources for the western continental US, western Canada, and Alaska using the finest scale mapped delineations available (for sources and detailed methodology, Roberts and Hamann 2011). In this ecosystem-based modelling approach, ecosystem classes serve as a dependent categorical variable that is predicted with climate variables using three different methods. The maps of projected ecosystem classes were subsequently converted to species habitat maps, where the probability of presence of a species was calculated as the proportion of the inventory plots within the ecosystem polygon where the species was present.

This general ecosystem-based approach was implemented with three different techniques that allow for a categorical response variable. Two methods have previously been used to project species distributions: discriminant analysis (Hamann and Wang 2006) and classification tree analysis implemented with the random forest software package (Breiman 2001) for the R programming environment (R Development Core Team); for details see Mbogga et al. (2010). The third implementation of the ecosystem-based approach is an analogue-based inference method similar to those described by Overpeck et al. (1992) and Williams et al. (2001). Our implementation subjects average values of climate variables for each ecosystem to a principal component analysis implemented with PROC PRINCOMP of the SAS statistical software package (SAS Inst.). Subsequently, we calculated a matrix of Euclidean distances between the climate scores of each ecosystem average (in columns) versus climate scores of predicted surfaces (in rows). The classification was then carried out based on the minimum distance (i.e. the ecosystem climatically most similar to the grid cell to be classified). For the distance calculations, that are roughly equivalent to a Mahalanobis distance measure, the number of principal components was limited to those that explain a relevant amount of variance in the dataset, subjectively determined with a scree plot (the first 5 components in this case).

From the output of all twelve species distribution models, we also generated ensemble projections based on the mean and median of predicted probabilities of presence.

Species and climate data

All individual species-based models directly predict probability of presence using presence/absence data from forest inventory plots as training data. In the case of ecosystem-based habitat models that predict an ecosystem class, we derived a probability of presence value by substituting the ecosystem class with a ration of species' presences over the total number of plot samples that fall within the ecosystem delineation. Our 55 743 plot samples for western North America were compiled from the British Columbia Ministry of Forests (Hamann and Wang 2006), Sustainable Resources Development of the Government of Alberta (2004), and the United States Forest Service (Betchtold and Patterson 2005). The database also contained 3273 non-forested plot sites.

Predictor variables were interpolated climate data for the 1961–1990 reference period, generated at 1 km resolution with a software package that we make freely available (<www.ualberta.ca/~ahamann/climate.html> or <www.genetics.forestry.ubc.ca/cfcg/climate-models.html>) (Wang et al. 2006, Mbogga et al. 2009). For a continental-scale modelling effort, we use relatively high-resolution climate grids to avoid over-estimating climate change effects in mountainous areas. With coarse-resolution grids, climate envelopes of species or ecosystems would be too narrowly defined with smaller temperature ranges than in reality (Hamann and Wang 2005). From the available climate variables, we used a principal component analysis to select 10 variables with the lowest collinearity: mean annual precipitation, the mean temperature of the warmest month, the mean temperature of the coldest month, the difference between January and July temperature as a measure of continentality, May to September (growing season) precipitation, the number of frost-free days, the number of growing degree days above 5°C, and summer and annual dryness indices according to Hogg (1997). Past climate reconstructions for the periods 6000, 11 000, 14 000, 16 000, and 21 000 calendar years before present were derived with previously-run simulations of the community climate model (CCM1) general circulation model (Kutzbach et al. 1998). The coarse-resolution (7.5° longitude by 4.5° latitude) CCM1 data were overlaid on high resolution modern climate data as deviations from the 1961–1990 reference period using the software package described above.

Model evaluation

Model projections were evaluated in four different ways: 1) using all the sample plot data for training and for evaluation (referred to as all-data); 2) with a random data-split, using 67% of the sample plots for training and the remaining 33% for evaluation (out-of-bag); 3) using a regional extrapolation where models were trained with plot data from Canada and Alaska and projections were evaluated with plot data from the continental United States (north-to-south); and 4) using all the present sample plot data for training and evaluation with fossil and pollen records from four time periods since the last glacial maximum (past-periods). For this past-periods model evaluation, we use palaeoecological data comprised of 931 fossil pollen and plant macrofossils records from 737 unique sampling sites compiled by the North American Pollen Database (COHMAP 1988), Thompson and Anderson (2000), and Dyke (2005). Of the 24 western North American tree species considered, six species were omitted from the

past-periods evaluation due to a lack of records in the fossil data (n < 10) and five species were omitted from the north-to-south regional evaluation due to a lack of records in either the north or south data split (Table 2 for details).

We report model sensitivity (calculated as TP/(TP + FN), where TP = true positives and FN = false negatives), model specificity (calculated as TN/(TN + FP), where TN = true negatives and FP = false positives). Sensitivity and specificity values represent an integrated measure for a range of thresholds between zero and one, calculated with the ROCR package (Sing et al. 2005) for the R programming environment (R Development Core Team). The area under the curve (AUC) of the receiver operating characteristic (Fawcett 2006), also calculated with the ROCR package (Sing et al. 2005), is a useful summary statistic of model accuracy as it is a threshold-independent evaluation of the rate of true presences vs false presences for all output probabilities simultaneously. The AUC of the receiver operating characteristic balances the ability of the model to detect a species when it is present (sensitivity) against its ability to not predict a species when it is absent (specificity). AUC values range from 0 to 1, where 1 indicates perfect model accuracy, 0.5 represents a prediction expected by random chance, and 0 indicates that all predictions are false.

In order to quantify the relative contribution of modelling methods and species' ecological and biogeographic attributes, we also carried out a variance partitioning analysis using AUC values as the dependent variable, implemented by PROC VARCOMP of the SAS Statistical software package (SAS Inst.), using the restricted maximum likelihood method (option REML).

Results

Independent model validation

We find that model accuracy substantially declines across all techniques and all species when subjected to independent validations (Fig. 1). Mean AUC values across all model techniques are represented by vertical lines in Fig. 1; median AUC values that are less influenced by outliers are 0.90, 0.78, and 0.75 for out-of-bag, regional, and palaeoecological validations, respectively. This comparison excludes species that did not have sufficient palaeoecological records or sample plots north and south of 49° latitude, as indicated in Table 2. Considering that the expected AUC value for a random classifier is 0.5, the reduction in accuracy is substantial. The individual AUC values for each species, model technique, and validation scenario are provided in the Supplementary material Appendix 1, Table A1–A4. Standard errors of the mean AUC values represented as symbols in Fig. 1 were on average 0.01, 0.02, and 0.02 for out-of-bag, regional, and temporal validations, respectively. Standard deviations that provide a measure of variation among species (rather than statistical accuracy of the mean) were 0.04, 0.10, and 0.11, respectively.

Even though we find substantial reductions in AUC values between non-independent and independent validation scenarios, these reductions are consistent in magnitude across all methods that we investigated. A completely non-

independent evaluation where all sample plots were used is virtually identical to out-of-bag validations. We further observe high correlations between the AUC values (inverse-transformed for normality) of the non-independent out-of-bag and the independent regional validations (r = 0.70, p = 0.012) or palaeoecological validations (r = 0.89, p < 0.001). Notably, methods that have very high AUC values in non-independent validations, which could indicate over-parameterisation, also rank as most accurate in independent tests (Table 3).

In Table 3, we also report AUC values for ensemble projections, where the predicted probabilities of presence for species are represented by the mean or median across multiple model techniques. Ensemble projections outperformed all individual methods in independent evaluations. While more complex ensemble methods are available that weigh contributions of individual techniques by various statistics (for methodologies, Thuiller et al. 2009), we found that the simplest methods based on measures of central tendency (mean and median) yielded amongst the highest AUC values in both dependent and independent evaluations (Supplementary material Appendix 1). Removing the poor-performing individual models from the ensemble calculations did not improve the ensemble projections. Even the inclusion in the ensembles of the surface range envelope outputs (with AUC values only slightly above random chance) served to either increase or not affect the AUC values of the ensembles (Table 3).

Biogeographic and ecological characteristics

The influence of biogeographic and ecological characteristics of western North American trees species on model accuracy is summarised in Fig. 2. This comparison is based on average values from the two independent evaluations (regional and palaeoecological validations), and we aggregate the results further by groups of modelling techniques used in Fig. 1 and described in Table 1. Standard deviations for the mean AUC values represented by symbols in Fig. 2 ranged between 0.01 and 0.08 (mean of 0.04). The individual AUC value for each species, modelling technique, and validation scenario is provided in Supplementary material Appendix 1. In general, modelling techniques do not show interactions with biogeographic or ecological characteristics of species, but perform consistently well or consistently poorly across all ecological or biogeographic criteria. There are, however, some moderate main effects of biogeographic or ecological characteristics on overall model accuracy.

