A Journal of Macroecology

Global Ecology and Biogeography

**RESEARCH
PAPER**

Predicting potential climate change impacts with bioclimate envelope models: a palaeoecological perspective

David R. Roberts* and Andreas Hamann

Department of Renewable Resources, University of Alberta, 751 General Services Building, Edmonton, AB T6G 2H1, Canada

ABSTRACT

Aim We assess the realism of bioclimate envelope model projections for anticipated future climates by validating ecosystem reconstructions for the late Quaternary with fossil and pollen data. Specifically, we ask: (1) do climate conditions with no modern analogue negatively affect the accuracy of ecosystem reconstructions? (2) are bioclimate envelope model projections biased towards under-predicting forested ecosystems? (3) given a palaeoecological perspective, are potential habitat projections for the 21st century within model capabilities?

Location Western North America.

Methods We used an ensemble classifier modelling approach (RandomForest) to spatially project the climate space of modern ecosystem classes throughout the Holocene (at 6000, 9000, 11,000, 14,000, 16,000, and 21,000 YBP) using palaeoclimate surfaces generated by two general circulation models (GFDL and CCM1). The degree of novel arrangement of climate variables was quantified with the multivariate Mahalanobis distance to the nearest modern climatic equivalent. Model projections were validated against biome classifications inferred from 1460 palaeoecological records.

Results Model accuracy assessed against independent palaeoecology data is generally low for the present day, increases for 6000 YBP, and then rapidly declines towards the last glacial maximum, primarily due to the under-prediction of forested biomes. Misclassifications were closely correlated with the degree of climate dissimilarity from the present day. For future projections, no-analogue climates unexpectedly emerged in the coastal Pacific Northwest but were absent throughout the rest of the study area.

Main conclusions Bioclimate envelope models could approximately reconstruct ecosystem distributions for the mid- to late-Holocene but proved unreliable in the Late Pleistocene. We attribute this failure to a combination of no-analogue climates and a potential lack of niche conservatism in tree species. However, climate dissimilarities in future projections are comparatively minor (similar to those of the mid-Holocene), and we conclude that no-analogue climates should not compromise the accuracy of model predictions for the next century.

Keywords

Bioclimate envelope models, climate change, climate dissimilarity, ecological niche models, extinction risk, Holocene, niche conservatism, no-analogue climate, RandomForest, western North America.

*Correspondence: David R. Roberts, Department of Renewable Resources, University of Alberta, 751 General Services Building, Edmonton, AB T6G 2H1, Canada. E-mail: drr3@ualberta.ca

INTRODUCTION

Natural fluctuations of global climate have occurred throughout earth's history, but in the coming centuries, anthropogenic factors may force global climate into conditions unseen for millions of years (IPCC, 2007). It has also been suggested that anticipated climatic conditions may include novel combinations of climate variables that do not exist in the present day nor have existed for millennia or longer (Crowley, 1990; Williams *et al.*, 2007; Salzmann *et al.*, 2009). Such 'no-analogue' climates could result in ecological communities that also lack modern analogues (Overpeck *et al.*, 1992; Williams *et al.*, 2001). It has therefore been questioned whether it is possible to predict a biological response (e.g. altered growth rates or demographic change) to future climate conditions with modelling approaches that are essentially correlative and based on currently observed spatial or temporal climate variation (Jackson & Williams, 2004; Williams & Jackson, 2007; Fitzpatrick & Hargrove, 2009; Van der Wal *et al.*, 2009).

A widely used class of models to predict potential species habitat under projected climate changes are bioclimate envelope models – also referred to as niche models or species distribution models. These models correlate environmental predictor variables such as climate with species occurrence data via statistical or machine learning procedures (e.g. Guisan & Zimmermann, 2000). This class of models has a number of important limitations that need to be considered when interpreting the results. For example, species interactions such as competition are not modelled in a direct way, but they are indirectly accounted for because bioclimate envelope models predict the realised niche rather than the fundamental niche. The models also rely on a number of assumptions such as the constancy of species' niches over time, genetic homogeneity among populations within a species, and the assumption of equilibrium of species distributions with current climate conditions (see reviews by Pearson & Dawson, 2003; Guisan & Thuiller, 2005; Araújo & Guisan, 2006).

While some of these assumptions may be violated, it is also widely understood that 'all models are wrong' (Box & Draper, 1987) and that their value lies in capturing relevant predictor variables and ignoring factors that have minor or no influence on the results at the scale of interest, which is often continental or global. To evaluate the potential realism of bioclimate envelope models, various statistical techniques exist for assessing accuracy and robustness of predictions. Most of these accuracy statistics rely on some form of cross-validation, where a subset of the data is used to build the predictive model and the remaining data is used to evaluate model accuracy. However, spatial autocorrelations in biological census data can substantially inflate the apparent accuracy of species distribution models that rely on cross-validation techniques (Segurado *et al.*, 2006). For this reason, model evaluation with truly independent data – for example, validation of species back-predictions using fossil and pollen data – has been proposed (Araújo *et al.*, 2005; Botkin *et al.*, 2007a). Back-predictions can also be used to test the validity of various assumptions underlying bioclimate envelope

model projections (Araújo *et al.*, 2005; Botkin *et al.*, 2007b; Nogués-Bravo, 2009).

While back-predicting species with climate envelopes is not a new idea (e.g. Prentice *et al.*, 1991), the field has seen rapid recent developments to identify causes of extinction (e.g. Rodríguez-Sánchez & Arroyo, 2008), to reconstruct migration routes and glacial refugia (e.g. Svenning *et al.*, 2008; Van der Wal *et al.*, 2009), and to help understand the evolutionary processes of geographic isolation, genetic differentiation, and speciation (e.g. Yesson & Culham, 2006; Carstens & Richards, 2007). In this paper, we contribute a new approach that projects ecosystem climate envelopes to more generally assess bioclimate envelope model capabilities at a continental scale. We focus on independent model validation and the issue of no-analogue climates. Additionally, we address the issue that bioclimate envelope models may over-estimate climate change threats to tree species, which tend to have large fundamental niches when mature and high within- and among-population genetic diversity (Hamrick, 2004). This could lead to underestimating either tree species' adaptability or their capability to persist in micro sites (e.g. Loehle & Leblanc, 1996; Morin *et al.*, 2008; Morin & Thuiller, 2009; Chen *et al.*, 2010).

Here, we carry out bioclimate envelope model-based back-predictions of ecosystems between the present and the last glacial maximum, for the periods 6000, 9000, 11,000, 14,000, 16,000, and 21,000 years before present (YBP). We evaluate the results with biome reconstructions based on fossil and pollen data at 1460 western North American study sites. Novel arrangement of climate variables is quantified with the multivariate Mahalanobis distance to the nearest modern equivalent. Our working hypothesis is that the emergence of no-analogue climates will increase model misclassification rates for palaeoecological records. We expect the predictive model to misclassify many fossil and pollen sites that represent forested ecosystems with high species diversity as too cold to sustain such communities. Lastly we aim at providing a palaeoecological perspective on whether ecosystem and species habitat projections for the 21st century are generally within model capabilities.

METHODS

To address these objectives in a broad way, we have carried out a relatively general analysis: rather than modelling individual species distributions, we used a bioclimate envelope model technique that uses ecosystem classes as the dependent variable. The model predicts several hundred fine-scale ecosystem classes which we summarise for broader ecosystem classifications. For these summaries we adopt the same biome classification that was used by Dyke (2005) to characterise fossil and pollen records (Table 1). For concise reporting at an even higher level, ecosystem projections were summarised into three categories: those which support forest communities (*Forested*) and those which do not support forest communities either due to heat/moisture constraints (*Dry*) or low temperature constraints (*Cold*) (also indicated in Table 1).

