We are now faced with the fact that tomorrow is today. We are confronted with the fierce urgency of now. In this unfolding conundrum of life and history there is such a thing as being too late… We may cry out desperately for time to pause in her passage, but time is deaf to every plea and rushes on. Over the bleached bones and jumbled residue of numerous civilizations are written the pathetic words: "Too late."

- Martin Luther King, Jr.

I wouldn't have seen it if I hadn't believed it.

- Marshall McLuhan

University of Alberta

Biogeographic histories and genetic diversity of western North American tree species: implications for climate change

by

David Roxburgh Roberts

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Conservation Biology

Department of Renewable Resources

Fall 2013
Edmonton, Alberta

For Bridget,
with whom I can see it clear, out to the end.

For Finlay,
who inspires me to just go and play.

And for Richard,
for your enthusiasm: here's to blue skies, no flat tires, and no blisters.

Abstract

Over the last two million years, the evolution of North American tree species, subspecies, and genetic varieties has taken place in a constantly changing landscape often dominated by extensive ice sheets and restricted temperate climate environments. Here, I approximately reconstruct post-glacial vegetation histories and glacial refugia of western North American trees using species distribution models to test biogeographic hypotheses regarding the existence of glacial refugia in Beringia, the evolution of subspecies in widespread conifers, the origin of Pacific Northwest inland rainforests, and levels of modern genetic diversity.

The first two chapters have a methodological emphasis, where I select and test a variety of species distribution models for their accuracy and robustness. Validations against 3,571 pollen and fossil records from 835 study sites indicated fair accuracy for most techniques (AUCs around 0.75). Ecosystem-based modelling approaches outperformed in specificity statistics and robustness against extrapolations far beyond training data, suggesting that they are well suited to reconstruct historical biogeographies. Evaluations at the biome level indicated that species distribution models could approximately reconstruct biomes for the mid- to late-Holocene but became unreliable in the Late Pleistocene due to the emergence of no-analogue climates. However, the limitation applied primarily to non-forested biomes.

Using a set of three robust species distribution modelling techniques, I investigate how modern genetic diversity and genetic structure was shaped by refugial history, using published estimates of allelic richness and expected heterozygosity for 473 populations of 22 tree species. Species with strong genetic differentiation into subspecies and varieties had widespread and large glacial refugia, while species with restricted refugia showed no differentiation and little genetic diversity. In a regression tree analysis, 66% of allelic richness could be explained by the total size of glacial refugia, while expected heterozygosity was best explained by the number of glacial refugia. A comparison of projected past and future habitats showed that future migration requirements were much faster than, but highly correlated with, past migration requirements across 24 tree species. Populations in certain localities may be at risk of maladaptation due to the inability of gene flow to keep pace with the migration of climate habitats.

Acknowledgements

First and foremost, I would like to thank my supervisor Andreas Hamann. He has been the most supportive and down-to-earth “boss” I have ever worked with. Andreas has provided me with direction, mentorship, a fair dose of reality when I needed it, the occasional kick in the pants, and a wealth of personal and professional experience for which I am very grateful. As a new graduate student, fresh out of a professional career that would otherwise have left me with a comfortable salary and early retirement, Andreas provided a fresh perspective on life and research, as well as a welcome introduction to the scientific and research process. Andreas also had a knack for encouraging my mistakes, half-cooked investigations, and certain-to-fail analyses, and for that I am surprisingly grateful. I learned a great deal about research and patience through these experiences, so thank you Andreas for both your support and your challenges along the way.

My gratitude also goes to Scott Nielsen, the second member of my supervisory committee and my pseudo-co-supervisor. While Scott’s deep knowledge of things ecological, biometrical, and statistical could be a source of intimidation, his light-hearted nature and approachable demeanor make him instead a supportive and welcome collaborator and co-supervisor. Thank you Scott for the opportunities which you have provided to me and thank you for your input on this research.

Andrew Bush, the third member of my supervisory committee, was a valuable source of knowledge on all things climate-related. His knowledge of climate systems and his clear passion for the mathematical explanation of the winds, rains, and energy of our planet is inspiring. Thank you Andy for sharing some of your knowledge with me, and thank you for being friendly and enthusiastic when you did so.

Other members of faculty and graduate students at the department of Renewable Resources also deserve my gratitude. Thank you Vic Lieffers for your early contributions on my committee. Thank you Xianli Wang for your Jedi-like knowledge of R and SAS. Thank you Rick Pelletier for the laughs, for the philosophy, and for fixing all the stuff I broke. The entire lab would grind to a halt without your help. Thank you Glen Armstrong for your kind encouragement and for your support of graduate students in our department.

Research requires more than just good people, however. It takes money and data, and I have been fortunate for the generosity of others in regards to both.

Government funding for this research was provided by the Government of Canada, Natural Sciences and Engineering Research Council (NSERC) and the Government of Al-

berta, Alberta Innovates Technology Futures program. Financial support also came from industrial partners, including Ainsworth Engineered Canada LP, Alberta-Pacific Forest Industries Inc., Daishowa-Marubeni International Ltd., and Weyerhaeuser Company, Ltd. I have also personally been supported through various scholarships and bursaries, generously awarded by the Faculty of Graduate Studies and Research at the University of Alberta and the Graduate Students Association at the University of Alberta.

Data for this project was personally provided by Andrew Bush (GFDL model data) and Art Dyke (palaeoecological data sets). The public posting of other data, including climate model data, palaeoecological data, and forestry data, is also greatly appreciated. Thanks Laura Gray for being my partner in modelling crime for many years. Thanks Elisabeth for making me feel welcome and for making the lab feel like home. Thanks also to all my other lab-mates for their help and company along the way. I also thank all the hard-working and typically under-appreciate people working in the main office at the Department of Renewable Resources. Thanks for helping me navigate the bureaucracy.

Last, but certainly not least by any stretch imaginable, I would like to thank my family and friends (the latter being a category into which all those mentioned above also fit). I have been blessed with the company of wonderfully supportive people throughout my tenure in Edmonton and at the University of Alberta. As a lonely pair of displaced Calgarians, my wife and I both found a warm welcome among new friends here. Thanks Stefan for being a good friend and for always recognizing when I most needed one. Thanks Shari for reminding me that most things in life and work really are worth it, coffee included. And thanks for never letting me off the hook. Thanks Andrea for knowing when someone needs a hug, which is always. Thanks Simmon for being genuine in a world where falsity can dominate. Thanks Josh for your integrity and character. For just a young kid, you taught me a lot. Thanks Becky for showing me how to keep perspective. Thanks Lesley for your gentle enthusiasm for all the great stuff that life and friends have to offer. And thanks to all the other Rising Stars: Joe, Travis, Pat, the Adams, Pow, DMACK, Emily, Robin, Mike and Steph, Greg and Heather, Jan and Lenka, and all the other friends I can't list here. I love all you guys. You helped make these years as some of the best of my life.

Thanks to whoever showed up for the long rides in the river valley, the punishing swim sets, the ab-crushing core workouts, the eyelashes-frozen-together ski laps at Goldbar, the nipple-cracking open water swims in April, and the four-season outdoor running adventures. Sanity was preserved and research problems were solved between intervals and hill climbs. If your company helped push me out the door for any of these, a part of this thesis belongs to you.

My family and my parents in particular deserve my deepest gratitude for their support of my life pursuits in all the directions they have taken me. I have had the benefit for most of my life of having four parents, all of whom gave their support whenever it was needed. Especially mom and dad, thanks for always pushing me to follow my dreams and thanks for your support along the way. Thanks also for making academic careers seem highly rewarding and just sexy enough to pursue. Your encouragement and sacrifices for my benefit have provided me with experiences and opportunities that I cannot measure and cannot begin to repay. Though I will try.

Thanks to Richard for not just telling me, but showing me that adventures are worth pursuing and risks are worth taking. You were right. This adventure was rewarding and this risk was worthwhile. I miss you.

The greatest debt of gratitude I owe to my wife Bridget. She has willingly and enthusiastically followed me on my sinuous path, both physically and emotionally. Without her support, this thesis would never have been started, let alone completed. She has been our stable foundation through the frustrations, the failures, and the tears, but also my soul mate to share the miracles, the victories, the smiles, and the laughs. She is the person to whom I owe every success and accomplishment, both in my work and in everything else in my life. A million times over, thank you, Bridget. I am the luckiest.

Table of Contents

List of Tables

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

Predicting potential climate change impacts with species distribution models: a palaeoecological perspective

The number and size of glacial refugia predict modern genetic diversity in western North American tree species

The velocity of species migration in complex landscapes: western North American trees under past and future climate change

List of Figures

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

Predicting potential climate change impacts with species distribution models: a palaeoecological perspective

The number and size of glacial refugia predict modern genetic diversity in western North American tree species

The velocity of species migration in complex landscapes: western North American trees under past and future climate change

Chapter 1

General introduction

1.1 Historic climate variability

While we may perceive our present-day interglacial climate as “normal”, for roughly 80,000 of the last 100,000 years the earth has existed in a glacial state in which global mean annual temperatures were about 3°C to 5°C colder than present (IPCC, 2007). In western North America, estimates of mean annual temperatures near the end of the last glacial period from the Global Fluid Dynamics Lab (GFDL) climate model were between 4-6°C colder along the British Columbia coast and as much as 20°C colder in the northern interior (Bush & Philander, 1999). As a consequence, large tracts of uninterrupted ice covered much of present-day Canada. The Cordilleran and Laurentide ice sheets were two adjoining continental-scale glaciers, up to three and a half kilometres thick, extending from about the 49^{th} parallel northward to the Arctic Ocean and from the Pacific to Atlantic coast (Dyke *et al.*, 2002; Peltier, 1994). Left unglaciated through this period was much of present-day Alaska (Beringia) as well as some areas along the Pacific coast (Lacourse *et al.*, 2005). South of the continental ice sheets, the estimated mean annual temperature difference from the present day in the GFDL model was somewhat milder, ranging from about 4-6°C colder in Mexico and along the Pacific coast to 10-14°C colder in the vicinity of the continental ice (Bush & Philander, 1999).

Long-term changes in global temperatures are largely a consequence of changes in the orbital parameters of the earth, producing millennial-scale variation in the intensity of solar radiation received by the planet. Earth’s elliptical orbit changes in eccentricity over ca. 100,000 years, the obliquity of its rotational axis swings back and forth over ca. 42,000 years, and the orientation of its rotational axis cycles over ca. 23,000 years (Milanković, 1941). Known collectively as Milankovich Cycles, these gradual changes in insolation are considered the primary drivers of global-scale climate change in the past (Hays *et al.*, 1976): shifts between glacial periods, when much of earth lies covered in ice, and interglacial periods, when prevailing warmer temperatures allow plants and animals to spread into formerly glaciated areas.

Adding a layer of ice several hundred to thousand of metres thick to the continental landscape also has implications for air circulation. On a local scale, katabatic winds (flowing downslope) off the ice sheets would have had a strong cooling effect at the periphery of the ice (Bromwich *et al.*, 2005). On a continental scale, the ice produces two notable phenomena: one from the cooling effect of the ice itself, the other from the orographic effect of the ice thickness. First, differences in surface albedo produced extremely cold surface temperatures over the Laurentide ice sheet adjacent to much warmer temperatures over the ice-free land. The result of this temperature gradient is the creation of an anticyclonic (clockwise) airflow over the continental ice (COHMAP, 1988; Manabe & Broccoli, 1985). Second, the topography of the very thick continental ice served to split the North American jet stream around the ice sheet itself with the stronger airflow along the southern edge of the continental ice (farther south than at present) and the weaker airflow northward around the Laurentide ice sheet (Bush & Philander, 1999; COHMAP, 1988; Manabe & Broccoli, 1985). Together, these air circulation phenomena change glacial-period storm paths and frequencies, generally altering precipitation regimes by bringing more moisture to the southwestern United States (Bromwich *et al.*, 2005; Bush & Philander, 1999; COHMAP, 1988).

Dramatic differences in temperature, precipitation patterns, and ice coverage at the height of the last glacial maximum, 21,000 years ago, made the North American continent climatically, topographically, and ecologically, a very different place than North America of today. Cold and wet conditions have dominated earth's surface for most of the last 3 million years (Webb & Bartlein, 1992), and recently persisted for more than 100,000 years through the last glacial period. Only in the last 20,000 years of the current interglacial period has the climate warmed such that continental ice sheets have receded and habitat in northern glaciated areas has become suitable for flora and fauna. For the duration of the last glacial period, until global temperatures began to warm through the Holocene, flora and fauna in North America were generally restricted to smaller ranges (due simply to the loss of area to the continental ice sheets), typically south of their present ranges where temperatures were warmer (Comes & Kadereit, 1998; Jackson *et al.*, 1997).

1.2 Reconstructing past vegetation patterns

Knowledge of the climate conditions and vegetation distributions of the past has been attempted in many different ways. Historical climate measurements are available from the past few centuries, as are records of past plant and animal distributions (Brazdil *et al.*, 2005). However, further in the past the main data sources are macrofossil and fossil pollen records (Dyke, 2005; Webb, 1981). These fossil records, usually recovered and carbon dated from core samples of lake sediments, provide a chronological record of pollen abundances (fossil pollen), or a record of the immediate local communities (macrofossils) for a specific location. Despite the high effort required to obtain these samples and the difficulty in identifying some pollen beyond the level of genus, many samples exist from North America covering the time period since the last glacial maximum, providing a proxy for vegetation patterns through the Holocene. However, most North American records are concentrated in the eastern half of the continent and such data are largely unavailable in drier regions like the US southwest where lakes are scarce (e.g. Thompson & Anderson, 2000). Furthermore, while samples may provide a complete chronology for a given location, pollen productivity and dispersal in plants is highly variable, often making specific inferences about local biodiversity problematic without the inclusion of rarer macrofossils in the samples (Odgaard, 1999).

Similar to fossil pollen data, packrat (genus *Neotoma*) middens are preserved nests that contain plant material collected and stored by small rodents that may be used to infer local plant community structure from the past. Middens often contain material dated to tens of thousands of years ago and are more widely available in drier regions where a scarcity of water bodies make fossil pollen records rare. Advantages of packrat midden data over pollen data are their reflection of truly local plant communities (within approximately 50m of nest locations), their tendency to hold material identifiable to the species level, and their quality of preservation (Betancourt *et al.*, 1990; Thompson *et al.*, 2004). However, there are potential biases in these records related to contamination of datable material, selection biases of packrats, and difficulties relating nest material abundance with local plant abundance (Dial & Czaplewski, 1990). Despite these challenges, these records provide valuable information on plant communities in locations or environmental conditions that may be underrepresented or missing in the fossil pollen record.

Where paleoecological data is available, including locations of fossil pollen, macrofossil, and packrat midden data, local biogeographic histories can be inferred with some detail. For example, vegetation patterns for the Pacific Northwest have been reconstructed largely from fossil pollen records. These reconstructions suggest that coastal and adjacent forest communities have developed from drier-adapted communities in the last glacial period to the wet-adapted rainforest communities that we see today while subalpine forest communities have moved upward in elevation through the Holocene (Gavin *et al.*, 2001; Whitlock & Bartlein, 1997; Whitlock & Brunelle, 2007). Pollen records for the northern interior show a development from colder and drier steppe-dominated ecosystems in the last glacial period to the modern forest communities, dominated by ponderosa pine and Douglas-fir in dry regions or by hemlock, western white pine, and fir in more mesic areas (Whitlock & Bartlein, 1997). In the Beringian region of North America (modern-day Alaska), numerous fossil pollen records show strong evidence for the persistence of many tree species through the last glacial period, most notably species of pine, spruce, aspen, and birch (Brubaker *et al.*, 2005). In the southern USA interior at the last glacial maximum, combinations of fossil pollen and packrat data suggests more widespread conifer woodlands than at present as well as a general replacement of modern conifer and mixedwood forests by steppe grasslands (Thompson & Anderson, 2000). In addition to local vegetation histories, several continental-scale biome-level reconstructions of vegetation distributions have been undertaken using a wide breadth of palaeoecological data (e.g. Dyke, 2005; Whitmore *et al.*, 2005; Thompson & Anderson, 2000). While these reconstructions provide a general overview of past species distributions, they must be inferred from local palaeoecological data records which may be geographically biased towards conditions favouring the preservation of palaeoecological data.

1.3 Genetic clues and genetic drivers

To compliment past vegetation reconstructions based on palaeoecological data, information about past species distributions and migration patterns can also be inferred through investigations of species' genetic diversity. Known as "phylogeography," the study of geographic patterns in the diversity of neutral genetic markers in modern populations of species can provide insight into where species were located through the long glacial period and how species have moved across the landscape through Holocene climate warming (Avise, 2009). Certain geographic features of past species distributions can leave unique patterns on the structure of modern genetic diversity when measured with neutral genetic markers such as allozymes, SNPs, or other polymorphisms. While genetic analyses cannot provide information about precisely where and when a species was present on the landscape, past geographic isolation of sub-populations can easily be inferred, and sometimes, past range restrictions detected.

The limited ranges in which species persisted through the glacial period are known as glacial “refugia”, and there are many ways in which these range dynamics shape the genetic diversity of populations. For example, species’ range contractions, such as those that occurred through the glacial period, result in population bottlenecks which can reduce genetic diversity due to decreased population sizes and increased influence of genetic drift, even within a single generation (Nei *et al.*, 1975). Further, if a species is divided into multiple geographically-isolated populations for long time spans, these allopatric communities will diverge genetically, leaving a genetic fingerprint on the species even if populations are later recombined. Because of the multi-generational requirements of genetic changes through gene flow, genetic admixture, and mutations, the recovery of genetic diversity tends to be a slow process (Petit & Hampe, 2006). Similar to range contractions, rapid expansion into new range area can also have the effect of limiting genetic diversity, as colonist populations may be established by only a few individuals (Hewitt, 2004).

These patterns in neutral (non-adaptive) genetic markers are not to be confused with adaptive genetic traits which respond to selection pressures. The composition of adaptive genetic traits can also change very quickly within a population. While organisms such as trees may be long-lived and we may therefore think of them as slow-changing, they produce frequent large numbers of offspring subject to selection pressure. This enables relatively quick evolutionary adaptation and, in the process, changes population genetic structures on shorter timescales (Aitken *et al.*, 2008; Petit & Hampe, 2006). However, as neutral genetic markers are by definition non-adaptive, they are not subject to such selection pressures, and only change through random mutation, genetic drift, or the influx of new genetic material. For this reason, neutral markers are ideal to provide information about long-term population structures, as they are not affected by short-term selection pressures.

This slow and biogeographically-sensitive nature of neutral marker traits means that, in order to gain a complete understanding of the sources of neutral genetic structure of trees in North America, we must consider species distributions through the “normal” glacial conditions that persisted for over 100 millennia prior to our current interglacial “exception”. Even in this context, while patterns of genetic markers may allow us to make general inferences about the historical biogeography of a species, they provide limited information about geographic specifics. Such general trends as estimates of refugia locations, potential migration routes, or the likelihood of allopatric populations may be inferred, but the genetic data cannot tell us where and when a species was specifically present.

1.4 Using models to project species distributions

To achieve more specificity in time and space, modelling applications are increasingly being used to reconstruct past species habitats as well as forecast future species habitats. Traditionally, reconstructing the historical biogeography of species has been the objective of the fields of paleoecology (e.g. Dyke, 2005; Williams *et al.*, 2000) and phylogeography (see review by Jaramillo-Correa *et al.*, 2009). While these data can assist our understanding of past species ranges and movements, they are often sparse or incomplete (as with fossil data) or unable to provide specific spatial or geographic information (as with phylogeography). Species distribution models, also known as bioclimate envelope models or ecological niche models, are empirical models that statistically relate environmental variables with observed ecosystem or species occurrence data. These models are effective tools for reconstructing complete species distribution maps from limited input data (Guisan & Zimmermann, 2000), making them important tools for studying potential plant responses to changes in the environment and to climate in particular (Elith & Leathwick, 2009). Modelled hindcasts have been employed previously 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. Carstens & Richards, 2007; Yesson & Culham, 2006). Unfortunately, back-prediction studies are often restricted to a small number of species (e.g. Bradshaw *et al.*, 2000; Giesecke *et al.*, 2007; Gignac *et al.*, 2002; Svenning *et al.*, 2008), or they are geographically limited (e.g. Kaplan *et al.*, 2003; Pearman *et al.*, 2008).

Species distribution models often incorporate high-resolution climate data, such as those from global climate model projections, to generate projections of suitable species habitat under projected climate change (e.g. Hamann & Wang, 2006) or to hindcast past species distributions from palaeoclimate reconstructions (e.g. Rodríguez-Sánchez & Arroyo, 2008; Wilson & Pitts, 2012). Coupled with abundant present-day species distribution data, the environmental niches of species may be characterized statistically. These statistical relationships may then be used to predict species occurrences, or the probabilities thereof, from new environmental data. To generate maps of projected species habitats, however, species distribution models require environmental predictor data both from the present day and from a different time period (either the past or future). Climate projections for the future, or palaeoclimate reconstructions for the past, are most often generated at a global scale using coupled oceanic-atmospheric general circulation models (GCMs). These models are three-dimensional computer-based reconstructions of the earth's atmospheric and oceanic circulation, grounded in the mathematical laws of thermodynamics and chemical interactions that drive global circulation and energy fluxes. Typically, they account for insolation rates (determined by orbital parameters), atmospheric chemistry, and include feedbacks from the oceans, ice, land cover, and even vegetation. GCMs have the ability to recreate modern climate, predict future climates, and reconstruct past climates based on the chemical and physical properties of the atmosphere and oceans of the time (Gignac *et al.*, 2002). GCMs allow us to generate seamless maps of species habitat at continental or global scales, for the past, present, and future.

These habitat projections are powerful tools, but their use is based on many biological assumptions and the associated limitations of these models must be recognized (Araújo & Peterson, 2012; Guisan & Thuiller, 2005; Hampe, 2004; Pearson & Dawson, 2003). First, species distribution models are built on the premise that species have a definable environmental niche which represents the suite of abiotic conditions in which the species can survive and reproduce (Hutchinson, 1957). When niches are characterized using species occurrence and environmental data within a model, it must be assumed that species are in equilibrium with their environment, which, largely due to migration constraints on the species, may not always be the case (see review by Araújo & Peterson, 2012). A similar assumption exists for model projections also: that the species being projected will be (in the future) or was (in the past) in equilibrium with the environment. While this assumption may hold true over longer time scales in which species are able to migrate across the landscape to fill suitable environmental habitat (Webb, 1986), it is certain to be violated over short time spans with fast-paced environmental change, such as anthropogenic climate change. In these cases, model projections must be interpreted differently, and conclusions should only be drawn about the location of suitable habitat for species and not interpreted as projected migrations of species themselves.

Over long time scales, other modelling assumptions become problematic. In order for species projections to retain validity, they must assume that species niches remain constant over time. Despite constant adaptive evolution in response to selection pressures, it has been shown that species environmental tolerances are largely stable, suggesting that preservation of the fundamental niche (i.e. niche conservatism) is likely over millennial timescales for long-lived organisms such as trees (Ackerly, 2003; Araújo & Peterson, 2012; Peterson *et al.*, 1999; Wiens & Graham, 2005). But the realized niche could change given changes in the biotic community (i.e. exposure to new competitors). For this reason, we must be cautious about projecting habitat for no-analogue climate conditions (climate conditions without similar representation in the model training data), as these may result in new arrangements of communities, new competitive regimes, and thus new realized niche space (Williams & Jackson, 2007).

Biotic effects on niche space are of great importance to species distribution models, as the models themselves do not include mechanistic processes such as species interactions and dispersal limitations. While this helps maintain model simplicity, it could also make range projections problematic (Hampe, 2004). In one sense, biotic interactions are considered indirectly through the use of modern species presence/absence data for model training, which effectively characterizes the realized niche of the species (limited by competition, diseases, pests and other biotic factors) rather than the fundamental niche of the species (containing the full breadth of abiotic environmental conditions in which the species could persist). While biotic factors may be critically important to defining species ranges at a local scale (Davis *et al.*, 1998), it has been shown that, at larger regional to continental scales, environmental drivers such as climate dominate the general presence or absence of a species in a broader region (but see Beale *et al.*, 2008; Pearson & Dawson, 2003).

As with any biometrical tool, ecological niche models are only useful if they are built and used appropriately, recognizing important assumptions and limitations while ignoring others (Araújo & Guisan, 2006; Araújo & Peterson, 2012; Botkin *et al.*, 2007). Models of all kinds are necessary simplifications of more complex systems, making them only representative of reality. However, their purpose is not to be wholly correct, but rather useful for the intended application (Box & Draper, 1987). The most fundamental quality of a useful model is its ability to provide accurate predictions. In the case of species distribution models, validating predictions of species range is relatively straightforward: one only needs to compare modelled distributions to known distributions. For projections of the present day, the most common type of model evaluation involves splitting a single data set into two subsets: one used for model training/building; the other used for model validation. Conversely, independent validations use completely different data to evaluate model projections than those used to train the model, often going beyond just different modern data sources by using projections and data from different time periods or different geographical areas (Araújo *et al.*, 2005). Validation based on data splitting may be quicker and easier, but these evaluations lack true independence because of spatial auto-correlations in species census data, and often produce overly-optimistic assessments of model accuracy (Araújo *et al.*, 2005).

Many critiques of species distribution models have focused on ways to improve modelling accuracy, refine model validation techniques, or increase the ecological relevance of modelling outputs (see reviews by Araújo & Guisan, 2006; Botkin *et al.*, 2007; Elith & Graham, 2009; Elith & Leathwick, 2009; Pearson & Dawson, 2003). But the underlying message of this literature is often similar: Like all models, species distribution models simplify reality by making biologically unrealistic assumptions and characterizing complex natural systems often with simple statistical formulae. However, when constructed, evaluated, and interpreted appropriately, these models represent a useful tool, especially for continental-scale research.

1.5 Links to future climate change

The development and application of species distribution models in ecology has increased greatly over the past decade, driven largely by studies of future climate change and its effect on species habitats (Dawson *et al.*, 2011). Without context, however, the magnitude of the projected future effects of climate change is difficult to assess. For this reason, looking back through the past to see how communities have dealt with prior environmental changes is also an integral part of future assessments of plant vulnerability (Jackson *et al.*, 2009; Jackson & Hobbs, 2009). However, there are critical differences between the climate variability of the past and that which is expected over the coming decades and centuries (IPCC, 2007).

While orbital drivers of climate may play a role in slow, long-term climate trends, other parameters such as atmospheric chemistry can force climates to greater extents and at faster speeds (Viau *et al.*, 2006). At the last glacial maximum, atmospheric CO_2 is estimated to have been approximately 180ppm, then increasing to approximately 265ppm in the early Holocene, and remaining at that level until human CO_2 emissions became relevant (IPCC, 2007). The addition to the atmosphere of anthropogenic CO_2, primarily from fossil fuel burning, has increased atmospheric CO_2 levels to approximately 390ppm in 2010, increasing at a rate of approximately 2ppm per year over the last three decades. This additional CO_2 causes infrared radiation to be re-emitted back to earth by the atmosphere (the greenhouse effect) resulting in a general warming trend since industrialization. With ever-increasing levels of atmospheric greenhouse gasses, the earth is poised to continue this warming trend at least through the next century (IPCC, 2007). A comparison of projected warming over the next century (around 4°C warmer at the scale of decades) with the reconstructed global mean temperature changes (4-6°C colder at the scale of millennia) suggests that anthropogenic changes in atmospheric chemistry have the potential to drive earth's climate systems far more profoundly than natural forces, such as the Milankovich cycles.

Comparing the extent of past and future temperature change provides context for the environmental changes associated with anthropogenic global warming. Holocene warming, while fast from a geological perspective, was sufficiently slow for tree species to find suitable habitat to persist on the landscape. That said, there was also much variability in Holocene warming rates. At times, for example, combinations of orbital parameters and atmospheric chemistry lead to periods of rapid warming while cooling from large influxes of freshwater from melting continental ice sheets (melt water pulses) at times helped offset warming signals (Clark *et al.*, 2009; Shakun *et al.*, 2012). This variability is reflected in rapid warming and cooling events such as the Younger Dryas and Bølling-Allerød, which may have featured temperature changes up to several degrees per century, analagous with future projections (Clark *et al.*, 2012; Willis & MacDonald, 2011).

By comparison, anticipated future climate warming, driven by increases in atmospheric greenhouse gasses, is projected to be similar to these periods of rapid change, with temperatures rising several degrees over decades or centuries, not millennia (IPCC, 2007). To cope with these accelerated changes, tree populations must accomplish one of two things, depending on their location within the species range. At the leading edge of the range, trees must physically migrate into regions previously uninhabited by the species, requiring the dispersal of seed via wind, mammals, or birds and the subsequent germination and survival to maturity of the seedling. Unfortunately, most migrating tree populations in the future will be moving into habitat currently occupied by other species (north or upslope of their current ranges), undermining ability of seedlings to germinate and survive. By comparison, in post-glacial warming, retreating continental ice sheets left abundant vacant habitat for primary succession. Few species today, with the exception of some northern boreal populations, have this same advantage.

While physical migration may be required at the leading edge of species ranges, throughout the majority of the species current habitat, populations may survive *in situ* by adapting to new conditions through selection. Populations of high genetic diversity, common among tree species, may have adequate genetic breadth to locally contain well-adapted genotypes. If not, the admixture of genetic material from other populations better-adapted to the new local conditions is required to alter environmental tolerances. In this respect, trees have the advantage of producing large amounts of pollen which can travel extensively with the wind, allowing outcrossing with diverse and far away populations (Petit & Hampe, 2006). Whether these migratory and adaptive capabilities are sufficient for trees to keep pace with future climate change is a critical question which requires three pieces of information to answer: 1) the potential rate of movement of required species migration under future warming scenarios (e.g. Iverson *et al.*, 2004; Loarie *et al.*, 2009; Malcolm *et al.*, 2002), 2) estimations of past species migration rates through the most recent warming analogue of the Holocene (Davis, 1981; Huntley, 1991; King & Herstrom, 1997), and 3) the expected rate of future climate change. While several assessments of these questions have been undertaken (Pearson, 2006), they almost exclusively investigate genetic variation at the species level, whereas changes in local climate are more likely to have effects at the population level, wherein plants are specifically adapted to local environmental conditions (Aitken *et al.*, 2008).

Global climate change will place survival challenges on populations of all species, but particularly those without the ability to move or those without the genetic diversity to adapt. Management and conservation efforts to address climate change may be possible, but they require an assessment of vulnerabilities of species and locally adapted populations within species, as well as a thorough understanding of the inherent capacities of tree species and populations to respond to changing environments through migration and adaptation. Reconstructions of past species ranges and migration capacities observed in the past can be an important contribution to identify species that may be vulnerable to anthropogenic climate change (Jackson, 2007; Jackson *et al.*, 2009).

1.6 Thesis outline

In this thesis, I develop and employ species distribution models to reconstruct biogeographical histories of 24 tree species in western North America (west of 100° latitude) to the last glacial maximum. I investigate if the number and size of reconstructed glacial refugia can partially explain modern genetic diversity and genetic structure, including the existence of species variants and subspecies. Further, I quantify required rates of gene flow and migration under past and projected future climate change. This thesis is structured into four research chapters. The first two chapters have a methodological emphasis, developing effective approaches to modelling species habitats of the past and future and testing model assumptions and model limitations. The third and fourth chapters focus on biological questions, where I address the historical biogeography of trees and the effect of changing climates on genetic diversity.

All chapters make use of similar methodological approaches and similar data sources: I used tree frequency data from 55,744 forest inventory plots, assembled from Canadian and American databases (Govt. of Alberta, 2004; Smith, 2002). As a representation of current climate, I used interpolated weather station data for the 1961-1990 period (Wang *et al.*, 2012). For past climate reconstructions I use data from the Geophysical Fluid Dynamics Laboratory (GFDL) for the periods 21,000, 16,000, 9,000, and 6,000 years ago (Bush & Philander, 1999), and from the Community Climate Model Version 1 (CCM1) from the National Center for Atmospheric Research for the periods 21,000, 16,000, 14,000, 11,000, and 6,000 years ago (Kutzbach *et al.*, 1998). I validated all species range reconstructions against 3,571 palaeoecological records from 835 unique study sites, comprising fossil pollen, macrofossil, and packrat midden data (Dyke, 2005; Strickland *et al.*, 2001). For model-based biome projections, I validated against 1,395 records of reconstructed biomes from 1,224 unique sites (Thompson & Anderson, 2000). Finally for genetic analysis, I used reported rates of allelic richness and expected heterozygosity for 477 populations of 23 tree species from the published literature.