Among the biogeographic and ecological characteristics, the continentality of the distribution of western North American tree species accounts for most variation in model accuracy (Table 4), with interior species having generally lower accuracies than coastal species (Fig. 2). An ANOVA using the classes 'Coast', 'Interior', or 'Both' as predictor variable and AUC values across all modelling techniques as response variable confirms a significant main effect (p = 0.002). While common but range restricted species appear to have higher model accuracies (Fig. 2), this effect is not significant in an equivalent ANOVA (p = 0.775) and does not account for any meaningful amount of the variance in AUC (Table 4). Mean AUC values declined from species that are

Table 2. The 24 western North American tree species included in the modelling and their respective biological categories (based on Burns et al. 1990). The total number of presence records in the modern sample plot data is provided as well as the number of presences for each species included in each of the evaluation data sets (in the out-of-bag (OOB) and north-to-south (N2S) data-splits). The total sum of species presences in the fossil record for 6000, 11 000, 14 000, 16 000, and 21 000 yr ago is also provided (TMP). An asterisk (*) indicates that the species was removed from the evaluation either due to a lack of presences in either the north or south data split (in the regional evaluation) or due to a lack of records (n < 10) in the fossil data (in the temporal evaluation).

| Species name	Taxon/group	Total plot presences	Number of validation plots OOB	Number of validation plots N2S	Number of validation plots TMP	Range size (10³ km²)	Type	Distribution Range	Distribution Elevation	Shade tolerance
Abies amabilis (Pacific silver fir)	Abietoideae	1615	526	269	1*	272	restricted	coastal	intermediate	very tolerant
Abies lasiocarpa (subalpine fir)	Abietoideae	10804	3486	1715	46	1957	widespread	interior	intermediate	tolerant
Abies procera (noble fir)	Abietoideae	82	30	82*	1*	44	restricted	coastal	highly restricted	tolerant
Acer macrophyllum (bigleaf maple)	Angiosperm	437	145	301	5*	382	widespread	coastal	restricted	very tolerant
Alnus rubra (red alder)	Angiosperm	715	236	369	19	491	widespread	coastal	highly restricted	intolerant
Betula papyrifera (paper birch)	Angiosperm	3926	1349	68	14	10251	widespread	interior	restricted	intolerant
Calocedrus decurrens (incense cedar)	Cupressaceae	561	187	561*	1*	134	restricted	coastal	restricted	intermediate
Chamaecyparis nootkatensis (yellow cedar)	Cupressaceae	707	223	24	23	392	widespread	coastal	restricted	tolerant
Larix occidentalis (western larch)	Laricoideae	821	281	463	3*	217	restricted	interior	highly restricted	very intolerant
Picea engelmannii (Engelman spruce)	Piceoideae	6223	1994	1733	44	1002	widespread	interior	unrestricted	tolerant
Picea glauca (white spruce)	Piceoideae	7115	2398	22	55	10320	widespread	interior	intermediate	intermediate
Picea mariana (black spruce)	Piceoideae	2922	1005	0*	48	10446	widespread	interior	restricted	tolerant
Picea sitchensis (Sitka spruce)	Piceoideae	1016	338	85	31	482	widespread	coastal	highly restricted	tolerant
Pinus albicaulis (whitebark pine)	Pinoideae	1038	347	412	59	559	restricted	interior	intermediate	intermediate
Pinus contorta (lodgepole pine)	Pinoideae	11275	3722	1971	163	2458	widespread	both	intermediate	very intolerant
Pinus edulis (pinyon pine)	Pinoideae	2836	977	2836*	13	280	restricted	interior	highly restricted	intolerant
Pinus monticola (western white pine)	Pinoideae	820	289	307	18	429	restricted	both	unrestricted	intermediate
Pinus ponderosa (Ponderosa pine)	Pinoideae	3967	1325	3372	25	884	widespread	interior	unrestricted	intolerant
Populus tremuloides (trembling aspen)	Angiosperm	7241	2400	1090	14	11481	widespread	interior	unrestricted	very intolerant
Pseudotsuga menziesii (Douglas-fir)	Laricoideae	8808	2992	4438	174	1445	widespread	both	unrestricted	intermediate
Sequoia sempervirens (giant sequoia)	Cupressaceae	90	32	90*	0*	14	restricted	coastal	highly restricted	intolerant
Thuja plicata (western redcedar)	Cupressaceae	3798	1235	409	29	601	widespread	both	intermediate	very tolerant
Tsuga heterophylla (western hemlock)	Abietoideae	4860	1619	707	90	714	widespread	both	restricted	very tolerant
Tsuga mertensiana (mountain hemlock)	Abietoideae	1136	401	241	66	437	restricted	both	unrestricted	tolerant
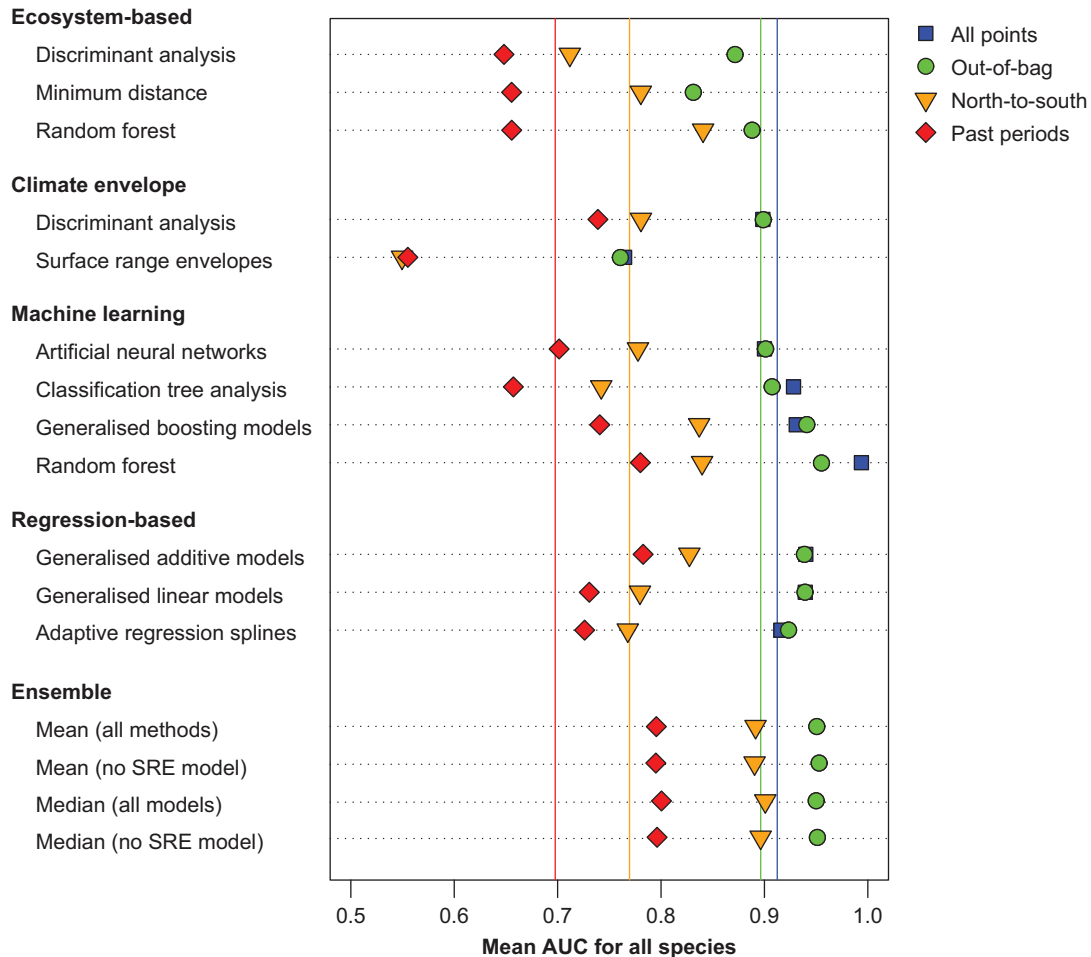
Figure 1. Model accuracy for 12 individual and 4 simple ensemble techniques, evaluated by the area under the curve (AUC) of the receiver operating characteristic. Non-independent evaluations include training and validation data being the same (all points), a random data split of 67% for training and 33% for evaluation (out-of-bag). Independent validations include a 49° latitude data split, extrapolating south for validation (north-to-south) and projections based on palaeoclimate data using fossil data from 6000, 11 000, 14 000, 16 000, and 21 000 yr before present for validation (past periods). Vertical lines represent the mean AUC value across all species and methods (excluding ensembles).

highly restricted in their elevation range to species that we classified as unrestricted (p < 0.001). However, the overall variance in AUC explained by elevation category, despite this apparent linear relationship was minimal (Table 4). It should be noted that this elevation relationship does not depend on absolute values: range restricted species may be found at high elevations (e.g. *Pinus edulis*) as well as at low elevations (e.g. *Picea sitchensis*). We also observe a weaker trend toward higher model accuracies for shade tolerant species (p = 0.041). In total, 29.7% of the variance in AUC was explained by the ecological and biogeographic traits of species, as compared to 16.4% explained by the modelling method: together explaining just less than half of the total variance in AUC (Table 4).