Table 1 List of western North American biome classes used in predictive modelling and inferred from pollen and fossil data by Dyke (2005). The dominant tree species or genera for each forested biome are described in parentheses.

Biome	Description
Non-forested dry	
Desert (DES)	Hot and dry areas with poor soil development, mainly devoid of vegetation.
Steppe (STE)	Bunchgrass with sagebrush and some woody shrubs.
Grassland (GRA)	Tall grasslands dominated by graminoid species.
Forested	
Savannah (SAV)	Open coniferous canopy with steppe, grass and shrub components (juniper and pinion pine with some lodgepole pine, Douglas-fir, and oaks).
Deciduous Parkland (DPK)	Transition between boreal forest and grasslands; large tracts of grasses with localised forest stands (aspen, poplar).
Interior Conifer Forest (DCO)	Almost exclusively conifer forest with semi-open or open canopies; includes extensive steppe, grass or shrublands (Douglas-fir, ponderosa pine).
Sub-Boreal Mixedwood (SBM)	Conifer-dominated mixedwood (white spruce, Douglas-fir, subalpine fir, lodgepole pine, aspen, poplars and birches).
Coastal Dry Mixedwood (CDM)	Mixed forest cover, largely non-boreal species; includes chaparral communities (Douglas-fir, redcedar, oak, grand fir, arbutus, red alder, maple).
Wet Temperate Forest (WTF)	Wet, diverse, and largely conifer-dominated; confined to coastal and areas of heavy orographic precipitation (hemlock, redcedar, Douglas-fir, grand fir).
Sub-Alpine Forest (SAF)	Transition zone between denser, lower-elevation forests and treeless alpine (Engelmann spruce, subalpine fir, mountain hemlock, larch, lodgepole pine, Douglas-fir).
Boreal Forest (BOR)	Cover of conifer and mixedwood forest; coldest and driest forest ecosystem in North America (white & black spruce, lodgepole pine, aspen, tamarack, poplar, birch).
Boreal Sub-Arctic (BSA)	Transition between boreal and arctic; stunted and widely spaced boreal trees; includes grass and shrublands (stunted white spuce, black spruce, birch and aspen).
Non-forested cold	
Alpine Tundra (ALT)	Treeless alpine meadows, barren land.
Arctic Tundra (ARC)	High-latitude tundra, largely devoid of trees and dominated by shrubs and lichens.
Glaciers / Ice (ICE)	Climate conditions favourable for year-round ice coverage.

Bioclimate envelope modelling

Ecosystem projections were carried out with a classification tree analysis, which can use a class variable as the dependent variable. This approach has been shown to be effective even for the prediction of species distributions, which can subsequently be inferred from known species frequencies for projected ecosystem classes (Hamann & Wang, 2006; Mbogga et al., 2010). For the dependent variable we used 770 mapped ecosystem classes covering western North America to 100°W longitude. The ecosystem delineations were compiled using six sources: the 'Ecosystems of Alaska' (Joint Federal-State Land Use Planning Commission for Alaska, 1991), the 'Biogeoclimatic Ecosystem Classification System' of British Columbia (Pojar & Meidinger, 1991), 'Natural Regions and Subregions' of Alberta (Govt. of Alberta, 2005), the 'National Ecological Framework' for the remaining western Canadian provinces (Govt. of Canada, 1999), 'Potential Natural Vegetation Maps' for California and Arizona (Kuchler, 1993, 1996), and 'Ecoregions of the Continental United States' for the remaining western states (Omernik, 2003).

Although we selected the highest resolution datasets available, we had to refine some delineations in mountainous areas so that certain ecosystem classes were characterised by a narrower climate envelope. The alpine ecosystem delineations for British Columbia and Alaska were subdivided by major mountain ranges and classified as 'Alpine Tundra', 'Barren/Rock', and 'Glacier/Ice' within each mountain range using 30 m resolution remotely sensed landcover data for the US (Homer et al., 2007) and Canada (Wulder et al., 2008). In addition, lower-montane ecosystem classes in the Yukon Territory, Northwest Territories, and Washington were removed from the dataset because their delineations were too coarse to be useful. Nearby finer-scale delineations in Alaska and British Columbia with similar climatology were available to accurately describe these climate envelopes.

Predictions were made for a 1km resolution digital elevation model of North America that we generated in Lambert Conformal Conic projection from 90 m resolution data of the Shuttle Topographic Mission (Farr et al., 2007). North of 60°N latitude, where these data were not available, we used re-projected etopo30 elevation data (Verdin & Greenlee, 1996). To build classification trees, we randomly sampled 100 grid cells within each of the 770 ecosystem delineations (i.e. 77,000 grid cells from a total of approximately 10 million grid cells of the digital elevation model). These sample points were climatically characterised and used as 'training data' for classification tree analysis implemented with the RandomForest software package v.4.5 (Breiman, 2001) for the open-source R programming environment (R Development Core Team, 2009). RandomForest has been shown to be a robust ensemble classifier and a useful technique for bioclimate envelope modelling (e.g. Lawler et al., 2006).

Past and future climate data

For past and future climatic characterisation we used General Circulation Model (GCM) projections overlaid as anomalies

(deviation from the 1961–90 reference climate) on high resolution interpolated climate normal data. The first set of back-predictions is based on the coupled oceanic-atmospheric GCM developed by the Geophysical Fluid Dynamics Laboratory (Anderson *et al.*, 2004) at Princeton University (Bush & Philander, 1999) for 6000, 9000, 16,000, and 21,000 Years Before Present (YBP). The second GCM is the Community Climate Model version 1 (CCM1) developed by the National Center for Atmospheric Research (NCAR) (Kutzbach *et al.*, 1998) for 6000, 11,000, 14,000, 16,000, and 21,000 YBP. For future climate projections we used individual and ensemble projections for four main emission scenario families A1FI, B1, A2, and B2 implemented by the following GCMs: CGCM2, HADCM3, ECHAM4, and CSIRO2 (Mitchell *et al.*, 2004).

All spatial climate data processing was carried out with a custom software package that we make freely available (Wang *et al.*, 2006; Mbogga *et al.*, 2009)[1], which uses 1961–90 climate normal grids for Canada and the United States generated by Daly *et al.* (2008) as present day climate representation. In addition, this software package estimates biologically-relevant climate variables according to Wang *et al.* (2006). Of all the available climate variables, ten of the least correlated variables were identified with a principal component analysis and selected as predictors: mean annual temperature, mean annual precipitation, the mean temperature of the warmest month, 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 two dryness indices according to Hogg (1997): an annual climate moisture index and a summer climate moisture index.