I begin in my **first research chapter** (thesis Chapter 2) with a methodological assessment, in which I evaluate and compare different statistical techniques used in ecological niche modelling. Multiple statistical approaches are available with which to build species distribution models and many have been used successfully to project habitats for different species in different locations. I compare these approaches to determine the most effective approach for projection of North American tree habitats. I also consider ensemble modelling approaches, which combine numerous individual techniques into a single projection, as these methods have been previously shown to be effective (Araújo & New, 2007). Further, because more techniques have been developed to predict habitat for specific species occurrences (probability of presence output) rather than ecosystems or communities (class variable output), this chapter also focuses on evaluating the differences between the outputs of these approaches. A new class-based modelling approach, which uses the smallest Euclidean distance of principal components of predictive variables as modelling criteria, is also tested against more established modelling methods.

In this chapter, I also consider methods of model evaluation. Traditionally, model accuracy has been validated by withholding a subset of the available data for model development, and subsequently evaluating the accuracy of model predictions against this withheld dataset (a type of cross-validation). However, Araújo et al. (2005) and Segurado et al. (2006) point out that spatial autocorrelations in biological census data disproportionately favour models that are prone to over-parameterization and generally result in overly-optimistic model evaluations. Here, I carry out independent tests of model performance, building models with present-day species data for North American trees and using pollen and fossil data for validation. Other pseudo-independent validations, such as extrapolating to different geographic areas, are also tested, as suggested by Araújo et al. (2005).

In my **second research chapter** (thesis Chapter 3), I address an important criticism of species distribution models, which is of particular importance when projecting models into the future. If future climates contain conditions or arrangements for which there exist no modern analogues (i.e. different arrangements of climate variables or different extreme values), model projections represent extrapolations beyond data coverage and may therefore lack realism (Williams & Jackson, 2007). To better understand how no-analogue climates may compromise future projections, I investigate how back-predictions of species habitats to the last glacial maximum correspond to palaeoecological data, given palaeoclimate conditions that have no modern analogue. In this chapter, I use a machine-learning technique (Random Forest) to reconstruct historical distribution of biomes (a categorical variable), a Mahalanobis distance measure to identify no-analogue climates, and a database of pollen and fossil data to evaluate model performance for well-matched versus no-analogue climates.

My **third research chapter** (thesis Chapter 4) builds on the previous methodology sections, using preferred modelling methods to simulate the historical biogeography of major forest tree species of western North America since the last glacial maximum. Hindcasted species habitats are developed and interpreted in the context of three biogeographical mysteries in the region: 1) the presence of many genetically diverse subspecies, adapted to highly divergent climates; 2) the history of disjunct population of temperate rainforest species in the interior, only some of which exhibit genetic differentiation from coastal populations; and 3) the impossibly high migration rates calculated for species recolonising post-glacial areas from southern refugia. Further, quantitative summaries of reconstructed species ranges at the last glacial maximum are compared to data on modern genetic diversity of species to investigate the implications of refugial isolation, testing the conceptual hypothesis that species which experienced more range bottlenecks at the last glacial maximum continue to exhibit less present-day genetic diversity, as measured by neutral markers.

My **fourth research chapter** (thesis Chapter 5) considers climate change at a more local level, examining potential future migration and gene flow demands on local populations. Particular focus is placed on migration rates required for tree populations to keep pace, either through physical migration or gene flow, with suitable climates as they move across the landscape. Future within-range migration demands on local populations of western North American trees are estimated, as are post-glacial migration rates as a basis for comparison, using a k-sample nearest neighbour algorithm measuring to locations of similar climate conditions in different time periods. Calculated rates for the past and future are compared in magnitude and pattern, with the aim of determining whether future climate changes may result in qualitatively different migration demands on populations.

1.7 References

Ackerly, D. D. 2003. Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Sciences*, **164**(3), S165–S184.

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T. L., & Curtis-McLane, S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**(1), 95–111.

Araújo, M. B., & Guisan, A. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**(10), 1677–1688.

Araújo, M. B., & New, M. 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**(1), 42–47.

Araújo, M. B., & Peterson, A. T. 2012. Uses and misuses of bioclimatic envelope modeling. *Ecology*, **93**(7), 1527–1539.

Araújo, M. B., Pearson, R. G., Thuiller, W., & Erhard, M. 2005. Validation of species-climate impact models under climate change. *Global Change Biology*, **11**(9), 1504–1513.

Avise, J. C. 2009. Phylogeography: retrospect and prospect. *Journal of Biogeography*, **36**(1), 3–15.

Beale, C. M., Lennon, J. J., & Gimona, A. 2008. Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(39), 14908–14912.

Betancourt, J. L., Van Devender, T. R., & Martin, P. S. 1990. *Packrat Middens: The Last 40,000 Years of Biotic Change.* Tucson: University of Arizona Press.

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. 2007. Forecasting the effects of global warming on biodiversity. *Bioscience*, **57**(3), 227–236.

Box, G. E. P., & Draper, N. R. 1987. *Empirical Model-Building and Response Surfaces.* Wiley Series in Probability and Statistics. New York: John Wiley & Sons.

Bradshaw, R. H. W., Holmqvist, B. H., Cowling, S. A., & Sykes, M. T. 2000. The effects of climate change on the distribution and management of Picea abies in southern Scandinavia. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **30**(12), 1992–1998.

Brazdil, R., Pfister, C., Wanner, H., Von Storch, H., & Luterbacher, J. 2005. Historical climatology in Europe - The state of the art. *Climatic Change*, **70**(3), 363–430.

Bromwich, D. H., Toracinta, E. R., Oglesby, R. J., Fastook, J. L., & Hughes, T. J. 2005. LGM summer climate on the southern margin of the Laurentide Ice Sheet: Wet or dry? *Journal of Climate*, **18**(16), 3317–3338.

Brubaker, L. B., Anderson, P. M., Edwards, M. E., & Lozhkin, A. V. 2005. Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *Journal of Biogeography*, **32**(5), 833–848.

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**(D20), 24509–24525.

Carstens, B. C., & Richards, C. L. 2007. Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**(6), 1439–1454.

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., Mitrovica, J. X., Hostetler, S. W., & McCabe, A. M. 2009. The Last Glacial Maximum. *Science*, **325**(5941), 710–714.

Clark, P. U., Shakun, J. D., Baker, P. A., Bartlein, P. J., Brewer, S., Brook, E., Carlson, A. E., Cheng, H., Kaufman, D. S., Liu, Z. Y., Marchitto, T. M., Mix, A. C., Morrill, C., Otto-Bliesner, B. L., Pahnke, K., Russell, J. M., Whitlock, C., Adkins, J. F., Blois, J. L., Clark, J., Colman, S. M., Curry, W. B., Flower, B. P., He, F., Johnson, T. C., Lynch-Stieglitz, J., Markgraf, V., McManus, J., Mitrovica, J. X., Moreno, P. I., & Williams, J. W. 2012. Global climate evolution during the last deglaciation. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(19), E1134–E1142.

COHMAP. 1988. Climatic Changes of the Last 18,000 Years - Observations and Model Simulations. *Science*, **241**(4869), 1043–1052.

Comes, H. P., & Kadereit, J. W. 1998. The effect of quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**(11), 432–438.

Davis, A. J., Jenkinson, L. S., Lawton, J. H., Shorrocks, B., & Wood, S. 1998. Making mistakes when predicting shifts in species range in response to global warming. *Nature*, **391**(6669), 783–786.

Davis, M. B. 1981. *Quaternary history and the stability of forest communities.* New York: Springer-Verlag. Chap. 10, pages 132–153.

Dawson, T. P., Jackson, S. T., House, J. I., Prentice, I. C., & Mace, G. M. 2011. Beyond Predictions: Biodiversity Conservation in a Changing Climate. *Science*, **332**(6025), 53–58.

Dial, K. P., & Czaplewski, N. J. 1990. *Do woodrat middens accurately represent the animals' environments and diets? The Woodhouse Mesa study.* Tuscon: University of Arizona Press. Chap. 4, pages 43–58.

Dyke, A. S. 2005. Late Quaternary vegetation history of northern North America based on pollen, macrofossil, and faunal remains. *Gographie physique et Quaternaire*, **59**(2-3), 52.

Dyke, A. S., Andrews, J. T., Clark, P. U., England, J. H., Miller, G. H., Shaw, J., & Veillette, J. J. 2002. The Laurentide and Innuitian ice sheets during the Last Glacial Maximum. *Quaternary Science Reviews*, **21**(1-3), 9–31.

Elith, J., & 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**(1), 66–77.

Elith, J., & Leathwick, J. R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution and Systematics*, **40**, 677–697.

Gavin, D. G., McLachlan, J. S., Brubaker, L. B., & Young, K. A. 2001. Postglacial history of subalpine forests, Olympic Peninsula, Washington, USA. *Holocene*, **11**(2), 177–188.

Giesecke, T., Hickler, T., Kunkel, T., Sykes, M. T., & Bradshaw, R. H. W. 2007. Towards an understanding of the Holocene distribution of Fagus sylvatica L. *Journal of Biogeography*, **34**(1), 118–131.

Gignac, L. D., Halsey, L. A., Bush, A. B. G., & Vitt, D. H. 2002. Evaluating GCM outputs for past climates in North America based on the distribution of Sphagnum peatlands. *Ecoscience*, **9**(4), 489–502.

Govt. of Alberta. 2004. *Ecological Site Information System (ESIS).* Government of Alberta, Sustainable Resource Development (SRD). Available: http://www.srd.gov.ab.ca/lands/geographicinformation/resourcedataproductcatalogue/esis.aspx.

Guisan, A., & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**(9), 993–1009.

Guisan, A., & Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**(2-3), 147–186.

Hamann, A., & Wang, T. L. 2006. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. *Ecology*, **87**(11), 2773–2786.

Hampe, A. 2004. Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography*, **13**(5), 469–471.

Hays, J. D., Imbrie, J., & Shackleton, N. J. 1976. Variations in Earths Orbit - Pacemaker of Ice Ages. *Science*, **194**(4270), 1121–1132.

Hewitt, G. M. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**(1442), 183–195.

Huntley, B. 1991. How Plants Respond to Climate Change - Migration Rates, Individualism and the Consequences for Plant-Communities. *Annals of Botany*, **67**, 15–22.

Hutchinson, G. E. 1957. Population Studies - Animal Ecology and Demography - Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.

IPCC. 2007. *Synthesis Report*.

Iverson, L. R., Schwartz, M. W., & Prasad, A. M. 2004. How fast and far might tree species migrate in the eastern United States due to climate change? *Global Ecology and Biogeography*, **13**(3), 209–219.

Jackson, S. T. 2007. Looking forward from the past: history, ecology, and conservation. *Frontiers in Ecology and the Environment*, **5**(9), 455–455.

Jackson, S. T., & Hobbs, R. J. 2009. Ecological Restoration in the Light of Ecological History. *Science*, **325**(5940), 567–569.

Jackson, S. T., Overpeck, J. T., Webb, T., Keattch, S. E., & Anderson, K. H. 1997. Mapped plant-macrofossil and pollen records of late quaternary vegetation change in eastern North America. *Quaternary Science Reviews*, **16**(1), 1–70.

Jackson, S. T., Betancourt, J. L., Booth, R. K., & Gray, S. T. 2009. Ecology and the ratchet of events: Climate variability, niche dimensions, and species distributions. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 19685–19692.

Jaramillo-Correa, J. P., Beaulieu, J., Khasa, D. P., & Bousquet, J. 2009. Inferring the past from the present phylogeographic structure of North American forest trees: seeing the forest for the genes. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **39**(2), 286–307.

Kaplan, J. O., Bigelow, N. H., Prentice, I. C., Harrison, S. P., Bartlein, P. J., Christensen, T. R., Cramer, W., Matveyeva, N. V., McGuire, A. D., Murray, D. F., Razzhivin, V. Y., Smith, B., Walker, D. A., Anderson, P. M., Andreev, A. A., Brubaker, L. B., Edwards, M. E., & Lozhkin, A. V. 2003. Climate change and Arctic ecosystems: 2. Modeling, paleodata-model comparisons, and future projections. *Journal of Geophysical Research-Atmospheres*, **108**(D19), –.

King, G. A., & Herstrom, A. A. 1997. *Holocene tree migration rates objectively determined from follis pollen data.* New York: Springer-Verlag. Pages 91–101.

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**(6-7), 473–506.

Lacourse, T., Mathewes, R. W., & Fedje, D. W. 2005. Late-glacial vegetation dynamics of the Queen Charlotte Islands and adjacent continental shelf, British Columbia, Canada. *Palaeogeography Palaeoclimatology Palaeoecology*, **226**(1-2), 36–57.

Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B., & Ackerly, D. D. 2009. The velocity of climate change. *Nature*, **462**(7276), 1052–U111.

Malcolm, J. R., Markham, A., Neilson, R. P., & Garaci, M. 2002. Estimated migration rates under scenarios of global climate change. *Journal of Biogeography*, **29**(7), 835–849.

Manabe, S., & Broccoli, A. J. 1985. The Influence of Continental Ice Sheets on the Climate of an Ice-Age. *Journal of Geophysical Research-Atmospheres*, **90**(Nd1), 2167–2190.

Milanković, Milutin. 1941. *Canon of insolation and the ice-age problem (Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeitenproblem).* Belgrade: Israel Program for Scientific Translations (1969).

Nei, M., Maruyama, T., & Chakraborty, R. 1975. Bottleneck Effect and Genetic-Variability in Populations. *Evolution*, **29**(1), 1–10.

Odgaard, B. V. 1999. Fossil pollen as a record of past biodiversity. *Journal of Biogeography*, **26**(1), 7–17.

Pearman, P. B., Randin, C. F., Broennimann, O., Vittoz, P., van der Knaap, W. O., Engler, R., Le Lay, G., Zimmermann, N. E., & Guisan, A. 2008. Prediction of plant species distributions across six millennia. *Ecology Letters*, **11**(4), 357–369.

Pearson, R. G. 2006. Climate change and the migration capacity of species. *Trends in Ecology & Evolution*, **21**(3), 111–113.

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**(5), 361–371.

Peltier, W. R. 1994. Ice-Age Paleotopography. *Science*, **265**(5169), 195–201.

Peterson, A. T., Soberon, J., & Sanchez-Cordero, V. 1999. Conservatism of ecological niches in evolutionary time. *Science*, **285**(5431), 1265–1267.

Petit, R. J., & Hampe, A. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution and Systematics*, **37**, 187–214.

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**(6), 685–695.

Segurado, P., Araújo, M. B., & Kunin, W. E. 2006. Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, **43**(3), 433–444.

Shakun, J. D., Clark, P. U., He, F., Marcott, S. A., Mix, A. C., Liu, Z. Y., Otto-Bliesner, B., Schmittner, A., & Bard, E. 2012. Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature*, **484**(7392), 49–+.

Smith, W. B. 2002. Forest inventory and analysis: a national inventory and monitoring program. *Environmental Pollution*, **116**, S233–S242.

Strickland, L. E., Thompson, R. S., & Anderson, K. H. 2001. *USGS/NOAA North American packrat midden database. Available: http://gec.cr.usgs.gov/data/midden/.*

Svenning, J. C., Normand, S., & Kageyama, M. 2008. Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *Journal of Ecology*, **96**(6), 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**(3), 555–584.

Thompson, R. S., Shafer, S., Strickland, L. E., Van De Water, P. K., & Anderson, K. H. 2004. *Quaternary vegetation and climate change in the western United States: Developments, perspectives, and prospects.* Elsevier. Pages 403–426.

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**(2), 291–301.

Viau, A. E., Gajewski, K., Sawada, M. C., & Fines, P. 2006. Millennial-scale temperature variations in North America during the Holocene. *Journal of Geophysical Research-Atmospheres*, **111**(D9), 12 p.

Wang, T. L., Hamann, A., Spittlehouse, D. L., & Murdock, T. Q. 2012. ClimateWNA-High-Resolution Spatial Climate Data for Western North America. *Journal of Applied Meteorology and Climatology*, **51**(1), 16–29.

Webb, T. 1981. The Past 11,000 Years of Vegetational Change in Eastern North-America. *Bioscience*, **31**(7), 501–506.

Webb, T. 1986. Is Vegetation in Equilibrium with Climate - How to Interpret Late-Quaternary Pollen Data. *Vegetatio*, **67**(2), 75–91.

Webb, T., & Bartlein, P. J. 1992. Global Changes during the Last 3 Million Years - Climatic Controls and Biotic Responses. *Annual Review of Ecology and Systematics*, **23**, 141–173.

Whitlock, C., & Bartlein, P. J. 1997. Vegetation and climate change in northwest America during the past 125 kyr. *Nature*, **388**(6637), 57–61.

Whitlock, C., & Brunelle, A. 2007. *Pollen records from northwestern North America.* Amsterdam: Elsevier.

Whitmore, J., Gajewski, K., Sawada, M., Williams, J. W., Shuman, B., Bartlein, P. J., Minckley, T., Viau, A. E., Webb, T., Shafer, S., Anderson, P., & Brubaker, L. 2005. Modern pollen data from North American and Greenland for multi-scale paleoenvironmental applications. *Quaternary Science Reviews*, **24**(16-17), 1828–1848.

Wiens, J. J., & Graham, C. H. 2005. Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution and Systematics*, **36**, 519–539.

Williams, J. W., & Jackson, S. T. 2007. Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, **5**(9), 475–482.

Williams, J. W., Webb, T., Richard, P. H., & Newby, P. 2000. Late Quaternary biomes of Canada and the eastern United States. *Journal of Biogeography*, **27**(3), 585–607.

Willis, K. J., & MacDonald, G. M. 2011. Long-Term Ecological Records and Their Relevance to Climate Change Predictions for a Warmer World. *Annual Review of Ecology, Evolution, and Systematics, Vol 42*, **42**, 267–287.

Wilson, J. S., & Pitts, J. P. 2012. Identifying Pleistocene refugia in North American cold deserts using phylogeographic analysis and ecological niche modelling. *Diversity and Distributions*.

Yesson, C., & Culham, A. 2006. Phyloclimatic modeling: combining phylogenetics and bioclimatic modeling. *Systematic Biology*, **55**(5), 785–802.

Chapter 2

Method selection for species distribution modelling: are temporally or spatially independent evaluations necessary?[1]

2.1 Summary

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 6,000, 11,000, 14,000, 16,000 and 21,000 years before present. In addition, we evaluate regional extrapolations based on geographic splits of more than 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.

---

[1] A version of this chapter has been accepted for publication. Roberts, D. R. and A. Hamann. 2012. Method selection for species distribution modelling: are temporally or spatially independent evaluations necessary? *Ecography* **21**, 792-802.

2.2 Introduction

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 (e.g. Araújo & Williams, 2000). Over the last decade, such models have been extensively applied to project future species habitat under anticipated climate change (e.g. reviewed by Elith & Leathwick, 2009). A search with ISI Web of Knowledge^{SM} 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 (e.g. Botkin *et al.*, 2007; Elith & 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.*, 2005b; 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.*, 2005b).

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), Flojgaard et al. (2009), and Morueta 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 & Peterson, 2006), or by using modern census data for training and palaeoecological data for validation (Giesecke *et al.*, 2007; Martinez-Meyer *et al.*, 2004; Pearman *et al.*, 2008; Roberts & Hamann, 2012; Rodriguez-Sánchez & Arroyo, 2008). 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 apply these techniques to predict the habitat of 24 major tree species of western North America based on more than 55,000 present-day sample plots and more than 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. We conclude by discussing the suitability of available techniques for different objectives in species distribution modelling.

2.3 Methods

2.3.1 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 for five periods of the Holocene and Late Pleistocene (6,000, 11,000, 14,000, 16,000, 21,000 years 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 2.1). All techniques, except the ecosystem-based approaches and Discriminant Analysis were implemented using the *BIOMOD* package (Thuiller, 2003) for the R programming environment (R Core Team, 2009). *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 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 Institute, 2008), 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). 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 (see Roberts & Hamann, 2012, for sources and detailed methodology). 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 & Wang, 2006) and classification tree analysis implemented with the random forest software package (Breiman, 2001) for the R programming environment (R Core Team, 2009); 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 Institute, 2008). 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.

2.3.2 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 ratio of species' presences over the total number of plot samples that fall within the ecosystem delineation. The 55,743 plot samples for western North America were compiled from the British Columbia Ministry of Forests (Hamann & Wang, 2006), Sustainable Resources Development of the Government of Alberta (Govt. of Alberta, 2004), and the United States Forest Service (Betchtold & Patterson, 2005). The database also contained 3,273 non-forested plot sites.

Predictor variables were interpolated climate data for the 1961–1990 reference period, generated at 1km resolution with a software package that is freely available (Mbogga *et al.*, 2009; Wang *et al.*, 2006). For a continental-scale modelling effort, we used 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 reality (Hamann & 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 6,000, 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.

2.3.3 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 Dyke (2005) and the North American Pollen Database (COHMAP, 1988). 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 (see Table 2.2 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 Core Team, 2009). 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 Institute, 2008), using the restricted maximum likelihood method (option *REML*).

2.4 Results

2.4.1 Independent model validation

We find that model accuracy substantially declines across all techniques and all species when subjected to independent validations (Figure 2.1). Mean AUC values across all model techniques are represented by vertical lines in Figure 2.1; median AUC values that are less influenced by outliers are 0.90, 0.78, and 0.75 for out-of-bag (random sub-sample), 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 2.2. 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 Tables 2.5-2.8. Standard errors of the mean AUC values represented as symbols in Figure 2.1 were on average 0.011, 0.028, and 0.024 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 0.053, 0.14, and 0.12, 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 2.3).

In Table 2.3, 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, see 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 (see Tables 2.5-2.8). 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 2.3).

2.4.2 Biogeographic and ecological characteristics

The influence of biogeographic and ecological characteristics of western North American trees species on model accuracy is summarised in Figure 2.2. 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 Figure 2.1 and described in Table 2.1. Standard deviations for the mean AUC values represented by symbols in Figure 2.2 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 Tables 2.5-2.8.

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 2.4), with interior species having generally lower accuracies than coastal species (Figure 2.2). 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 (Figure 2.2), 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 2.4). Mean AUC values declined from species that are 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 2.4). 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. pinyon pine, *Pinus edulis*) as well as at low elevations (e.g. Sitka spruce, *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 2.4).

2.4.3 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 2.3). 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 Douglas-fir (*Pseudotsuga meziesii*) (Figure 2.3). Here, we compared the ecosystem- and species-based modelling approach, relying on the same technique, to the random forest ensemble classifier. The ecosystem-based model run (Figure 2.3, bottom row) has better defined species absences in the two extrapolations, whereas the species-based models (Figure 2.3, top row) show large areas of high probability values well outside the species range. Another notable observation is that the 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 years 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.

2.5 Discussion

2.5.1 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 (Figure 2.1). In general terms, an AUC greater than 0.90 is considered excellent, between 0.80 and 0.90 good, between 0.70 to 0.80 fair, and less than 0.70 poor (e.g. 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 (e.g. Araújo *et al.*, 2005b; Dobrowski *et al.*, 2011; Rodriguez-Sánchez & Arroyo, 2008), while closely comparable or higher than others (e.g. 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 2,500 km, and temporally up to 21,000 years into the past.

The somewhat poorer performance of the 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 & Jackson, 2007). 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 Figure 2.3, 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.

2.5.2 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 (Figure 2.1, Table 2.3). 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 (e.g. Araújo *et al.*, 2005b). However, we should note that a recent study by Dobrowski et al. (2011), incorporating a broader range of trees and shrub species, did not a find strong relationship among independent and non-dependent evaluations.

Like others (e.g. Dobrowski *et al.*, 2011; Guisan *et al.*, 2007), we found that ecological and biogeographic traits had an influence on accuracy (Figure 2.2). 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 2.4), we did not find that any of these attributes favour certain modelling methods for trees (Figure 2.2). 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 (e.g. Araújo *et al.*, 2005a; 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 (e.g. 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.

2.5.3 Ecosystem-based models

Modelling approaches that incorporate community data or species assemblages have been implemented and evaluated before (see review by Ferrier & Guisan, 2006). These approaches use a species composition (a collection of species and their particular frequencies) as the dependent variable in the model (e.g. Baselga & Araújo, 2009; Elith *et al.*, 2006), 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 Arajo (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 years ago (Figure 2.3) 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 (e.g. 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 & 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, see Gray *et al.*, 2011; Gray & Hamann, 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.

2.6 References

Araújo, M. B., & Williams, P. H. 2000. Selecting areas for species persistence using occurrence data. *Biological Conservation*, **96**(3), 331–345.

Araújo, M. B., Whittaker, R. J., Ladle, R. J., & Erhard, M. 2005a. Reducing uncertainty in projections of extinction risk from climate change. *Global Ecology and Biogeography*, **14**(6), 529–538.

Araújo, M. B., Pearson, R. G., Thuiller, W., & Erhard, M. 2005b. Validation of species-climate impact models under climate change. *Global Change Biology*, **11**(9), 1504–1513.

Baselga, A., & Araújo, M. B. 2009. Individualistic vs community modelling of species distributions under climate change. *Ecography*, **32**(1), 55–65.

Beaumont, L. J., & Hughes, L. 2002. Potential changes in the distributions of latitudinally restricted Australian butterfly species in response to climate change. *Global Change Biology*, **8**(10), 954–971.

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

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. 2007. Forecasting the effects of global warming on biodiversity. *Bioscience*, **57**(3), 227–236.

Breiman, L. 2001. Random forests. *Machine Learning*, **45**(1), 5–32.

Burns, R. M., Honkala, H., & tech coords. 1990. *Silvics of North America: 1. Conifers; 2. Hardwoods.* U.S. Department of Agriculture, Forest Service.

COHMAP. 1988. Climatic Changes of the Last 18,000 Years - Observations and Model Simulations. *Science*, **241**(4869), 1043–1052.

De'ath, G., & Fabricius, K. E. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, **81**(11), 3178–3192.

Dobrowski, B., Thorne, J.H., Greenberg, J. A., Safford, H.D., Mynsberge, A. R., Crimmins, S. M., & Swanson, A.K. 2011. Modeling plant ranges over 75 years of climate change in California, USA: temporal transferability and species traits. *Ecological Monographs*, **81**(2), 241–257.

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

Elith, J., & 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**(1), 66–77.

Elith, J., & Leathwick, J. R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution and Systematics*, **40**, 677–697.

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberon, J., Williams, S., Wisz, M. S., & Zimmermann, N. E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**(2), 129–151.

Fawcett, T. 2006. An introduction to ROC analysis. *Pattern Recognition Letters*, **27**(8), 861–874.

Ferrier, S., & Guisan, A. 2006. Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, **43**(3), 393–404.

Fløjgaard, C., Normand, S., Skov, F., & Svenning, J. C. 2009. Ice age distributions of European small mammals: insights from species distribution modelling. *Journal of Biogeography*, **36**(6), 1152–1163.

Franklin, J. 2009. *Mapping Species Distributions.* Cambridge, UK: Cambridge University Press.

Giesecke, T., Hickler, T., Kunkel, T., Sykes, M. T., & Bradshaw, R. H. W. 2007. Towards an understanding of the Holocene distribution of Fagus sylvatica L. *Journal of Biogeography*, **34**(1), 118–131.

Govt. of Alberta. 2004. *Ecological Site Information System (ESIS).* Government of Alberta, Sustainable Resource Development (SRD), Available: http://www.srd.gov.ab.ca/lands/geographicinformation/resourcedataproductcatalogue/esis.aspx.

Gray, L. K., Gylander, T., Mbogga, M. S., Chen, P. Y., & Hamann, A. 2011. Assisted migration to address climate change: recommendations for aspen reforestation in western Canada. *Ecological Applications*, **21**(5), 1591–1603.

Gray, L.K., & Hamann, A. 2011. Strategies for reforestation under uncertain future climates: guidelines for Alberta, Canada. *PLoS One*, **6**(8), e22977. doi:10.1371/journal.pone.0022977.

Grenouillet, G., Buisson, L., Casajus, N., & Lek, S. 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. *Ecography*, **34**(1), 9–17.

Guisan, A., & Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**(2-3), 147–186.

Guisan, A., Zimmermann, N. E., Elith, J., Graham, C. H., Phillips, S., & Peterson, A. T. 2007. What matters for predicting the occurrences of trees: Techniques, data, or species' characteristics? *Ecological Monographs*, **77**(4), 615–630.

Hamann, A., & Wang, T. L. 2005. Models of climatic normals for genecology and climate change studies in British Columbia. *Agricultural and Forest Meteorology*, **128**(3-4), 211–221.

Hamann, A., & Wang, T. L. 2006. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. *Ecology*, **87**(11), 2773–2786.

Hamann, A., Gylander, T., & Chen, P. Y. 2011. Developing seed zones and transfer guidelines with multivariate regression trees. *Tree Genetics & Genomes*, **7**(2), 399–408.

Hogg, E. H. 1997. Temporal scaling of moisture and the forest-grassland boundary in western Canada. *Agricultural and Forest Meteorology*, **84**(1-2), 115–122.

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**(6-7), 473–506.

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., & Taylor, P. 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series*, **321**, 267–281.

Li, P., & Adams, W. T. 1989. Range-Wide Patterns of Allozyme Variation in Douglas-Fir (Pseudotsuga-Menziesii). *Canadian Journal of Forest Research/Revue Canadienne de Recherche Forestiere*, **19**(2), 149–161.

Lobo, J. M., Jimenez-Valverde, A., & Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, **17**(2), 145–151.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. 2009. Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, **15**(1), 59–69.

Martinez-Meyer, E., & Peterson, A. T. 2006. Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *Journal of Biogeography*, **33**(10), 1779–1789.

Martinez-Meyer, E., Townsend Peterson, A., & Hargrove, W. W. 2004. Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. *Global Ecology and Biogeography*, **13**(4), 305–314.

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**(5), 881–890.

Mbogga, M. S., Wang, X. L., & Hamann, A. 2010. Bioclimate envelope model predictions for natural resource management: dealing with uncertainty. *Journal of Applied Ecology*, **47**(4), 731–740.

Morueta-Holme, N., Flojgaard, C., & Svenning, J. C. 2010. Climate Change Risks and Conservation Implications for a Threatened Small-Range Mammal Species. *PLoS One*, **5**(4), –.

Muller, M. P., McGeer, A. J., Hassan, K., Marshall, J., Christian, M., & Dis, Toronto Invasive Bacterial. 2010. Evaluation of Pneumonia Severity and Acute Physiology Scores to Predict ICU Admission and Mortality in Patients Hospitalized for Influenza. *PLoS One*, **5**(3), –.

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**(12), 1071–1074.

Pearman, P. B., Randin, C. F., Broennimann, O., Vittoz, P., van der Knaap, W. O., Engler, R., Le Lay, G., Zimmermann, N. E., & Guisan, A. 2008. Prediction of plant species distributions across six millennia. *Ecology Letters*, **11**(4), 357–369.

Prasad, A. M., Iverson, L. R., & Liaw, A. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems*, **9**(2), 181–

199.

R Core Team. 2009. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Randin, C. F., Dirnbock, T., Dullinger, S., Zimmermann, N. E., Zappa, M., & Guisan, A. 2006. Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**(10), 1689–1703.

Roberts, D. R., & Hamann, A. 2012. Predicting potential climate change impacts with bioclimate envelope models: a palaeoecological perspective. *Global Ecology and Biogeography*, **21**(2), 121–133.

Rodriguez-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**(6), 685–695.

SAS Institute. 2008. *SAS/STAT 9.2 User's Guide.* SAS Institute Inc., Cary, NC.

Segurado, P., & Araújo, M. B. 2004. An evaluation of methods for modelling species distributions. *Journal of Biogeography*, **31**(10), 1555–1568.

Segurado, P., Arajo, M. B., & Kunin, W. E. 2006. Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology*, **43**(3), 433–444.

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. 2005. ROCR: visualizing classifier performance in R. *Bioinformatics*, **21**(20), 3940–3941.

Thuiller, W. 2003. BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, **9**(10), 1353–1362.