Model sensitivity versus specificity

Beyond the AUC statistic we also considered model sensitivity (the ability to detect a species when it is present) and model specificity (the ability not to predict a species when it is absent) for evaluation. Across all species and methods, model specificity is generally higher than model sensitivity (Table 3). The best performing techniques tend to have high values for both sensitivity and specificity with the exception of the random forest ensemble classifier, which appears to have higher specificity than sensitivity values in both the ecosystem- and species-based implementations. For the independent evaluations, ecosystem-based methods have significantly higher model specificity values (p = 0.002) and generally, but not significantly, lower sensitivities (p = 0.09) when compared to species-based approaches. Unlike AUC values, sensitivity and specificity values for the out-of-bag versus independent evaluations were not significantly correlated. Only sensitivity values for the out-of-bag versus regional extrapolations were significantly correlated (r = 0.89, p < 0.001).

The ability of ecosystem-based models to better predict species absence is apparent in an example for *Pseudotsuga meziesii* (Douglas-fir) (Fig. 3). Here, we compared the ecosystem- and species-based modelling approach relying on the same technique, the random forest ensemble classifier. The ecosystem-based model run (Fig. 3, bottom row) has better defined species absences in the two extrapolations, whereas the species-based models (Fig. 3, top row) show large areas of high probability values well outside the species range. Another notable observation is that the

Table 3. Mean AUC values (AUC), mean sensitivity (Sens.), and mean specificity (Spec.) across all species for 1) the non-independent out-of-bag evaluation; 2) the independent north-to-south regional evaluation; and 3) the independent past-periods temporal evaluation. The rank of each modelling technique within the independent evaluation scenarios and the sum of these ranks (Sum) is given. Ecosystem-based methods are shown in italics. The results of the mean and median of all methods as ensemble projections are also included, as are values of the ensembles with the worst-performing model (SRE) and ecosystem-based methods (Eco) removed.

Modelling method	Out-of-bag (OOB)			North-to-south (N2S)			Past periods (TMP)			Method rank		
	AUC	Sens.	Spec.	AUC	Sens.	Spec.	AUC	Sens.	Spec.	N2S	TMP	Sum
Random forest (Sp)	0.95	0.59	0.94	0.84	0.57	0.76	0.78	0.46	0.83	2	2	4
Generalised additive models	0.94	0.85	0.82	0.83	0.74	0.68	0.78	0.38	0.89	4	1	5
Generalised boosting models	0.94	0.88	0.83	0.84	0.75	0.70	0.74	0.36	0.86	3	3	6
Discriminant analysis (Sp)	0.90	0.88	0.54	0.78	0.78	0.52	0.74	0.73	0.51	5	4	9
Random forest (Eco)	*0.89*	*0.61*	*0.90*	*0.84*	*0.56*	*0.85*	*0.66*	*0.27*	*0.90*	*1*	*9*	*10*
Generalised linear models	0.94	0.85	0.84	0.78	0.66	0.72	0.73	0.34	0.90	7	5	12
Adaptive regression splines	0.92	0.46	0.93	0.77	0.46	0.79	0.73	0.39	0.82	9	6	15
Artificial neural networks	0.90	0.58	0.90	0.78	0.40	0.86	0.70	0.35	0.87	8	7	15
Minimum distance (Eco)	*0.83*	*0.53*	*0.90*	*0.78*	*0.46*	*0.88*	*0.66*	*0.28*	*0.90*	*6*	*10*	*16*
Classification trees	0.91	0.79	0.72	0.74	0.64	0.68	0.66	0.31	0.84	10	8	18
Discriminant analysis (Eco)	*0.87*	*0.59*	*0.90*	*0.71*	*0.46*	*0.79*	*0.65*	*0.28*	*0.90*	*11*	*11*	*22*
Surface range envelopes	0.76	0.54	0.63	0.55	0.37	0.66	0.56	0.39	0.65	12	12	24
Ensemble: mean	0.95	0.92	0.59	0.89	0.86	0.58	0.80	0.78	0.52	–	–	–
Ensemble: median	0.95	0.76	0.83	0.90	0.81	0.59	0.80	0.53	0.79	–	–	–
Mean (SRE and Eco removed)	0.95	0.92	0.59	0.89	0.86	0.58	0.79	0.78	0.52	–	–	–
Median (SRE and Eco removed)	0.95	0.74	0.81	0.90	0.82	0.56	0.80	0.53	0.78	–	–	–

species-based model run often indicates high probability of presence outside the area of available validation points (Mexico in the north-to-south extrapolation and in the areas of continental ice cover in the run for 16 000 yr ago). These differences between ecosystem- and species-based approaches for Douglas-fir are typical for other species and modelling methods also, where individual-species based techniques regularly show relatively high probability of presence values in spatial and temporal extrapolations beyond validation data coverage (maps not shown).
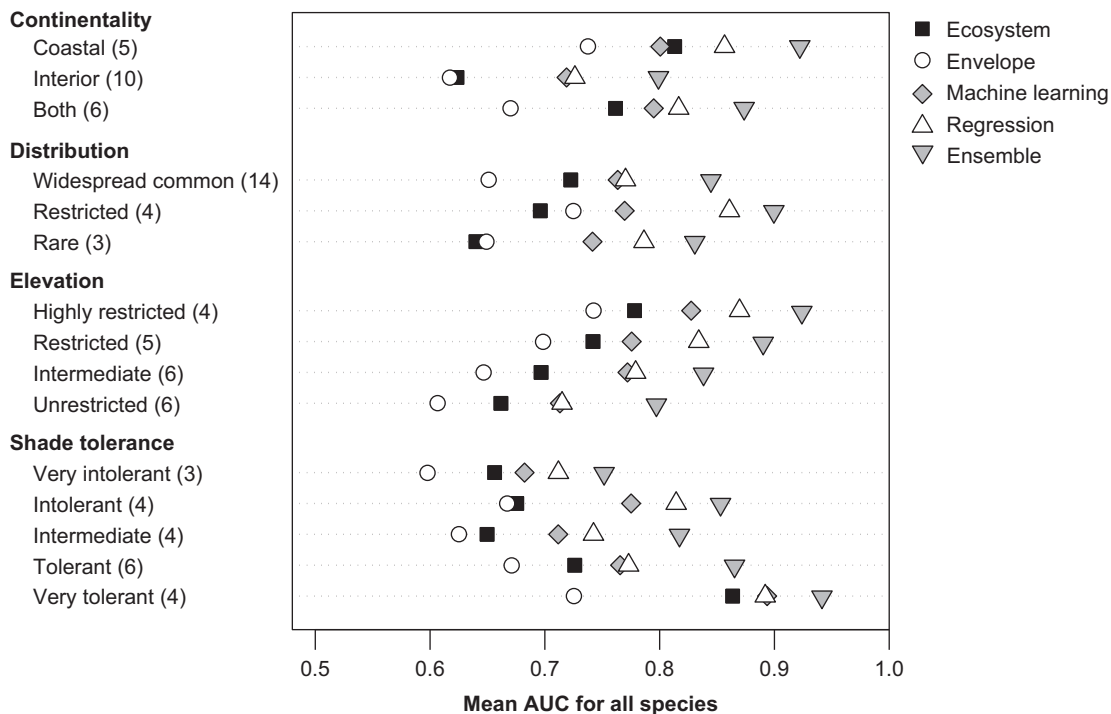
Figure 2. Model accuracy as a function of biogeographic and ecological characteristics of species, evaluated by the area under the curve (AUC) of the receiver operating characteristic. The values represent an average of the two independent validations (regional and temporal extrapolations), and are aggregated by the five model categories used in Fig. 1 and explained in Table 1. The number of species in each category is noted in parentheses (Table 2). Individual AUC values for all species, methods, and validation techniques are provided in Supplementary material Appendix 1.

Table 4. The variance in model accuracy (AUC) explained by the modelling method as well as by the various biogeographic and ecological characteristics of species (as listed in Table 2).

Component	Variance explained
Modelling method	16.4%
Distribution range	14.7%
Shade tolerance	6.5%
Taxon/group	5.8%
Elevation range	2.7%
Distribution type	0.0%
Total	46.1%
Error	53.9%

Discussion

Are species distribution models accurate?

Measured by area under the curve of the receiver operating characteristic, the species distribution models we evaluated in this paper performed reasonably well with the exception of the surface range envelope method, which we exclude from all subsequent discussion and summary statistics (Fig. 1). In general terms, an AUC > 0.90 is considered excellent, between 0.80 and 0.90 good, between 0.70 to 0.80 fair, and < 0.70 poor (Muller et al. 2010). On this scale, ecosystem-based methods showed a good model fit in out-of-bag validations with a mean AUC of 0.86, while individual species-based models showed excellent predictive accuracy. For the independent evaluations (regional and temporal), both species-based models and ecosystem-based techniques had fair predictive capabilities with mean AUC values of 0.74 and 0.72, respectively, values which are slightly lower than some hindcast studies for plant species (Araújo et al. 2005a, Rodríguez-Sánchez and Arroyo 2008, Dobrowski et al. 2011), while closely comparable or higher than others (Giesecke et al. 2007, Pearman et al. 2008). We should note, however, that our independent model evaluations are demanding in that they extrapolate spatially up to 2500 km, and temporally up to 21 000 yr into the past. We have previously shown that climate change expected for western North America would not include climates as novel as observed during the late Quaternary (Roberts and Hamann 2011).