Analysis

Novel combinations of these climate variables based on past or future GCM projections were determined with the multivariate Mahalanobis distance measure (Mahalanobis, 1936). This distance measure is a normalised Euclidean distance that weighs individual variables according to their collinearity with all other variables. Variables that are perfectly correlated are weighted as a single variable in distance calculations, while the Mahalanobis distance for completely independent variables would equal the Euclidean distance. The Mahalanobis distance to the closest modern equivalent was determined with a distance matrix between all past and current climate grid cells. Since it is not feasible to calculate a distance matrix that large (approximately 10^{14} values), we calculated a reduced distance matrix. We retained all projected grid cells for past and future projections, but we summarised current climate conditions as 770 ecosystem climate averages (resulting in a distance matrix with just 7.7×10^9 values). The smallest Mahalanobis distance in each row of this matrix therefore reflects the distance to the nearest modern ecosystem climate average, which we displayed on maps to iden-

tify no-analogue climate conditions in the future and past. All distance calculations were performed with PROC DISTANCE and PROC PRINCOMP in the SAS statistical software package (Example 30,662 in SAS Institute, 2007).

For model evaluation, we used palaeoecological data comprised of fossil pollen and plant macrofossils (compiled by Thompson & Anderson, 2000; Dyke, 2005). Duplicates as well as mammal records were removed (to retain purely vegetation-based data) for a total of 1460 sites used in this analysis. Modern classifications from the last 1000 years were available for most sites. Approximately 500 sites had records for 6000 YBP, which declined to 300 sites for 9000 YBP. For time periods approaching the last glacial maximum, records become fairly scarce with 150 sites for 16,000 YBP and 100 sites for 21,000 YBP for western North America. The palaeoecological records were already classified into biomes and we adopted the same classification system for predicted ecosystems. Minor differences arise because we model ice and barren landcover (for which there are no pollen and fossil records). We also lacked climate data for the very northern herb-tundra biome of Dyke (2005). Lastly, we separated Dyke's 'Interior Forest' into Dry Coastal Mixedwood, Sub-Boreal Mixedwood, and Dry Interior Conifer Forest because we perceived those as climatically and ecologically distinct ecosystems.

Because we evaluate the accuracy of a multi-level classifier, we use error-of-confusion matrices and report the numbers and ratios of correct and incorrect classifications at the biome level as well as for forested and non-forested classes. To maintain the highest possible data accuracy, modelled biome classifications were made based on climate values for the location and elevation of the palaeoecological records, estimated by our software package described above, rather than using a classification made for a nearby 1 km grid cell.

RESULTS

Independent model evaluation

The model outputs for the present, based on the modern day 1961–90 reference climate, visually conform to both the mapped ecosystem distributions that were used to train the model and also to the approximate delineations by Dyke (2005) (Fig. 1). Misclassification error rates of predicted biomes against independent fossil and pollen data are shown in Table 2. The percentage of correct classifications tends to be quite low with rates per biome ranging from 0 to 67%. Some of the low match rates can be attributed to small sample sizes, but nevertheless the overall percentage of correct classifications with independent data is just 46%. Misclassifications often occur among adjacent biomes and often in spatially complex landscapes (e.g. high-resolution inset in Fig. 1). In addition, we find misclassification rates for pollen and fossil data representing the boreal forest ecosystem as the adjacent boreal subarctic, which has essentially the same species composition (Table 2).

For conciseness, we do not report full misclassification matrices for 14 biomes for ecosystem predictions based on past
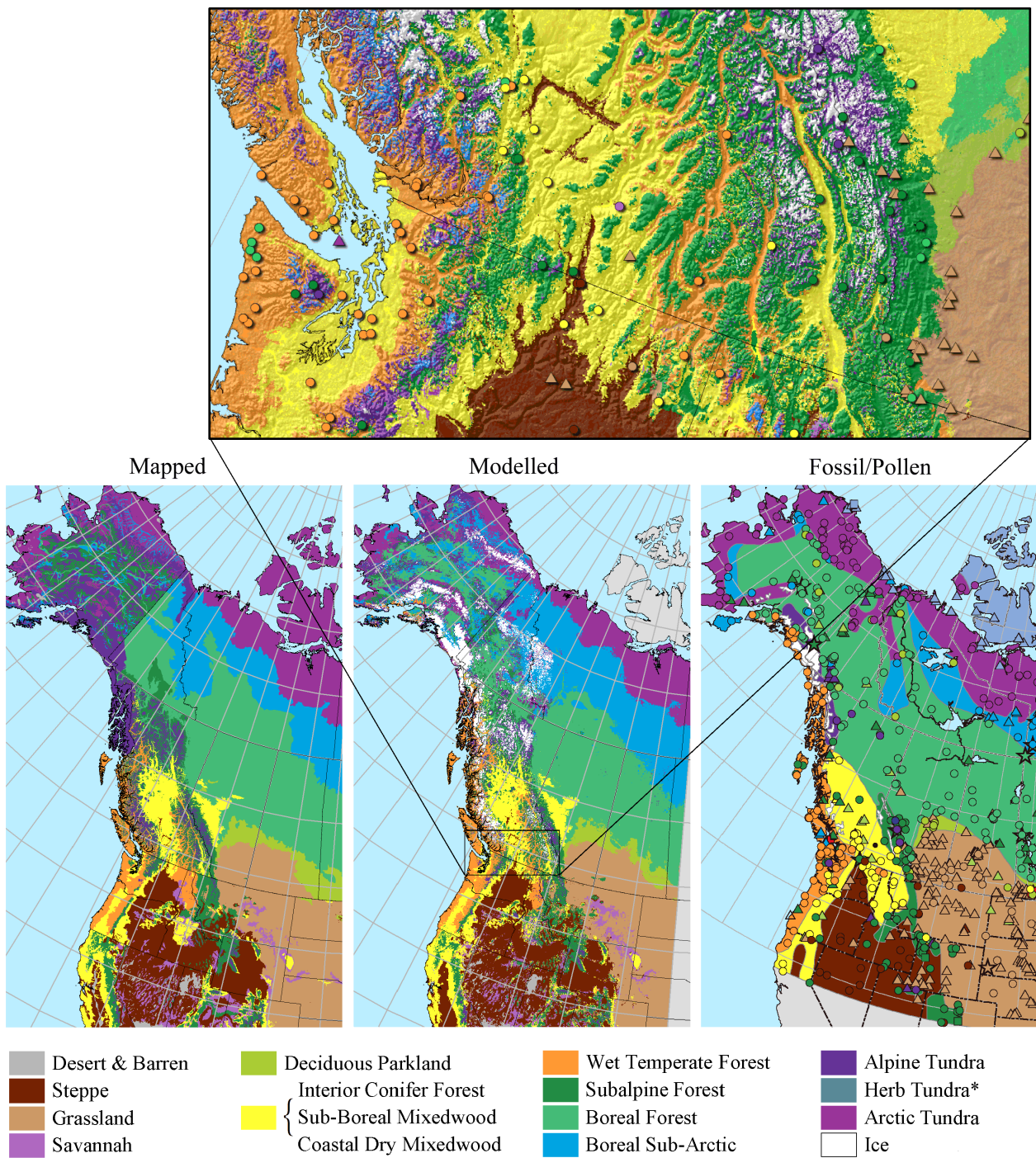
Figure 1 Present day mapped biomes, modelled biomes, and biome classes inferred from pollen (circles), macrofossil (stars) and mammal (triangles) records according to Dyke (2005), reproduced with permission. We did not model Dyke's herb tundra biome but we distinguish three types of interior forest (yellow). The inset map provides a detailed comparison between the modelled ecosystems and the fossil/pollen sites for a mountainous area of southern British Columbia. Note that mammal points (triangles) were not used in the model evaluation calculations.

climates. Rather, we provide a higher level summary using the afore-mentioned three categories of ecosystems: *Forested*, moisture-restricted *Dry*, and temperature-restricted *Cold*. Rates of misclassification are generally very similar for projections

based on CCM1 and GDFL climate reconstructions (Table 3). Misclassifications increase abruptly for pollen and fossil sites that are classified as *Forested* and *Dry* during the cooler early-Holocene between 9000 to 14,000 YBP (Table 3, Fig. 2). In

Table 2 Misclassification rates between biomes predicted with climate envelope models for the 1961–90 climate normal period and biome reconstructions from fossil and pollen samples for the last millennium. Correct classifications are highlighted in bold, and also reported as percentage of fossil and pollen points correctly classified. Cohen's Kappa statistic representing correct classifications minus randomly expected matches is also reported.