Thuiller, W., Albert, C., Araújo, M. B., Berry, P. M., Cabeza, M., Guisan, A., Hickler, T., Midgely, G. F., Paterson, J., Schurr, F. M., Sykes, M. T., & Zimmermann, N. E. 2008. Predicting global change impacts on plant species' distributions: Future challenges. *Perspectives in Plant Ecology Evolution and Systematics*, **9**(3-4), 137–152.

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. 2009. BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**(3), 369–373.

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**(3), 383–397.

Williams, J. W., & Jackson, S. T. 2007. Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, **5**(9), 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**(12), 3346–3362.

Zarnetske, P. L., Edwards, T. C., & Moisen, G. G. 2007. Habitat classification modeling with incomplete data: Pushing the habitat envelope. *Ecological Applications*, **17**(6), 1714–1726.

Table 2.1: Descriptions of each species distribution modelling technique evaluated in the study, grouped by similar methodological approaches (based on groups described by Elith and Leathwick (2009) and Franklin (2009).

Method	Description
Ecosystem-based methods	
Discriminant analysis (DAC)	A standard multivariate approach to classification, using an ecosystem class as dependent class variable and climate data as predictor variables (Hamann & Wang, 2006).
Minimum distance (MDC)	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 (RFC)	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 (DAS)	A standard multivariate approach to classification, using a binary response variable (presence or absence).
Surface range envelopes (SRE)	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 & Hughes, 2002).
Machine learning techniques	
Artificial neural networks (ANN)	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 & Araújo, 2004).
Classification tree analysis (CTA)	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 & Fabricius, 2000).
Generalised boosting model (GBM)	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 (RFS)	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).

Continued on next page

Table 2.1 – continued from previous page

Method	Description
Regression-based techniques	
Generalised additive model (GAM)	Generalised linear models for individual predictor variables are combined additively, using smoothing equations to generalise the data and fit to local data subsets (Guisan & Zimmermann, 2000).
Generalised linear model (GLM)	An extension of the general linear model for binomial data capable of capturing non-linear relationships (Guisan & Zimmermann, 2000).
Multivariate adaptive regression splines (MRS)	Fits splines to distinct but unequal intervals of the predictors before pruning excess spline connections through a stepwise analysis (Prasad *et al.*, 2006).

Table 2.2: 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 (N plots) as well as the number of presences for each species (N validation) 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 6,000, 11,000, 14,000, 16,000, and 21,000 years 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	N plots	N validation			Range Size (10^3 km^2)	Distribution			Shade Tolerance
			OOB	N2S	TMP		Type	Range	Elevation	
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	1,957	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	10,251	Restricted	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

Continued on next page

Table 2.2 – continued from previous page

Species Name	Taxon/Group	N plots	N validation			Range Size (10^3 km^2)	Distribution			Shade Tolerance
			OOB	N2S	TMP		Type	Range	Elevation	
Picea engelmannii (Engelman spruce)	Piceoideae	6223	1994	1733	44	1,002	Widespread	Interior	Unrestricted	Tolerant
Picea glauca (white spruce)	Piceoideae	7115	2398	22	55	10,320	Widespread	Interior	Intermediate	Intermediate
Picea mariana (black spruce)	Piceoideae	2922	1005	0*	48	10,446	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	2,458	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	11,481	Widespread	Interior	Unrestricted	Very intolerant
Pseudotsuga menziesii (Douglas-Fir)	Laricoideae	8808	2992	4438	174	1,445	Widespread	Both	Unrestricted	Intermediate
Sequoia sempervirens (coast redwood)	Cupressaceae	90	32	90*	0*	14	Restricted	Coastal	Highly restricted	Intolerant

Continued on next page

Table 2.2 – continued from previous page

Species Name	Taxon/Group	N plots	N validation			Range Size (10^3 km^2)	Distribution			Shade Tolerance
			OOB	N2S	TMP		Type	Range	Elevation	
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

Table 2.3: 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.

	Out-of-bag (OOB)			North-to-south (N2S)			Past periods (TMP)			Method rank		
Modelling method	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 & Eco removed)	0.95	0.92	0.59	0.89	0.86	0.58	0.79	0.78	0.52	-	-	-
Median (SRE & Eco removed)	0.95	0.74	0.81	0.90	0.82	0.56	0.80	0.53	0.78	-	-	-

Table 2.4: 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 2.2).

Component	Variance Explained
Modelling Method	0.16
Distribution Range	0.15
Shade Tolerance	0.07
Taxon/Group	0.06
Elevation Range	0.03
Distribution Type	0.00
Total Explained	0.46
Error Variance	0.54

Table 2.5: 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 italics. Model abbreviations correspond to those list in Table 2.1. Ensemble methods include a Kappa-weighted (K), arithmetic mean ($\bar{X}$), median (Med), receiver operatic characteristic-weighted (ROC), true skill statistic-weighted (TSS), and a weighted mean ($\bar{X}$w). Two ensembles without the SRE model were also evaluated: arithmetic mean ($\bar{X}$) and a median (Med).

	Individual methods										Emsembles							
	Species-based										All Methods						No SRE	
Species	CTA	DAS	GAM	GBM	GLM	MRS	ANN	RFS	SRE	Avg.	K	$\bar{X}$	Med	ROC	TSS	$\bar{X}w$	$\bar{x}$	Med.
A. amabilis	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	0.90
A. lasiocarpa	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	0.90
A. procera	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	0.90
A. macrophyllum	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	0.90
A. rubra	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	0.90
B. papyrifera	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	0.90
C. decurrens	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	0.90
C. nootkatensis	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	0.90
L. occidentalis	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	0.90
P. engelmannii	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	0.90
P. glauca	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	0.90
P. mariana	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	0.90
P. sitchensis	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	0.90
P. albicaulis	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	0.90
P. contorta	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	0.90
P. edulis	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	0.90
P. monticola	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	0.90
P. ponderosa	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	0.90
P. tremuloides	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	0.90

Continued on next page

Table 2.5 – continued from previous page

	Individual methods										Emsembles							
	Species-based										All Methods						No SRE	
Species	CTA	DAS	GAM	GBM	GLM	MRS	ANN	RFS	SRE	Avg.	K	$\bar{X}$	Med	ROC	TSS	$\bar{X}w$	$\bar{x}$	Med.
P. menziesii	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	0.90
S. sempervirens	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	0.90
T. plicata	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	0.90
T. heterophylla	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	0.90
T. mertensiana	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	0.90
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 2.6: 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 italics. Model abbreviations correspond to those list in Table 2.1. Ensemble methods include a Kappa-weighted (K), arithmetic mean ($\bar{X}$), median (Med), receiver operatic characteristic-weighted (ROC), true skill statistic-weighted (TSS), and a weighted mean ($\bar{X}$w). Two ensembles without the SRE model were also evaluated: arithmetic mean ($\bar{X}$) and a median (Med).

	Individual methods													Ensembles								
	Ecosys-based			Species-based										All methods							No SRE	
Species	DAC	MDC	RFC	CTA	DAS	GAM	GBM	GLM	MRS	ANN	RFS	SRE	*Avg*	$\bar{X}_E$	K	$\bar{X}_S$	Med	ROC	TSS	$\bar{X}_W$	$\bar{X}$	Med
A. amabilis	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
A. lasiocarpa	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
A. procera	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
A. macrophyllum	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
A. rubra	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
B. papyrifera	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
C. decurrens	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
C. nootkatensis	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
L. occidentalis	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
P. engelmannii	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
P. glauca	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
P. mariana	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
P. sitchensis	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
P. albicaulis	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
P. contorta	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
P. edulis	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
P. monticola	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
P. ponderosa	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
P. tremuloides	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

Continued on next page

Table 2.6 – continued from previous page

	Individual methods													Ensembles								
	Ecosys-based			Species-based										All methods							No SRE	
Species	DAC	MDC	RFC	CTA	DAS	GAM	GBM	GLM	MRS	ANN	RFS	SRE	*Avg*	$\bar{X}_E$	K	$\bar{X}_S$	Med	ROC	TSS	$\bar{X}_W$	$\bar{X}$	Med
P. menziesii	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
S. sempervirens	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
T. plicata	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
T. heterophylla	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
T. mertensiana	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 2.7: 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 italics. Species without records in both the north and south data sets are denoted with hyphens and were excluded from the analysis. Model abbreviations correspond to those list in Table 2.1. Ensemble methods include a Kappa-weighted (K), arithmetic mean ($\bar{X}$), median (Med), receiver operatic characteristic-weighted (ROC), true skill statistic-weighted (TSS), and a weighted mean ($\bar{X}$w). Two ensembles without the SRE model were also evaluated: arithmetic mean ($\bar{X}$) and a median (Med).

	Individual methods													Ensembles								
	Ecosys-based			Species-based										All methods							No SRE	
Species	DAC	MDC	RFC	CTA	DAS	GAM	GBM	GLM	MRS	ANN	RFS	SRE	*Avg*	$\bar{X}_E$	K	$\bar{X}_S$	Med	ROC	TSS	$\bar{X}_W$	$\bar{X}$	Med
A. amabilis	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
A. lasiocarpa	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
A. procera	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
A. macrophyllum	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
A. rubra	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
B. papyrifera	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
C. decurrens	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
C. nootkatensis	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
L. occidentalis	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
P. engelmannii	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
P. glauca	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
P. mariana	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
P. sitchensis	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
P. albicaulis	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
P. contorta	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
P. edulis	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
P. monticola	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
P. ponderosa	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

Continued on next page

Table 2.7 – continued from previous page

	Individual methods													Ensembles								
	Ecosys-based			Species-based										All methods							No SRE	
Species	DAC	MDC	RFC	CTA	DAS	GAM	GBM	GLM	MRS	ANN	RFS	SRE	*Avg*	$\bar{X}_E$	K	$\bar{X}_S$	Med	ROC	TSS	$\bar{X}_W$	$\bar{X}$	Med
P. tremuloides	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
P. menziesii	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
S. sempervirens	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
T. plicata	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
T. heterophylla	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
T. mertensiana	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 2.8: AUC values by species and modelling method for the fossil/pollen temporal evaluation for all records from 6,000, 11,000, 14,000, 16,000, and 21,000 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 a hyphen and were excluded from the analysis. Model abbreviations correspond to those list in Table 2.1. Ensemble methods include a Kappa-weighted (K), arithmetic mean ($\bar{X}$), median (Med), receiver operatic characteristic-weighted (ROC), true skill statistic-weighted (TSS), and a weighted mean ($\bar{X}$w). Two ensembles without the SRE model were also evaluated: arithmetic mean ($\bar{X}$) and a median (Med).

	Individual methods													Ensembles								
	Ecosys-based			Species-based										All methods							No SRE	
Species	DAC	MDC	RFC	CTA	DAS	GAM	GBM	GLM	MRS	ANN	RFS	SRE	*Avg*	$\bar{X}_E$	K	$\bar{X}_S$	Med	ROC	TSS	$\bar{X}_W$	$\bar{X}$	Med
A. amabilis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A. lasiocarpa	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
A. procera	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A. macrophyllum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A. rubra	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
B. papyrifera	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
C. decurrens	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C. nootkatensis	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
L. occidentalis	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P. engelmannii	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
P. glauca	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
P. mariana	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
P. sitchensis	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
P. albicaulis	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
P. contorta	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
P. edulis	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
P. monticola	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
P. ponderosa	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

Continued on next page

Table 2.8 – continued from previous page

	Individual methods													Ensembles								
	Ecosys-based			Species-based										All methods							No SRE	
Species	DAC	MDC	RFC	CTA	DAS	GAM	GBM	GLM	MRS	ANN	RFS	SRE	*Avg*	$\bar{X}_E$	K	$\bar{X}_S$	Med	ROC	TSS	$\bar{X}_W$	$\bar{X}$	Med
P. tremuloides	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
P. menziesii	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
S. sempervirens	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
T. plicata	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
T. heterophylla	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
T. mertensiana	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

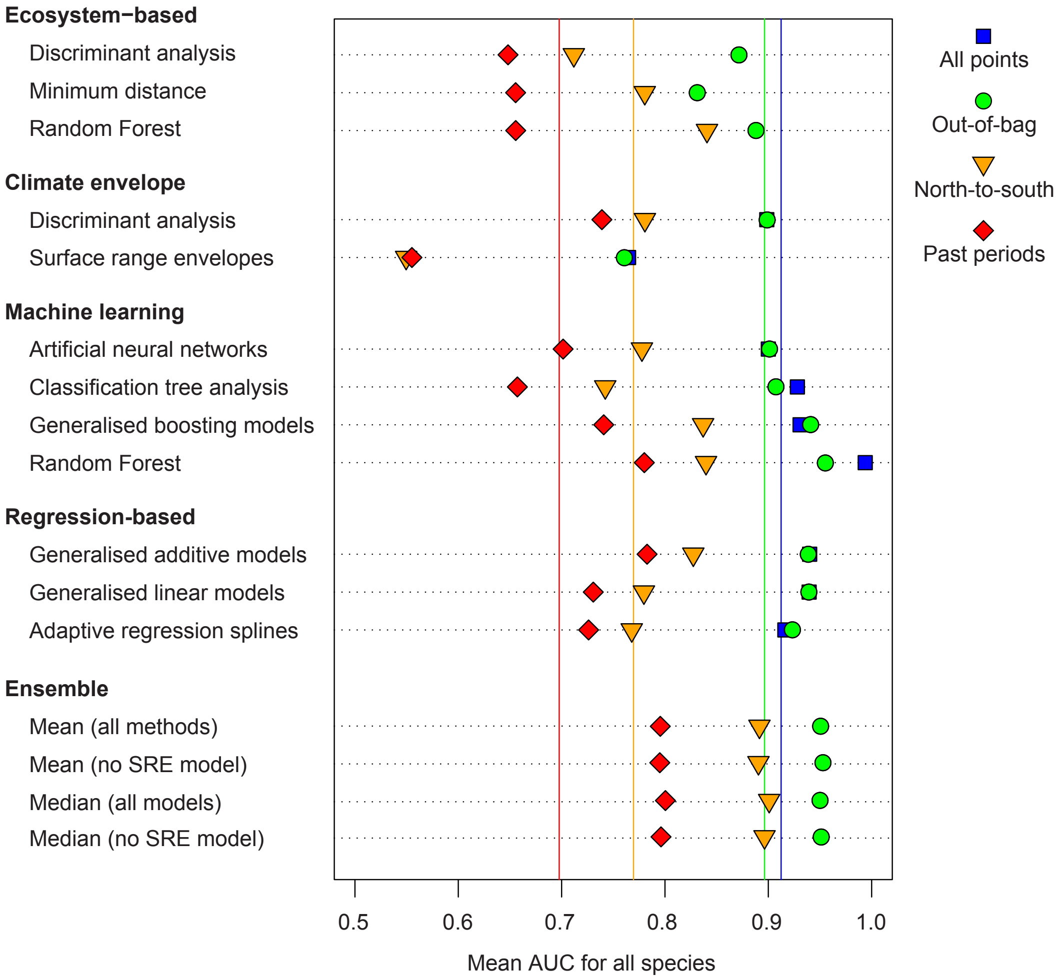

Figure 2.1: Model accuracy for 12 individual and 4 simple ensemble techniques, evaluated by the area under the curve (AUC) of the receiver operating characteristic. Non-independent evaluations include training and validation data being the same (All points), a random data split of 67% for training and 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 6,000, 11,000, 14,000, 16,000, and 21,000 years before present for validation (Past periods). Vertical lines represent the mean AUC value across all species and methods (excluding ensembles).

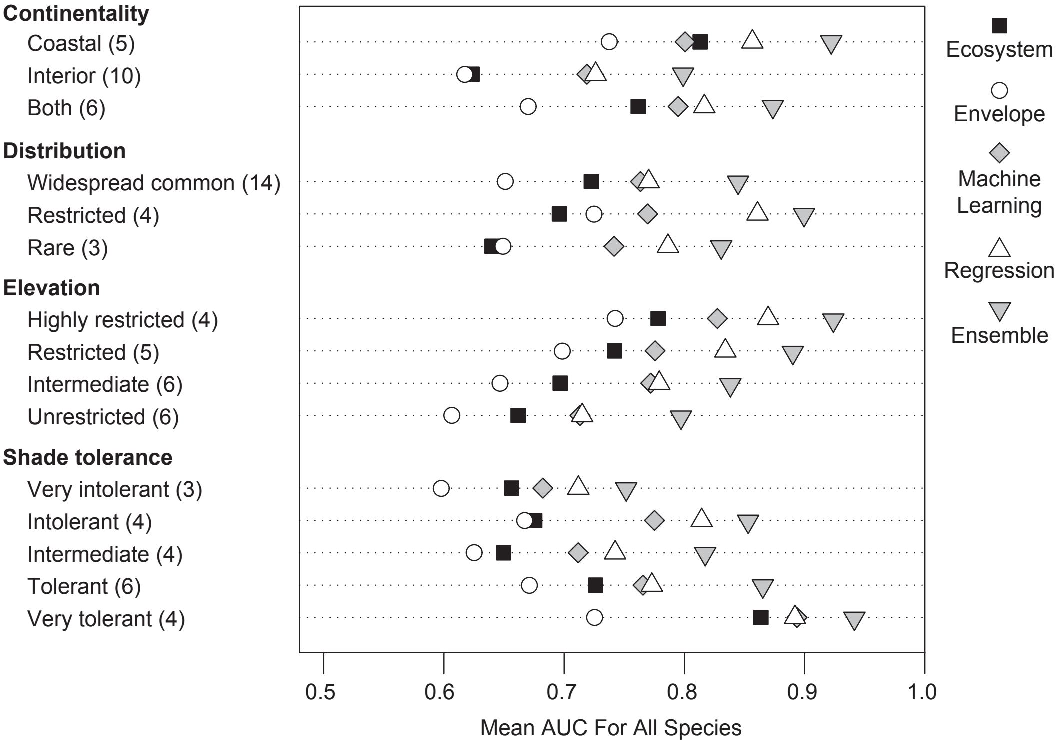

Figure 2.2: Model accuracy as a function of biogeographic and ecological characteristics of species, evaluated by the area under the curve (AUC) of the receiver operating characteristic. The values represent an average of the two independent validations (regional and temporal extrapolations), and are aggregated by the five model categories used in Figure 2.1 and explained in Table 2.1. The number of species in each category is noted in parentheses (Table 2.2). Individual AUC values for all species, methods, and validation techniques are provided in Tables 2.5-2.8.

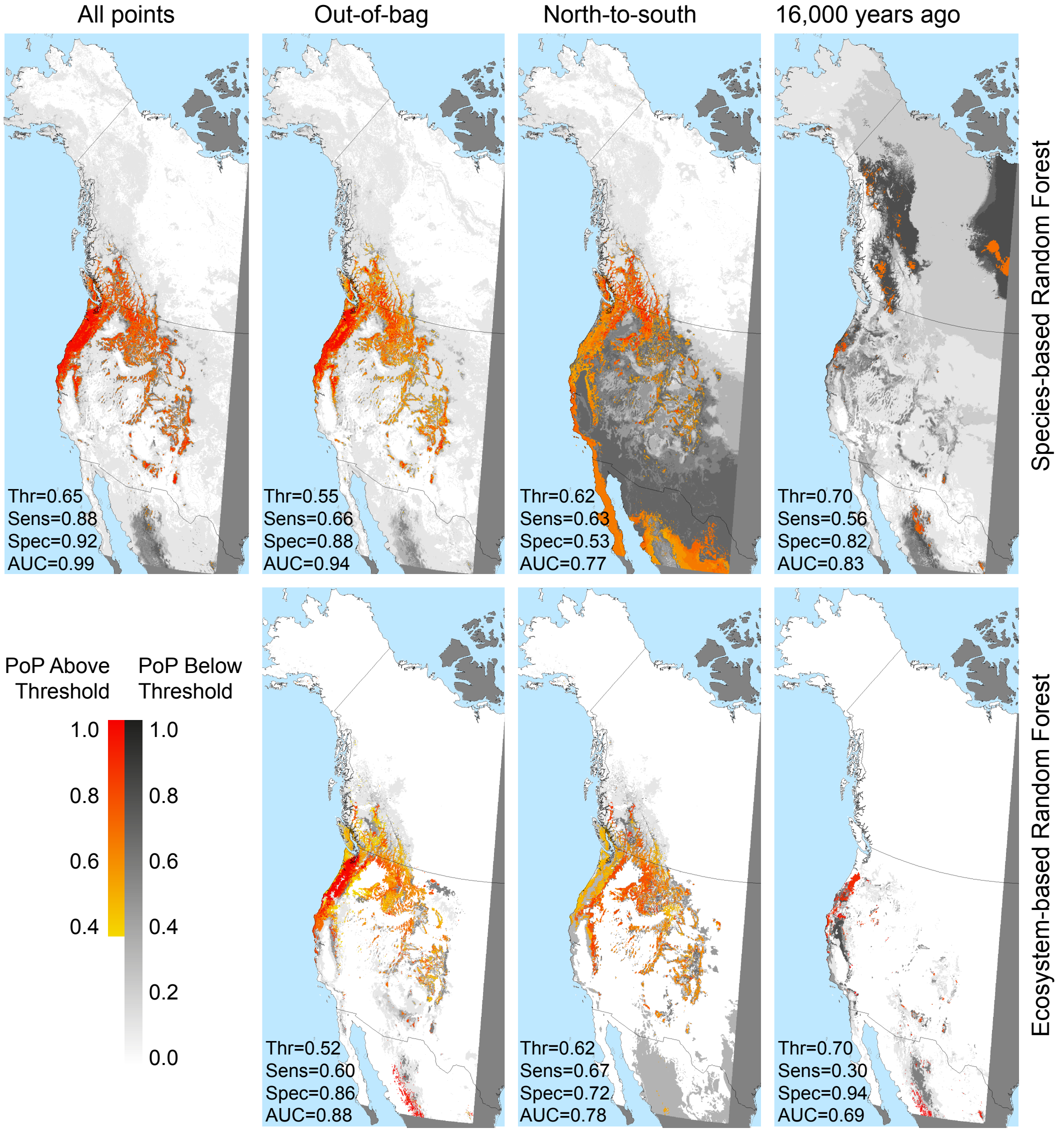

Figure 2.3: Projected probability of presence (PoP) for Douglas-fir (*Pseudotsuga meziesii*) using the species- and ecosystem-based Random Forest ensemble classifier under four model training and validation scenarios: training and validation based on the entire dataset (all points), a random data split 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.

Chapter 3

Predicting potential climate change impacts with bioclimate envelope models: a palaeoecological perspective.[1]

3.1 Summary

We assess the realism of species distribution 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 species distribution model projections biased towards under-predicting forested ecosystems? (3) given a palaeoecological perspective, are potential habitat projections for the 21^{st} century within model capabilities? We used an ensemble classifier modelling approach to spatially project the climate space of modern ecosystem classes throughout the Holocene (at 6,000, 9,000, 11,000, 14,000, 16,000, and 21,000 years ago) 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 1,460 palaeoecological records. Model accuracy assessed against independent palaeoecology data is generally low for the present day, increases until 6,000 years ago, 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

---

[1] A version of this chapter has been accepted for publication. Roberts, D. R. and A. Hamann. 2012. Predicting potential climate change impacts with bioclimate envelope models: a palaeoecological perspective. *Global Ecology and Biogeography* **21**, 121-133.

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. Species distribution 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.

3.2 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; Salzmann *et al.*, 2009; Williams *et al.*, 2007). Such "no-analogue" climates could result in ecological communities that also lack modern analogues (Williams *et al.*, 2001; Overpeck *et al.*, 1992). 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 (Fitzpatrick & Hargrove, 2009; Williams & Jackson, 2007; Jackson & Williams, 2004; Van der Wal *et al.*, 2009).

A widely used class of models to predict potential species habitat under projected climate changes are species distribution models (also referred to as bioclimate envelope models, or niche 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 species distribution 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 Araújo & Guisan, 2006; Guisan & Thuiller, 2005; Pearson & Dawson, 2003).

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 species distribution 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.*, 2007). Back-predictions can also be used to test the validity of various assumptions underlying species distribution model projections (Araújo *et al.*, 2005; Botkin *et al.*, 2007; Nogues-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. Rodriguez-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 (Carstens & Richards, 2007; Yesson & Culham, 2006). In this paper, we contribute a new approach that projects ecosystem climate envelopes to more generally assess species distribution 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 species distribution 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 (Chen *et al.*, 2010; Loehle & LeBlanc, 1996; Morin & Thuiller, 2009; Morin *et al.*, 2008).

Here, we carry out species distribution model-based back-predictions of ecosystems between the present and the last glacial maximum, for the periods 6,000, 9,000, 11,000, 14,000, 16,000, and 21,000 years ago. We evaluate the results with biome reconstructions based on fossil and pollen data at 1,460 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 to provide a palaeoecological perspective on whether ecosystem and species habitat projections for the 21^{st} century are generally within model capabilities.

3.3 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 species distribution 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 3.1). 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 3.1).

3.3.1 Species distribution 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 1 km 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 *Random Forest* package v.4.5 (Breiman, 2001) for the open-source R programming environment (R Core Team, 2009). Random Forest has been shown to be a robust ensemble classifier and a useful technique for species distribution modelling (e.g. Lawler *et al.*, 2006) and was consistently the best individual ecosystem-based method in our previous evaluations (Roberts & Hamann, 2012).

3.3.2 Past and future climate data

For past and future climatic characterisation we used general circulation model (GCM) projections overlaid as anomalies (deviation from the 1961-1990 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 (GFDL) at the National Oceanic and Atmospheric Administration (NOAA) (Bush & Philander, 1999) for 6,000, 9,000, 16,000, and 21,000 years ago. The second is the Community Climate Model version 1 (CCM1) developed by the National Center for Atmospheric Research (NCAR) (Kutzbach *et al.*, 1998) for 6,000, 11,000, 14,000, 16,000, and 21,000 years ago. For future climate projections we used individual and ensemble projections for four main emission scenario families A1FI, B1, A2, and B2 (Nakicenovic *et al.*, 2000), 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 is freely available (Wang *et al.*, 2006; Mbogga *et al.*, 2009) , which uses 1961-1990 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.

3.3.3 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 1,014 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 7.7x10^9 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 identify 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 (SAS Institute, 2007).

For model evaluation, we used palaeoecological data comprised of fossil pollen and plant macrofossils (compiled by Dyke, 2005; Thompson & Anderson, 2000). Duplicates as well as mammal records were removed (to retain purely vegetation-based data) for a total of 1,460 sites used in this analysis. Modern classifications from the last 1,000 years were available for most sites. Approximately 500 sites had records for 6,000 years ago, which declined to 300 sites for 9,000 years ago. For time periods approaching the last glacial maximum, records become fairly scarce with 150 sites for 16,000 years ago and 100 sites for 21,000 years ago 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 tables of classification and mis-classification (i.e. 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 the software package described above, rather than using a classification made for a nearby 1 km grid cell.

3.4 Results

3.4.1 Independent model evaluation

The model outputs for the present, based on the modern day 1961-1990 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) (Figure 3.1). Misclassification error rates of predicted biomes against independent fossil and pollen data are shown in Table 3.2. 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 Figure 3.1). 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 3.2).

For conciseness, we do not report full misclassification matrices for 14 biomes for ecosystem predictions based on past 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 3.3). Misclassifications increase abruptly for pollen and fossil sites that are classified as Forested and Dry during the cooler early-Holocene between 9,000 to 14,000 years ago (Table 3.3, Figure 3.2). In contrast, near the last glacial maximum (16,000 to 21,000 years ago), pollen and fossil points representing Cold ecosystems are almost always predicted correctly, indicating an increasing bias towards under-predicting Forested and Dry ecosystem (Figure 3.2). This is reflected in Table 3.3, 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 6,000 years ago 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 (Figure 3.2), which holds true for both CCM1 and GFDL based predictions.

3.4.2 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 3.4, Figure 3.3). For example, high climate dissimilarities emerge in the United States Rocky Mountains during the mid-Holocene warm period at 6,000 years ago. These climates are characterised by drier, cooler summers and warmer winters, 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 species distribution model. Climates without modern equivalents also appear in the area immediately south of the ice sheets at 21,000 years ago (Figure 3.3). 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 species distribution 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 3.4). 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 3.4). 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 species distribution 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 (Figure 3.3). This also holds true for predictions based on the GFDL general circulation model (data not shown).

3.4.3 Future projections

Climate projections for future periods result in dissimilarities roughly on par with those observed for the 6,000 to 11,000 years ago back-predictions (Table 3.4). 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 (Figure 3.4). 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 3.4). 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 3.4). The intermediate scenario that assumes slow population growth and regionally fragmented economic growth (A2) has larger climate dissimilarity values equivalent to 6,000 to 9,000 years ago, which still do not imply very high misclassification rates due to no-analogue climates (Figure 3.2).

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) (Figures 3.1 and 3.4). Alaska gains landscape level diversity of habitat conditions, comparable to British Columbia at present (Figure 3.4, 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 Figure 3.4 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).

3.5 Discussion

3.5.1 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 species distribution models (Fitzpatrick & Hargrove, 2009; Williams & Jackson, 2007; Jackson & Williams, 2004; 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 3.4). 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 anomalies with no modern analogue emerge (Figure 3.4). 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 and we therefore conclude that their projections should not be generally compromised by extrapolating into no-analogue climate space. Conversely it is clear that regional-scale species distribution 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.

3.5.2 Violation of species distribution model assumptions

In addition to misclassifications, we also showed bias in species distribution 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, models based on 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.

Misclassifications due to no-analogue climates should not introduce bias since there is an equal probability of misclassifications into all classes (in this 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 species distribution 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 is 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* suggest that these genera 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-hour 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 species distribution 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, they are a cause for concern, as climatic stresses on locally adapted populations may compromise forest productivity and forest health (Allen *et al.*, 2010).

3.6 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**(4), 660–684.

Araújo, M. B., & Guisan, A. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**(10), 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**(9), 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. *In:* Christie, R. L., & McMillan, N. J. (eds), *Tertiary fossil forests of the Geodetic Hills, Axel Heiberg Island, Arctic Archipelago.* Geologic Survey of Canada, Bulletin 403.

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. 2007. Forecasting the effects of global warming on biodiversity. *Bioscience*, **57**(3), 227–236.

Box, G. E. P., & Draper, N. R. 1987. *Empirical Model-Building and Response Surfaces.* Wiley Series in Probability and Statistics. New York: John Wiley & Sons.

Breiman, L. 2001. Random forests. *Machine Learning*, **45**(1), 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**(D20), 24509–24525.

Carstens, B. C., & Richards, C. L. 2007. Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**(6), 1439–1454.

Chen, P. Y., Welsh, C., & Hamann, A. 2010. Geographic variation in growth response of Douglas-fir to interannual climate variability and projected climate change. *Global Change Biology*, **16**(12), 3374–3385.

Cowling, S. A. 1999. Simulated effects of low atmospheric CO2 on structure and composition of North American vegetation at the Last Glacial Maximum. *Global Ecology and Biogeography*, **8**(2), 81–93.

Crowley, T. J. 1990. Are there any satisfactory geologic analogs for a future greenhouse warming? *Journal of Climate*, **3**(11), 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**(15), 2031–2064.

Davis, M. B., & Shaw, R. G. 2001. Range shifts and adaptive responses to Quaternary climate change. *Science*, **292**(5517), 673–679.

Dyke, A. S. 2005. Late Quaternary vegetation history of northern North America based on pollen, macrofossil, and faunal remains. *Géographie 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**(2), 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**(8), 2255–2261.

Govt. of Alberta. 2005. *Seed Zones of Alberta, Digital Vector Data in ARC/INFO format.* Government of Alberta, Sustainable Resource Development (SRD), Available: http://srd.alberta.ca/lands/geographicinformation/resourcedataproductcatalogue/ seedzonesalberta.aspx.

Govt. of Canada. 1999. *A National Ecological Framework for Canada.* Government of Canada, Agriculture and Agri-Food Canada, Available: 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**(9), 993–1009.

Guisan, A., & Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**(2-3), 147–186.

Hamann, A., & Wang, T. L. 2006. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. *Ecology*, **87**(11), 2773–2786.

Hamrick, J. L. 2004. Response of forest trees to global environmental changes. *Forest Ecology and Management*, **197**(1-3), 323–335.

Hogg, E. H. 1997. Temporal scaling of moisture and the forest-grassland boundary in western Canada. *Agricultural and Forest Meteorology*, **84**(1-2), 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**(4), 337–341.