The somewhat poorer performance of ecosystem-based approach may be due to the constraint of modelling fixed species communities, where species assemblages are limited to compositions represented on the modern landscape. This is conceptually problematic for reconstructions from past time periods and spatial extrapolations far beyond data coverage that may require species assemblages without analogues in the training data (Williams and Jackson 2007, Roberts and Hamann 2011). From a methodological standpoint, the area under the curve of the receiver operating characteristic has limitations as an evaluative metric, as outlined by Lobo et al. (2008). As with all single statistics that summarise a model's accuracy, it is important to also closely investigate the model projections in detail (Lobo et al. 2008). As illustrated in Fig. 3, we often find large, obvious errors in the model output (e.g. extensive projected presences under the continental ice or extensive Douglas-fir habitat in Mexico) that go largely undetected by the AUC calculation due to a lack validation points in these locations. While conceptually

problematic, we find that ecosystem-based techniques are less prone to produce false positives for habitat projections in areas that lack data coverage for statistical evaluations.

How should we select techniques?

Our results show little evidence of model over-parameterisation (model over-fit) among the techniques. Methods with high AUC values based on non-independent validations, such as the random forest ensemble classifier, generalised boosting models, and generalised additive models also performed well in independent tests (Fig. 1, Table 3). Somewhat surprisingly, relatively high correlations in AUC values between non-independent and independent validations suggests that simple, out-of-bag evaluations can be used for comparing and selecting modelling techniques. This also suggests that the relative quality of model projections into new geographic space or different time periods can reasonably be inferred from a non-independent evaluation, even though the absolute values may imply over-optimistic accuracies, as has been shown previously (Araújo et al. 2005a). However, we should note that a recent study by Dobrowski et al. (2011), incorporating a broader range of trees and shrub species, did not find strong relationship among independent and non-dependent evaluations.

Like others (Guisan et al. 2007, Dobrowski et al. 2011), we found that ecological and biogeographic traits had an influence on accuracy (Fig. 2). For example, habitat of elevation-restricted species appears more accurately modelled. A straight-forward explanation is that elevation limitations reflect a temperature optimum where a species is most competitive. Other types of habitat specialisations may not be described well by our available predictor variables. However, despite finding greater variation explained by our suite of ecological and biogeographic attributes of species as compared to modelling methods (Table 4), we did not find that any of these attributes favour certain modelling methods for trees (Fig. 2). We are inclined to conclude that, at least for trees, model selection based on life history or biogeographic traits is not necessary.

That said, with the development of software packages like BIOMOD, there is little reason to select individual techniques rather than relying on ensemble (or consensus) projections (Araújo et al. 2005b, Marmion et al. 2009). We find that the highest AUC, sensitivity, and specificity statistics could be consistently achieved with simple mean or median probability of presences from all techniques, which confirm other recent findings (Grenouillet et al. 2011). Unexpectedly, including even the poorest performing individual technique, surface range envelope, served to increase the predictive accuracy of the ensemble projections.

Ecosystem-based models

Modelling approaches that incorporate community data or species assemblages have been implemented and evaluated before (review by Ferrier and Guisan 2006). These approaches use a species composition (a collection of species and their particular frequencies) as the dependent variable in the model (Elith et al. 2006, Baselga and Araújo 2009),
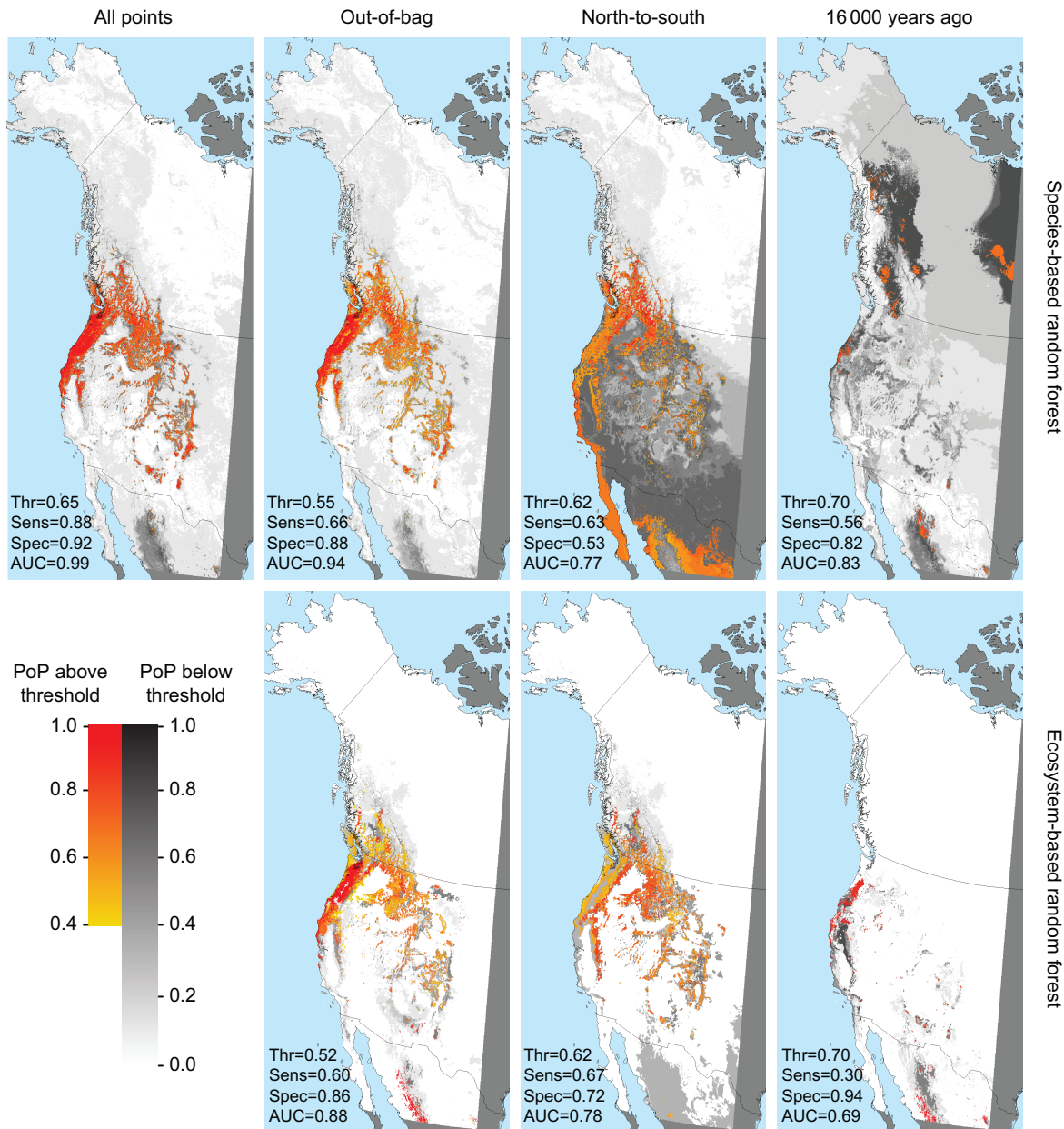
Figure 3. Projected probability of presence (PoP) for *Pseudotsuga meziesii* (Douglas-fir) using the species- and ecosystem-based random forest ensemble classifier under four model training and validation scenarios: training and validation based on the entire dataset (all points), a random data split of 67% for training and 33% for evaluation (out-of-bag), a 49° latitude data split, extrapolating south for validation (north-to-south) and projections based on palaeoclimate data using pollen and fossil data from the end of the last ice age. The area under the curve (AUC), model sensitivity (Sens) and model specificity (Spec) represent the accuracy of projections. Threshold probabilities (Thr) are determined by the AUC calculation and represent the PoP for which the evaluation error rate is minimal.

which is different to our approach of predicting a class variable that represents delineated ecological regions with a known species composition. We find that predictive accuracy of our ecosystem-based approach is somewhat inferior to individual species models, which is similar to an equivalent evaluation of community-based models by Baselga and Araújo (2009). It is further notable that the ecosystem-based methods behave poorly for the hindcasts toward the last glacial maximum, when different climate conditions drove major changes in species communities.

However, maps based on ecosystem-based models generally appear to produce better defined species range limits,

reflected by high specificity statistics, with probability of presence values rapidly approaching zero outside the actual species range. In spatial or temporal extrapolations far beyond training data, the ecosystem-based models may be more robust because the entire multivariate climate space of the study area is well defined by sampling all delineated ecosystems without bias. In fact, the example that we included for *Pseudotsuga meziesii* for 16 000 yr ago (Fig. 3) corresponds well to phylogeography studies based on genetic data. Li and Adams (1989) identify three genetically distinct populations of the species from Mexico, the interior Rocky Mountains, and the coast of western North America from which they

postulate three glacial refugia based on genetic data, which appear well-defined in the model projections. It therefore appears that in this example better robustness and specificity of ecosystem-based models outweigh the inferred benefits of higher statistical accuracy of individual species models.