Observed	Predicted Dry DES	STE	GRA	Forested SAV	DPK	DCO	SBM	CDM	WTF	SAF	BOR	BSA	Cold ALT	ARC	n	Match rate	Cohen's Kappa
Non-Forested Dry																	
Desert (DES)	**4**	7	0	0	0	0	0	0	0	0	0	0	0	0	11	36%	0.29
Steppe (STE)	1	**13**	3	2	0	2	0	0	0	2	0	0	0	0	23	57%	0.49
Grassland (GRA)	0	7	**54**	3	7	3	0	0	0	6	1	0	0	0	81	67%	0.59
Forested																	
Savannah (SAV)	0	2	0	**0**	0	0	0	0	0	1	0	0	0	0	3	0%	0.00
Deciduous Parkland (DPK)	0	0	1	0	**3**	0	0	0	0	1	4	4	1	2	16	19%	0.11
Interior Forest (DCO, SBM, CDM)	2	7	0	0	0	**8**	2	**0**	3	6	1	1	2	2	34	29%	0.22
Wet Temperate Forest (WTF)	0	1	0	0	0	7	2	12	**68**	7	0	2	12	2	113	60%	0.52
Sub-Alpine Forest (SAF)	0	2	2	2	0	2	3	3	4	**32**	0	0	16	0	66	48%	0.41
Boreal Forest (BOR)	0	1	12	0	6	7	10	0	0	31	**136**	105	69	67	444	31%	0.23
Boreal Sub-Arctic (BSA)	0	0	0	0	0	0	0	0	0	1	2	**5**	2	2	12	42%	0.34
Non-Forested Cold																	
Alpine Tundra (ALT)	0	1	0	0	0	0	0	0	0	4	0	2	**9**	2	18	50%	0.42
Arctic Tundra (ARC)	0	0	0	0	0	0	0	1	0	7	21	83	59	**234**	405	58%	0.50

contrast, near the last glacial maximum (16,000 to 21,000 YBP), pollen and fossil points representing *Cold* ecosystems are almost always predicted correctly, indicating an increasing bias towards under-predicting *Forested* and *Dry* ecosystem (Fig. 2). This is reflected in Table 3, as most errors are located on the upper right side of the diagonal for these time periods. Interestingly, overall model accuracy is higher for the mid-Holocene warm period at 6000 YBP than for the present day. This is driven by increased accuracy in the prediction of pollen and fossil sites that represent *Forested* and *Dry* ecosystem classes (Fig. 2), which holds true for both CCM1 and GFDL based predictions.

No-analogue climates in the past

Climatic reconstructions based on the general circulation models CCM1 and GFDL reveal novel combinations of climate variables in western North America for all time periods (Table 4, Fig. 3). For example, high climate dissimilarities emerge in the United States Rocky Mountains during the mid-Holocene warm period at 6000 YBP. These climates are characterised by drier, cooler summers and warmer winters (data not shown), conditions that have no modern equivalent. Fossil and pollen records for these areas indicate forested ecosystems, but they were classified as steppe or grassland by the bioclimate envelope model. Climates without modern equivalents also appear in the area immediately south of the ice sheets at 21,000 YBP (Fig. 3). These areas were characterised in the data by notably colder annual and mean warmest month temperatures, resulting in a shortened frost-free period while farther south in eastern Oregon and northern Nevada, no-analogue climates were driven less by

overall cooling than by differences in seasonal temperature variables (data not shown). The climatology in both areas was classified by the bioclimate envelope model as supporting alpine or arctic tundra, although there is no modern climatic equivalent.

Summary statistics for the study area, broken down by *Forested*, *Dry*, and *Cold* biomes show associations between novel climates and erroneous classifications as well (Table 4). For forested biomes, misclassification rates increase as average climate dissimilarity increases towards the last glacial maximum (e.g. 0.44 toward 1.3 distance units versus 74% toward 13% correct classifications for CCM1) (Table 4). In contrast, biomes that are classified as too cold to support forested ecosystems have lower misclassification rates (e.g. 0.17 toward 4.84 distance units versus 55% toward 100% correct classifications for CCM1). However, this latter association simply reflects, with increasing confidence, that extremely cold (and therefore novel) environments are correctly classified as too cold to support forested ecosystems.

For subsequent interpretation of the causes of biome misclassifications, it is also important to point out that the bioclimate envelope model predicts the southern extent of the continental ice sheets with remarkable accuracy, even though the northern portion of the ice sheet is not correctly represented (Fig. 3). This also holds true for predictions based on the GFDL general circulation model (data not shown).

Future projections

Climate projections for future periods result in dissimilarities roughly on par with those observed for the 6000 to 11,000 YBP

Table 3 Misclassifications between biome groups inferred from fossil and pollen samples and biome groups independently predicted with bioclimate envelope modelling for the same periods, based on the general circulation models CCM1 and GFDL. Correct classifications are highlighted in bold and the total number of pollen and fossil samples available for each time period is given in parentheses (*n*).

	Predicted					
	Present Day (*n* = 1226)					
Observed	Dry	Forested	Cold			
Dry	**89**	26	0			
Forested	30	**481**	177			
Cold	1	118	**304**			
	CCM1 Model			GFDL Model		
	6,000 YBP (*n* = 554)			6,000 YBP (*n* = 554)		
	Dry	Forested	Cold	Dry	Forested	Cold
Dry	**58**	7	0	**55**	10	0
Forested	53	**285**	46	21	**323**	40
Cold	1	46	**58**	1	38	**66**
	11,000 YBP (*n* = 275)			9,000 YBP (*n* = 376)		
	Dry	Forested	Cold	Dry	Forested	Cold
Dry	**27**	6	0	**19**	16	6
Forested	26	**103**	8	16	**147**	85
Cold	2	67	**36**	1	24	**62**
	14,000 YBP (*n* = 179)					
	Dry	Forested	Cold			
Dry	**9**	6	3			
Forested	7	**33**	24			
Cold	5	32	**60**			
	16,000 YBP (*n* = 129)			16,000 YBP (*n* = 129)		
	Dry	Forested	Cold	Dry	Forested	Cold
Dry	**5**	4	1	**0**	0	10
Forested	5	**8**	21	2	**3**	29
Cold	3	8	**74**	1	1	**83**
	21,000 YBP (*n* = 89)			21,000 YBP (*n* = 89)		
	Dry	Forested	Cold	Dry	Forested	Cold
Dry	**1**	2	13	**0**	5	11
Forested	1	**3**	20	0	**9**	15
Cold	0	0	**49**	0	4	**45**

back-predictions (Table 4). Areas of high dissimilarity are primarily restricted to the coast mountains of the Pacific Northwest, where combinations of very high precipitation and high summer temperatures emerge that have no modern equivalent (Fig. 4). The most pessimistic 'business as usual' CO_2 emission scenario (A1FI) also results in a prediction of hot and dry climatic conditions in the southern United States that have no equivalent in the present day study area (maps not shown, but reflected in high average dissimilarities for the *Dry* biome type in Table 4). The most optimistic emission scenarios assume less resource intensive service economies (B1) and environmentally sustainable economic and population growth (B2). These yield climate dissimilarities roughly equivalent to values of the mid-Holocene warm period, which had the highest accuracy of all time periods in the independent model evaluation above (Table 4). The intermediate scenario that assumes slow population growth and regionally fragmented economic growth (A2) has larger climate dissimilarity values equivalent to 6000 to 9000 YBP, which still do not imply very high misclassification rates due to no-analogue climates (Fig. 2).