IPCC. 2007. Synthesis Report. *Page 104 of: Climate Change 2007: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.*

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: http://agdc.usgs.gov/data/usgs/erosafo/ecosys/metadata/ecosys.html.

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, Available: http://ceos.cnes.fr:8100/cdrom-00b/ceos1/casestud/ecoreg/datasets/b13/ek.htm.

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: http://www.ngdc.noaa.gov/ecosys/ged.shtml.

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**(6-7), 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**(8), 1568–1584.

Loehle, C., & LeBlanc, D. 1996. Model-based assessments of climate change effects on forests: A critical review. *Ecological Modelling*, **90**(1), 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**(5), 881–890.

Mbogga, M. S., Wang, X. L., & Hamann, A. 2010. Bioclimate envelope model predictions for natural resource management: dealing with uncertainty. *Journal of Applied Ecology*, **47**(4), 731–740.

Mitchell, T.D., Carter, T.R., & Jones, P. 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 for Climate Change Research.

Morin, X., & Thuiller, W. 2009. Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology*, **90**(5), 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**(4), 784–794.

Nakicenovic, N., Alcamo, J., Davis, G., de Vries, B., Fenhann, J., Gaffin, S., Gregory, K., Grbler, A., Jung, T.Y., Kram, T., La Rovere, E.L., Michaelis, L., Mori, S., Morita, T., Pepper, W., Pitcher, H., Price, L., Raihi, K., Roehrl, A., Rogner, H.-H., Sankovski, A., Schlesinger, M., Shukla, P., Smith, S., Swart, R., van Rooijen, S., Victor, N., & Dad, Z. 2000. *Emissions Scenarios. A Special Report of Working Group III of the Intergovernmental Panel on Climate Change.* Cambridge, U.K.: Cambridge University Press.

Nogues-Bravo, D. 2009. Predicting the past distribution of species climatic niches. *Global Ecology and Biogeography*, **18**(5), 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: http://www.epa.gov/wed/pages/ecoregions/level_iv.htm.

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**(12), 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**(5), 361–371.

Pojar, J., & Meidinger, D. V. 1991. *Ecosystems of British Columbia.* B.C. Ministry of Forests, Special Report Series 6, Available: http://www.for.gov.bc.ca/hfd/pubs/Docs/Srs/Srs06.htm.

Prentice, I. C., Bartlein, P. J., & Webb, T. 1991. Vegetation and Climate Change in Eastern North-America since the Last Glacial Maximum. *Ecology*, **72**(6), 2038–2056.

R Core Team. 2009. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Roberts, D. R., & Hamann, A. 2012. Method selection for species distribution modelling: are temporally or spatially independent evaluations necessary? *Ecography*, **35**, 792–802.

Rodriguez-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**(6), 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**(1886), 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 Institute 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**(3), 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**(6), 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**(3), 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**(2), 291–301.

Verdin, K. L., & Greenlee, S. K. 1996. Development of continental scale digital elevation models and extraction of hydrographic features. *In: Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling.*

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**(3), 383–397.

Williams, J. W., & Jackson, S. T. 2007. Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, **5**(9), 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**(12), 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 of the United States of America*, **104**(14), 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**(6), 549–562.

Yesson, C., & Culham, A. 2006. Phyloclimatic modeling: combining phylogenetics and bioclimatic modeling. *Systematic Biology*, **55**(5), 785–802.

Table 3.1: 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 the 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 (ARC)	High-latitude tundra, largely devoid of trees and dominated by shrubs and lichens.
Glacier & ice (ICE)	Climate conditions favourable for year-round ice coverage.

Table 3.2: Misclassification rates between biomes predicted with climate envelope models for the 1961-1990 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 the total number (N) of fossil and pollen points correctly classified (MR). Cohen's Kappa (K) statistic representing correct classifications minus randomly expected matches is also reported.

	Predicted																
	Dry			Forested									Cold				
Observed	DES	STE	GRA	SAV	DPK	DCO	SBM	CDM	WTF	SAF	BOR	BSA	ALT	ARC	N	MR	K
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

Table 3.3: Misclassifications between biome groups inferred from fossil and pollen samples and biome groups independently predicted with species distribution modelling for the same periods (noted in calendar years before the present, YBP), 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**			
	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**

Table 3.4: 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	

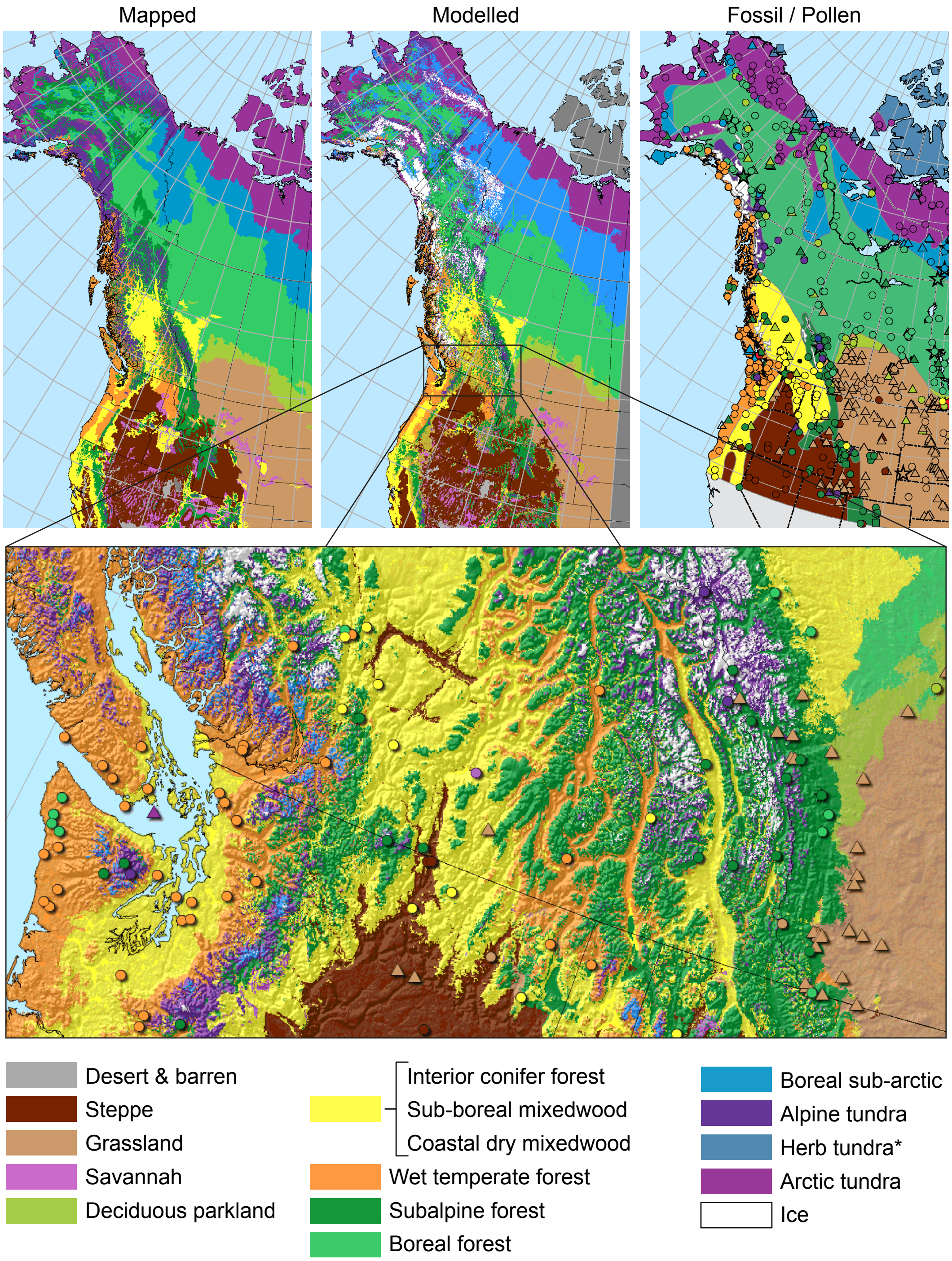

Figure 3.1: Present day mapped biomes, modelled biomes, and biome classes inferred from pollen (circles), macrofossil (stars) and mammal (triangles) records according to Dyke (2005), reproduced with permission. We did not model Dyke's herb tundra biome but we distinguish three types of interior forest (all shown in 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 were not used in the model evaluation calculations.

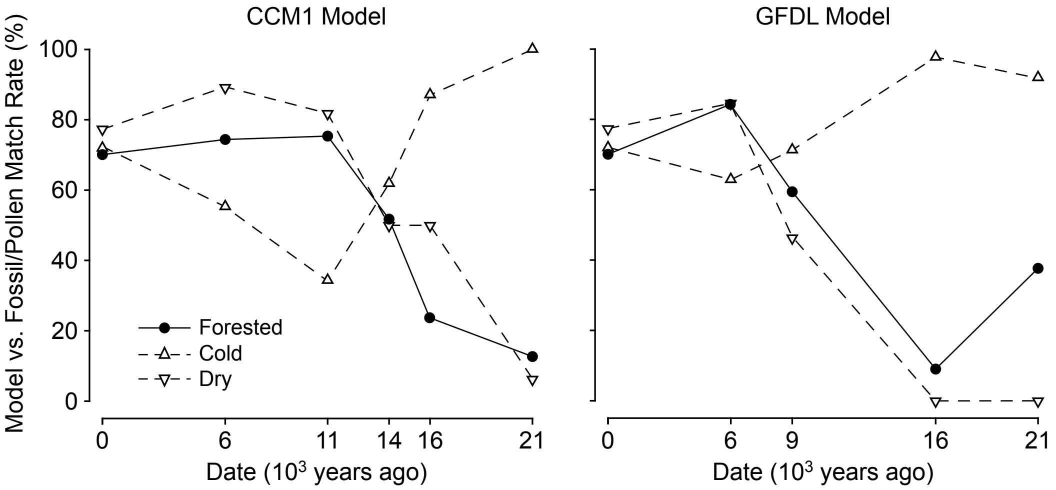

Figure 3.2: Match rates marked as bold in Table 4 expressed as percentage and plotted over time for paleoclimate projections of two general circulation models CCM1 (left) and GFDL (right).

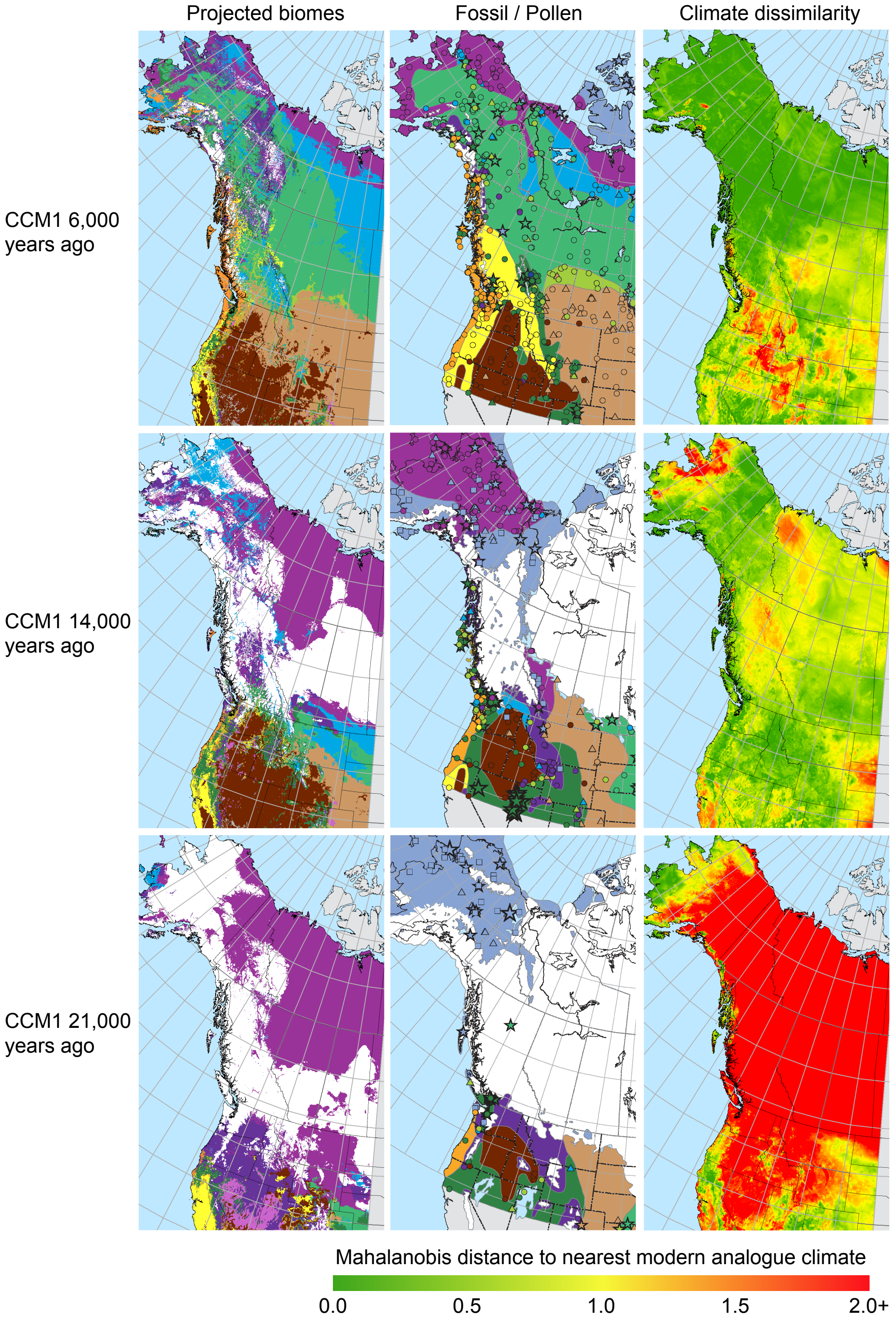

Figure 3.3: Predicted biome classes, biome reconstructions from pollen and fossil data, and climate dissimilarities measured as multivariate Mahalanobis distance to the nearest modern climate space. Green indicates climate arrangements analogous to those witnessed in the present day and red indicates increasing diversion from any modern climate conditions in the study. Summary statistics for additional model runs are given in Tables 3.3 and 3.4. Legend of biome classes is shown in Figure 3.1.

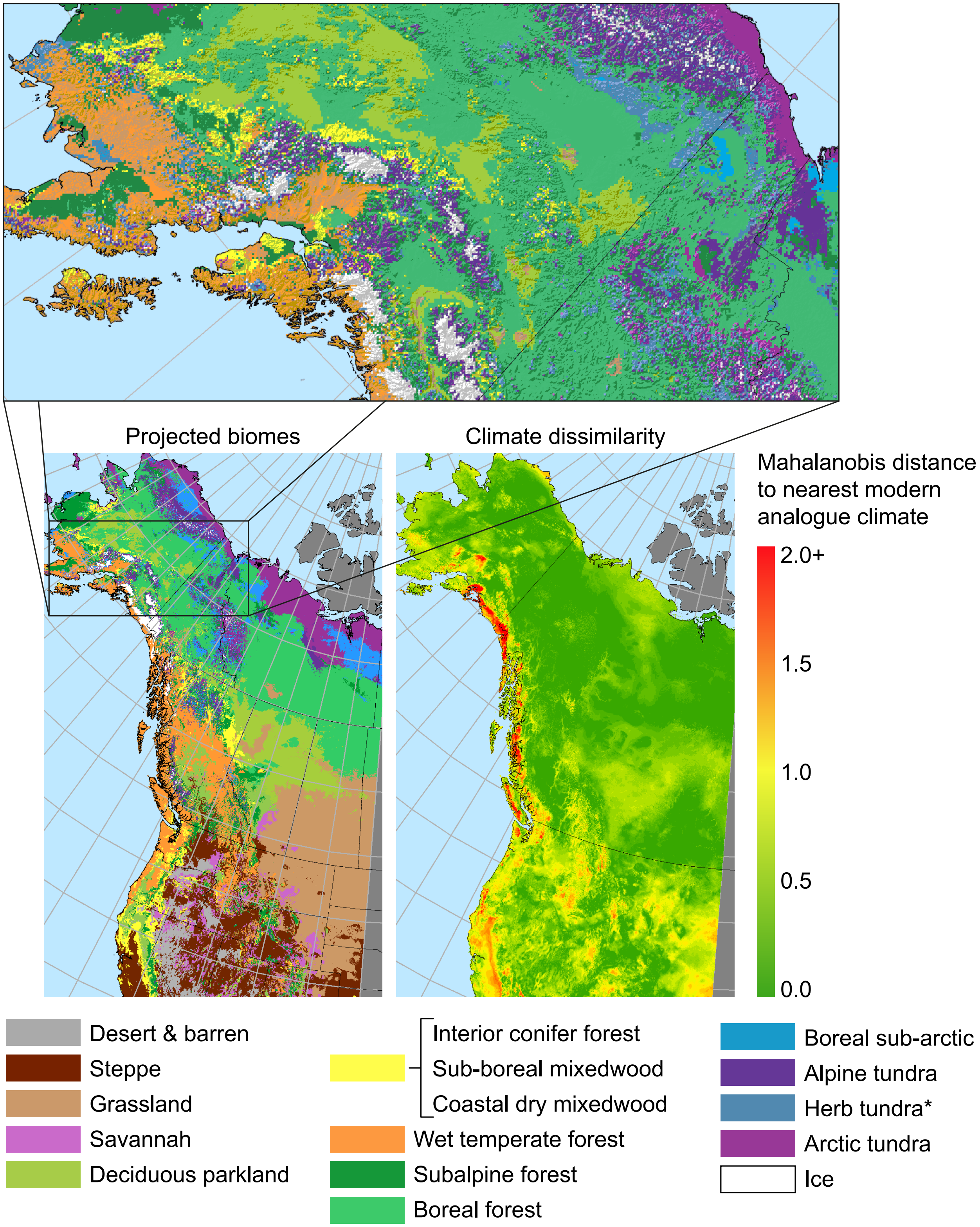

Figure 3.4: Predicted biome classes and climate dissimilarity according to an ensemble projection of the A2 emissions scenario from five general circulation models for the 2080s. The inset map provides a detailed image of the modelled ecosystems in Alaska and the Yukon Territory.

Chapter 4

Glacial refugia and modern genetic diversity of western North American tree species.[1]

4.1 Summary

Interglacial warm periods, such as the current Holocene, are the exception to the Quaternary norm. The evolution of North American tree species, subspecies, and genetic varieties has therefore taken place in a landscape with extensive continental ice and restricted temperate climate environments. Here, we reconstruct historical biogeographies and glacial refugia of western North American trees using species distribution models, validated against 3,571 fossil pollen and packrat midden records from 835 study sites. We investigate how modern genetic diversity and genetic structure was shaped by refugial history, using published estimates of allelic richness and expected heterozygosity for 473 populations of 22 tree species. We find that species with strong genetic differentiation into subspecies and varieties had widespread and large glacial refugia. In contrast, species with restricted refugia show no differentiation and little genetic diversity, despite being common over a wide range of environments today. In a regression tree analysis, 66% of allelic richness could be explained by the total size of glacial refugia. Expected heterozygosity was best explained by the number of glacial refugia (27% variance explained). Finally, reconstructed vegetation histories are used to evaluate the merit of biogeographic hypotheses regarding the existence of glacial refugia in Beringia and along the Pacific coast, the evolution of subspecies in widespread conifers, and the origin of Pacific Northwest inland rainforests.

---

[1] A version of this chapter has been submitted for publication.

4.2 Introduction

The current Holocene represents one of many relatively short, warm interglacial periods of the Quaternary, while cold temperatures and extensive continental ice have dominated North America for at least the last million years of the Pleistocene (Lisiecki & Raymo, 2005). Evolutionary processes leading to geographic differentiation of tree populations, genetic varieties, and in some cases sub-species, have therefore taken place in landscape quite different from today (Comes & Kadereit, 1998; Hewitt, 2000). The evolution of distinct genetic varieties or sub-species is often attributed to presumed disjunct populations during glacial periods, where geographic isolation in combination with different selection pressures or genetic drift would allow sub-species to form (Hewitt, 2004). Similarly, low levels of modern genetic diversity and particular signatures in allele frequency distributions are attributed to species being subjected to historical periods of constricted population sizes, referred to as genetic bottlenecks (Leberg, 1992).

Important evolutionary events and processes that have shaped species can therefore be better understood in the light of their historical biogeographies, which have traditionally been inferred by two types of research approaches. Dated fossil and pollen records have been used by paleoecologists to infer glacial refugia and reconstruct post-glacial migration routes (e.g. Thompson & Anderson, 2000; Webb, 1981). However, such records are scarce and typically restricted to lake sediments or peat deposits, which limit their usefulness to comprehensively reconstruct vegetation histories. The second approach, phylogeography, infers historical isolation of populations by analyzing present-day geographic patterns of neutral genetic markers. Modern genetic population structure can be screened comprehensively with a moderate research effort, but evolutionary events inferred from genetic markers cannot be linked to a specific time and location in the past. Nevertheless, phylogeographic studies allow formulating and testing hypotheses about refugia and migration routes, summarised in several review papers for North America (Carstens *et al.*, 2005; Jaramillo-Correa *et al.*, 2009; Shafer *et al.*, 2010; Soltis *et al.*, 1997)

These reviews highlight a number of recurring patterns in the post-glacial vegetation histories of western North American tree species. Several widespread tree species have genetically distinct varieties or subspecies in climatically distinct coastal and interior regions, e.g.

lodgepole pine (*Pinus contorta*), ponderosa pine (*Pinus ponderosa*), Douglas-fir(*Pseudotsuga menziesii*). These could have emerged through east-west separation of populations by the coastal Cascade Mountains throughout cycles of glaciations and de-glaciation (Carstens *et al.*, 2005; Fazekas & Yeh, 2006; Godbout *et al.*, 2008). On the other hand, many temperate tree species of the Pacific Northwest with disjunct coastal and interior populations show little or no genetic differentiation, e.g. western hemlock (*Tsuga heterophylla*) and western redcedar (*Thuja plicata*), suggesting post-glacial recolonization from a single refugium (Ally *et al.*, 2000; O'Connell *et al.*, 2008). For some species, north-south genetic splits can be observed in southern Oregon or northern California, with several possible recolonization paths from different refugia (Soltis *et al.*, 1997). Finally, Beringian refugia, refugia west of the continental ice, and nunatak refugia (where mountaintops emerge from the ice) have been proposed (e.g. Provan & Bennett, 2008). There is some fossil and genetic evidence that refugia west of the continental ice contained tree species, allowing for rapid recolonization of northern coastal areas (Fazekas & Yeh, 2006; Godbout *et al.*, 2008; Hamann *et al.*, 1998). Similarly, some boreal tree species may have found habitat in ice-free Beringia, allowing southward post-glacial recolonization routes (Anderson *et al.*, 2006; Jaramillo-Correa *et al.*, 2009; Shafer *et al.*, 2010).

In recent years a third approach, species distribution modelling, has emerged to complement information from paleoecological or phylogeographic studies (Carstens & Richards, 2007; Richards *et al.*, 2007). Examples include the reconstruction of post-glacial species migrations to help confirm phylogeographic inferences (e.g. Carstens & Richards, 2007; Gugger *et al.*, 2011; Rebelo *et al.*, 2012; Waltari *et al.*, 2007), but also the delineation of present-day habitat to interpret recently evolved genetic lineages (e.g. Graham *et al.*, 2004; Rissler & Apodaca, 2007). Species distribution models are based on statistical approaches that correlate species census data with the environments (often just climate conditions) in which they occur. Species distribution models rely on several assumptions that we know to be false or at least problematic (Araújo & Guisan, 2006; Guisan & Thuiller, 2005). However, the approach tends to be useful for applications at continental scales and long time frames, where local demographic processes and biological interactions that these models ignore may not have noticeable effects (Pearson & Dawson, 2003).

In this study, we contribute comprehensive reconstruction of glacial refugia and post-glacial migration histories for 22 western North American forest trees (Table 4.1). We build on two previous methodological investigations (Roberts & Hamann, 2012a,b), where we selected and optimised species distribution modelling techniques specifically for this task, and subsequently tested the accuracy and limitations of the chosen techniques with respect to equilibrium issues, niche constancy, and no-analogue climates. Here, we present ensemble projections from three selected methods for paleoclimate reconstructions from 6,000, 9,000, 11,000, 14,000, 16,000 and 21,000 years before the present from two coupled atmospheric-ocean general circulation models (GCMs). Our aim is to aid the interpretation of modern genetic population structure in western North American trees, including the evolution of sub-species, and to test inferences from phylogeographic research. Secondly, we investigate if overall levels of genetic diversity in neutral genetic markers can be understood in a historical biogeographic context. As measures of genetic diversity, we compiled or calculated the number of alleles per locus (A) and expected heterozygosity (H_e) for 473 populations of 22 tree species from published allozyme marker studies. We investigate if the number and size of glacial refugia or other life history attributes, such as dispersal and pollination type, best explain modern genetic diversity in western North American tree species.

4.3 Materials and Methods

We used high resolution maps of ecosystems to characterise the climate space of a dependent class variable. Ecosystem delineations for the continental United States and Canada west of 100° Latitude, were compiled from various public data sources (as described in Roberts & Hamann, 2012a,b). Species frequencies and probabilities of presence were calculated for each ecological region based on forest inventory plot records within the ecoregion boundaries. Forest inventory data consisted of a total of 55,743 plot samples from throughout western North America, compiled from multiple sources from different regions (Betchtold & Patterson, 2005; Govt. of Alberta, 2004; Hamann & Wang, 2006). Ecoregions in forested locations lacking sample plots were removed from the model. Samples in non-forested locations (e.g. arctic, deserts, grasslands) were included in the model to better define the limits of the climatic niche of forested ecosystems.

Modern climate data were generated for the 1961-1990 normal period based on a 1km resolution digital elevation model with the publically available software package ClimateWNA (Mbogga *et al.*, 2009; Wang *et al.*, 2012). Ten climate variables which correlated least were selected via principal component analysis as predictor variables: 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 (as calculated by Hogg, 1997). Paleoclimate data was generated by overlaying on the modern data temperature and precipitation anomalies generated with two general circulation models (GCMs): the Community Climate Model (CCM1) for the periods 6,000, 11,000, 14,000, 16,000, and 21,000 years ago (Kutzbach *et al.*, 1998) and the Geophysical Fluid Dynamics Laboratory model (GFDL) (Bush & Philander, 1999) for the periods 6,000, 9,000, 14,000, 16,000, and 21,000 years ago.

We generated ecoregion class projections using three species distribution modelling methods that permit a categorical dependent variable: (1) Random Forest (Breiman, 2001), which is an ensemble classification and regression tree technique, (2) standard discriminant analysis (Hamann & Wang, 2006), and (3) minimum Mahalanobis distance method (Roberts & Hamann, 2012a,b). Climate space of ecosystems were projected for each time period (6,000, 11,000, 14,000, 16,000, and 21,000) available from the two GCMs. For each model, we linked species frequencies and probabilities of presence based on the projected ecoregion class variable to produce reconstructed vegetation histories. Ensemble modelling methods, which create a single projection based on outputs from many individual methods, have been shown to produce more accurate model outputs than individual methods alone (Araújo & New, 2007). Accordingly, we produced a single ensemble model output for each tree species by averaging frequency or probability of presence for all data points of all three models within each GCM and period.

This probability of presence projections for each species, each model run, and ensemble summaries for the past were validated against 3,571 fossil pollen, macrofossil, and packrat midden records from 835 study sites, compiled by Dyke (2005), and reported in the North American Pollen Database (COHMAP, 1988) and the North American packrat midden database (Strickland *et al.*, 2001). The area under the curve of the receiver operating characteristic (AUC) was calculated for the past time periods as well as for the modern projections using a cross-validation. AUC values, averaged across all species for the modern period, ranged from 0.79 to 0.89 for individual modelling methods and 0.92 for the ensemble projections. Values for past periods, excluding validations with fewer than 10 validation points, were generally lower, with average AUC values of 0.65 to 0.71 for the individual methods, and averages of 0.73 and 0.75 for the ensemble method for the CCM1 and GFDL models, respectively. AUC values for all species and all methods are provided in Table 4.2.

Data on modern genetic diversity of the tree species included in the study (see Table 4.1) were collected or calculated from allele frequency tables in the published literature, as discovered via searches with ISI Web of Knowledge$^{\mathrm{SM}}$. For consistency, all available studies of allozyme diversity for the included species, which are common investigations into neutral genetic markers, were considered. A complete list of data compiled from all references, listed by sampled population, is provided in Table S4.1. The number of alleles per locus (allelic richness) and expected heterozygosity were chosen as metrics of modern genetic diversity because of their widespread and consistent reporting in the allozyme literature. Allelic richness represents the number of different expressions of a given gene whereas expected heterozygosity represents the evenness of representation of each gene expression (where higher heterozygosity is more even). While sample size has been shown to affect estimates of allelic diversity (Leberg, 2002), we were unable to account for this due to an inability to subsample all populations not reported in the literature and have thus made no adjustment to the reported averages of number of alleles per locus. Furthermore, we find no relationship (r=0.18, p=0.446) between the average number of alleles reported and the number of populations sampled (as a proxy for sampling intensity).

Maps and tables, as well as quantitative range statistics, were calculated for the modern period and the last glacial maximum using all data points where the ensemble model indicated a species frequency greater than zero. The *SDMTools* package (Van der Wal *et al.*, 2012) for the R programming environment (R Core Team, 2009) was used to calculate landscape metrics, including total area, contiguity, and the number of discrete habitat units (polygons). Contiguity was calculated as a ratio of total area to total perimeter. Landscape metrics were calculated for each individual model run, and were reported as estimates of the mean with standard errors. Life history traits were assigned based on descriptions in the USDA Silvics of North America manual (Burns *et al.*, 1990) to match the categories used by Hamrick et al. (1992). Species names are according to the Flora of North America Editorial Committee (1993).

We partitioned the variance in expected heterozygosity and allelic richness explained using a regression tree procedure with the *mvpart* package (Therneau & Atkinson, 2011) for the R programming environment (R Core Team, 2009). This technique iteratively splits response data (metrics of genetic diversity) by maximising variance explained by several continuous or class-based predictor variables. These predictors included measures of both present day and last glacial maximum range area, range contiguity, and number of habitat patches, as well as the successional stage and dispersal mechanism of each species.

4.4 Results

4.4.1 Reconstruction of glacial refugia

A summary of projected range reconstructions for all species suggest a number of general climate regions where tree species may have found refuge during the last glacial maximum (Figure 4.1, Table 4.2). Today's subalpine tree species, such as subalpine fir (*Abies lasiocarpa*), whitebark pine (*Pinus albicaulis*) and Engelmann spruce (*Picea engelmannii*), found equivalent climate habitat primarily in the northern basins, from the Williamette Valley in the northwest to the New-Mexico plateau in the southeast. Boreal species, such as white spruce (*Picea glauca*), black spruce (*P. mariana*), lodgepole pine and trembling aspen (*Populus tremuloides*), find climate habitat equivalent to today's boreal conditions near the eastern limits of the study area, including the High Plains and Southwest Tablelands of Colorado and New Mexico. Species that occur in interior lowlands under relatively dry conditions, including western larch (*Larix occidentalis*), ponderosa pine and pinyon pine (*P. edulis*), find suitable habitat in the southern mountain ranges. Coastal species of the Pacific Northwest were often restricted to relatively small areas along the California coast during the last glacial maximum. For a species-level breakdown of projected habitat corresponding to the regions in Figure 4.1, refer to Table 4.3 and Supplementary Figures S4.1A to S4.1V.

Figure 4.1 also conveys that much of the land area south of the ice did not serve as refugia for more than a few tree species considered in this study. The low-frequency yellow areas in Figure 4.1 tend to be occupied by species adapted to xeric conditions (e.g. ponderosa pine, pinyon pine, and interior Douglas-fir). Today's nearest climate equivalents to these regions are open savannas rather than closed canopy forests. We also find areas of very high species richness (Figure 4.1, purple), where a large majority of all species modelled in this study was predicted to find suitable habitat conditions. This includes the Sierra Nevada, the California Coast Mountains, and the Sierra Madre.