A final desirable characteristic of the ecosystem-based modelling approach is that within-population genetic diversity can be integrated into species distribution models, which has been previously proposed (Botkin et al. 2007, Thuiller et al. 2008). The use of delineated ecoregions as training units within the model allows for the division of a species range into small, genetically homogenous populations, which in turn facilitates the subsequent tracking of individual populations under climate change projections (Gray and Hamann 2011). Simply, it is possible to determine the geographic location where the habitat (climate niche, in this case) of a species in a future model projection originated in the present day. If genetic data on adaptive differentiation of populations is available, ecosystem-based models can guide assisted migration efforts at the population level, rather than at the species level (for a detailed discussion, Gray and Hamann 2011, Gray et al. 2011, Hamann et al. 2011). Furthermore, in addition to species frequencies, any other ecosystem attribute, including those applicable to management prescriptions and conservation objectives (e.g. disturbance regimes), can potentially be matched to anticipated future climates. This makes ecosystem-based methods useful as effective decision-making tools for climate-informed conservation and resource management applications.

References

Araújo, M. B. and Williams, P. H. 2000. Selecting areas for species persistence using occurrence data. – Biol. Conserv. 96: 331–345.

Araújo, M. B. et al. 2005a. Validation of species-climate impact models under climate change. – Global Change Biol. 11: 1504–1513.

Araújo, M. B. et al. 2005b. Reducing uncertainty in projections of extinction risk from climate change. – Global Ecol. Biogeogr. 14: 529–538.

Baselga, A. and Araújo, M. B. 2009. Individualistic vs community modelling of species distributions under climate change. – Ecography 32: 55–65.

Beaumont, L. J. and Hughes, L. 2002. Potential changes in the distributions of latitudinally restricted Australian butterfly species in response to climate change. – Global Change Biol. 8: 954–971.

Betchtold, W. A. and Patterson, P. L. 2005. The enhanced forest inventory and analysis national sample design and estimation procedures. – Gen. Tech. Rep. SRS-80, United States Dept of Agriculture, Forest Service, Southern Research Station, Asheville, NC.

Botkin, D. B. et al. 2007. Forecasting the effects of global warming on biodiversity. – Bioscience 57: 227–236.

Breiman, L. 2001. Random forests. – Machine Learn. 45: 5–32.

Burns, R. M. et al. 1990. Silvics of North America: 1. conifers; 2. hardwoods. – Agricultural Handbook 654, U.S. Dept of Agriculture, Forest Service.

COHMAP 1988. Climatic changes of the last 18,000 years – observations and model simulations. – Science 241: 1043–1052.

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

Dobrowski, B. et al. 2011. Modeling plant ranges over 75 years of climate change in California, USA: temporal transferability and species traits. – Ecol. Monogr. 81: 241–257.

Dyke, A. S. 2005. Late Quaternary vegetation history of northern North America based on pollen, macrofossil, and faunal remains. – Geogr. Phys. Quat. 59: 52.

Elith, J. and Graham, C. H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. – Ecography 32: 66–77.

Elith, J. and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. – Annu. Rev. Ecol. Evol. Syst. 40: 677–697.

Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – Ecography 29: 129–151.

Fawcett, T. 2006. An introduction to ROC analysis. – Pattern Recognit. Lett. 27: 861–874.

Ferrier, S. and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. – J. Appl. Ecol. 43: 393–404.

Fløjgaard, C. et al. 2009. Ice age distributions of European small mammals: insights from species distribution modelling. – J. Biogeogr. 36: 1152–1163.

Franklin, J. 2009. Mapping species distributions. – Cambridge Univ. Press.

Giesecke, T. et al. 2007. Towards an understanding of the Holocene distribution of *Fagus sylvatica* L. – J. Biogeogr. 34: 118–131.

Govt. of Alberta 2004. Ecological Site Information System (ESIS). – Sustainable Resource Development, Edmonton, AB.

Gray, L. K. and Hamann, A. 2011. Strategies for reforestation under uncertain future climates: guidelines for Alberta, Canada. – PLoS One 6: e22977.

Gray, L. K. et al. 2011. Assisted migration to address climate change: recommendations for aspen reforestation in western Canada. – Ecol. Appl. 21: 1591–1603.

Grenouillet, G. et al. 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. – Ecography 34: 9–17.

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

Guisan, A. et al. 2007. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? – Ecol. Monogr. 77: 615–630.

Hamann, A. and Wang, T. L. 2005. Models of climatic normals for genecology and climate change studies in British Columbia. – Agric. For. Meteorol. 128: 211–221.

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

Hamann, A. et al. 2011. Developing seed zones and transfer guidelines with multivariate regression trees. – Tree Genet. Genom. 7: 399–408.

Hogg, E. H. 1997. Temporal scaling of moisture and the forest-grassland boundary in western Canada. – Agric. For. Meteorol. 84: 115–122.

Kutzbach, J. et al. 1998. Climate and biome simulations for the past 21,000 years. – Quat. Sci. Rev. 17: 473–506.

Leathwick, J. R. et al. 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. – Mar. Ecol. Prog. Ser. 321: 267–281.

Li, P. and Adams, W. T. 1989. Range-wide patterns of allozyme variation in Douglas-fir (*Pseudotsuga menziesii*). – Can. J. For. Res. 19: 149–161.

Lobo, J. M. et al. 2008. AUC: a misleading measure of the performance of predictive distribution models. – Global Ecol. Biogeogr. 17: 145–151.

Marmion, M. et al. 2009. Evaluation of consensus methods in predictive species distribution modelling. – Divers. Distrib. 15: 59–69.

Martinez-Meyer, E. and Peterson, A. T. 2006. Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. – J. Biogeogr. 33: 1779–1789.

Martinez-Meyer, E. et al. 2004. Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. – Global Ecol. Biogeogr. 13: 305–314.

Mbogga, M. S. et al. 2009. Historical and projected climate data for natural resource management in western Canada. – Agric. For. Meteorol. 149: 881–890.

Mbogga, M. S. et al. 2010. Bioclimate envelope model predictions for natural resource management: dealing with uncertainty. – J. Appl. Ecol. 47: 731–740.

Morueta-Holme, N. et al. 2010. Climate change risks and conservation implications for a threatened small-range mammal species. – PLoS One 5: e10360.

Muller, M. P. et al. 2010. Evaluation of pneumonia severity and acute physiology scores to predict ICU admission and mortality in patients hospitalized for influenza. – PLoS One 5: e9563.

Overpeck, J. T. et al. 1992. Mapping eastern North-American vegetation change of the past 18 Ka: no-analogs and the future. – Geology 20: 1071–1074.

Pearman, P. B. et al. 2008. Prediction of plant species distributions across six millennia. – Ecol. Lett. 11: 357–369.

Prasad, A. M. et al. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. – Ecosystems 9: 181–199.

Randin, C. F. et al. 2006. Are niche-based species distribution models transferable in space? – J. Biogeogr. 33: 1689–1703.

Roberts, D. R. and Hamann, A. 2011. Predicting potential climate change impacts with bioclimate envelope models: a palaeoecological perspective. – Global Ecol. Biogeogr. in press.

Rodríguez-Sánchez, F. and Arroyo, J. 2008. Reconstructing the demise of Tethyan plants: climate-driven range dynamics of *Laurus* since the Pliocene. – Global Ecol. Biogeogr. 17: 685–695.

Segurado, P. and Araújo, M. B. 2004. An evaluation of methods for modelling species distributions. – J. Biogeogr. 31: 1555–1568.

Segurado, P. et al. 2006. Consequences of spatial autocorrelation for niche-based models. – J. Appl. Ecol. 43: 433–444.

Sing, T. et al. 2005. ROCR: visualizing classifier performance in R. – Bioinformatics 21: 3940–3941.

Thompson, R. S. and Anderson, K. H. 2000. Biomes of western North America at 18,000, 6000 and 0 C-14 yr BP reconstructed from pollen and packrat midden data. – J. Biogeogr. 27: 555–584.

Thuiller, W. 2003. BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. – Global Change Biol. 9: 1353–1362.

Thuiller, W. et al. 2008. Predicting global change impacts on plant species' distributions: future challenges. – Perspect. Plant Ecol. Evol. Syst. 9: 137–152.

Thuiller, W. et al. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. – Ecography 32: 369–373.

Wang, T. et al. 2006. Development of scale-free climate data for western Canada for use in resource management. – Int. J. Climatol. 26: 383–397.

Williams, J. W. and Jackson, S. T. 2007. Novel climates, no-analog communities, and ecological surprises. – Front. Ecol. Environ. 5: 475–482.

Williams, J. W. et al. 2001. Dissimilarity analyses of late-Quaternary vegetation and climate in eastern North America. – Ecology 82: 3346–3362.

Zarnetske, P. L. et al. 2007. Habitat classification modeling with incomplete data: pushing the habitat envelope. – Ecol. Appl. 17: 1714–1726.

Validating species distribution models with fossil, pollen, and modern census data

David R. Roberts* & Andreas Hamann

Department of Renewable Resources, University of Alberta. 751 General Services Building, University of Alberta, Edmonton, Alberta, Canada, T6G2H1.