According to this intermediate emission scenario (A2), biome climate envelopes for the 2080s change most notably in the higher latitudes, where the warming signal is strongest (IPCC, 2007) (Figs 1 and 4). Alaska gains landscape level diversity of habitat conditions, comparable to British Columbia at present (Fig. 4, inset). Changes in British Columbia are driven by increased precipitation leading to climate envelopes that support wet temperate forest types. The Canadian Plains of Alberta and Saskatchewan lose substantial area with climate conditions suitable for boreal forests. Areas of minimal change at biome-level climate conditions are projected for the southern latitudes, with some expansion of desert and steppe climate envelopes. It should be noted that the projection in Fig. 4 is based on an ensemble of multiple individual GCM implementations of the A2 scenario. Notable differences in projections arise from model runs of individual GCMs (Hamann & Wang, 2006; Mbogga *et al.*, 2010).

DISCUSSION

Model accuracy and no-analogue climates

For back-predictions towards the last glacial maximum our results confirm that no-analogue climates are indeed prevalent. We further demonstrated that no-analogue climates compromise accuracy of biome classifications based on palaeoclimatic predictions, which has been previously discussed as a potential limitation of bioclimate envelope models (Jackson & Williams, 2004; Williams & Jackson, 2007; Fitzpatrick & Hargrove, 2009; Van der Wal *et al.*, 2009). At the same time, we provide a perspective for the magnitude of novel climates expected under projected anthropogenic climate change (Table 4). The degree of climate dissimilarity expected for the coming century would not imply significant effects on misclassification rates, except perhaps for isolated areas in the Pacific Northwest Cordillera, where high precipitation and temperature
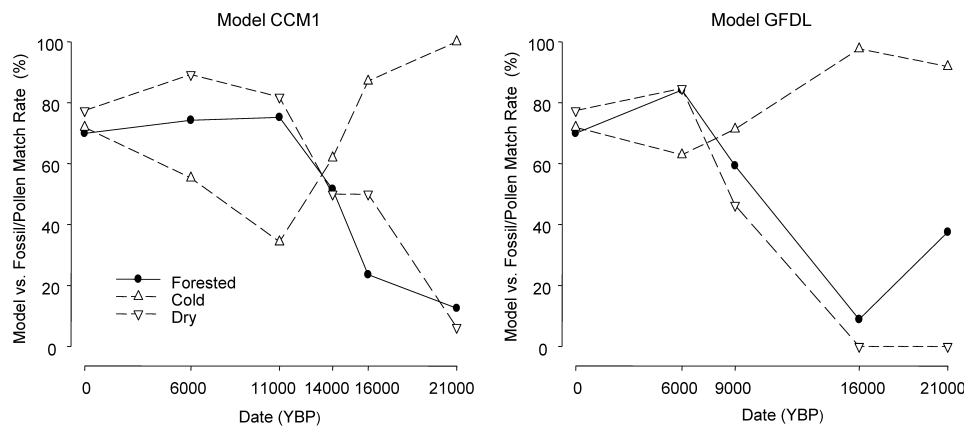
Figure 2 Match rates marked as bold in Table 4 expressed as percentage and plotted over time for paleoclimate projections of two general circulation models CCM1 (left) and GFDL (right).

Table 4 Mean climate dissimilarity values for forested (*Forested*), non-forested cold (*Cold*), and non-forested dry (*Dry*) biome groups. Climate dissimilarity is quantified as the Mahalanobis distance to the nearest present-day equivalent found in the study area. The corresponding percentages of correct model classifications of fossil and pollen data are shown in parentheses.

	Biome Type					
Period	Dry		Forested		Cold	
Present Day	0	(77%)	0	(70%)	0	(72%)
GFDL Model						
6,000 YBP	0.23	(85%)	0.20	(84%)	0.19	(63%)
9,000 YBP	0.27	(46%)	0.32	(59%)	0.45	(71%)
16,000 YBP	0.26	(0%)	1.27	(9%)	2.79	(98%)
21,000 YBP	0.30	(0%)	1.35	(38%)	2.91	(92%)
CCM1 Model						
6,000 YBP	0.62	(89%)	0.44	(74%)	0.17	(55%)
11,000 YBP	1.15	(82%)	0.46	(75%)	0.44	(34%)
14,000 YBP	0.66	(50%)	0.55	(52%)	0.76	(62%)
16,000 YBP	0.73	(50%)	1.01	(24%)	2.42	(87%)
21,000 YBP	0.46	(6%)	1.30	(13%)	4.84	(100%)
Future Projections						
2080-A1FI	0.63		0.53		0.42	
2080-A2	0.42		0.37		0.30	
2080-B1	0.28		0.26		0.23	
2080-B2	0.31		0.26		0.23	

anomalies with no modern analogue emerge (Fig. 4). Even though we use a different spatial resolution, a different set of climate variables, and a different similarity metric, our results generally coincide with those of Williams *et al.* (2007) who also found a low risk of novel climates at high latitudes of North America.

The common notion that we are headed towards unknown climatic futures caused by greenhouse gas emissions may be true at a local scale, but at the sub-continental scale of this study, truly novel combinations of climate conditions in this region are

the exception, as this and other studies have shown. Sub-continental scales are typically used for the development of species distribution models; therefore, we conclude that their projections should not be generally compromised by extrapolating into no-analogue climate space. Conversely it is clear that regional-scale bioclimate envelope projections are less useful. For example, if we had developed a model just for Alaska, we would find high rates of no-analogue climates for 'unknown' biomes that are currently only found in British Columbia.

Our results are broadly applicable, not only for the classification tree approach that we use to project ecosystems, but to any species distribution model. The measure of climate dissimilarity is independent of any particular model technique. It is the correlational nature of the niche modelling approach in general, rather than any specific mathematical or statistical procedure, that is susceptible to confounding by no-analogue climates.