Another general observation is that species, which are widespread today, were typically confined to rather restricted areas during the last glacial maximum. With the exception of whitebark pine, all species had more restricted climate habitat during the last glaciation than at present. In particular, coastal species appear to be very restricted compared to their current ranges, e.g. Pacific silver fir (*Abies amabilis*, Figure S4.1A), noble fir (*Abies procera*, Figure S4.1C), red alder (*Alnus rubra*, Figure S4.1E), yellow cedar (*Chamaecyparis nootkatensis*, Figure S4.1G), Sitka spruce (*Picea sitchensis*, Figure S4.1L), and western redcedar (Figure S4.1T). Some coastal species also had small projected refugia west of the continental ice (Figure 4.1, Haida Gwaii and Vancouver Island). Boreal and sub-boreal species, e.g. white spruce, black spruce, lodgepole pine, and aspen, were generally least restricted during the last ice age, and also had projected glacial refugia along the northern Pacific coast and in Beringia.

4.4.2 Species genetic diversity

For species with strong modern genetic differentiation into subspecies and ecotypes, we generally find widespread and diverse refugia. Ponderosa pine, for example, consists of several subspecies today, had widespread coastal and interior refugia at the last glacial maximum, including the California coast, Arizona Mountains, Southwest Tablelands of New Mexico and Colorado, and the Sierra Madre (Figure S4.1Q). For species that today have strong genetic differentiation between coastal and interior populations such as lodgepole pine (*pinus contorta var. contorta* versus *var. latifolia*), and Douglas-fir (*Pseudotsuga menziesii var. menziesii* versus *var. glauca*) we find large glacial refugia east and west of the ice-covered Sierra Nevada (Figures S4.1N and S4.1S1S). In contrast, species that today share the same dual interior and coastal distributions, red alder, yellow cedar, western white pine (*Pinus monticola*), western redcedar, mountain hemlock, and western hemlock, but that do not have distinct varieties or subspecies, showed no continuous interior habitat (Figure S4.1E, Figure S4.1G, Figure S4.1U) or only very restricted interior refugia (Figure S4.1P, Figure S4.1T, Figure S4.1V).

Modern genetic diversity measured as allelic richness in neutral genetic markers (Table 4.1) reveals strong associations with modelled species ranges and number of refugia at the last glacial maximum. In a regression tree analysis, the best explanatory variable for allelic richness is the total area of the projected species range at the last glacial maximum (Figure 4.2A). This can also be visualised with a simple linear regression, which explains 55% of the variation in allelic richness among species (Figure 4.3). Note that white and black spruce were excluded from this graph because their glacial ranges extended much farther east than the area covered by this study (thus their glacial range size was underestimated in this study). Aspen was left in this analysis despite an easward-extending range, as it also featured extensive western refugia. Additional explanatory variables for allelic richness were contiguity of glacial refugia (a measure of landscape fragmentation) and seed dispersal mechanism. For species with small total refugial area, fragmentation of those refugia was associated with particularly low allelic richness. For species with large glacial refugia, wind-dispersed species had particularly high allelic richness (Figure 4.2A).

Landscape metrics and life history traits accounted for a smaller portion of the variance in expected heterozygosity (Figure 4.2B). The best explanatory variable was the number of discrete glacial refugia, accounting for 27% of the variance in allozyme heterozygosity in a regression tree analysis. The size of modern and projected glacial range size accounted for another 21% and 16%, respectively. It should be noted that results from regression tree analysis, like all correlative analytical techniques, must consider auto-correlations among predictor variables. For the second regression tree, an almost equally strong predictor for expected heterozygosity was contiguity of refugia during the last glacial maximum, accounting for 25%, and thus very close to the first choice (27%). The highest expected heterozygosities were therefore associated with high numbers of refugia, greater fragmentation of refugia or both. A comprehensive list of alternative splits for both allelic richness and expected heterozygosity regression trees are provided in Table 4.4.

4.5 Discussion

4.5.1 Vegetation history of temperate inland rainforests

Many phylogeographic hypotheses for western North America have focused on the temperate rainforest flora of the Pacific Northwest (e.g. Brunsfeld *et al.*, 2001; Carstens *et al.*, 2005; Shafer *et al.*, 2010; Soltis *et al.*, 1997), typically interpreting observations of a north-south genetic division centred in Oregon and inferring the origin of disjunct temperate rainforest communities of the Washington and B.C. interior. To explain these patterns, three hypotheses were proposed by Brunsfeld et al. (2001). First, the ancient vicariance hypothesis postulates that species were split into coastal and interior refugia, divided by the dry Columbia Basin throughout the Pleistocene. Second, the inland dispersal from the north hypothesis proposes that today's interior ranges were populated post-glacially by migration from refugia in the Cascades to the west. Third, a similar inland dispersal from the south hypothesis posits recolonization of interior distributions from glacial refugia in central Oregon.

Model hindcasts from this study generally support the inland dispersal hypotheses, implying late colonization of inland rainforest communities due to the lack of stable interior habitat. Species with modern disjunct coastal and inland rainforest distributions, e.g. red alder, yellow cedar, western white pine, western redcedar, western and mountain hemlock, showed none or very sporadic appearances of suitable interior habitat towards the last glacial maximum. This is further supported by fossil data for western and mountain hemlock, which appears in the interior dating back only 3,500 to 4,500 years (Mehringer, 1996; Rosenberg *et al.*, 2003). Further, published data suggests little genetic differentiation between coastal and interior populations (Ally *et al.*, 2000; Benowicz & El-Kassaby, 1999). Plausible paths of recolonization differ among species, with red alder and yellow cedar requiring long-distance dispersal events to reach inland habitats (Figure S4.1E, S1G), and western white pine and western redcedar showing connectivity of habitat inland in northern Washington or southern British Columbia. This path is also supported by the genetic similarity of inland and Washington populations (O'Connell *et al.*, 2008). Finally, western and mountain hemlock show a possible alternative path through northern British Columbia for recolonization of inland refugia in the GFDL-based model as early as 9,000 years ago (Figure S4.1U, Figure

S4.1V).

4.5.2 Widespread trees with sub-species structure

Both ponderosa pine and Douglas-fir feature present-day coastal and interior subspecies, divided by topographic barriers such as the Sierra Nevada and Cascade mountains (Gugger & Sugita, 2010; Li & Adams, 1989; Norris *et al.*, 2006). Their paleoecological and phylogeographic data also suggests a common history: disjunct southern coastal and interior refugia at the last glacial maximum and allopatric postglacial migration northward (Gugger & Sugita, 2010; Norris *et al.*, 2006). Their vegetation histories reconstructed in this study support separate continuous inland and coastal populations, with the ice-coverd Sierra Nevada serving as a topographic barrier during glacial periods (Figure S4.1Q and Figure S4.1S). Modelled refugia for both species appear in the Sierra Nevada and Klamath Mountains on the coast and through the Arizona Mountains and Southwest Tablelands in the interior, locations which are also confirmed by paleoecological data (Gugger & Sugita, 2010; Norris *et al.*, 2006). Neither ponderosa pine nor Douglas-fir projections suggest stable habitat north or west of the continental ice.

The phylogeography of interior and coastal subspecies of lodgepole pine is somewhat more complex, with genetic data suggesting additional refugia either in Beringia or along the Pacific coast, possibly in the area of Haida Gwaii (Fazekas & Yeh, 2006; Godbout *et al.*, 2008). The potential for refugia west or north of the continental ice is supported in our GFDL model projections in particular, which show extensive lodgepole pine habitat along the Pacific Coast and in Beringia, the most stable of which appears in Haida Gwaii and western Alaska. It should be noted that while our models hindcast suitable habitat back to the last glacial maximum, paleoecological data places lodgepole pine in Haida Gwaii only as early as 15,600 to 13,800 years ago (Lacourse *et al.*, 2005) or 10,000 years ago (Peteet, 1991). A northern refugium for lodgepole pine in the Yukon, as suggested by Wheeler et al. (Wheeler & Guries, 1982) but since questioned (Fazekas & Yeh, 2006), has little support in paleoecological data nor in model hindcasts. Notably, present day Yukon populations exhibit rare alleles shared by both adjacent interior and coastal populations (Fazekas & Yeh, 2006; Wheeler & Guries, 1982). Fazekas and Yeh (Fazekas & Yeh, 2006) have hypothesised northward migration from the south interior as well as inland from the west coast to the Yukon via deep fjords and river valleys, a pattern very well supported in our model reconstructions (Figure S4.1N).

4.5.3 Refugia west and north of the ice

Glacial refugia in Beringia and Haida Gwaii may also explain very high calculated post-glacial migration rates, a phenomenon known as "Reid's Paradox" (Clark *et al.*, 1998). For example, the expansion of black spruce into northern Canada and Alaska evident from the fossil record requires migration rates of 1.5-2 km yr^{-1}, necessitating mechanisms such as repeated long distance dispersal events (Anderson *et al.*, 2006). The existence of a glacial refugium for white and black spruce in Alaska, for which there is moderate genetic evidence (Anderson *et al.*, 2006; Jaramillo-Correa *et al.*, 2004) but weaker paleoecological evidence (Brubaker *et al.*, 2005), much reduces required migration rates. Our hindcasts based on the GFDL model suggest possible refugia in Beringia both for white spruce (Figure S4.1J) and black spruce (Figure S4.1K). Other species where both GFDL and CCM1 models suggest refugial habitat in Beringia include aspen (Figure S4.1R), and lodgepole pine (Figure S4.1N), discussed above.

Because arctic currents were blocked from entering the Pacific by the Beringian land bridge at the last glacial maximum, climate conditions along the Pacific coast were relatively mild. Low sea levels and isostatic rise of the coastal areas due to the weight of the continental ice allowed for refugia west of the continental ice that could have sustained tree populations in areas around Vancouver Island, Haida Gwaii and the Alaskan Panhandle, including areas now submerged (Lacourse *et al.*, 2005). Conclusive paleoecological evidence of glacial tree populations in the Haida Gwaii area is still lacking, with the earliest lodgepole pine macrofossil evidence dating to ca. 16,600 years ago, and Sitka spruce, mountain hemlock, and red alder dating to ca. 13,000 years ago (Lacourse *et al.*, 2005). However, allelic richness of modern Haida Gwaii populations of several species is relatively high, suggesting that these populations were not established from repeated long distance dispersal events (Wellman, 2004; Wheeler & Guries, 1982). Genetic differentiation in neutral markers support glacial refugia in this area for red alder, yellow cedar, western redcedar, and white spruce (Anderson *et al.*, 2006; Barnosky *et al.*, 1987; Godbout *et al.*, 2008; Hamann *et al.*, 1998).

Our model hindcasts suggest suitable habitat along the Pacific coast, notably in Haida Gwaii, for many temperate rainforest species including western redcedar, western and mountain hemlock, yellow cedar, Pacific silver fir, Sitka spruce, and red alder. However, the weak representation of tree species at the last glacial maximum should serve as a note of caution. Further, the species Pacific silver fir, Sitka spruce, western redcedar, and western and mountain hemlock have little genetic support for cryptic northern refugia (Ally *et al.*, 2000; Benowicz & El-Kassaby, 1999; Holliday *et al.*, 2010; Yeh & Hu, 2005).

4.5.4 Drivers of modern genetic diversity

We find surprisingly strong relationships between modern genetic diversity and landscape metrics that describe species distributions during the last glacial maximum, complementing analysis of life history traits as drivers of genetic diversity by Hamrick et al. (Hamrick *et al.*, 1992). Our observation that allelic richness is better explained than expected heterozygosity by past vegetation history corresponds to previous studies (e.g. Comps *et al.*, 2001) and fits population genetic theory. Rare alleles, which strongly influence measures of allelic richness, are likely to be lost in population bottlenecks (Leberg, 1992), and their recovery over time is less affected by the rate of population growth, which would vary between species (e.g. Nei *et al.*, 1975). By contrast, rare alleles should contribute little to expected heterozygosity, which is instead driven by the frequency and evenness of common alleles that are not readily lost even in very small populations (Nei *et al.*, 1975). Instead, large numbers of disjunct refugia that have persisted over long periods of time during the Pleistocene should favor genetic differentiation of past populations and high levels of expected heterozygosity in modern populations, which is what we found in our analysis.

4.6 References

Ally, D., El-Kassaby, Y. A., & Ritland, K. 2000. Genetic diversity, differentiation and mating systems in mountain hemlock (Tsuga mertensia) across British Columbia. *Forest Genetics*, **7**(2), 97–108.

Anderson, L. L., Hu, F. S., Nelson, D. M., Petit, R. J., & Paige, K. N. 2006. Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, **103**(33), 12447–12450.

Araújo, M. B., & Guisan, A. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**(10), 1677–1688.

Araújo, M. B., & New, M. 2007. Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**(1), 42–47.

Barnosky, K. W., Anderson, P. M., & Bartlein, P. J. 1987. *The northwestern U.S. during deglaciation; Vegetational history and paleoclimatic implications.* Vol. Vol. K-3. Boulder, CO: The Geological Society of America.

Benowicz, A., & El-Kassaby, Y. A. 1999. Genetic variation in mountain hemlock (Tsuga mertensiana Bong.): quantitative and adaptive attributes. *Forest Ecology and Management*, **123**(2-3), 205–215.

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

Breiman, L. 2001. Random forests. *Machine Learning*, **45**(1), 5–32.

Brubaker, L. B., Anderson, P. M., Edwards, M. E., & Lozhkin, A. V. 2005. Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *Journal of Biogeography*, **32**(5), 833–848.

Bruederle, L. P., Tomback, D. F., Kelly, K. K., & Hardwick, R. C. 1998. Population genetic structure in a bird-dispersed pine, Pinus albicaulis (Pinaceae). *Canadian Journal of Botany-Revue Canadienne De Botanique*, **76**(1), 83–90.

Brunsfeld, S. J., Sullivan, J., Soltis, D. E., & Soltis, P. S. 2001. Comparative phylogeography of northwestern North America: A synthesis. *Pages 319–339 of:* Silvertown, J., & Antovics, J. (eds), *Integrating ecological and evolutionary processes in a spatial context.* Oxford: Blackwell Science.

Burns, R. M., Honkala, H., & tech coords. 1990. *Silvics of North America: 1. Conifers; 2. Hardwoods.* U.S. Department of Agriculture, Forest Service.

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**(D20), 24509–24525.

Carstens, B. C., & Richards, C. L. 2007. Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**(6), 1439–1454.

Carstens, B. C., Brunsfeld, S. J., Demboski, J. R., Good, J. M., & Sullivan, J. 2005. Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: Hypothesis testing within a comparative phylogeographic framework. *Evolution*, **59**(8), 1639–1652.

Cheliak, W. M., & Dancik, B. P. 1982. Genic diversity of natural populations of a clone-forming tree Populus tremuloides. *Canadian Journal of Genetics and Cytology*, **24**(5), 611–616.

Clark, J. S., Fastie, C., Hurtt, G., Jackson, S. T., Johnson, C., King, G. A., Lewis, M., Lynch, J., Pacala, S., Prentice, C., Schupp, E. W., Webb, T., & Wyckoff, P. 1998. Reid's paradox of rapid plant migration - Dispersal theory and interpretation of paleoecological records. *Bioscience*, **48**(1), 13–24.

COHMAP. 1988. Climatic Changes of the Last 18,000 Years - Observations and Model Simulations. *Science*, **241**(4869), 1043–1052.

Comes, H. P., & Kadereit, J. W. 1998. The effect of quaternary climatic changes on plant distribution and evolution. *Trends in Plant Science*, **3**(11), 432–438.

Comps, B., Gomory, D., Letouzey, J., Thiebaut, B., & Petit, R. J. 2001. Diverging trends between heterozygosity and allelic richness during postglacial colonization in the European beech. *Genetics*, **157**(1), 389–397.

Copes, D. L. 1981. Isoenzyme Uniformity in Western Red Cedar Seedlings from Oregon and Washington. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **11**(2), 451–453.

Dancik, B. P., & Yeh, F. C. 1983. Allozyme Variability and Evolution of Lodgepole Pine (Pinus-Contorta Var Latifolia) and Jack Pine (Pinus-Banksiana) in Alberta. *Canadian Journal of Genetics and Cytology*, **25**(1), 57–64.

Davidson, R. H., & El Kassaby, Y. A. 1997. Genetic diversity and gene conservation of Pacific Silver Fir (Abies amabilis) on Vancouver Island, British Columbia. *Forest Genetics*, **4**(2), 85–98.

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

Dyke, A. S., Andrews, J. T., Clark, P. U., England, J. H., Miller, G. H., Shaw, J., & Veillette, J. J. 2002. The Laurentide and Innuitian ice sheets during the Last Glacial Maximum. *Quaternary Science Reviews*, **21**(1-3), 9–31.

El-Kassaby, Y. A., & Ritland, K. 1996. Genetic variation in low elevation Douglas-fir of British Columbia and its relevance to gene conservation. *Biodiversity and Conservation*, **5**(6), 779–794.

El-Kassaby, Y. A., Russell, J., & Ritland, K. 1994. Mixed Mating in an Experimental Population of Western Red Cedar, Thuja-Plicata. *Journal of Heredity*, **85**(3), 227–231.

El-Kassaby, Y. A., Dunsworth, B. G., & Krakowski, J. 2003. Genetic evaluation of alternative silvicultural systems in coastal montane forests: western hemlock and amabilis fir. *Theoretical and Applied Genetics*, **107**(4), 598–610.

Ettl, G. J., & Peterson, D. L. 2001. Genetic variation of subalpine fir (Abies lasiocarpa (HOOK.) NUTT.) in the Olympic Mountains, WA, USA. *Silvae Genetica*, **50**(3-4), 145–153.

Fazekas, A. J., & Yeh, F. C. 2006. Postglacial colonization and population genetic relationships in the Pinus contorta complex. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **84**(2), 223–234.

Fins, L., & Seeb, L. W. 1986. Genetic variation in allozymes of western larch. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **16**(5), 1013–1018.

Furnier, G. R., Stine, M., Mohn, C. A., & Clyde, M. A. 1991. Geographic Patterns of Variation in Allozymes and Height Growth in White Spruce. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **21**(5), 707–712.

Godbout, J., Fazekas, A., Newton, C. H., Yeh, F. C., & Bousquet, J. 2008. Glacial vicariance in the Pacific Northwest: evidence from a lodgepole pine mitochondrial DNA minisatellite for multiple genetically distinct and widely separated refugia. *Molecular Ecology*, **17**(10), 2463–2475.

Godt, M. J. W., Hamrick, J. L., Edwards-Burke, M. A., & Williams, J. H. 2001. Comparisons of genetic diversity in white spruce (Picea glauca) and jack pine (Pinus banksiana) seed orchards with natural populations. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **31**(6), 943–949.

Govt. of Alberta. 2004. *Ecological Site Information System (ESIS).* Government of Alberta, Sustainable Resource Development (SRD), Available: http://www.srd.gov.ab.ca/lands/geographicinformation/resourcedataproductcatalogue/esis.aspx.

Graham, C. H., Ron, S. R., Santos, J. C., Schneider, C. J., & Moritz, C. 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, **58**(8), 1781–1793.

Gugger, P. F., & Sugita, S. 2010. Glacial populations and postglacial migration of Douglas-fir based on fossil pollen and macrofossil evidence. *Quaternary Science Reviews*, **29**(17-18), 2052–2070.

Gugger, P. F., Gonzalez-Rodriguez, A., Rodriguez-Correa, H., Sugita, S., & Cavender-Bares, J. 2011. Southward Pleistocene migration of Douglas-fir into Mexico: phylogeography, ecological niche modeling, and conservation of 'rear edge' populations. *New Phytologist*, **189**(4), 1185–1199.

Guisan, A., & Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**(9), 993–1009.

Hamann, A., & Wang, T. L. 2006. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. *Ecology*, **87**(11), 2773–2786.

Hamann, A., El-Kassaby, Y. A., Koshy, M. P., & Namkoong, G. 1998. Multivariate analysis of allozymic and quantitative trait variation in Alnus rubra: geographic patterns and evolutionary implications. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **28**(10), 1557–1565.

Hamrick, J. L., Mitton, J. B., & Linhart, Y. B. 1979. Levels of genetic variation in trees: influence of life history characteristics. *In:* Conkle, M. T. (ed), *Isozymes of North American Forest Trees and Forest Insects.* Berkeley, CA: USDA Forest Service, Berkeley, CA.

Hamrick, J. L., Godt, M. J. W., & Shermanbroyles, S. L. 1992. Factors Influencing Levels of Genetic Diversity in Woody Plant-Species. *Population Genetics of Forest Trees*, **42**, 95–124.

Harry, D. E. 1984. *Genetic structure of incense cedar (Calocedrus decurrens) populations.* Ph.D. thesis.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature*, **405**(6789), 907–913.

Hewitt, G. M. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**(1442), 183–195.

Hogg, E. H. 1997. Temporal scaling of moisture and the forest-grassland boundary in western Canada. *Agricultural and Forest Meteorology*, **84**(1-2), 115–122.

Holliday, J. A., Yuen, M., Ritland, K., & Aitken, S. N. 2010. Postglacial history of a widespread conifer produces inverse clines in selective neutrality tests. *Molecular Ecology*, **19**(18), 3857–3864.

Hyun, J. O., Rajora, O. P., & Zsuffa, L. 1987. Genetic-Variation in Trembling Aspen in Ontario Based on Isozyme Studies. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **17**(9), 1134–1138.

Iddrisu, M. N., & Ritland, K. 2004. Genetic variation, population structure, and mating system in bigleaf maple (Acer macrophyllum Pursh). *Canadian Journal of Botany-Revue*

Canadienne De Botanique, **82**(12), 1817–1825.

Isabel, N., Beaulieu, J., & Bousquet, J. 1995. Complete Congruence between Gene Diversity Estimates Derived from Genotypic Data at Enzyme and Random Amplified Polymorphic DNA Loci in Black Spruce. *Proceedings of the National Academy of Sciences of the United States of America*, **92**(14), 6369–6373.

Jaquish, B., & El-Kassaby, Y. A. 1998. Genetic variation of western larch in British Columbia and its conservation. *Journal of Heredity*, **89**(3), 248–253.

Jaramillo-Correa, J. P., Beaulieu, J., & Bousquet, J. 2004. Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (Picea mariana), a transcontinental North American conifer. *Molecular Ecology*, **13**(9), 2735–2747.

Jaramillo-Correa, J. P., Beaulieu, J., Khasa, D. P., & Bousquet, J. 2009. Inferring the past from the present phylogeographic structure of North American forest trees: seeing the forest for the genes. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **39**(2), 286–307.

Jelinski, D. E., & Cheliak, W. M. 1992. Genetic Diversity and Spatial Subdivision of Populus-Tremuloides (Salicaceae) in a Heterogeneous Landscape. *American Journal of Botany*, **79**(7), 728–736.

Jorgensen, S. M., & Hamrick, J. L. 1997. Biogeography and population genetics of whitebark pine, Pinus albicaulis. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **27**(10), 1574–1585.

Kim, M. S., Richardson, B. A., McDonald, G. I., & Klopfenstein, N. B. 2011. Genetic diversity and structure of western white pine (Pinus monticola) in North America: a baseline study for conservation, restoration, and addressing impacts of climate change. *Tree Genetics & Genomes*, **7**(1), 11–21.

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**(6-7), 473–506.

Lacourse, T., Mathewes, R. W., & Fedje, D. W. 2005. Late-glacial vegetation dynamics of the Queen Charlotte Islands and adjacent continental shelf, British Columbia, Canada. *Palaeogeography Palaeoclimatology Palaeoecology*, **226**(1-2), 36–57.

Leberg, P. L. 1992. Effects of Population Bottlenecks on Genetic Diversity as Measured by Allozyme Electrophoresis. *Evolution*, **46**(2), 477–494.

Leberg, P. L. 2002. Estimating allelic richness: Effects of sample size and bottlenecks. *Molecular Ecology*, **11**(11), 2445–2449.

Ledig, F. T., Hodgskiss, P. D., & Johnson, D. R. 2006. The structure of genetic diversity in Engelmann spruce and a comparison with blue spruce. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **84**(12), 1806–1828.

Li, P., & Adams, W. T. 1989. Range-Wide Patterns of Allozyme Variation in Douglas-Fir (Pseudotsuga-Menziesii). *Canadian Journal of Forest Research/Revue Canadienne de Recherche Forestiere*, **19**(2), 149–161.

Lisiecki, L. E., & Raymo, M. E. 2005. A Pliocene-Pleistocene stack of 57 globally distributed benthic delta O-18 records. *Paleoceanography*, **20**(1).

Liu, Z., & Furnier, G. R. 1993. Comparison of Allozyme, Rflp, and Rapd Markers for Revealing Genetic-Variation within and between Trembling Aspen and Bigtooth Aspen. *Theoretical and Applied Genetics*, **87**(1-2), 97–105.

Lund, S. T., Furnier, G. R., & Mohn, C. A. 1992. Isozyme Variation in Quaking Aspen in Minnesota. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **22**(4), 521–524.

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**(5), 881–890.

Mehringer, P. J. 1996. *Columbia River Basin Ecosystems: Late Quaternary Environments.* Department of Anthropology and Geology, Washington State University.

Nei, M., Maruyama, T., & Chakraborty, R. 1975. Bottleneck Effect and Genetic-Variability in Populations. *Evolution*, **29**(1), 1–10.

Niebling, C. R., & Conkle, M. T. 1990. Diversity of Washoe Pine and Comparisons with Allozymes of Ponderosa Pine Races. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **20**(3), 298–308.

Norris, J. R., Jackson, S. T., & Betancourt, J. L. 2006. Classification tree and minimum-volume ellipsoid analyses of the distribution of ponderosa pine in the western USA. *Journal of Biogeography*, **33**(2), 342–360.

O'Connell, L. M., Ritland, K., & Thompson, S. L. 2008. Patterns of post-glacial colonization by western redcedar (Thuja plicata, Cupressaceae) as revealed by microsatellite markers. *Botany-Botanique*, **86**(2), 194–203.

of North America Editorial Committee, Flora. 1993. *Flora of North America North of Mexico. 16 vols.*

O'Malley, D. M., Allendorf, F. W., & Blake, G. M. 1979. Inheritance of isozyme variation and heterozygosity in Pinus ponderosa. *Biochemical Genetics*, **17**(3/4), 233–250.

O'Reilly, G. J., Parker, W. H., & Cheliak, W. M. 1985. Isozyme Differentiation of Upland and Lowland Picea-Mariana Stands in Northern Ontario. *Silvae Genetica*, **34**(6), 214–221.

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**(5), 361–371.

Peteet, D. M. 1991. Postglacial Migration History of Lodgepole Pine near Yakutat, Alaska. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **69**(4), 786–796.

Premoli, A. C., Chischilly, S., & Mitton, J. B. 1994. Levels of Genetic Variation Captured by 4 Descendant Populations of Pinyon Pine (Pinus Edulis Engelm.). *Biodiversity and Conservation*, **3**(4), 331–340.

Provan, J., & Bennett, K. D. 2008. Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, **23**(10), 564–571.

R Core Team. 2009. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rajora, O. P., & Pluhar, S. A. 2003. Genetic diversity impacts of forest fires, forest harvesting, and alternative reforestation practices in black spruce (Picea mariana). *Theoretical and Applied Genetics*, **106**(7), 1203–1212.

Rebelo, H., Froufe, E., Brito, J. C., Russo, D., Cistrone, L., Ferrand, N., & Jones, G. 2012. Postglacial colonization of Europe by the barbastelle bat: agreement between molecular data and past predictive modelling. *Molecular Ecology*, **21**(11), 2761–2774.

Richards, C. L., Carstens, B. C., & Knowles, L. L. 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography*, **34**(11), 1833–1845.

Rissler, L. J., & Apodaca, J. J. 2007. Adding more ecology into species delimitation: Ecological niche models and phylogeography help define cryptic species in the black salamander (Aneides flavipunctatus). *Systematic Biology*, **56**(6), 924–942.

Ritland, C., Pape, T., & Ritland, K. 2001. Genetic structure of yellow cedar (Chamaecyparis nootkatensis). *Canadian Journal of Botany-Revue Canadienne De Botanique*, **79**(7), 822–828.

Roberts, D. R., & Hamann, A. 2012a. Method selection for species distribution modelling: are temporally or spatially independent evaluations necessary? *Ecography*, **35**, 792–802.

Roberts, D. R., & Hamann, A. 2012b. Predicting potential climate change impacts with bioclimate envelope models: a palaeoecological perspective. *Global Ecology and Biogeography*, **21**(2), 121–133.

Rogers, D L. 1994. *Spatial patterns of allozyme variation and clonal structure in coast redwood (Sequoia sempervirens).* Ph.D. thesis.

Rosenberg, S. M., Walker, I. R., & Mathewes, R. W. 2003. Postglacial spread of hemlock (Tsuga) and vegetation history in Mount Revelstoke National Park, British Columbia, Canada. *Canadian Journal of Botany-Revue Canadienne De Botanique*, **81**(2), 139–151.

Shafer, A. B. A., Cullingham, C. I., Cote, S. D., & Coltman, D. W. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology*, **19**(21), 4589–4621.

Soltis, D. E., Gitzendanner, M. A., Strenge, D. D., & Soltis, P. S. 1997. Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution*, **206**(1-4), 353–373.

Steinhoff, R. J., Joyce, D. G., & Fins, L. 1983. Isozyme Variation in Pinus-Monticola. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **13**(6), 1122–1132.

Strickland, L. E., Thompson, R. S., & Anderson, K. H. 2001. *USGS/NOAA North American packrat midden database. http://gec.cr.usgs.gov/data/midden/.*

Therneau, T. M., & Atkinson, E. J. 2011. *An introduction to recursive partitioning using the RPART routines, Available: http://r.789695.n4.nabble.com/attachment/3209029/0/zed.pdf.* Mayo Foundation, Technical Report #61.

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**(3), 555–584.

Van der Wal, J., Falconi, L., Januchowski, S., Shoo, L., & Storlie, C. 2012. *SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises (R package, version 1.1-13), Available: http://cran.r-project.org/web/packages/SDMTools.*

Waltari, E., Hijmans, R. J., Peterson, A. T., Nyari, A. S., Perkins, S. L., & Guralnick, R. P. 2007. Locating Pleistocene Refugia: Comparing Phylogeographic and Ecological Niche Model Predictions. *PLoS One*, **2**(7), –.

Wang, T. L., Hamann, A., Spittlehouse, D. L., & Murdock, T. Q. 2012. ClimateWNA-High-Resolution Spatial Climate Data for Western North America. *Journal of Applied Meteorology and Climatology*, **51**(1), 16–29.

Webb, T. 1981. The Past 11,000 Years of Vegetational Change in Eastern North-America. *Bioscience*, **31**(7), 501–506.

Wellman, H. F. 2004. *The genetics of selective breeding in western hemlock Tsuga heterophylla*. Ph.D. thesis.

Wheeler, N. C., & Guries, R. P. 1982. Population-Structure, Genic Diversity, and Morphological Variation in Pinus-Contorta Dougl. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **12**(3), 595–606.

Woods, J. H., Blake, G. M., & Allendorf, F. W. 1983. Amount and Distribution of Isozyme Variation in Ponderosa Pine from Eastern Montana. *Silvae Genetica*, **32**(5-6), 151–157.

Xie, C. Y., El-Kassaby, Y. A., & Ying, C. C. 2002. Genetics of red alder (Alnus Rubra Bong.) populations in British Columbia and its implications for gene resources management. *New Forests*, **24**(2), 97–112.

Yeh, F. C. 1988. Isozyme Variation of Thuja-Plicata (Cupressaceae) in British-Columbia. *Biochemical Systematics and Ecology*, **16**(4), 373–377.

Yeh, F. C., & El-Kassaby, Y. A. 1980. Enzyme Variation in Natural-Populations of Sitka Spruce (Picea-Sitchensis) .1. Genetic-Variation Patterns among Trees from 10 Iufro Provenances. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **10**(3), 415–422.

Yeh, F. C., & Hu, X. S. 2005. Genetic structure and migration from mainland to island populations in Abies procera Rehd. *Genome*, **48**(3), 461–473.

Yeh, F. C., & Layton, C. 1979. Organization of Genetic-Variability in Central and Marginal Populations of Lodgepole Pine Pinus-Contorta Spp Latifolia. *Canadian Journal of Genetics and Cytology*, **21**(4), 487–503.