* Email: drr3@ualberta.ca, Tel: (780) 492-2540, Fax: (780) 492-4323

Appendix 1:

Each table contains AUC values for all species and all methods for one of the four evaluation methods: all-points (Table S1), out-of-bag data split (Table S2), north-to-south regional evaluation (Table S3), and the past-periods temporal evaluation (Table S4). Average AUC values are given for each individual method (across all species) and for each individual species (across all methods, excluding the ensembles) for each evaluation scenario. Total average AUC for all species and all methods is shown in bold.

Table A1: AUC values by species and modelling method for the all-points evaluation. Averages are given for each species and each modelling methods. The overall average for all models and all species is shown in bold.

Table A2: AUC values by species and modelling method for the out-of-bag data split evaluation. Averages are given for each species and each modelling methods. The overall average for all models and all species is shown in bold.

Table A3: AUC values by species and modelling method for the north-to-south regional evaluation. Averages are given for each species and each modelling method. The overall average for all models and all species is shown in bold. Species without records in both the north and south data sets are denoted with dashes (-) and were excluded from the analysis.

Table A4: AUC values by species and modelling method for the fossil/pollen temporal evaluation for all records from 6000, 11000, 14000, 16000, and 21000 years before the present. Averages are given for each species and each modelling methods. Species with fewer than 10 observations in the record are denoted with dashes (-) and were excluded from the analysis.

Table A1: AUC values by species and modelling method for the all-points evaluation. Averages are given for each species and each modelling methods. The overall average for all models and all species is shown in bold.

Species	Classification Tree Analysis	Discriminant Analysis	Generalised Additive	Generalised Boosting	Generalised Linear	Adaptive Reg. Splines	Artificial Neural Net.	randomForest	Surface Range Envelopes	*Average*	Species: Kappa	Species: Mean	Species: Median	Species: ROC	Species: TSS	Species: Weighted	All metods: Mean	All methods: Median
Abies amabilis (pacific silver fir)	0.97	0.95	0.98	0.97	0.98	0.96	0.97	1.00	0.79	*0.95*	0.99	0.99	0.99	0.99	0.99	1.00	0.99	0.99
Abies lasiocarpa (subalpine fir)	0.84	0.82	0.89	0.90	0.89	0.85	0.89	1.00	0.71	*0.86*	0.94	0.91	0.93	0.93	0.95	0.99	0.93	0.92
Abies procera (noble fir)	0.96	0.96	0.99	0.99	0.98	0.96	0.87	1.00	0.79	*0.94*	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
Acer macrophyllum (bigleaf maple)	0.98	0.97	0.98	0.98	0.98	0.98	0.50	1.00	0.81	*0.91*	1.00	0.99	0.99	0.99	0.99	1.00	0.99	0.99
Alnus rubra (red alder)	0.97	0.97	0.98	0.98	0.98	0.97	0.96	1.00	0.82	*0.96*	0.99	0.98	0.99	0.99	0.99	1.00	0.99	0.99
Betula papyrifera (paper birch)	0.90	0.83	0.90	0.88	0.89	0.89	0.88	0.98	0.73	*0.88*	0.97	0.93	0.95	0.94	0.95	0.99	0.94	0.93
Calocedrus decurrens (incense cedar)	0.98	0.97	0.99	0.99	0.99	0.98	0.99	1.00	0.80	*0.97*	1.00	0.99	1.00	1.00	1.00	1.00	1.00	0.99
Chamaecyparis nootkatensis (yellow cedar)	0.98	0.98	0.99	0.98	0.99	0.99	0.97	1.00	0.81	*0.96*	1.00	0.99	0.99	1.00	1.00	1.00	0.99	0.99
Larix occidentalis (western larch)	0.94	0.85	0.95	0.94	0.95	0.80	0.95	1.00	0.79	*0.91*	0.99	0.96	0.97	0.97	0.97	1.00	0.98	0.97
Picea engelmannii (Engelman spruce)	0.88	0.86	0.91	0.88	0.91	0.89	0.90	0.99	0.70	*0.88*	0.96	0.93	0.95	0.95	0.96	0.99	0.94	0.94
Picea glauca (white spruce)	0.89	0.86	0.90	0.89	0.90	0.83	0.89	0.98	0.74	*0.88*	0.95	0.92	0.94	0.94	0.95	0.99	0.94	0.93
Picea mariana (black spruce)	0.91	0.85	0.90	0.89	0.91	0.86	0.92	0.98	0.73	*0.88*	0.97	0.94	0.94	0.94	0.95	0.99	0.95	0.94
Picea sitchensis (sitka spruce)	0.97	0.98	0.99	0.98	0.99	0.98	0.97	1.00	0.83	*0.97*	1.00	0.99	0.99	1.00	1.00	1.00	0.99	0.99
Pinus albicaulis (whitebark pine)	0.93	0.91	0.95	0.93	0.95	0.90	0.93	1.00	0.74	*0.91*	0.99	0.96	0.98	0.97	0.97	1.00	0.97	0.96
Pinus contorta (lodgepole pine)	0.79	0.74	0.79	0.78	0.78	0.79	0.80	0.97	0.70	*0.79*	0.92	0.85	0.88	0.88	0.90	0.99	0.89	0.86
Pinus edulis (pinyon pine)	0.97	0.96	0.98	0.98	0.98	0.97	0.99	1.00	0.81	*0.96*	0.99	0.99	0.99	0.99	1.00	1.00	0.99	0.99
Pinus monticola (western white pine)	0.94	0.87	0.94	0.93	0.93	0.93	0.84	1.00	0.77	*0.90*	0.99	0.96	0.98	0.96	0.96	1.00	0.97	0.96
Pinus ponderosa (ponderosa pine)	0.93	0.88	0.95	0.92	0.94	0.89	0.89	1.00	0.76	*0.91*	0.98	0.97	0.97	0.97	0.98	1.00	0.97	0.97
Populus tremuloides (trembling aspen)	0.83	0.76	0.83	0.83	0.84	0.83	0.85	0.98	0.69	*0.82*	0.94	0.87	0.91	0.90	0.91	0.99	0.90	0.88
Pseudotsuga menziesii (Douglas-fir)	0.88	0.81	0.90	0.88	0.90	0.90	0.83	0.99	0.72	*0.87*	0.96	0.93	0.94	0.94	0.95	0.99	0.94	0.93
Sequoia sempervirens (giant sequoia)	0.99	1.00	1.00	1.00	1.00	1.00	0.95	1.00	0.79	*0.97*	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Thuja plicata (western redcedar)	0.94	0.90	0.94	0.92	0.94	0.94	0.94	1.00	0.76	*0.92*	0.98	0.96	0.98	0.97	0.98	0.99	0.97	0.96
Tsuga heterophylla (western hemlock)	0.96	0.94	0.97	0.96	0.97	0.96	0.96	1.00	0.77	*0.94*	0.99	0.98	0.98	0.98	0.98	0.99	0.98	0.98
Tsuga mertensiana (mountain hemlock)	0.97	0.95	0.97	0.97	0.98	0.97	0.96	1.00	0.80	*0.95*	0.99	0.98	0.99	0.99	0.99	1.00	0.99	0.98
Average	*0.93*	*0.90*	*0.94*	*0.93*	*0.94*	*0.92*	*0.90*	*0.99*	*0.76*	**0.91**	*0.98*	*0.96*	*0.97*	*0.97*	*0.97*	*1.00*	*0.97*	*0.96*

Table A2: AUC values by species and modelling method for the out-of-bag data split evaluation. Averages are given for each species and each modelling methods. The overall average for all models and all species is shown in bold.