Violation of bioclimate envelope model assumptions

In addition to misclassifications, we also showed bias in bioclimate envelope model results toward the last glacial maximum. We find that at the height of the last ice age and in early deglaciation, forested biomes are under-predicted by the model. We reject the possible alternate explanation that we have bias due to migrational lag (i.e. that a lack of ecosystem-climate equilibrium at this time promotes model misclassification). If this were the case, this discrepancy would be manifested as forested ecosystem *over*-prediction. Secondly, inaccurate paleoclimatic reconstructions (too cold) could be responsible for the bias. However, both GFDL and CCM1 predict the southern extent of the continental ice sheets with remarkable accuracy. It would therefore appear to be an unlikely explanation for the under-prediction of forested ecosystems. A third factor that might account for differences between observed and predicted ecosystem distribution is the effect of CO_2, due to lower concentrations of around 200 parts per million during the last glacial maximum. However, not accounting for CO_2 in our model should lead to an over-prediction of forests in the past (Cowling, 1999), which is also contrary to the under-prediction reported here.
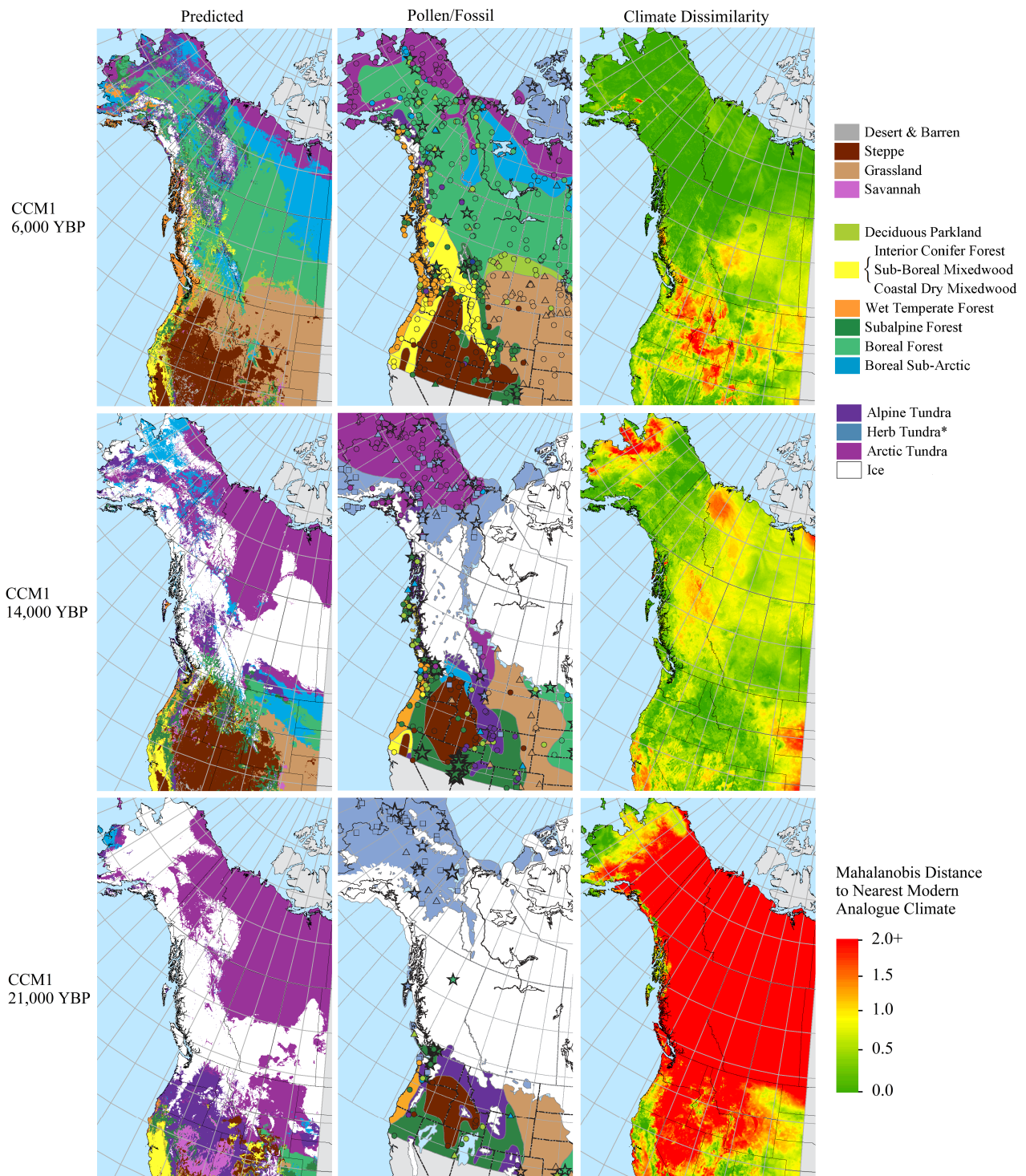
Figure 3 Predicted biome classes, biome reconstructions from pollen and fossil data, and climate dissimilarities measured as multivariate Mahalanobis distance to the nearest modern climate space. Green indicates climate arrangements analogous to those witnessed in the present day and red indicates increasing diversion from any modern climate conditions in the study. Summary statistics for additional model runs are given in Tables 3 and 4.
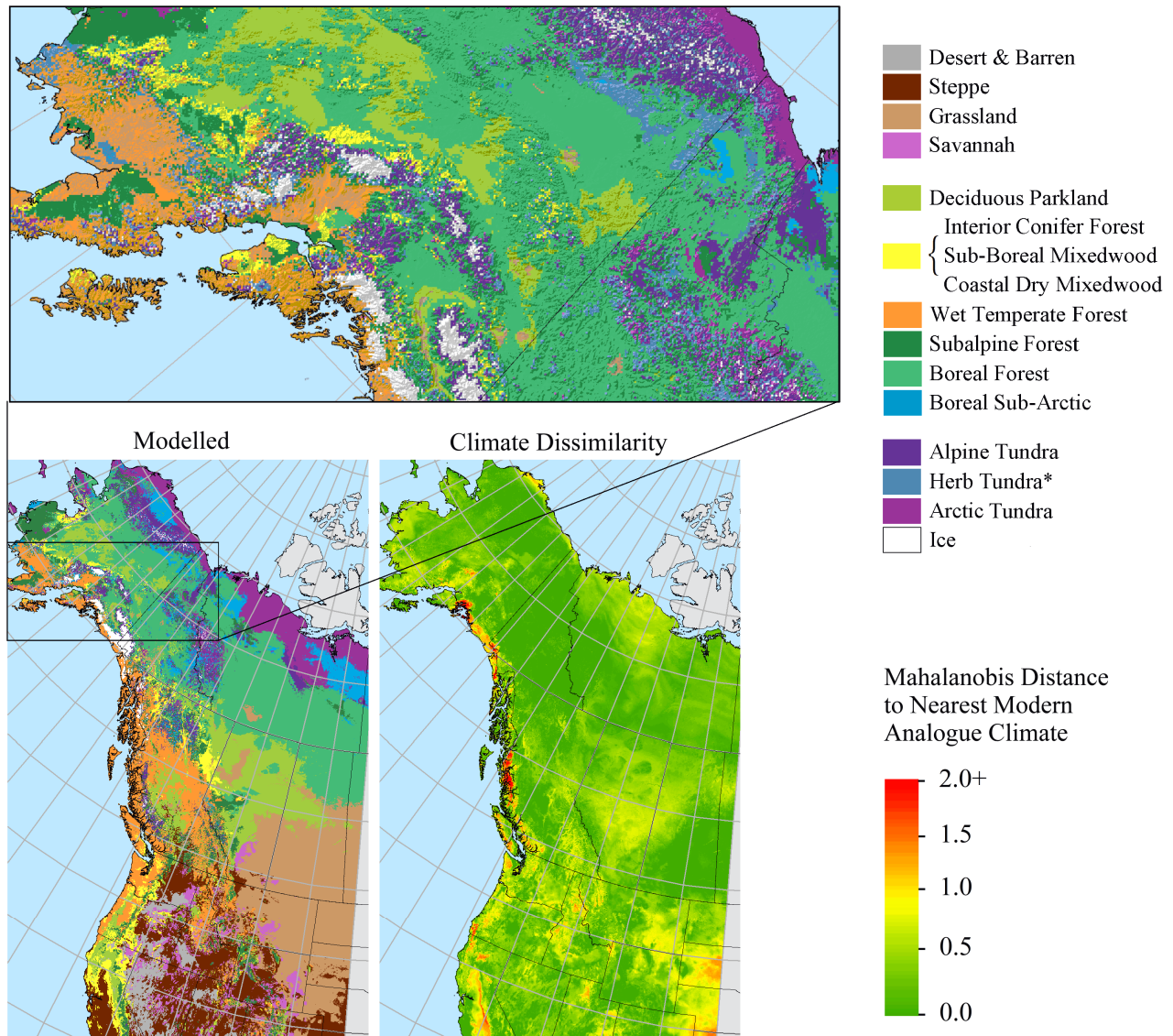
Figure 4 Predicted biome classes and climate dissimilarity according to an ensemble projection of the A2 emissions scenario from five general circulation models for the 2080s. The inset map provides a more detailed image of the modelled ecosystems in Alaska and the Yukon Territory.