Yeh, F. C. H., & O'Malley, D. 1980. Enzyme Variations in Natural-Populations of Douglas-Fir, Pseudotsuga Menziesii (Mirb) Franco, from British-Columbia .1. Genetic-Variation Patterns in Coastal Populations. *Silvae Genetica*, **29**(3-4), 83–92.

Table 4.1: List of species included in this study with allelic richness (A), expected heterozygosity (H_e) with standard error of the estimate (SE) in parentheses, and the number of populations sampled (N) in published studies that are the basis of these estimates. References and population-level genetic data are listed in Table S4.1. Also provided is the representation of individual species in refugia by broad regions to aid interpretation of Figure 4.1. A detailed breakdown for regions exactly corresponding to Figure 4.1 is provided in Table 4.3. Summary regions include: NC, northern coastal refugia west and north of the continental ice; SC, southern coastal areas with mesic climates; NIB, northern interior basins with subalpine climate equivalents; SIH, southern interior highlands with xeric conditions; HP, high plains with boreal conditions; and SM, Sierra Madre with boreal climate equivalents.

	Genetic information				Representation in refuiga					
Species	A	H_e	(SE)	N	NC*	SC	NIB	SIH	HP	SM
1. *Abies amabilis*	1.63	0.21	-	8	23	45	26	1	0	5
2. *Abies lasiocarpa*	1.92	0.13	(0.05)	9	12	8	45	27	7	1
3. *Abies procera*	1.23	0.23	(0.05)	21	0	90	0	3	0	7
4. *Acer macrophyllum*	1.71	0.15	(0.04)	8	0	88	0	0	0	12
5. *Alnus rubra*	1.51	0.10	(0.04)	19	0	89	0	3	0	8
6. *Calocedrus decurrens*	2.53	0.18	(0.04)	12	0	76	1	6	0	17
7. *Chamaecyparis nootkatensis*	1.68	0.15	(0.01)	17	44	45	0	0	0	11
8. *Larix occidentalis*	1.51	0.15	(0.04)	13	0	29	1	58	0	12
9. *Picea engelmannii*	2.39	0.26	(0.05)	16	6	4	10	70	8	2
10. *Picea glauca*	2.13	0.26	(0.05)	29	33	0	2	18	47	0
11. *Picea mariana*	1.61	0.28	(0.05)	31	53	0	1	2	44	0
12. *Picea sitchensis*	1.82	0.15	(0.04)	10	13	76	1	2	1	7
13. *Pinus albicaulis*	2.06	0.11	(0.04)	39	3	25	47	22	3	0
14. *Pinus contorta*	1.91	0.14	(0.04)	46	15	28	17	30	9	1
15. *Pinus edulis*	1.93	0.21	-	9	0	3	0	46	0	51
16. *Pinus monticola*	1.73	0.20	(0.05)	43	0	80	9	6	0	5
17. *Pinus ponderosa*	2.55	0.16	(0.04)	7	0	26	0	56	2	16
18. *Populus tremuloides*	2.48	0.29	(0.05)	31	5	3	4	64	20	4
19. *Pseudotsuga menziesii*	2.27	0.17	(0.04)	64	0	46	5	35	3	11
20. *Thuja plicata*	1.09	0.04	(0.02)	10	13	63	0	15	0	9
21. *Tsuga heterophylla*	1.69	0.14	(0.05)	12	16	65	8	2	0	9
22. *Tsuga mertensiana*	1.56	0.09	(0.03)	19	3	91	1	2	0	3

Table 4.2: The area under the curve (*AUC*) of the receiver operating characteristic is provided for each species and each modelling method for each of the two GCMs (*GFDL* and *CCM1*): discriminant analysis (*DA*), minimum distance (*MD*), Random Forests (*RF*), and a consensus ensemble method (*Cons*). AUCs are also noted for an averaged probability of presence from both GCMs (*Both GCMs*) as well as for the model projections based on the 1991-2000 climate normals (*Normals*). The number of palaeoecological records used for the validation is also reported (n). CCM1 projections were validated with data from 6000, 11000, 14000, 16000, and 21000 years ago. GFDL projections were validated with data from 6000, 9000, 16000, and 21000 years ago. Models were validated with fossil pollen and macrofossil data compiled by Dyke (2005), the North American Pollen Database (1988), and the North American Packrat Midden Database (Strickland *et al.*, 2001). Average AUC values for each method only include species for which more than 10 records exist.

	Normals		CCM1 Model					GFDL Model					Both GCMs	
Species	n	AUC	n	DA	MD	RF	Cons	n	DA	MD	RF	Cons	n	AUC
Abies amabilis (pacific silver fir)	1617	0.94	1*	0.98	0.95	0.47	0.46	1*	0.98	0.97	0.97	0.97	1*	0.93
Abies lasiocarpa (subalpine fir)	10814	0.86	46	0.58	0.60	0.57	0.56	47	0.66	0.62	0.61	0.66	64	0.63
Abies procera (noble fir)	82	0.86	1*	0.50	0.49	0.49	0.49	1*	1.00	1.00	1.00	1.00	1*	0.99
Acer macrophyllum (bigleaf maple)	440	0.96	5*	0.67	0.68	0.67	0.67	6*	0.81	0.80	0.80	0.80	7*	0.82
Alnus rubra (red alder)	720	0.96	19	0.79	0.70	0.75	0.77	17	0.82	0.79	0.81	0.81	26	0.84
Calocedrus decurrens (incense cedar)	562	0.98	1*	0.99	0.49	0.49	0.49	2*	1.00	0.48	0.99	0.99	2*	0.99
Chamaecyparis nootkatensis (yellow cedar)	748	0.95	23	0.56	0.58	0.62	0.63	24	0.75	0.56	0.53	0.51	26	0.62
Larix occidentalis (western larch)	824	0.87	3*	0.49	0.49	0.48	0.48	4*	0.48	0.74	0.87	0.86	4*	0.86
Picea engelmannii (Engelman spruce)	6243	0.86	44	0.57	0.60	0.52	0.54	41	0.64	0.58	0.58	0.62	58	0.61
Picea glauca (white spruce)	7166	0.89	55	0.63	0.58	0.68	0.67	92	0.62	0.56	0.66	0.66	100	0.64
Picea mariana (black spruce)	2960	0.88	48	0.63	0.60	0.73	0.71	60	0.71	0.60	0.69	0.69	66	0.71
Picea sitchensis (Sitka spruce)	1040	0.96	31	0.78	0.82	0.77	0.81	32	0.78	0.76	0.75	0.79	45	0.82
Pinus albicaulis (whitebark pine)	1050	0.84	59	0.56	0.48	0.54	0.54	39	0.61	0.58	0.65	0.66	76	0.60
Pinus contorta (lodgepole pine)	11338	0.83	163	0.62	0.61	0.59	0.63	148	0.59	0.65	0.67	0.68	223	0.69
Pinus edulis (pinyon pine)	2839	0.89	13	0.61	0.50	0.54	0.58	14	0.60	0.49	0.57	0.64	19	0.65
Pinus monticola (western white pine)	830	0.90	18	0.63	0.55	0.56	0.55	14	0.63	0.57	0.66	0.66	22	0.60

Continued on next page

Table 4.2 – continued from previous page

	Normals		CCM1 Model					GFDL Model					Both GCMs	
Species	n	AUC	n	DA	MD	RF	Cons	n	DA	MD	RF	Cons	n	AUC
Pinus ponderosa (ponderosa pine)	3986	0.91	25	0.57	0.60	0.64	0.65	26	0.63	0.67	0.67	0.72	36	0.73
Populus tremuloides (trembling aspen)	7255	0.83	14	0.54	0.54	0.50	0.60	9	0.50	0.53	0.49	0.55	17	0.60
Pseudotsuga menziesii (Douglas-fir)	8898	0.89	174	0.66	0.65	0.69	0.70	201	0.76	0.74	0.81	0.82	263	0.82
Thuja plicata (western redcedar)	3838	0.94	29	0.85	0.67	0.82	0.83	30	0.83	0.73	0.76	0.74	35	0.81
Tsuga heterophylla (western hemlock)	4934	0.95	90	0.82	0.73	0.79	0.80	93	0.83	0.78	0.83	0.83	123	0.84
Tsuga mertensiana (mountain hemlock)	1168	0.92	66	0.76	0.68	0.71	0.76	46	0.72	0.68	0.76	0.74	79	0.78
Average (n > 10)		*0.90*		*0.65*	*0.62*	*0.66*	*0.67*		*0.69*	*0.64*	*0.68*	*0.69*		*0.71*

Table 4.3: Percentage of representation (%) of each species for each identified hotspot in Figure 4.1 at 21,000 years before the present, as projected in the average of all model runs for both paleoclimate simulations. Percentages <1% have been removed. The total count of species projected as present in a given hotspot (of the 22 included species, where representation was ≥1%) is also listed (Species count).

	North coast				South coast				Northern interior basins					Southern interior basins					HP	Mex
Species	Central AK	Western AK	Haida Gwaii	Vancouver Isl.	CA Coast	Klamath Mtns.	Sierra Nevada	S. CA Mtns.	Columbia Plateau	Northern Basin	Snake R. Plains	Willamette	WY Basin	AZ Mtns.	Colorado Plateau	Central Basin	NM Plateau	SW Tablelands	High Plains	Mexico
Mesic: coastal																				
A. amabilis	-	-	16	7	3	28	11	3	6	4	4	8	4	-	-	-	1	-	-	5
A. procera	-	-	-	-	5	51	25	9	-	-	-	-	-	1	-	2	-	-	-	7
A. macrophyllum	-	-	-	-	27	24	25	12	-	-	-	-	-	-	-	-	-	-	-	12
A. rubra	-	-	-	-	46	25	14	4	-	-	-	-	-	1	1	1	-	-	-	8
C. decurrens	-	-	-	-	13	15	23	26	-	1	-	-	-	4	-	3	-	-	-	17
C. nootkatensis	-	-	37	7	15	25	2	3	-	-	-	-	-	-	-	-	-	-	-	11
P. sitchensis	-	-	9	4	51	21	3	1	-	-	-	-	1	-	-	-	1	1	1	7
Mesic: coastal & interior																				
P. monticola	-	-	-	-	12	30	26	12	2	3	1	2	1	3	1	2	-	-	-	5
P. menziesii	-	-	-	-	13	13	11	9	-	2	1	1	1	9	5	6	3	12	3	11
T. plicata	-	-	9	4	18	33	9	3	-	-	-	-	-	1	-	-	-	14	-	9
T. heterophylla	-	-	11	5	26	30	7	2	2	1	1	3	1	-	-	-	-	2	-	9
T. mertensiana	-	-	2	1	7	48	28	8	-	1	-	-	-	1	-	1	-	-	-	3

Continued on next page

Table 4.3 – continued from previous page

	North coast				South coast				Northern interior basins					Southern interior basins					HP	Mex
Species	Central AK	Western AK	Haida Gwaii	Vancouver Isl.	CA Coast	Klamath Mtns.	Sierra Nevada	S. CA Mtns.	Columbia Plateau	Northern Basin	Snake R. Plains	Willamette	WY Basin	AZ Mtns.	Colorado Plateau	Central Basin	NM Plateau	SW Tablelands	High Plains	Mexico
Boreal & sub-boreal																				
P. glauca	-	19	7	7	-	-	-	-	-	-	-	-	2	-	1	1	1	15	47	-
P. mariana	-	18	17	17	-	-	-	-	-	-	-	-	1	-	-	-	1	1	43	-
P. contorta	3	3	5	4	5	10	8	5	3	3	3	3	5	4	6	7	8	5	9	1
P. tremuloides	-	1	2	2	-	1	-	2	1	1	1	-	1	18	10	10	5	21	20	4
Sub-alpine																				
A. lasiocarpa	5	6	1	-	-	5	2	1	9	9	8	10	9	3	6	4	9	5	7	1
P. engelmannii	3	3	-	-	-	1	1	2	2	2	2	2	2	20	14	13	10	13	8	2
P. albicaulis	1	2	-	-	4	7	7	7	10	10	8	12	7	3	4	3	4	8	3	-
Xeric & sub-xeric																				
L.occidentalis	-	-	-	-	1	3	10	15	-	1	-	-	-	26	7	9	-	16	-	12
P. edulis	-	-	-	-	-	-	-	3	-	-	-	-	-	24	3	2	1	16	-	51
P. ponderosa	-	-	-	-	5	2	6	13	-	-	-	-	-	20	3	9	1	23	2	16
Species Count (rate $\leq$ 1)	*4*	*7*	*11*	*9*	*16*	*18*	*18*	*20*	*7*	*11*	*9*	*8*	*12*	*16*	*12*	*15*	*12*	*14*	*10*	*19*

Table 4.4: Alternative splits for each major node of the regression trees (see Figure 4.2), explaining allelic richness (A) and expected heterozygosity (H_e). Alternative splitting variables are listed (Variable), as is the percent variance explained at each alternative split (VarEx) and to which side of the regression tree each group splits at each node (to the left and to the right).

Variable	VarEx	To the left	To the right
Allelic richness (A)			
Split 1			
LGM range area	66.1%	<3.5e5 km^2	$\geq$ 3.5e5 km^2
LGM refugia count	47.4%	<21.9	$\geq$ 21.9
Modern range patch count	41.8%	<58	$\geq$ 58
Modern range area	37.5%	<8.4e5 km^2	$\geq$ 8.4e5 km^2
Modern range contiguity	32.3%	$\geq$ 0.86	<0.86
Split 2			
LGM range contiguity	8.6%	<0.85	$\geq$ 0.85
Modern range contiguity	8.3%	<0.86	$\geq$ 0.86
LGM refugia count	8.2%	$\geq$ 15	<15
LGM range area	7.4%	<1.4e5 km^2	$\geq$ 1.4e5 km^2
Modern range area	6.6%	$\geq$ 5.5e5 km^2	<5.5e5 km^2
Split 3			
Dispersal mechanism	3.7%	all others	wind
LGM refugia count	3.6%	<64	$\geq$ 64
LGM contiguity	2.6%	$\geq$ 0.84	<0.84
Modern range patch count	1.6%	<101	$\geq$ 101
LGM range area	1.6%	<9.7e5 km^2	$\geq$ 9.7e5 km^2
Expected heterozygosity (H_e)			
Split 1			
LGM refugia count	26.7%	<64	$\geq$ 64
LGM range contiguity	25.2%	$\geq$ 0.84	<0.84
Modern range contiguity	14.7%	$\geq$ 0.86	<0.86
LGM range area	12.3%	<2.1e5 km^2	$\geq$ 2.1e5 km^2
Modern range patch count	9.4%	<101	$\geq$ 101
Split 2			
Modern range area	20.8%	$\geq$ 4.3e5 km^2	<4.3e5 km^2
LGM range contiguity	13.1%	$\geq$ 0.86	<0.86
Modern range contiguity	7.3%	$\geq$ 0.85	<0.85
Modern range patch count	7.2%	$\geq$ 20	<20
LGM range area	4.1%	$\geq$ 0.8e5 km^2	<0.8e5 km^2

Continued on next page

Table 4.4 – continued from previous page

Variable	VarEx	To the left	To the right
Split 3			
LGM range area	16.3%	$<$1.4e5 km^2	$\geq$ 1.4e5 km^2
LGM refugia count	7.5%	$<$22	$\geq$ 22
Succession	7.1%	early, late	middle
LGM range contiguity	5.5%	$\geq$ 0.86	$\geq$ 0.86
Modern range patch count	5.2%	$<$36	$\geq$ 36

Table 4.5: The area under the curve (AUC) of the receiver operating characteristic is provided for each species and each modelling method for each of the two AOGCMs (GFDL and CCM1): discriminant analysis (DA), minimum distance (MD), Random Forests (RF), and the averaged ensemble method (Ens). AUCs are also noted for model projections based on the 1961-1990 observed climate (Modern). The number of paleoecological records used for the validation is also reported (N). CCM1 projections were validated with data from 6,000, 11,000, 14,000, 16,000, and 21,000 years ago. GFDL projections were validated with data from 6,000, 9,000, 16,000, and 21,000 years ago. The number of species presence records in the validation data is listed for the present day plots (N_{SP}) and for both date combinations of pollen data (N_{P1} and N_{P2}).

	Modern					CCM1 model					GFDL model				
	NSP	DA	MD	RF	Ens	NP1	DA	MD	RF	Ens	NP2	DA	MD	RF	Ens
Abies amabilis	1615	0.93	0.83	0.93	0.97	3	0.97	0.79	0.63	0.95	3	0.97	0.79	0.80	0.97
Abies lasiocarpa	10804	0.82	0.72	0.82	0.86	83	0.63	0.66	0.66	0.70	81	0.69	0.68	0.69	0.78
Abies procera	82	0.83	0.68	0.85	0.91	1	0.49	0.49	0.49	0.48	1	1.00	0.99	1.00	1.00
Acer macrophyllum	437	0.95	0.94	0.96	0.97	9	0.70	0.69	0.69	0.69	11	0.75	0.74	0.73	0.74
Alnus rubra	715	0.96	0.94	0.96	0.97	24	0.80	0.73	0.76	0.82	22	0.82	0.80	0.81	0.81
Calocedrus decurrens	561	0.96	0.88	0.98	0.99	2	0.99	0.74	0.74	0.99	3	1.00	0.65	0.99	1.00
Chamaecyparis nootkatensis	707	0.94	0.93	0.94	0.97	64	0.62	0.62	0.64	0.67	65	0.69	0.61	0.61	0.72
Larix occidentalis	821	0.86	0.67	0.87	0.92	6	0.66	0.57	0.65	0.73	7	0.63	0.70	0.84	0.91
Picea engelmannii	6223	0.80	0.69	0.81	0.86	87	0.62	0.60	0.63	0.70	79	0.66	0.60	0.67	0.76
Picea glauca	7115	0.87	0.83	0.88	0.90	108	0.68	0.64	0.72	0.73	145	0.66	0.61	0.69	0.69
Picea mariana	2922	0.88	0.80	0.89	0.89	87	0.70	0.66	0.76	0.77	99	0.73	0.65	0.73	0.75
Picea sitchensis	1016	0.94	0.90	0.95	0.98	55	0.74	0.80	0.76	0.83	56	0.74	0.77	0.74	0.81
Pinus albicaulis	1038	0.77	0.70	0.77	0.85	82	0.58	0.54	0.58	0.63	61	0.63	0.63	0.67	0.75
Pinus contorta	11275	0.81	0.72	0.82	0.83	234	0.64	0.63	0.63	0.69	219	0.62	0.66	0.69	0.73
Pinus edulis	2836	0.80	0.61	0.83	0.90	44	0.61	0.51	0.55	0.63	45	0.63	0.53	0.59	0.70
Pinus monticola	820	0.87	0.74	0.89	0.92	28	0.62	0.61	0.59	0.69	24	0.62	0.63	0.67	0.68
Pinus ponderosa	3967	0.83	0.76	0.88	0.91	55	0.68	0.59	0.68	0.72	54	0.67	0.62	0.67	0.71
Populus tremuloides	7241	0.82	0.78	0.83	0.85	33	0.57	0.55	0.55	0.58	24	0.60	0.59	0.60	0.61
Pseudotsuga menziesii	8808	0.85	0.79	0.88	0.90	306	0.66	0.66	0.68	0.72	320	0.73	0.72	0.77	0.80
Thuja plicata	3798	0.93	0.81	0.94	0.95	69	0.88	0.80	0.87	0.89	70	0.87	0.82	0.85	0.89
Tsuga heterophylla	4860	0.93	0.87	0.95	0.97	164	0.84	0.77	0.82	0.87	167	0.84	0.80	0.84	0.87
Tsuga mertensiana	1136	0.90	0.80	0.86	0.94	98	0.78	0.71	0.73	0.81	78	0.75	0.71	0.76	0.80
Average ($n \geq 10$)		0.87	0.79	0.89	0.92		0.69	0.65	0.68	0.73		0.71	0.68	0.71	0.75

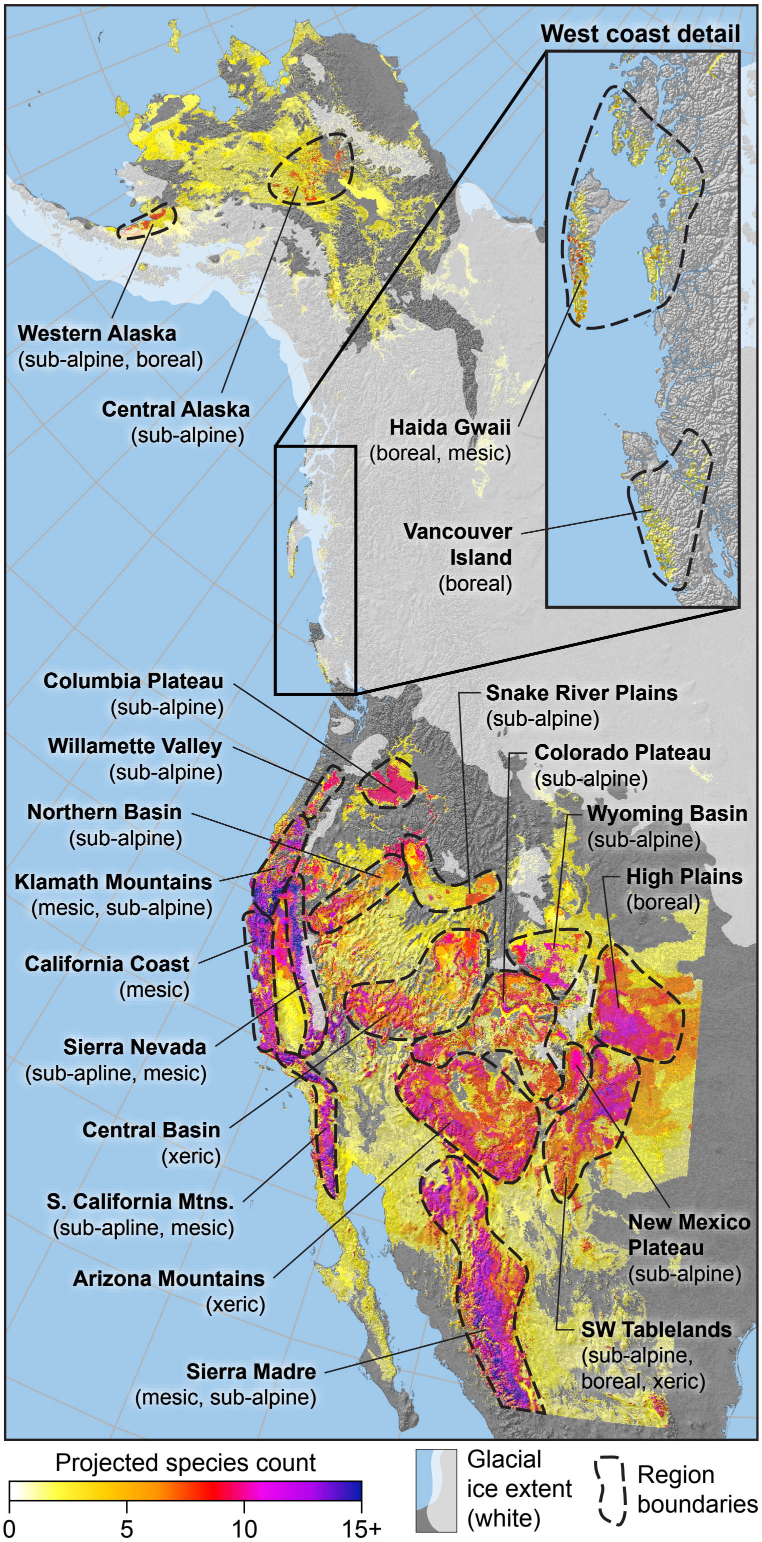

Figure 4.1: The number of tree species modelled as present at the last glacial maximum, averaged across six projections based on two GCM palaeoclimate simulations (GFDL and CCM1). Locations of higher species richness (i.e. hotspots) in the models have been labelled. For each noted hotspot, relative species occupancy is listed in Table 4.3. Semi-transparent white area represent glaciated area at the LGM, as per Dyke (2002) (omitted for clarity in the 'west coast detail').

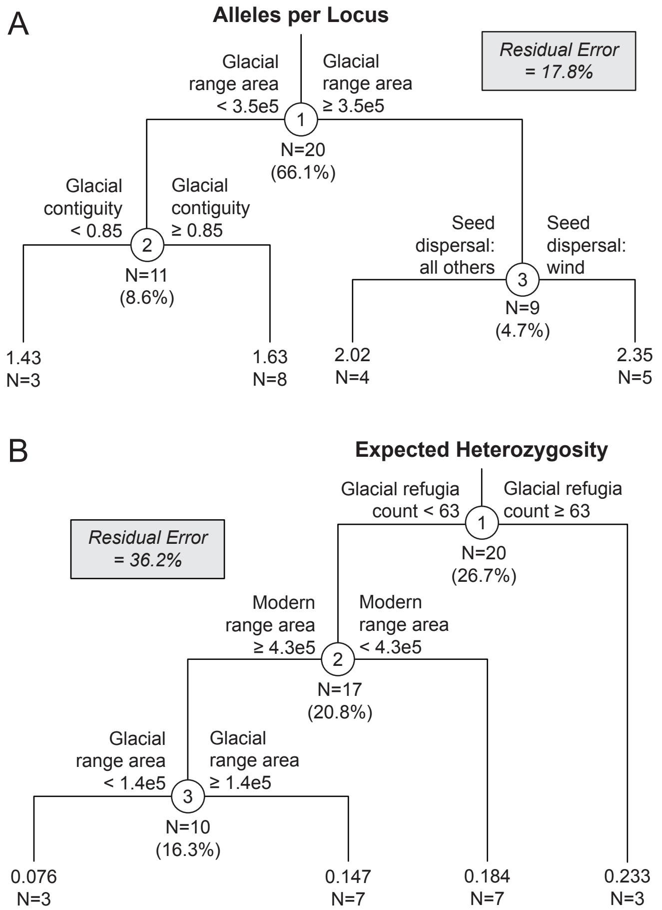

Figure 4.2: Regression tree analyses of species genetic diversity, as measured by **(A)** allelic richness and **(B)** expected heterozygosity. Three landscape metrics were included for the present day (Modern) and last glacial maximum (LGM): total range area, range contiguity (ratio of perimeter/area), and the number of habitat patches/refugia. We also included two categorical life history traits: successional stage and seed dispersal mechanism. The variance explained by each split is noted in the tree diagrams in parentheses. Mean values of A and H_e for each terminal group are noted at the terminal nodes, as is the number of species in each terminal group (N). Alternative splits for each node and corresponding variance explained are list in Table 4.4.

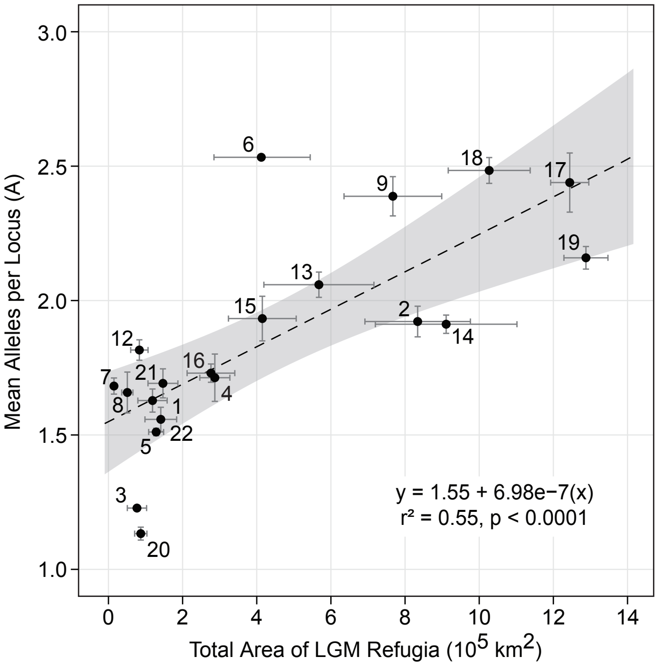

Figure 4.3: Allelic richness as a function of average modelled species range size at the last glacial maximum. Points represent the average modelled value of six projections based on two GCM-based models. Error bars represent standard errors and the shaded area shows the 95% confidence interval of the regression. Species are noted by numbers which correspond to Table 4.1.