Species	Ecosystem-based Discriminant Analysis	Minimum Distance	randomForest	Species-based Classification Tree Analysis	Discriminant Analysis	Generalised Additive	Generalised Boosting	Generalised Linear	Adaptive Reg. Splines	Artificial Neural Net.	randomForest	Surface Range Envelopes	Average	Ensemble Ecosystem: Mean	Species: Kappa	Species: Mean	Species: Median	Species: ROC	Species: TSS	Species: Weighted	All metods: Mean	All methods: Median
Abies amabilis (pacific silver fir)	0.93	0.88	0.93	0.95	0.95	0.98	0.98	0.98	0.96	0.96	0.99	0.80	*0.94*	0.97	0.98	0.98	0.96	0.97	0.98	0.99	0.98	0.98
Abies lasiocarpa (subalpine fir)	0.82	0.78	0.82	0.83	0.82	0.88	0.89	0.88	0.88	0.89	0.92	0.71	*0.84*	0.86	0.91	0.90	0.87	0.88	0.88	0.92	0.91	0.90
Abies procera (noble fir)	0.83	0.70	0.85	0.86	0.96	0.98	0.98	0.97	0.95	0.91	0.94	0.81	*0.90*	0.91	0.99	0.97	0.96	0.97	0.97	0.99	0.99	0.98
Acer macrophyllum (bigleaf maple)	0.95	0.93	0.96	0.96	0.97	0.98	0.98	0.98	0.97	0.85	0.99	0.78	*0.94*	0.97	0.99	0.98	0.98	0.98	0.98	0.99	0.98	0.98
Alnus rubra (red alder)	0.96	0.93	0.96	0.93	0.97	0.97	0.97	0.97	0.94	0.95	0.97	0.81	*0.94*	0.97	0.98	0.97	0.94	0.96	0.96	0.98	0.97	0.97
Betula papyrifera (paper birch)	0.87	0.83	0.89	0.88	0.83	0.89	0.89	0.90	0.89	0.90	0.93	0.73	*0.87*	0.90	0.91	0.90	0.88	0.89	0.90	0.92	0.91	0.91
Calocedrus decurrens (incense cedar)	0.96	0.93	0.98	0.96	0.97	0.99	0.99	0.99	0.97	0.96	0.99	0.79	*0.96*	0.99	0.99	0.99	0.97	0.99	0.99	0.99	0.99	0.99
Chamaecyparis nootkatensis (yellow cedar)	0.94	0.93	0.94	0.96	0.98	0.99	0.99	0.99	0.99	0.97	0.99	0.80	*0.95*	0.97	0.99	0.99	0.97	0.99	0.99	0.99	0.99	0.99
Larix occidentalis (western larch)	0.86	0.77	0.87	0.87	0.85	0.95	0.95	0.95	0.93	0.94	0.97	0.78	*0.89*	0.92	0.96	0.95	0.88	0.95	0.95	0.96	0.96	0.97
Picea engelmannii (Engelman spruce)	0.81	0.74	0.81	0.88	0.86	0.91	0.90	0.90	0.90	0.91	0.93	0.69	*0.85*	0.86	0.92	0.91	0.89	0.90	0.91	0.93	0.92	0.92
Picea glauca (white spruce)	0.87	0.85	0.88	0.86	0.86	0.90	0.90	0.90	0.83	0.90	0.93	0.74	*0.87*	0.90	0.91	0.90	0.88	0.89	0.89	0.92	0.91	0.91
Picea mariana (black spruce)	0.88	0.86	0.89	0.88	0.85	0.91	0.91	0.91	0.84	0.88	0.94	0.73	*0.87*	0.90	0.93	0.92	0.88	0.92	0.92	0.94	0.93	0.92
Picea sitchensis (sitka spruce)	0.94	0.91	0.95	0.97	0.98	0.99	0.99	0.99	0.98	0.50	0.99	0.81	*0.92*	0.97	0.99	0.99	0.97	0.98	0.99	0.99	0.99	0.99
Pinus albicaulis (whitebark pine)	0.78	0.78	0.77	0.89	0.92	0.95	0.94	0.94	0.92	0.92	0.95	0.74	*0.87*	0.85	0.95	0.94	0.88	0.94	0.94	0.96	0.95	0.95
Pinus contorta (lodgepole pine)	0.81	0.78	0.82	0.80	0.74	0.79	0.81	0.79	0.79	0.79	0.87	0.70	*0.79*	0.84	0.84	0.82	0.82	0.81	0.82	0.86	0.85	0.83
Pinus edulis (pinyon pine)	0.80	0.74	0.83	0.97	0.96	0.98	0.98	0.98	0.98	0.98	0.99	0.81	*0.92*	0.91	0.99	0.98	0.97	0.98	0.98	0.99	0.99	0.99
Pinus monticola (western white pine)	0.87	0.79	0.89	0.89	0.89	0.94	0.94	0.94	0.93	0.94	0.95	0.77	*0.89*	0.93	0.95	0.94	0.89	0.93	0.94	0.95	0.95	0.95
Pinus ponderosa (ponderosa pine)	0.83	0.80	0.88	0.92	0.87	0.94	0.95	0.94	0.93	0.95	0.96	0.75	*0.89*	0.92	0.95	0.95	0.93	0.94	0.95	0.96	0.96	0.96
Populus tremuloides (trembling aspen)	0.82	0.79	0.83	0.82	0.75	0.83	0.84	0.84	0.83	0.84	0.88	0.69	*0.81*	0.85	0.86	0.85	0.82	0.83	0.84	0.88	0.86	0.86
Pseudotsuga menziesii (Douglas-fir)	0.85	0.84	0.88	0.86	0.81	0.90	0.91	0.91	0.90	0.91	0.94	0.72	*0.87*	0.90	0.92	0.92	0.89	0.89	0.90	0.93	0.92	0.92
Sequoia sempervirens (giant sequoia)	0.80	0.76	0.96	0.99	1.00	1.00	1.00	1.00	0.98	0.95	1.00	0.80	*0.94*	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Thuja plicata (western redcedar)	0.93	0.87	0.94	0.93	0.90	0.94	0.95	0.94	0.93	0.93	0.97	0.76	*0.92*	0.95	0.96	0.95	0.91	0.94	0.95	0.96	0.96	0.96
Tsuga heterophylla (western hemlock)	0.94	0.90	0.95	0.95	0.94	0.97	0.97	0.97	0.96	0.96	0.98	0.76	*0.94*	0.97	0.97	0.97	0.94	0.96	0.96	0.98	0.97	0.98
Tsuga mertensiana (mountain hemlock)	0.90	0.87	0.86	0.93	0.95	0.97	0.97	0.97	0.96	0.92	0.97	0.78	*0.92*	0.94	0.97	0.97	0.93	0.96	0.96	0.98	0.97	0.98
Average	*0.87*	*0.83*	*0.89*	*0.91*	*0.90*	*0.94*	*0.94*	*0.94*	*0.92*	*0.90*	*0.95*	*0.76*	**0.90**	*0.92*	*0.95*	*0.94*	*0.92*	*0.94*	*0.94*	*0.96*	*0.95*	*0.95*

Table A3: AUC values by species and modelling method for the north-to-south regional evaluation. Averages are given for each species and each modelling method. The overall average for all models and all species is shown in bold. Species without records in both the north and south data sets are denoted with dashes (-) and were excluded from the analysis.

Species	Modelling method													Average	Ensemble							All methods: Mean	All methods: Median
	Ecosystem-based			Species-based																			
	Discriminant Analysis	Minimum Distance	randomForest	Classification Tree Analysis	Discriminant Analysis	Generalised Additive	Generalised Boosting	Generalised Linear	Adaptive Reg. Splines	Artificial Neural Net.	randomForest	Surface Range Envelopes		Ecosystem: Mean	Species: Kappa	Species: Mean	Species: Median	Species: ROC	Species: TSS	Species: Weighted		
Abies amabilis (pacific silver fir)	0.92	0.91	0.95	0.69	0.94	0.98	0.96	0.97	0.86	0.96	0.97	0.67	*0.90*	0.97	0.98	0.98	0.95	0.97	0.98	0.98	0.98	0.98
Abies lasiocarpa (subalpine fir)	0.83	0.88	0.92	0.84	0.93	0.95	0.87	0.85	0.78	0.92	0.93	0.54	0.85	0.93	0.91	0.93	0.86	0.84	0.88	0.91	0.93	0.95
Abies procera (noble fir)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Acer macrophyllum (bigleaf maple)	0.76	0.86	0.93	0.80	0.89	0.75	0.94	0.75	0.90	0.74	0.90	0.50	*0.81*	0.94	0.91	0.90	0.92	0.91	0.91	0.90	0.92	0.92
Alnus rubra (red alder)	0.94	0.95	0.94	0.91	0.94	0.96	0.96	0.96	0.75	0.85	0.88	0.54	*0.88*	0.96	0.96	0.97	0.94	0.96	0.96	0.97	0.97	0.96
Betula papyrifera (paper birch)	0.49	0.63	0.83	0.88	0.87	0.93	0.90	0.94	0.89	0.79	0.69	0.51	*0.78*	0.80	0.92	0.90	0.83	0.86	0.89	0.88	0.93	0.93
Calocedrus decurrens (incense cedar)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chamaecyparis nootkatensis (yellow cedar)	0.95	0.87	0.88	0.72	0.95	0.98	0.95	0.98	0.85	0.85	0.86	0.54	*0.86*	0.95	0.96	0.94	0.77	0.90	0.94	0.96	0.97	0.96
Larix occidentalis (western larch)	0.67	0.60	0.65	0.67	0.21	0.48	0.63	0.63	0.43	0.54	0.67	0.56	*0.56*	0.76	0.64	0.64	0.65	0.71	0.68	0.69	0.67	0.71
Picea engelmannii (Engelman spruce)	0.60	0.73	0.80	0.83	0.83	0.49	0.87	0.25	0.33	0.82	0.89	0.54	*0.67*	0.87	0.80	0.75	0.84	0.88	0.88	0.86	0.86	0.89
Picea glauca (white spruce)	0.34	0.55	0.79	0.34	0.90	0.95	0.53	0.90	0.95	0.82	0.90	0.50	*0.71*	0.42	0.90	0.88	0.45	0.45	0.48	0.90	0.89	0.98
Picea mariana (black spruce)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Picea sitchensis (sitka spruce)	0.74	0.85	0.97	0.89	0.99	0.99	0.91	0.98	0.72	0.56	0.99	0.50	*0.84*	0.97	0.99	0.99	0.92	0.98	0.97	0.99	0.99	0.99
Pinus albicaulis (whitebark pine)	0.48	0.83	0.81	0.72	0.93	0.86	0.93	0.35	0.90	0.91	0.89	0.63	*0.77*	0.84	0.88	0.89	0.82	0.92	0.91	0.87	0.91	0.93
Pinus contorta (lodgepole pine)	0.79	0.84	0.84	0.85	0.84	0.85	0.89	0.66	0.86	0.79	0.82	0.53	0.80	0.89	0.86	0.87	0.83	0.78	0.85	0.87	0.88	0.89
Pinus edulis (pinyon pine)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Pinus monticola (western white pine)	0.70	0.75	0.70	0.44	0.51	0.73	0.82	0.80	0.79	0.71	0.85	0.56	*0.70*	0.83	0.81	0.81	0.78	0.84	0.85	0.80	0.83	0.83
Pinus ponderosa (ponderosa pine)	0.63	0.56	0.65	0.61	0.50	0.58	0.57	0.66	0.67	0.61	0.58	0.51	*0.60*	0.71	0.66	0.65	0.68	0.70	0.70	0.67	0.68	0.67
Populus tremuloides (trembling aspen)	0.53	0.69	0.82	0.78	0.61	0.88	0.87	0.75	0.84	0.74	0.64	0.50	*0.72*	0.85	0.83	0.84	0.56	0.63	0.61	0.77	0.86	0.88
Pseudotsuga menziesii (Douglas-fir)	0.67	0.73	0.78	0.53	0.38	0.52	0.54	0.69	0.55	0.64	0.77	0.57	*0.61*	0.81	0.71	0.69	0.64	0.64	0.65	0.77	0.78	0.79
Sequoia sempervirens (giant sequoia)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Thuja plicata (western redcedar)	0.84	0.88	0.90	0.85	0.74	0.94	0.86	0.88	0.80	0.90	0.84	0.61	*0.84*	0.92	0.94	0.95	0.94	0.93	0.93	0.94	0.95	0.94
Tsuga heterophylla (western hemlock)	0.92	0.93	0.94	0.93	0.96	0.95	0.97	0.87	0.77	0.88	0.94	0.59	*0.89*	0.96	0.96	0.94	0.95	0.95	0.95	0.96	0.97	0.97
Tsuga mertensiana (mountain hemlock)	0.71	0.78	0.86	0.81	0.92	0.96	0.96	0.94	0.94	0.75	0.94	0.54	*0.84*	0.89	0.97	0.96	0.77	0.90	0.92	0.97	0.96	0.95
Average	*0.71*	*0.78*	*0.84*	*0.74*	*0.78*	*0.83*	*0.84*	*0.78*	*0.77*	*0.78*	*0.84*	*0.55*	**0.77**	*0.86*	*0.87*	*0.87*	*0.80*	*0.83*	*0.84*	*0.88*	*0.89*	*0.90*