Misclassifications due to no-analogue climates should not introduce bias as there is an equal probability of misclassification into all classes (in our case individual biomes or *Forested*, *Dry*, or *Cold* groups). However, if new niche space emerges on the landscape, species may genetically adapt and occupy newly available environmental space, which bioclimate envelope models cannot anticipate. Davis & Shaw (2001) have shown that the ecological niche space of tree species may not be constant over time. Adaptive traits with high genetic variability and heritability, which are common in tree species, may allow for occupation of new realised niche space (Hamrick, 2004), providing a potential explanation for the under-predictions of forested biomes observed in this study.

The relevance of evolutionary changes to the niche space of species is powerfully illustrated by palaeoecological studies that look beyond the Holocene. For example, fossil forest dating to the Eocene consisting of *Pseudotsuga*, *Larix*, *Sequoia*, and *Chamaecyparis* were found in the Canadian high arctic (Basinger, 1991). This fossil evidence includes giant stems that suggest temperate forest communities of similar appearance and composition to today's Pacific Northwest coastal forests. Trees must have adapted not only to a different climate but to the vastly different diurnal cycle of the arctic latitudes with 24 h of daylight during the summer and complete darkness in winter, as there were only minor continental shifts relative to the North Pole for this area at this time.

While niche constancy and no-analogue climates must have played an important role at evolutionary time scales, we do not think that these factors should effect bioclimate envelope model projections for the immediate future and we consider model

projections useful, if correctly interpreted. Projected ecosystems simply represent new equilibrium targets for ecological communities. Because of the long generation time of trees, forest communities that are resilient or resistant may not change at all over periods that are measured in decades. Nevertheless, discrepancies between current ecosystems and projected future habitat are of great concern. For example, we do not interpret Alaska's emerging landscape diversity as a cause for optimism. Rather, it is a cause for concern, as climatic stresses on locally adapted populations may compromise forest productivity and forest health (Allen *et al.*, 2010).

REFERENCES

Allen, C. D., Macalady, A. K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., Kitzberger, T., Rigling, A., Breshears, D. D., Hogg, E. H., Gonzalez, P., Fensham, R., Zhang, Z., Castro, J., Demidova, N., Lim, J. H., Allard, G., Running, S. W., Semerci, A. & Cobb, N. (2010) A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management*, **259**, 660–684.

Anderson, J. L., Balaji, V., Broccoli, A. J. *et al.* (2004) The new GFDL global atmosphere and land model AM2-LM2: evaluation with prescribed SST simulations. *Journal of Climate*, **17**, 4641–4673.

Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.

Araújo, M.B., Pearson, R.G., Thuiller, W. & Erhard, M. (2005) Validation of species-climate impact models under climate change. *Global Change Biology*, **11**, 1504–1513.

Basinger, J.F. (1991) The fossil forests of the Buchanan lake formation (early Tertiary), Axel Heiberg Island, Canadian high arctic: preliminary floristics and paleocliamte. *The fossil forests of tertiary age in the Candian Arctic Archipelago* (ed. by R.L. Christie and N.J. McMillan), pp. 39–65. Geological Survey of Canada Bulletin, No. 403. Canadian Geologic Survey in Ottawa, ON, Canada.

Botkin, D.B., Saxe, H., Araújo, M.B., Betts, R., Bradshaw, R.H.W., Cedhagen, T., Chesson, P., Dawson, T.P., Etterson, J.R., Faith, D.P., Ferrier, S., Guisan, A., Hansen, A.S., Hilbert, D.W., Loehle, C., Margules, C., New, M., Sobel, M.J. & Stockwell, D.R.B. (2007a) Forecasting the effects of global warming on biodiversity. *Bioscience*, **57**, 227–236.

Botkin, D.B., Saxe, H., Araújo, M.B., Betts, R., Bradshaw, R.H.W., Cedhagen, T., Chesson, P., Dawson, T.P., Etterson, J.R., Faith, D.P., Ferrier, S., Guisan, A., Hansen, A.S., Hilbert, D.W., Loehle, C., Margules, C., New, M., Sobel, M.J. & Stockwell, D.R.B. (2007b) Forecasting the effects of global warming on biodiversity. *Bioscience*, **57**, 227–236.

Box, G.E.P. & Draper, N.R. (1987) *Empirical model-building and response surfaces*. John Wiley & Sons, New York.

Breiman, L. (2001) Random forests. *Machine Learning*, **45**, 5–32.

Bush, A.B.G. & Philander, S.G.H. (1999) The climate of the Last Glacial Maximum: results from a coupled atmosphere-ocean general circulation model. *Journal of Geophysical Research-Atmospheres*, **104**, 24509–24525.

Carstens, B.C. & Richards, C.L. (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.

Chen, P., Welsh, C. & Hamann, A. (2010) Geographic variation in growth response of Douglas-fir to inter-annual climate variability and projected climate change. *Global Change Biology*, **16**, 3374–3385.

Cowling, S.A. (1999) Simulated effects of low atmospheric CO_2 on structure and composition of North American vegetation at the Last Glacial Maximum. *Global Ecology and Biogeography*, **8**, 81–93.

Crowley, T.J. (1990) Are there any satisfactory geologic analogs for a future greenhouse warming? *Journal of Climate*, **3**, 1282–1292.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J. & Pasteris, P.P. (2008) Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, **28**, 2031–2064.

Davis, M.B. & Shaw, R.G. (2001) Range shifts and adaptive responses to Quaternary climate change. *Science*, **292**, 673–679.

Dyke, A.S. (2005) Late Quaternary vegetation history of northern North America based on pollen, macrofossil, and faunal remains. *Geographie Physique et Quaternaire*, **59**, 211–262.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D. & Alsdorf, D. (2007) The shuttle radar topography mission. *Reviews of Geophysics*, **45**, RG2004.

Fitzpatrick, M.C. & Hargrove, W.W. (2009) The projection of species distribution models and the problem of non-analog climate. *Biodiversity and Conservation*, **18**, 2255–2261.

Govt. of Alberta (2005) Seed Zones of Alberta, Digital Vector Data in ARC/INFO format. Government of Alberta, Sustainable Resource Development (SRD).

Govt. of Canada (1999) *A national ecological framework for Canada* (ed. by I.B. Marshall and P.H. Schurt). Government of Canada, Agriculture and Sgri-Food Canada. Available at: http://sis.agr.gc.ca/cansis/nsdb/ecostrat/intro.html.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

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

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

Hamrick, J.L. (2004) Response of forest trees to global environmental changes. *Forest Ecology and Management*, **197**, 323–335.