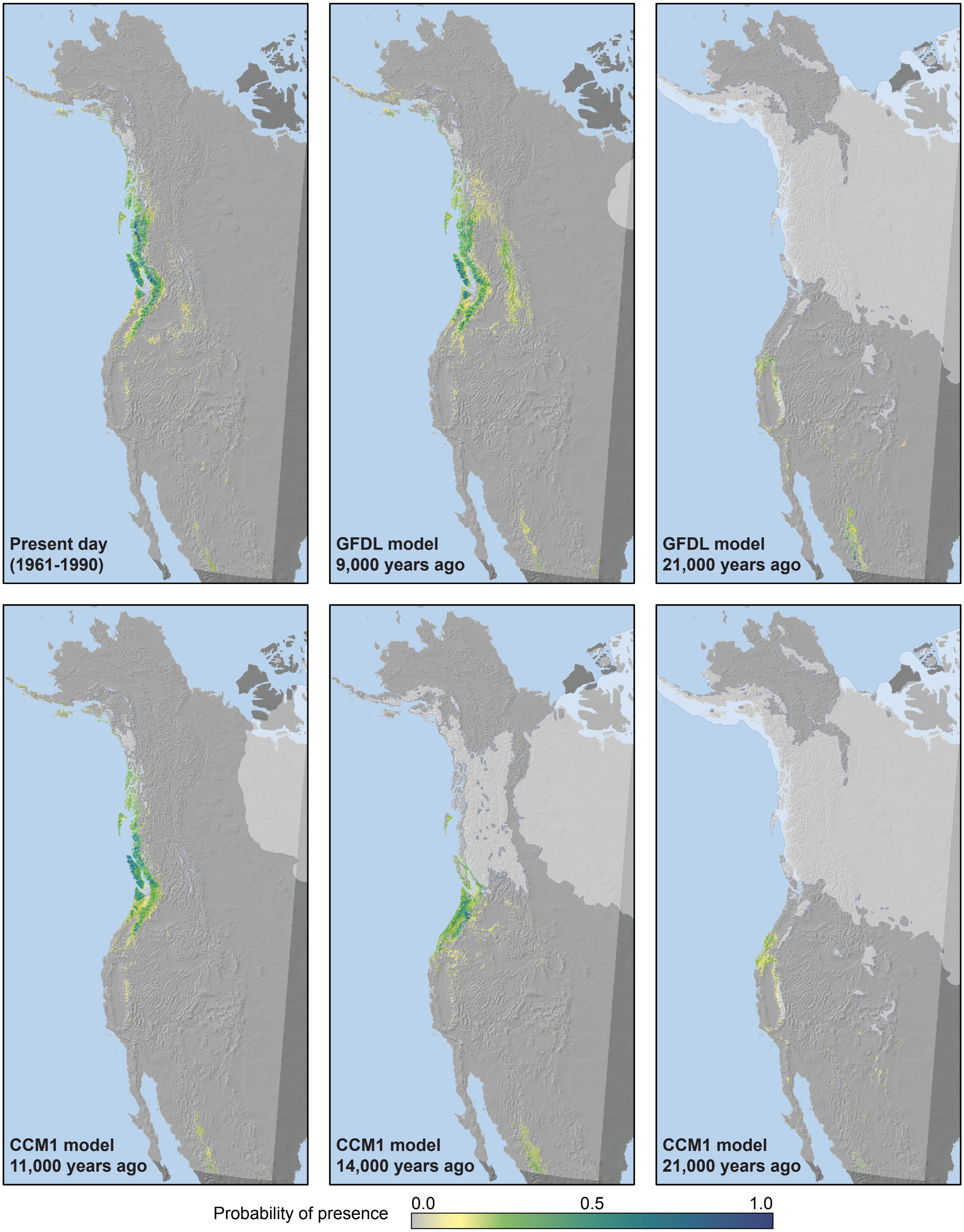

Figure S4.1A: *Abies amabilis* (Pacific silver fir) probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

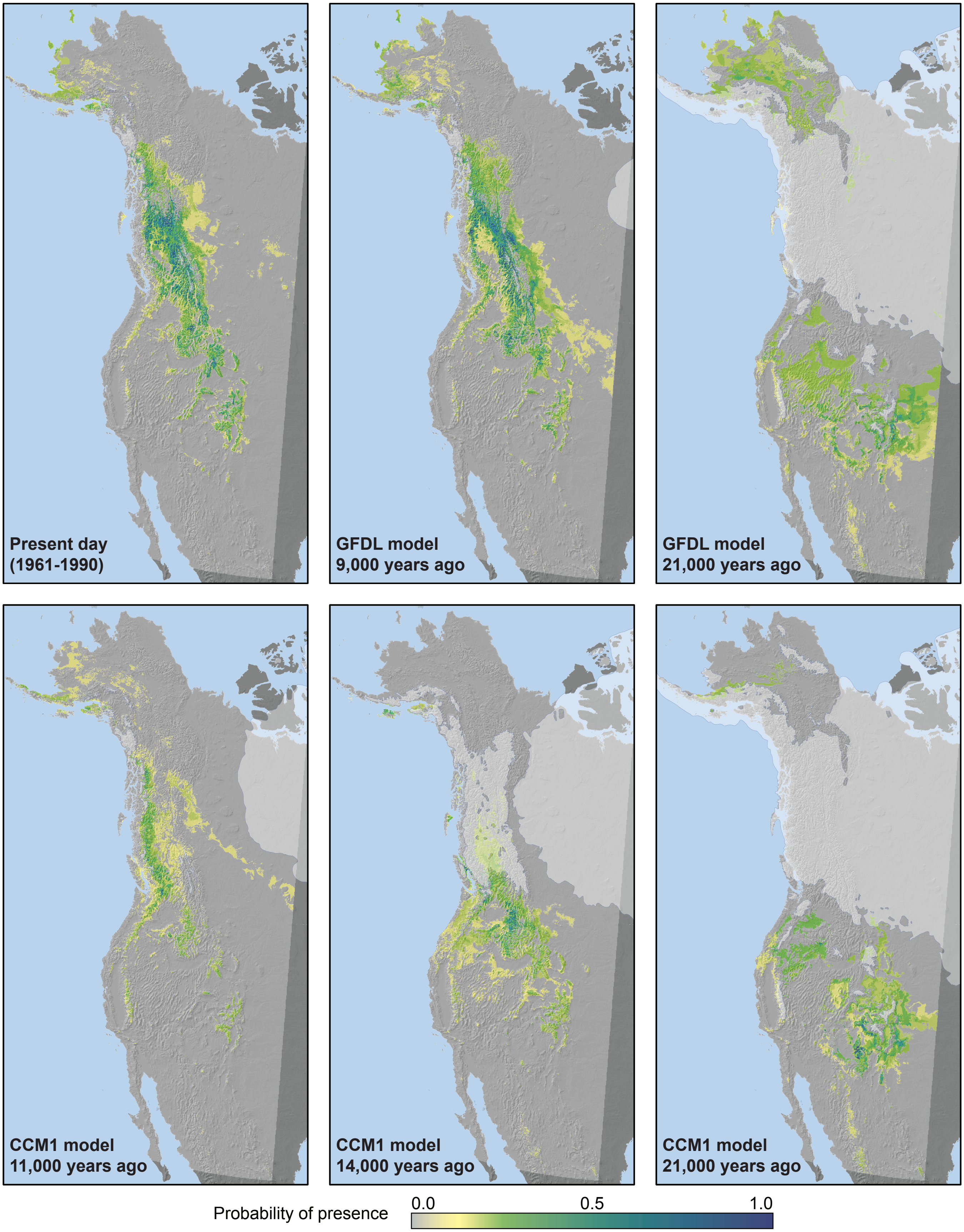

Figure S4.1B: ***Abies lasiocarpa*** **(subalpine fir)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

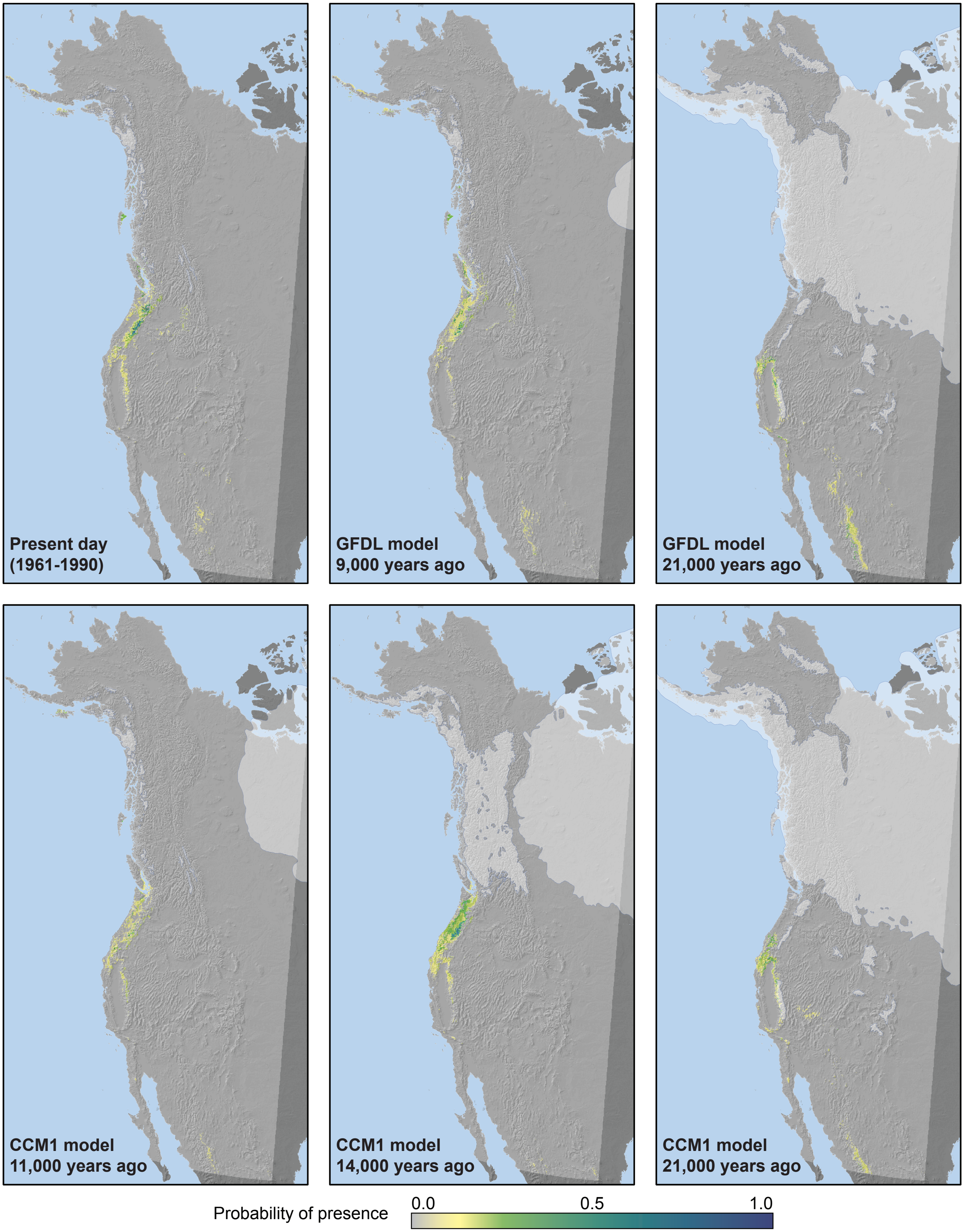

Figure S4.1C: ***Abies procera*** **(noble fir)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

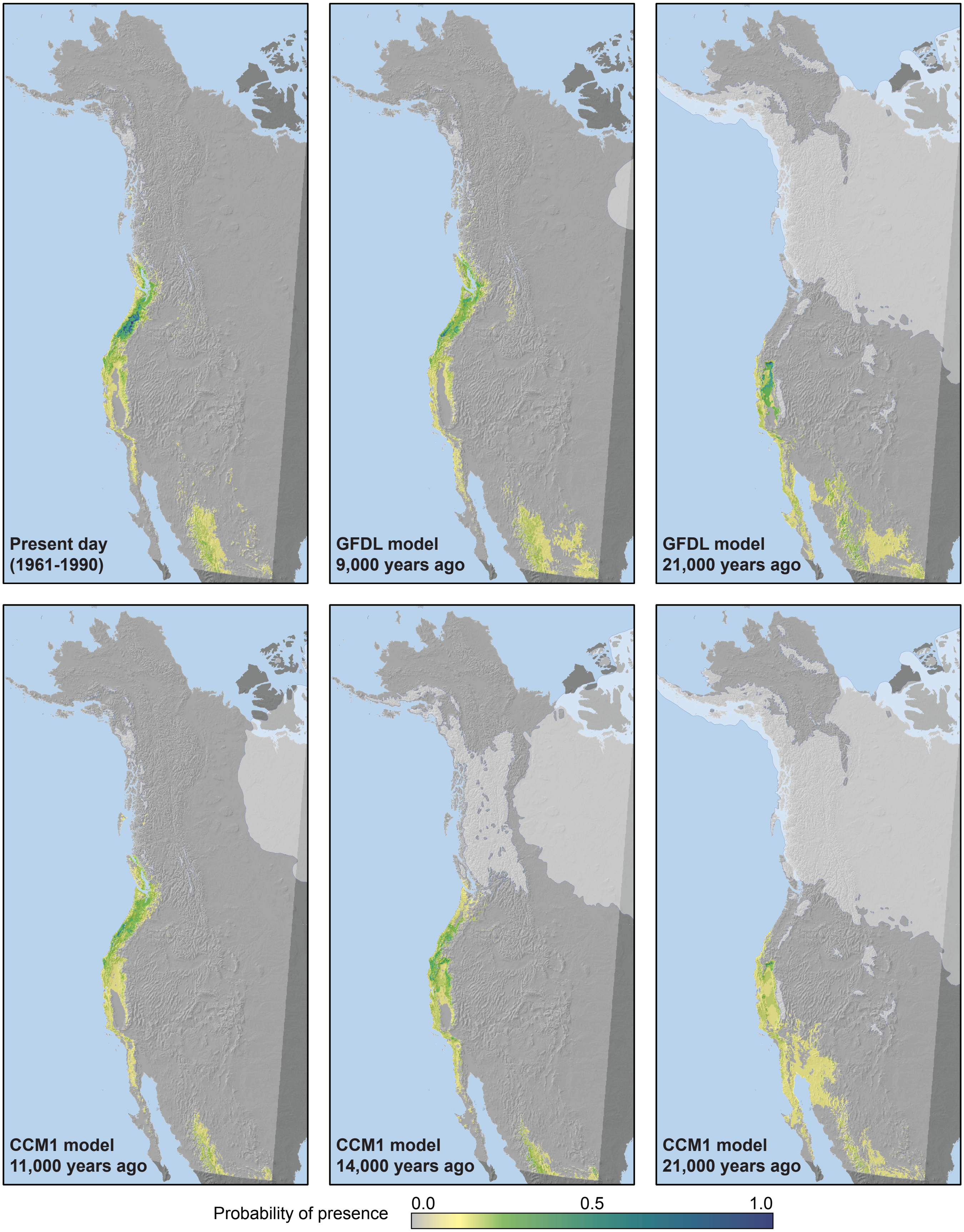

Figure S4.1D: ***Acer macrophyllum*** **(bigleaf maple)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

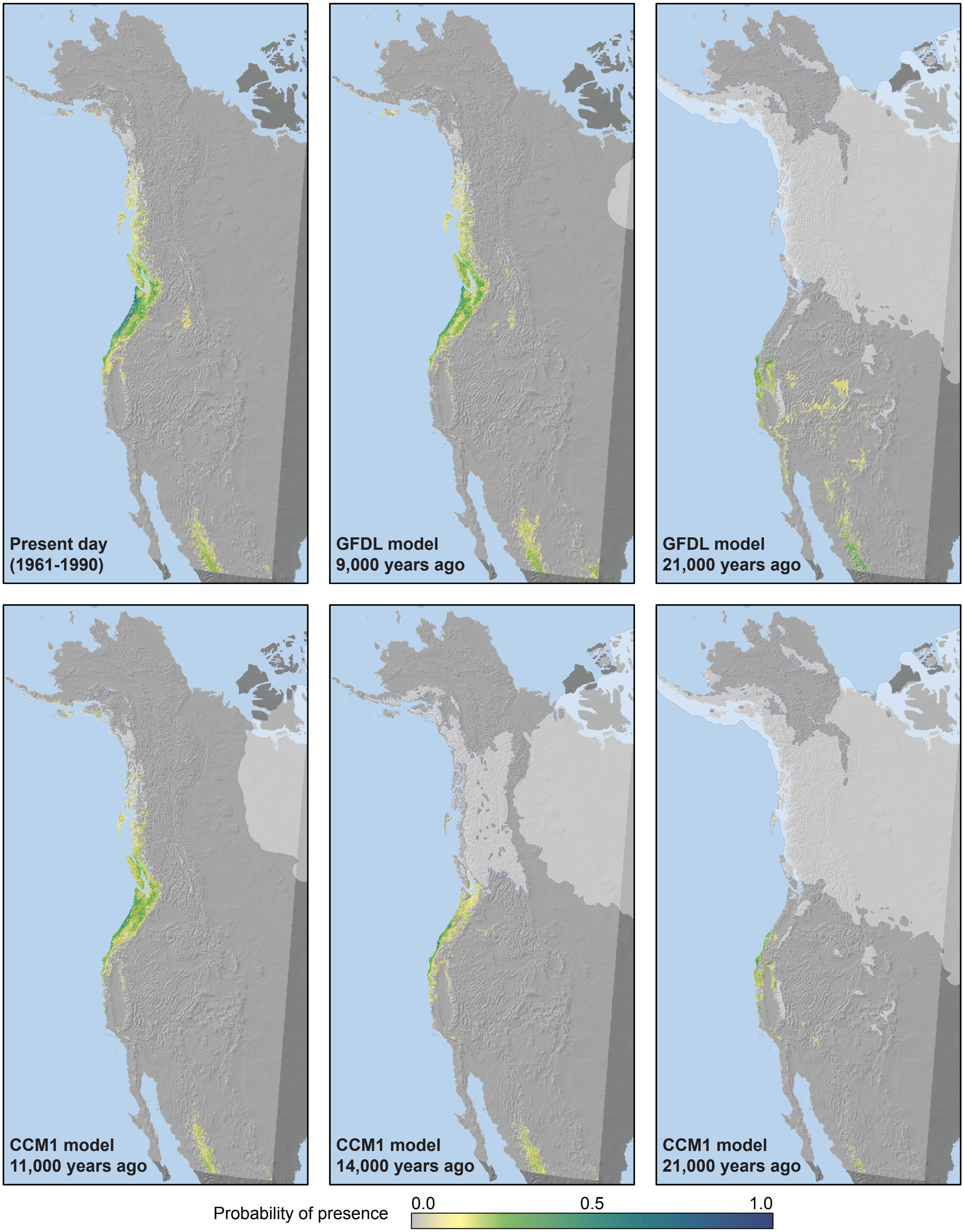

Figure S4.1E: *Alnus rubra* (red alder) probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

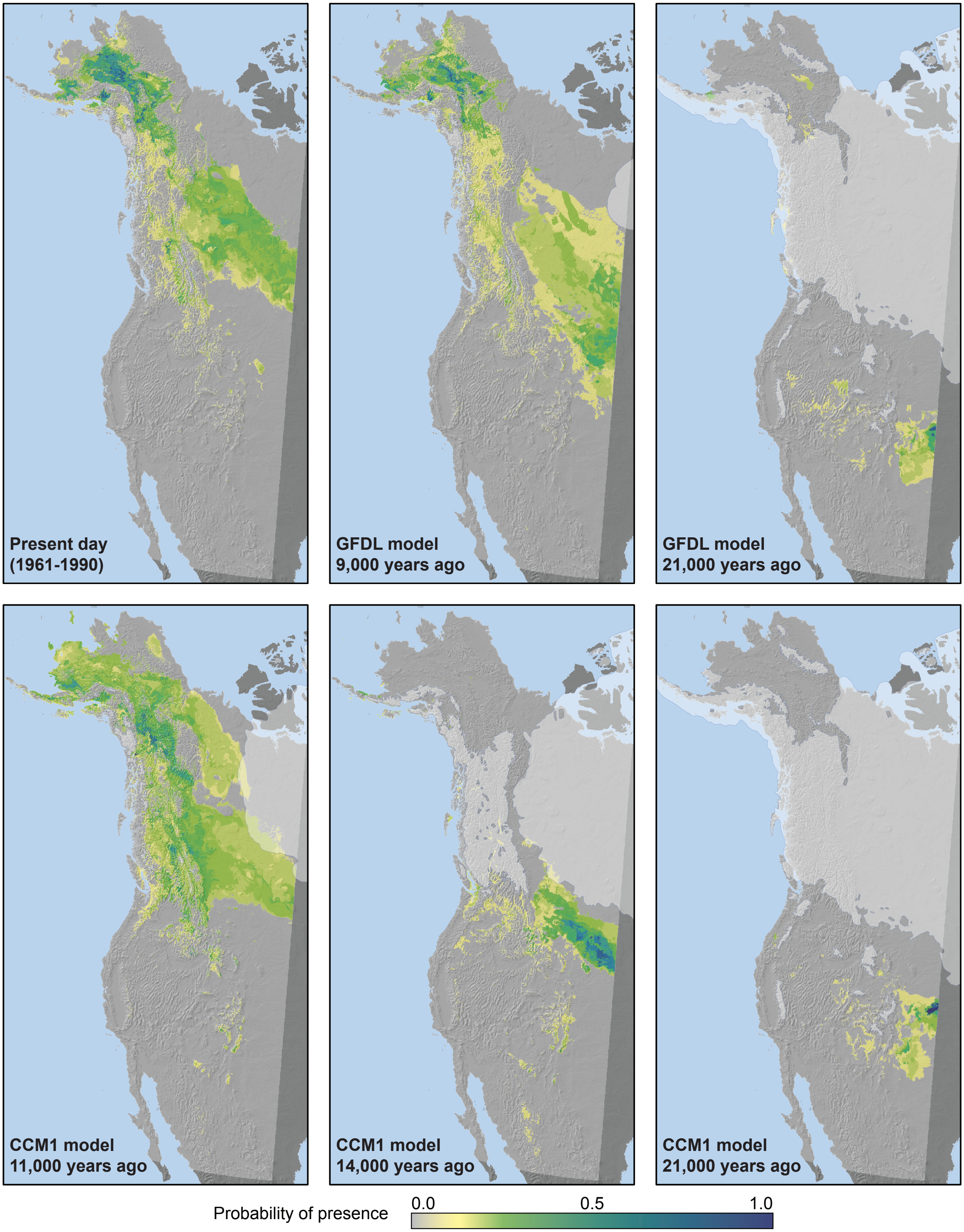

Figure S4.1F: ***Betula papyrifera*** **(paper birch)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

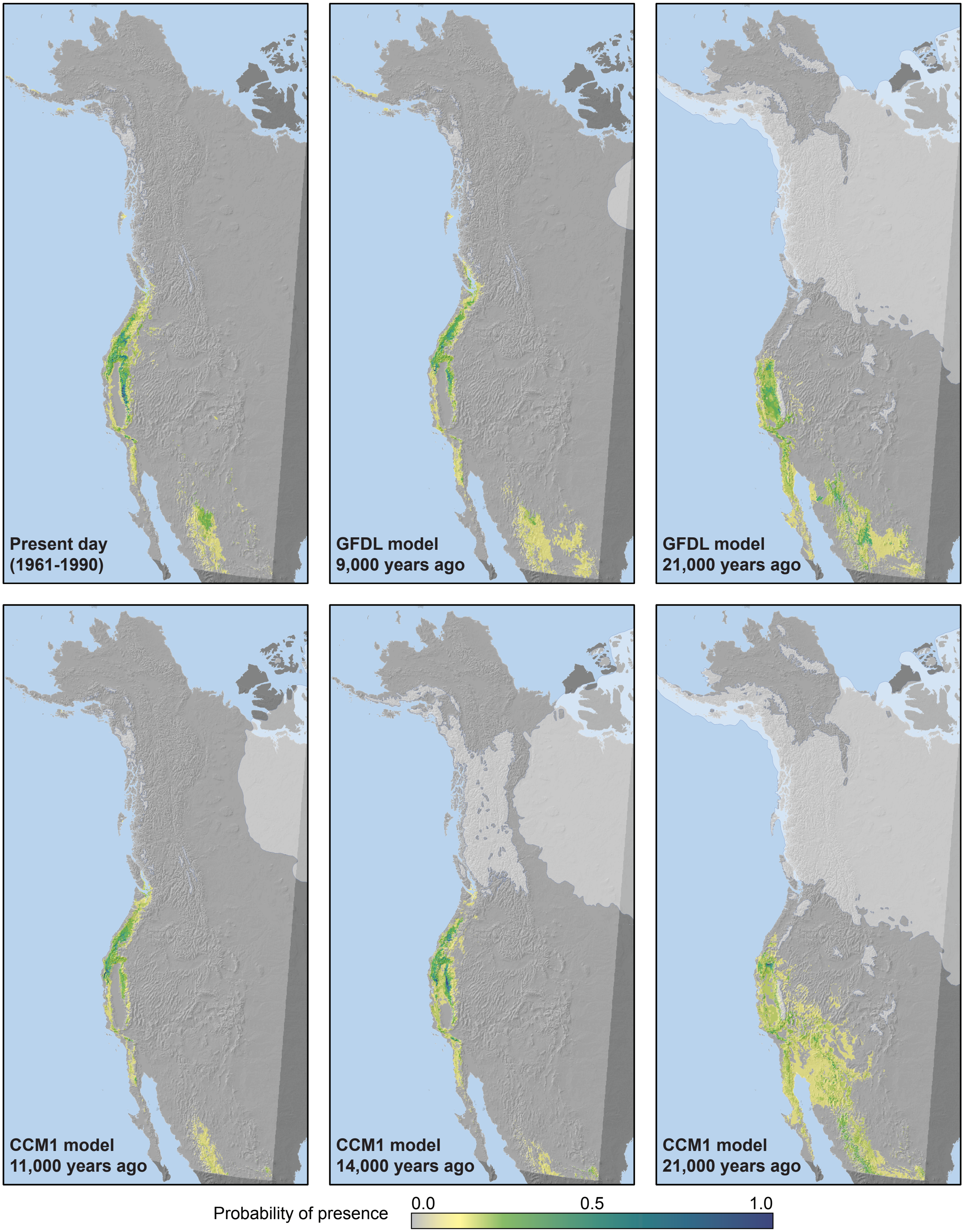

Figure S4.1G: *Calocedrus decurrens* (incense cedar) probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

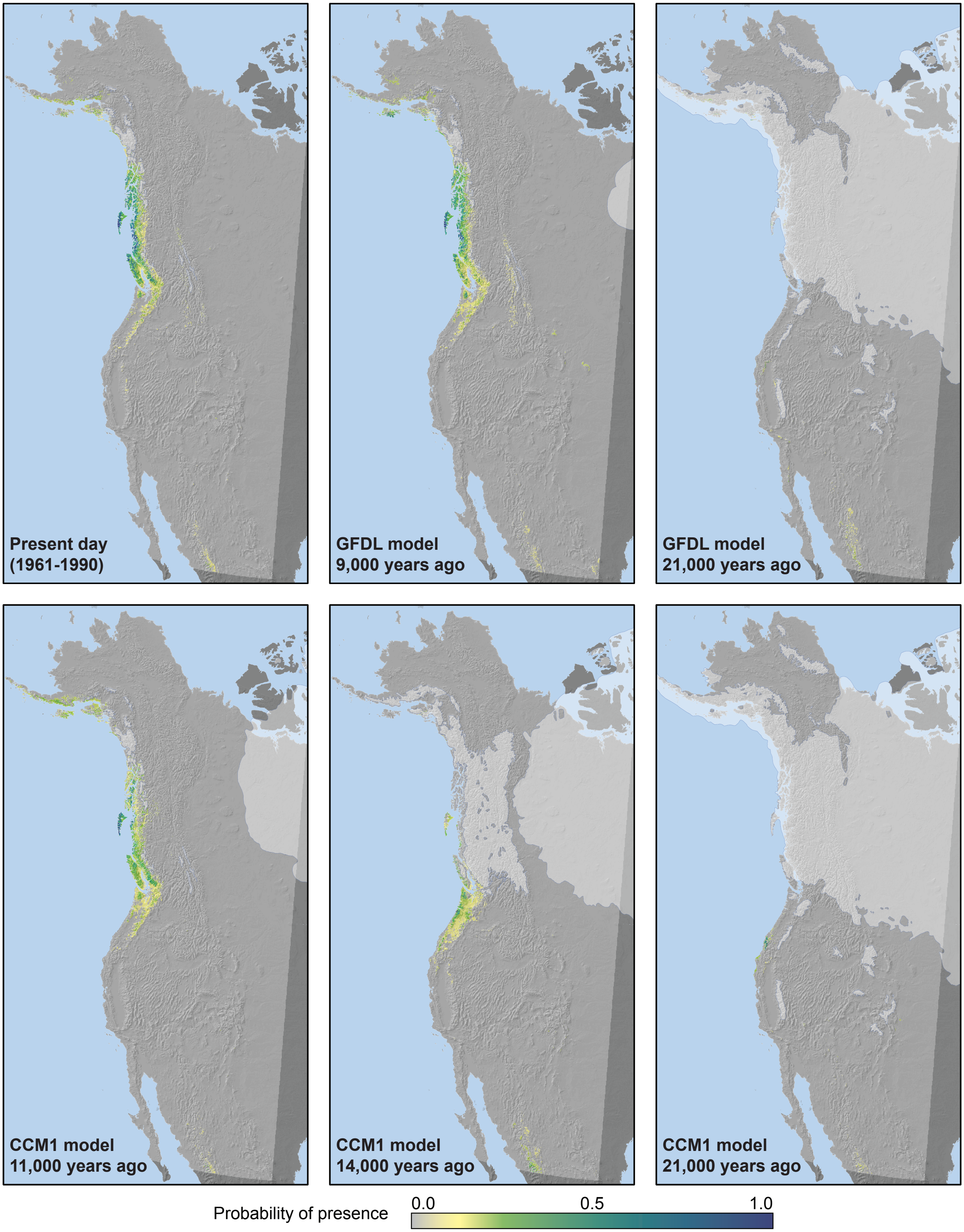

Figure S4.1H: ***Chamaecyparis nootkatensis*** **(Alaska-cedar)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

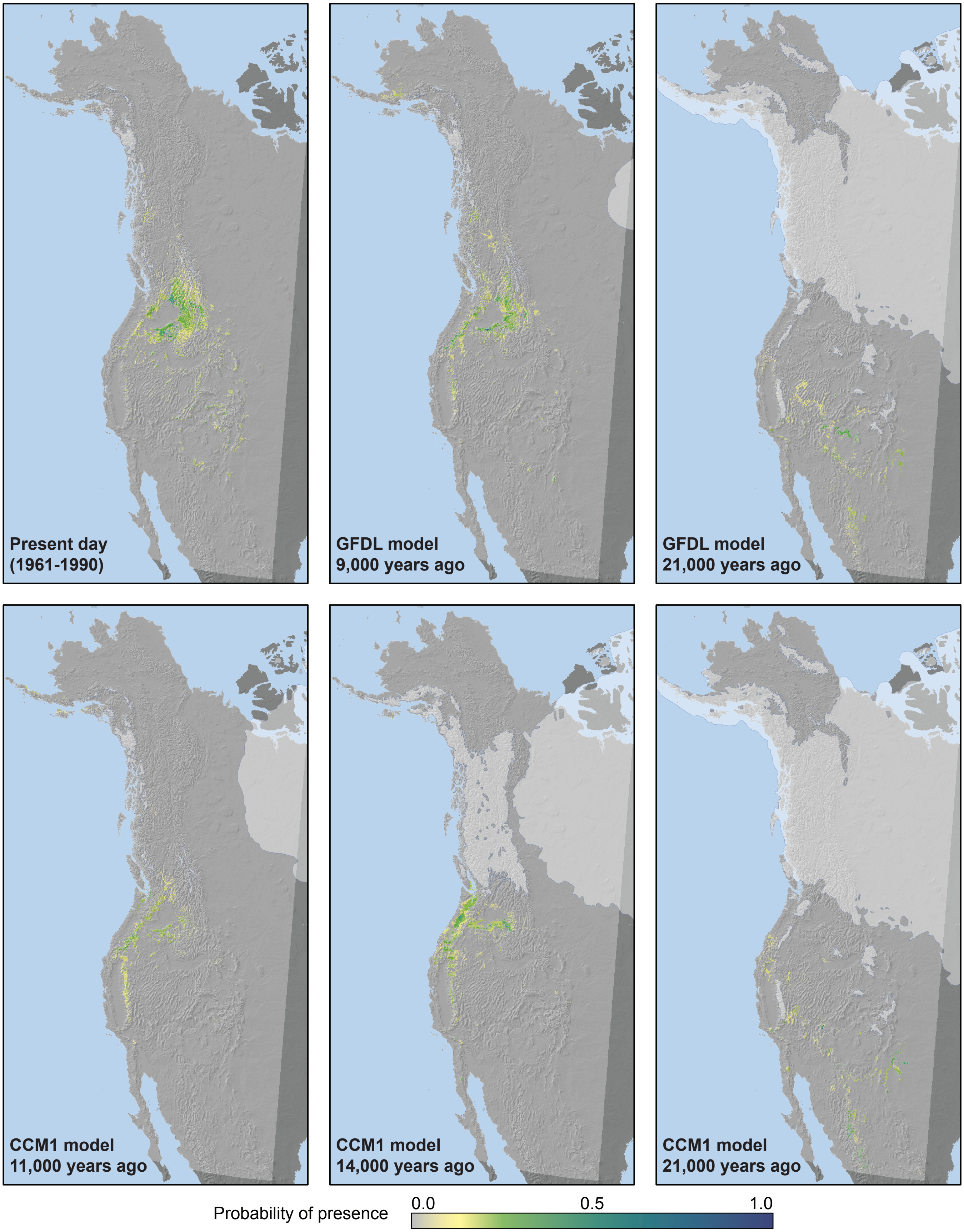

Figure S4.1I: ***Larix occidentalis* (western larch)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

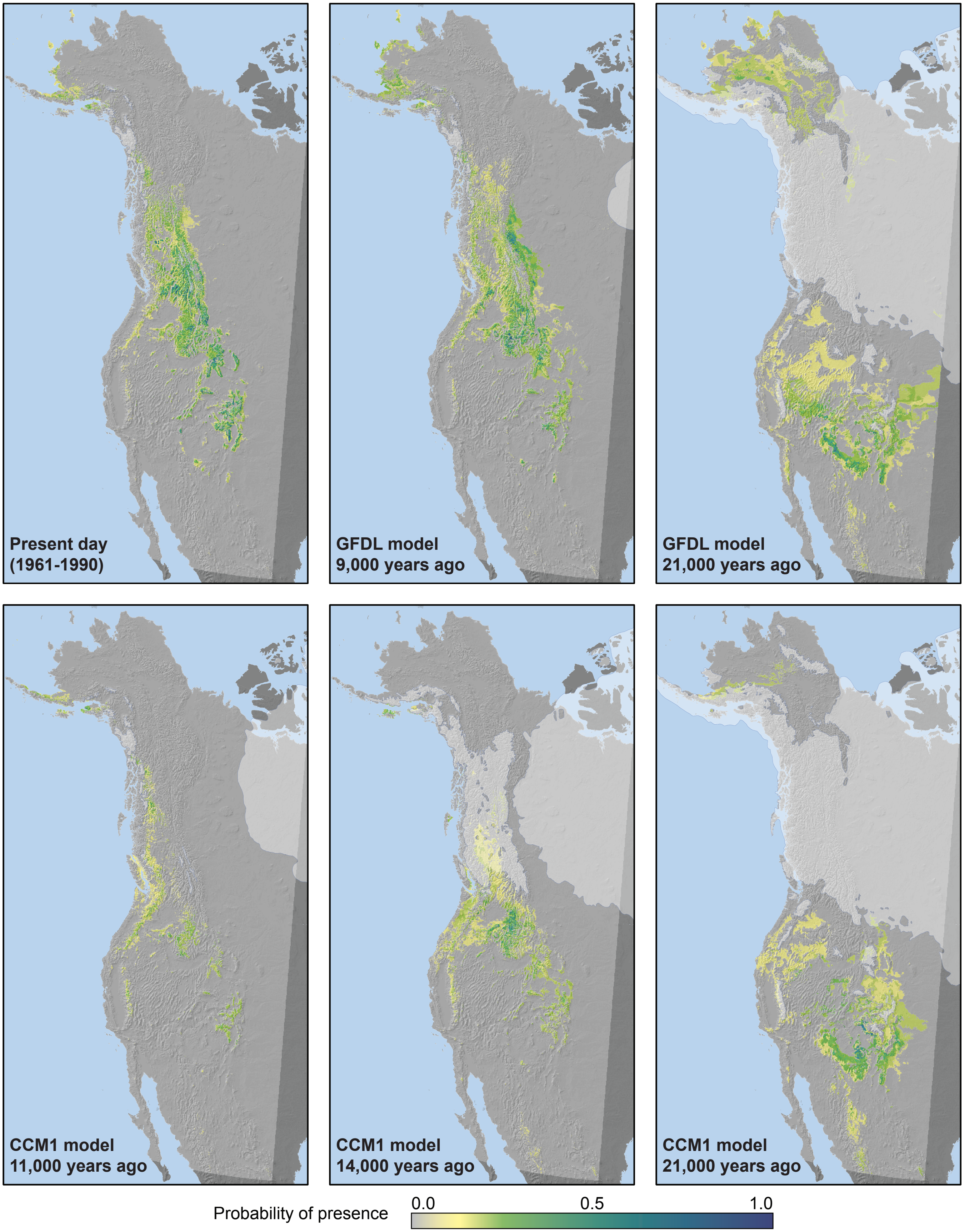

Figure S4.1J: *Picea engelmannii* (Engelmann spruce) probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