Table A4: AUC values by species and modelling method for the fossil/pollen temporal evaluation for all records from 6000, 11000, 14000, 16000, and 21000 years before the present. Averages are given for each species and each modelling methods. Species with fewer than 10 observations in the record are denoted with dashes (-) and were excluded from the analysis.

Species	Ecosystem-based			Species-based									Average	Ensemble								
	Discriminant Analysis	Minimum Distance	randomForest	Classification Tree Analysis	Discriminant Analysis	Generalised Additive	Generalised Boosting	Generalised Linear	Adaptive Reg. Splines	Artificial Neural Net.	randomForest	Surface Range Envelopes		Ecosystem: Mean	Species: Kappa	Species: Mean	Species: Median	Species: ROC	Species: TSS	Species: Weighted	All metods: Mean	All methods: Median
Abies amabilis (pacific silver fir)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Abies lasiocarpa (subalpine fir)	0.58	0.63	0.57	0.67	0.54	0.73	0.72	0.63	0.71	0.69	0.74	0.52	*0.64*	0.61	0.71	0.73	0.70	0.71	0.70	0.73	0.68	0.72
Abies procera (noble fir)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Acer macrophyllum (bigleaf maple)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Alnus rubra (red alder)	0.79	0.78	0.75	0.52	0.87	0.89	0.87	0.84	0.92	0.85	0.84	0.62	*0.79*	0.82	0.89	0.91	0.70	0.88	0.85	0.88	0.89	0.91
Betula papyrifera (paper birch)	0.48	0.56	0.78	0.74	0.71	0.85	0.75	0.79	0.80	0.70	0.78	0.62	*0.71*	0.75	0.82	0.84	0.81	0.83	0.82	0.81	0.83	0.83
Calocedrus decurrens (incense cedar)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Chamaecyparis nootkatensis (yellow cedar)	0.56	0.73	0.62	0.56	0.85	0.58	0.59	0.52	0.39	0.54	0.74	0.49	*0.60*	0.77	0.58	0.71	0.50	0.60	0.57	0.67	0.84	0.73
Larix occidentalis (western larch)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Picea engelmannii (Engelman spruce)	0.57	0.57	0.52	0.68	0.43	0.64	0.74	0.61	0.54	0.62	0.78	0.49	*0.60*	0.56	0.67	0.70	0.53	0.65	0.65	0.73	0.61	0.69
Picea glauca (white spruce)	0.63	0.64	0.68	0.66	0.63	0.80	0.73	0.81	0.67	0.71	0.73	0.55	*0.69*	0.72	0.76	0.78	0.75	0.78	0.78	0.76	0.77	0.78
Picea mariana (black spruce)	0.63	0.64	0.73	0.72	0.65	0.82	0.79	0.76	0.75	0.72	0.79	0.55	*0.71*	0.75	0.79	0.83	0.66	0.77	0.78	0.81	0.81	0.85
Picea sitchensis (sitka spruce)	0.78	0.76	0.77	0.83	0.83	0.90	0.94	0.87	0.84	0.91	0.94	0.73	*0.84*	0.81	0.93	0.93	0.85	0.92	0.92	0.94	0.93	0.92
Pinus albicaulis (whitebark pine)	0.56	0.56	0.54	0.46	0.58	0.67	0.56	0.53	0.66	0.38	0.68	0.50	*0.56*	0.65	0.64	0.61	0.51	0.61	0.61	0.67	0.61	0.68
Pinus contorta (lodgepole pine)	0.62	0.57	0.59	0.50	0.68	0.71	0.67	0.66	0.72	0.72	0.69	0.50	*0.64*	0.63	0.69	0.71	0.66	0.66	0.68	0.69	0.70	0.72
Pinus edulis (pinyon pine)	0.61	0.50	0.54	0.47	0.91	0.74	0.54	0.65	0.82	0.30	0.78	0.50	*0.61*	0.65	0.78	0.75	0.59	0.63	0.76	0.75	0.90	0.78
Pinus monticola (western white pine)	0.63	0.64	0.56	0.45	0.81	0.79	0.72	0.81	0.73	0.78	0.78	0.57	*0.69*	0.66	0.83	0.82	0.67	0.82	0.81	0.82	0.83	0.82
Pinus ponderosa (ponderosa pine)	0.57	0.54	0.64	0.82	0.68	0.87	0.87	0.85	0.52	0.81	0.86	0.58	*0.72*	0.64	0.82	0.84	0.79	0.78	0.79	0.85	0.79	0.86
Populus tremuloides (trembling aspen)	0.54	0.54	0.50	0.45	0.65	0.51	0.40	0.47	0.63	0.54	0.43	0.48	*0.51*	0.58	0.46	0.48	0.49	0.44	0.44	0.43	0.55	0.53
Pseudotsuga menziesii (Douglas-fir)	0.66	0.68	0.69	0.70	0.79	0.82	0.81	0.75	0.80	0.77	0.83	0.59	*0.74*	0.71	0.83	0.83	0.78	0.77	0.81	0.83	0.83	0.83
Sequoia sempervirens (giant sequoia)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Thuja plicata (western redcedar)	0.85	0.88	0.82	0.91	0.90	0.95	0.90	0.93	0.89	0.87	0.92	0.52	*0.86*	0.88	0.93	0.93	0.89	0.93	0.93	0.94	0.93	0.94
Tsuga heterophylla (western hemlock)	0.82	0.79	0.79	0.87	0.91	0.92	0.88	0.88	0.89	0.90	0.89	0.58	*0.84*	0.84	0.92	0.91	0.89	0.91	0.91	0.92	0.92	0.91
Tsuga mertensiana (mountain hemlock)	0.76	0.79	0.71	0.82	0.88	0.89	0.86	0.80	0.79	0.81	0.86	0.61	*0.80*	0.82	0.89	0.89	0.68	0.89	0.86	0.88	0.90	0.90
Average	*0.65*	*0.66*	*0.66*	*0.66*	*0.74*	*0.78*	*0.74*	*0.73*	*0.73*	*0.70*	*0.78*	*0.56*	**0.70**	*0.71*	*0.77*	*0.79*	*0.69*	*0.75*	*0.76*	*0.78*	*0.80*	*0.80*