Hogg, E.H. (1997) Temporal scaling of moisture and the forest-grassland boundary in western Canada. *Agricultural and Forest Meteorology*, **84**, 115–122.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., Mckerrow, A., Vandriel, J.N. & Wickham, J. (2007) Completion of the 2001 National Land Cover Database for the conterminous United States. *Photogrammetric Engineering and Remote Sensing*, **73**, 337–341.

IPCC (2007) *Synthesis report. Climate change 2007: contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change*. IPCC, Geneva.

Jackson, S.T. & Williams, J.W. (2004) Modern analogs in Quaternary paleoecology: here today, gone yesterday, gone tomorrow? *Annual Review of Earth and Planetary Sciences*, **32**, 495–537.

Joint Federal-State Land Use Planning Commission for Alaska (1991) Major Ecosystems of Alaska (1973). Digital Vector Data (digitized from 1:2,500,000 scale map) in ARC/INFO format. USGS, Available at: http://agdc.usgs.gov/data/usgs/erosafo/ecosys/metadata/ecosys.html (accessed 27 November 2008).

Kuchler, A.W. (1993) Potential Natural Vegetation of the Conterminous United States (1964). Digital Vector Data (digitized from 1:3,186,000 scale map) on an Albers Equal Area Conic polygon network in ARC/INFO format. EPA Environmental Research Laboratory.

Kuchler, A.W. (1996) Kuchler Vegetation Potential Map (1976) for California. Based on 1:1,000,000 scale map, with accompanying booklet, of the potential natural vegetation of California. U.S. Bureau of Reclamation, Mid-Pacific Region, MPGIS Service Center. Available at: http://www.ngdc.noaa.gov/ecosys/ged.shtml (accessed 15 October 2008).

Kutzbach, J., Gallimore, R., Harrison, S., Behling, P., Selin, R. & Laarif, F. (1998) Climate and biome simulations for the past 21,000 years. *Quaternary Science Reviews*, **17**, 473–506.

Lawler, J.J., White, D., Neilson, R.P. & Blaustein, A.R. (2006) Predicting climate-induced range shifts: model differences and model reliability. *Global Change Biology*, **12**, 1568–1584.

Loehle, C. & Leblanc, D. (1996) Model-based assessments of climate change effects on forests: a critical review. *Ecological Modelling*, **90**, 1–31.

Mahalanobis, P.C. (1936) On the generalised distance in statistics. *Proceedings of the National Institute of Science of India*, **12**, 49–55.

Mbogga, M.S., Hamann, A. & Wang, T.L. (2009) Historical and projected climate data for natural resource management in western Canada. *Agricultural and Forest Meteorology*, **149**, 881–890.

Mbogga, M.S., Wang, X. & Hamann, A. (2010) Bioclimate envelope modeling for natural resource management: dealing with uncertainty. *Journal of Applied Ecology*, **47**, 731–740.

Mitchell, T.D., Carter, T.R., Jones, P., Hulme, M. & New, M. (2004) A comprehensive set of high-resolution grids of monthly climate for Europe and the globe: the observed record (1901–2000) and 16 scenarios (2001–2100). *Tyndall Centre Working Paper No. 55*. Tyndall Centre for Climate Change Research, University of East Anglia, Norwich, UK.

Morin, X. & Thuiller, W. (2009) Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology*, **90**, 1301–1313.

Morin, X., Viner, D. & Chuine, I. (2008) Tree species range shifts at a continental scale: new predictive insights from a process-based model. *Journal of Ecology*, **96**, 784–794.

Nogués-Bravo, D. (2009) Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography*, **18**, 521–531.

Omernik, J.M. (2003) Level III and IV Ecoregions of the Continental United States, Digital Vector Data in ESRI/ARC format. United States Environmental Protection Agency (EPA), Western Ecology Division, Available at: http://www.epa.gov/wed/pages/ecoregions/level_iv.htm (accessed 12 August 2008).

Overpeck, J.T., Webb, R.S. & Webb, T. (1992) Mapping eastern North-American vegetation change of the past 18 Ka: no-analogs and the future. *Geology*, **20**, 1071–1074.

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

Pojar, J. & Meidinger, D.V. (1991) *Ecosystems of British Columbia*. B.C. Ministry of Forests, Victoria, BC, Canada.

Prentice, I.C., Bartlein, P.J. & Webb, T. (1991) Vegetation and climate change in Eastern North-America since the last glacial maximum. *Ecology*, **72**, 2038–2056.

R Development Core Team (2009) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. Available at: http://www.R-project.org.

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

Salzmann, U., Haywood, A.M. & Lunt, D.J. (2009) The past is a guide to the future? Comparing Middle Pliocene vegetation with predicted biome distributions for the twenty-first century. *Philosophical Transactions of the Royal Society A-Mathematical Physical and Engineering Sciences*, **367**, 189–204.

SAS Institute (2007) *SAS Knowledge Base Sample 30662: Mahalanobis distance: from each observation to the mean, from each observation to a specific observation, between all possible pairs*.

SAS Intstitute Inc., Cary, NC. Available at: http://support.sas.com/kb/30/662.html.

Segurado, P., Araújo, M.B. & Kunin, W.E. (2006) Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, **43**, 433–444.

Svenning, J.C., Normand, S. & Kageyama, M. (2008) Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *Journal of Ecology*, **96**, 1117–1127.

Thompson, R.S. & 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. *Journal of Biogeography*, **27**, 555–584.

Van der Wal, J., Shoo, L.P. & Williams, S.E. (2009) New approaches to understanding late Quaternary climate fluctuations and refugial dynamics in Australian wet tropical rain forests. *Journal of Biogeography*, **36**, 291–301.

Verdin, K.L. & Greenlee, S.K. (1996) Development of continental scale digital elevation models and extraction of hydrographic features. *Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling*. National Center for Geographic Information and Analysis, Santa Barbara, California, Santa Fe, New Mexico.

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.

Williams, J.W. & Jackson, S.T. (2007) Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, **5**, 475–482.

Williams, J.W., Shuman, B.N. & Webb, T. (2001) Dissimilarity analyses of late-Quaternary vegetation and climate in eastern North America. *Ecology*, **82**, 3346–3362.

Williams, J.W., Jackson, S.T. & Kutzbacht, J.E. (2007) Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences USA*, **104**, 5738–5742.

Wulder, M.A., White, J.C., Cranny, M., Hall, R.J., Luther, J.E., Beaudoin, A., Goodenough, D.G. & Dechka, J.A. (2008) Monitoring Canada's forests. Part 1: completion of the EOSD land cover project. *Canadian Journal of Remote Sensing*, **34**, 549–562.

Yesson, C. & Culham, A. (2006) Phyloclimatic modeling: combining phylogenetics and bioclimatic modeling. *Systematic Biology*, **55**, 785–802.

BIOSKETCHES

David R. Roberts is a PhD student at the University of Alberta. His research focuses on using statistical models to reconstruct Holocene migration routes and glacial refugia of forest trees in western North America with the goal of developing genetic population histories for North American tree species.

Andreas Hamann is an Assistant Professor at the University of Alberta in the field of ecological genetics. His research centres on how tree species and their populations are adapted to the environments in which they occur and how natural populations are affected by observed and projected climate change.

Editor: José Paruelo