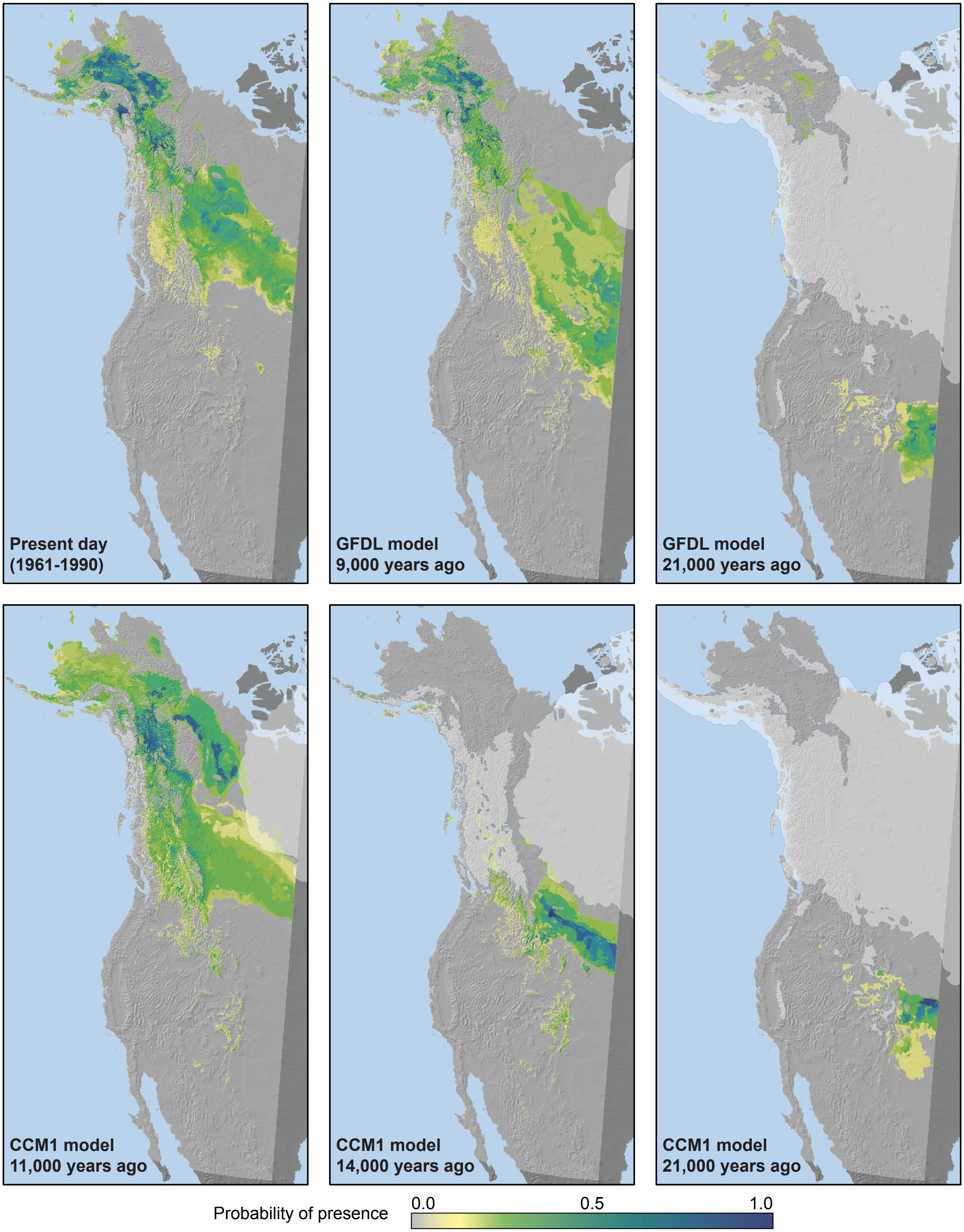

Figure S4.1K: *Picea glauca* (white spruce) probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

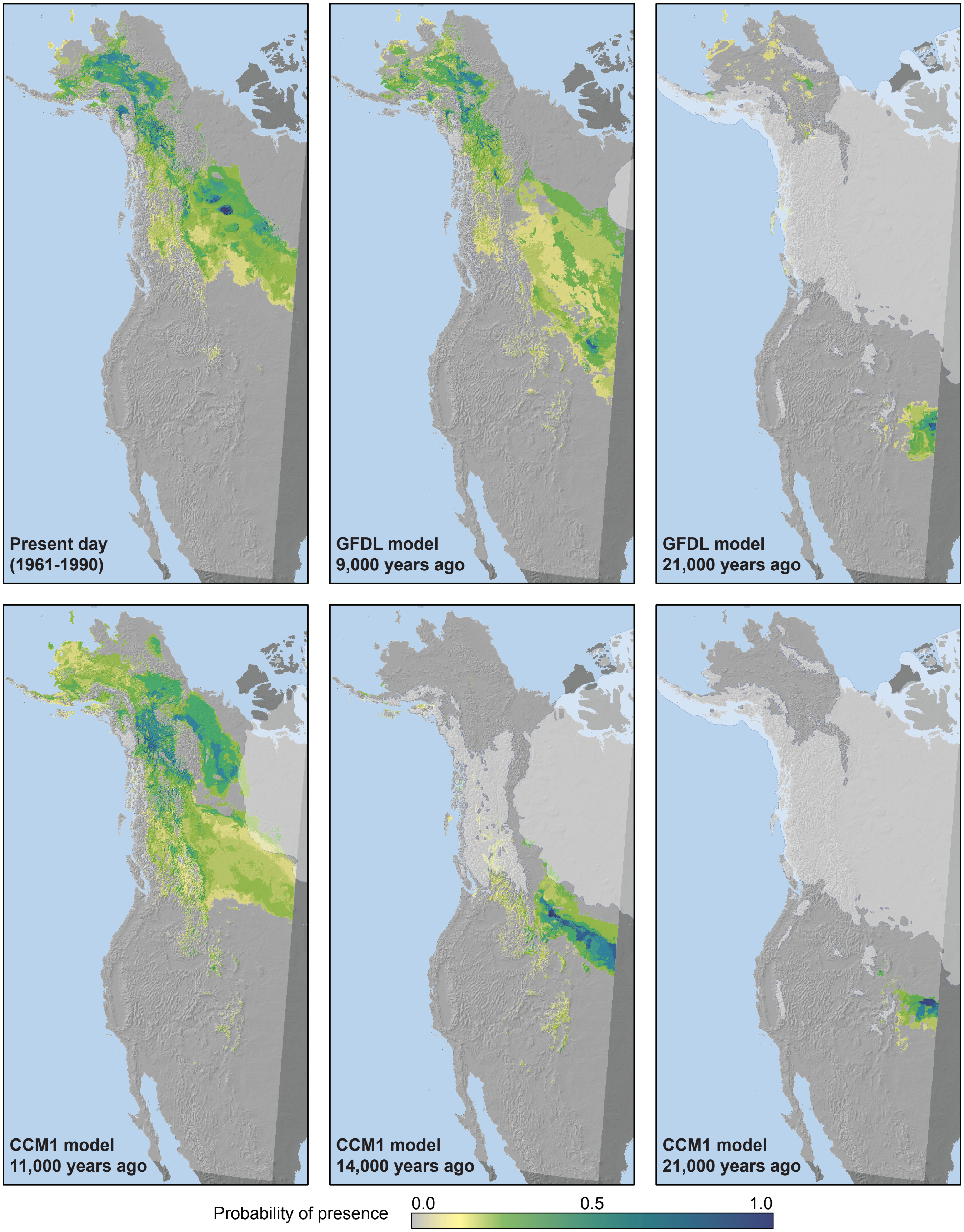

Figure S4.1L: *Picea mariana* (black spruce) probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

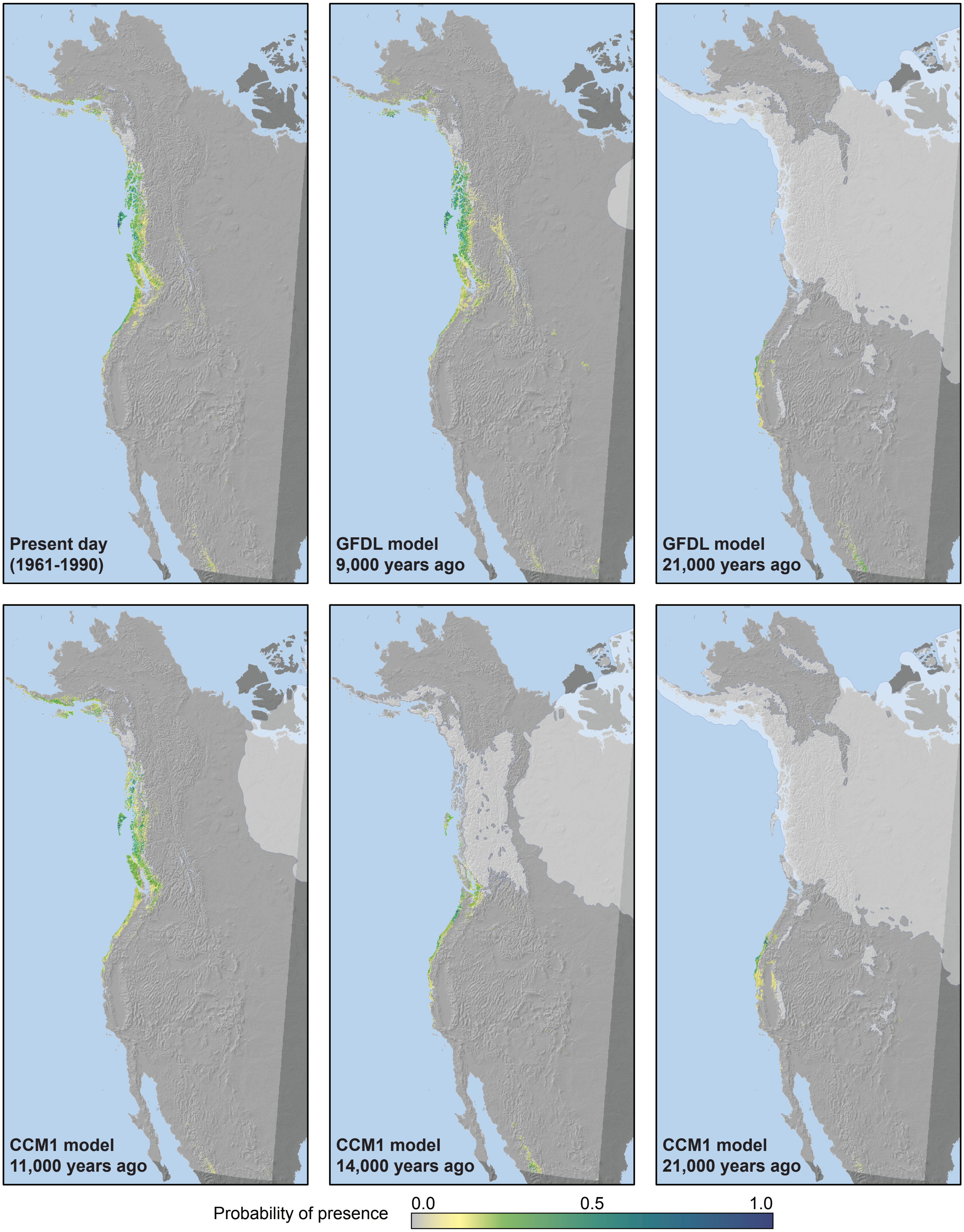

Figure S4.1M: ***Picea sitchensis*** **(Sitka spruce)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

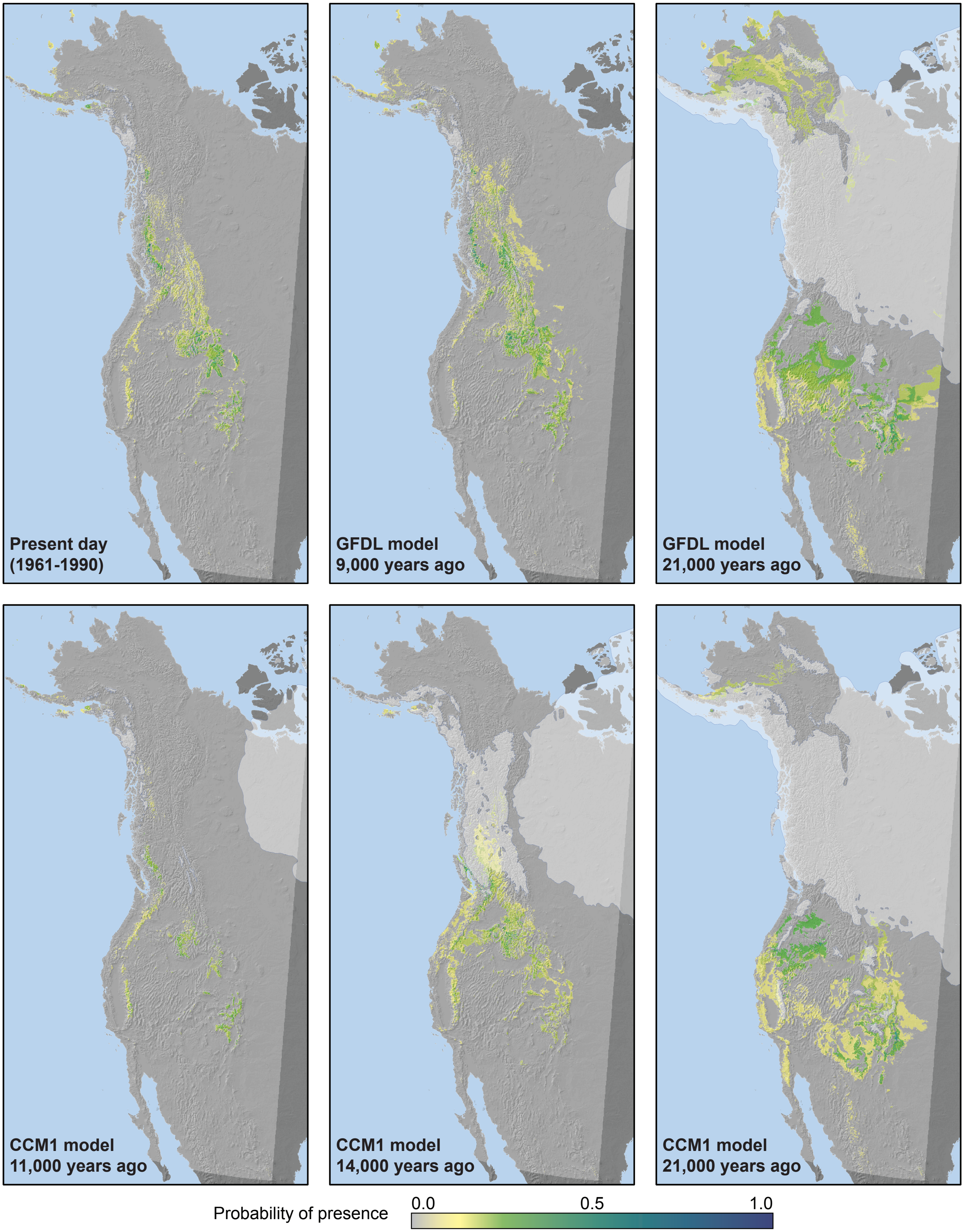

Figure S4.1N: *Pinus albicaulis* (whitebark pine) probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

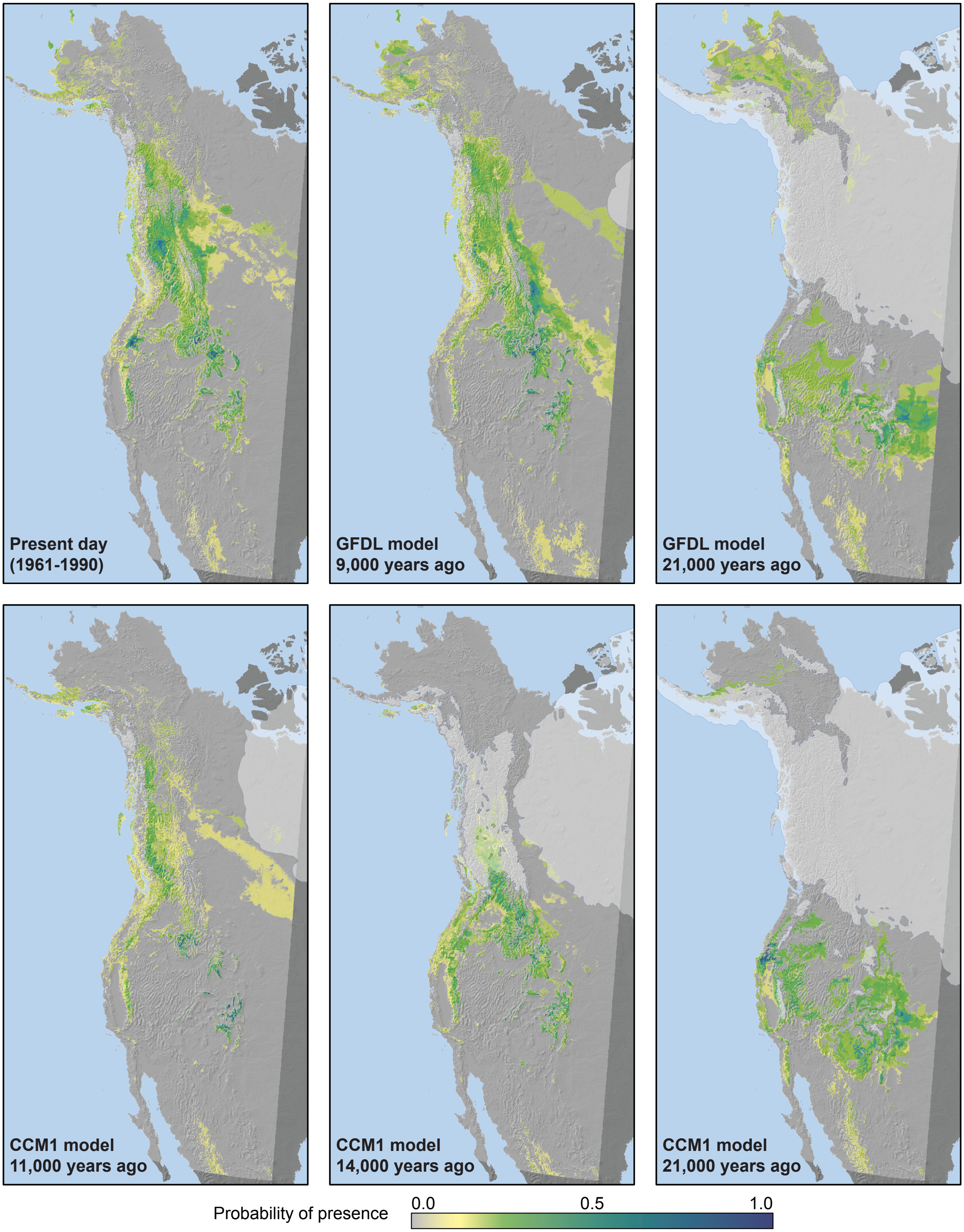

Figure S4.1O: ***Pinus contorta* (lodgepole pine)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

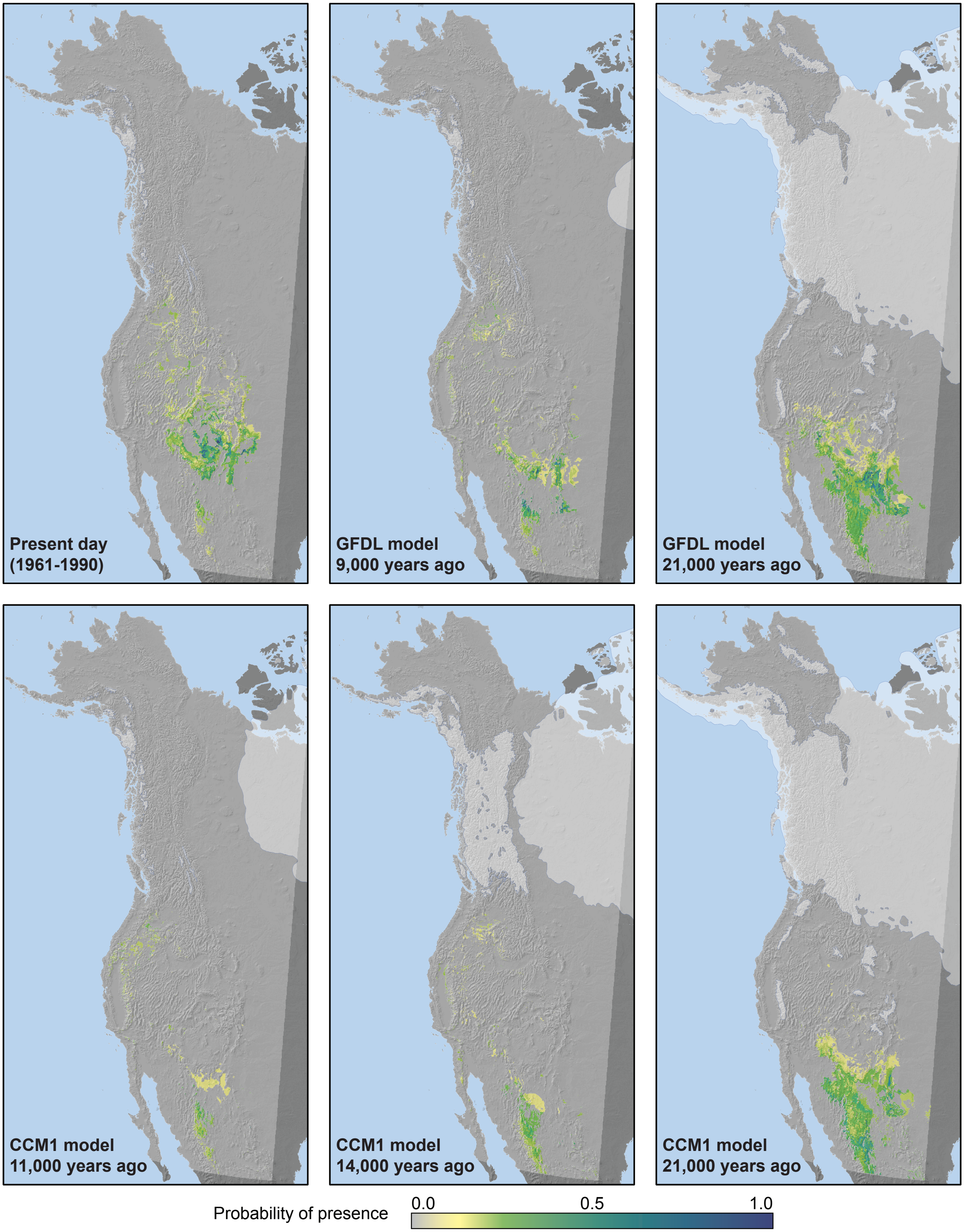

Figure S4.1P: *Pinus edulis* (pinyon pine) probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

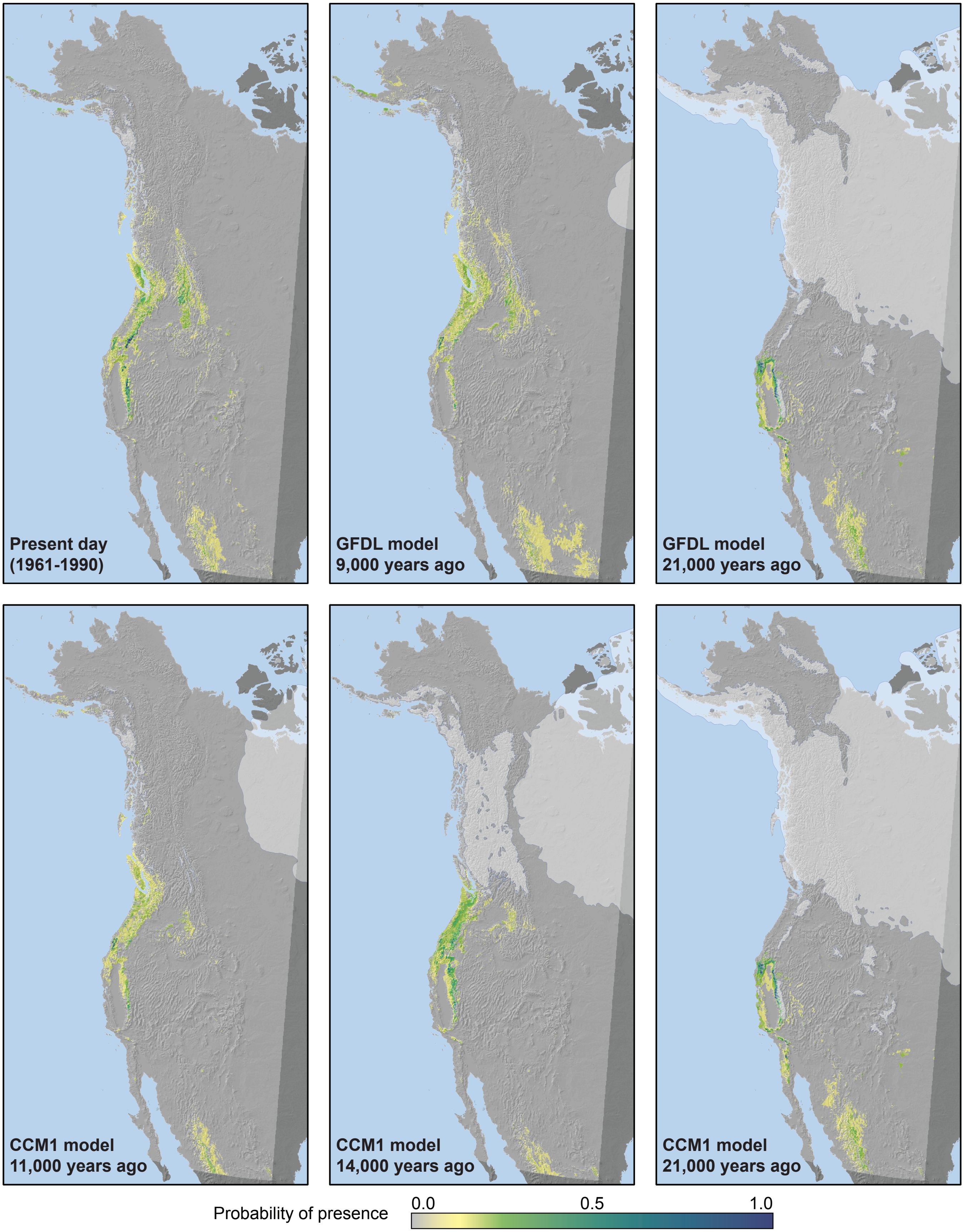

Figure S4.1Q: ***Pinus monticola* (western white pine)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

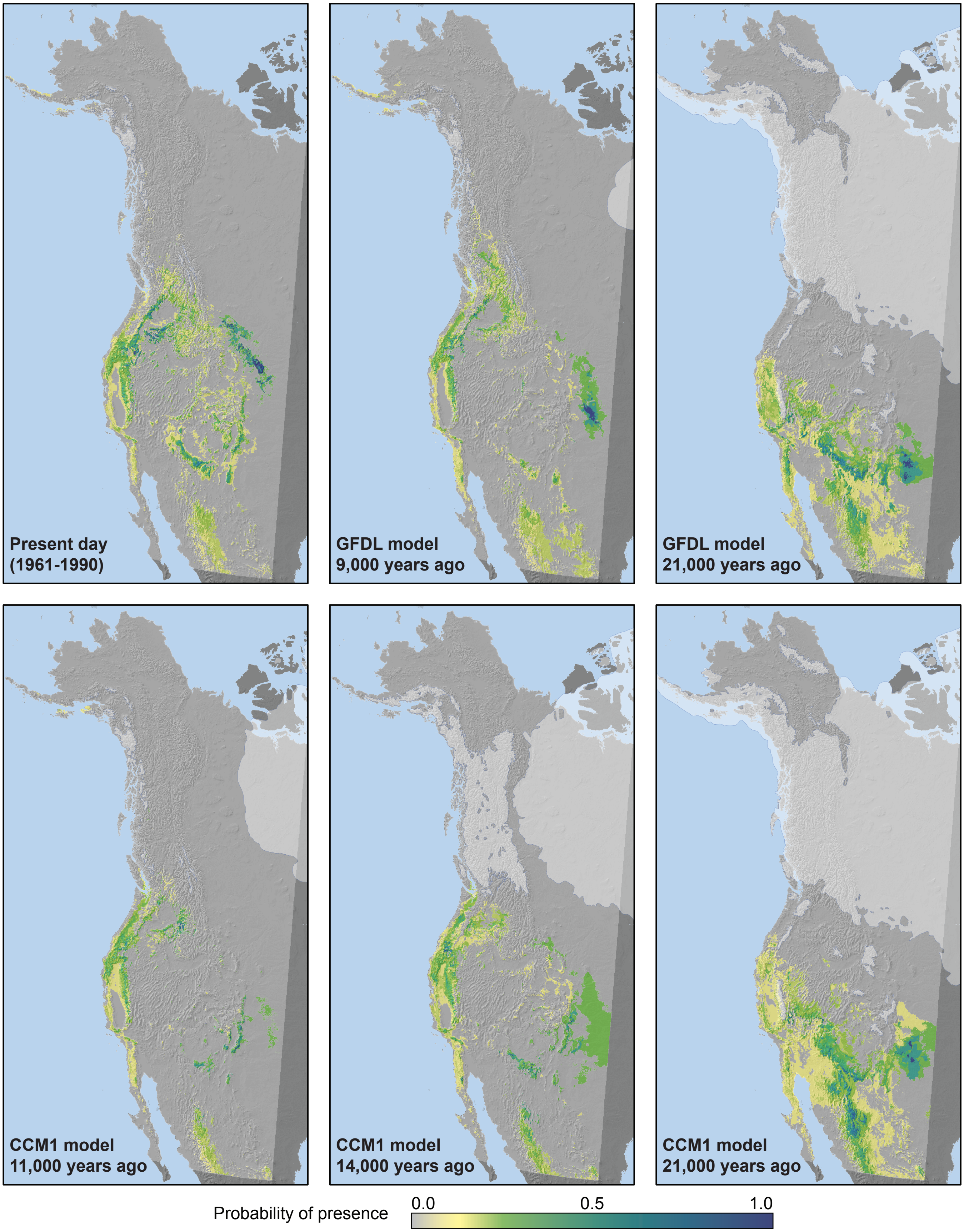

Figure S4.1R: ***Pinus ponderosa*** **(ponderosa pine)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

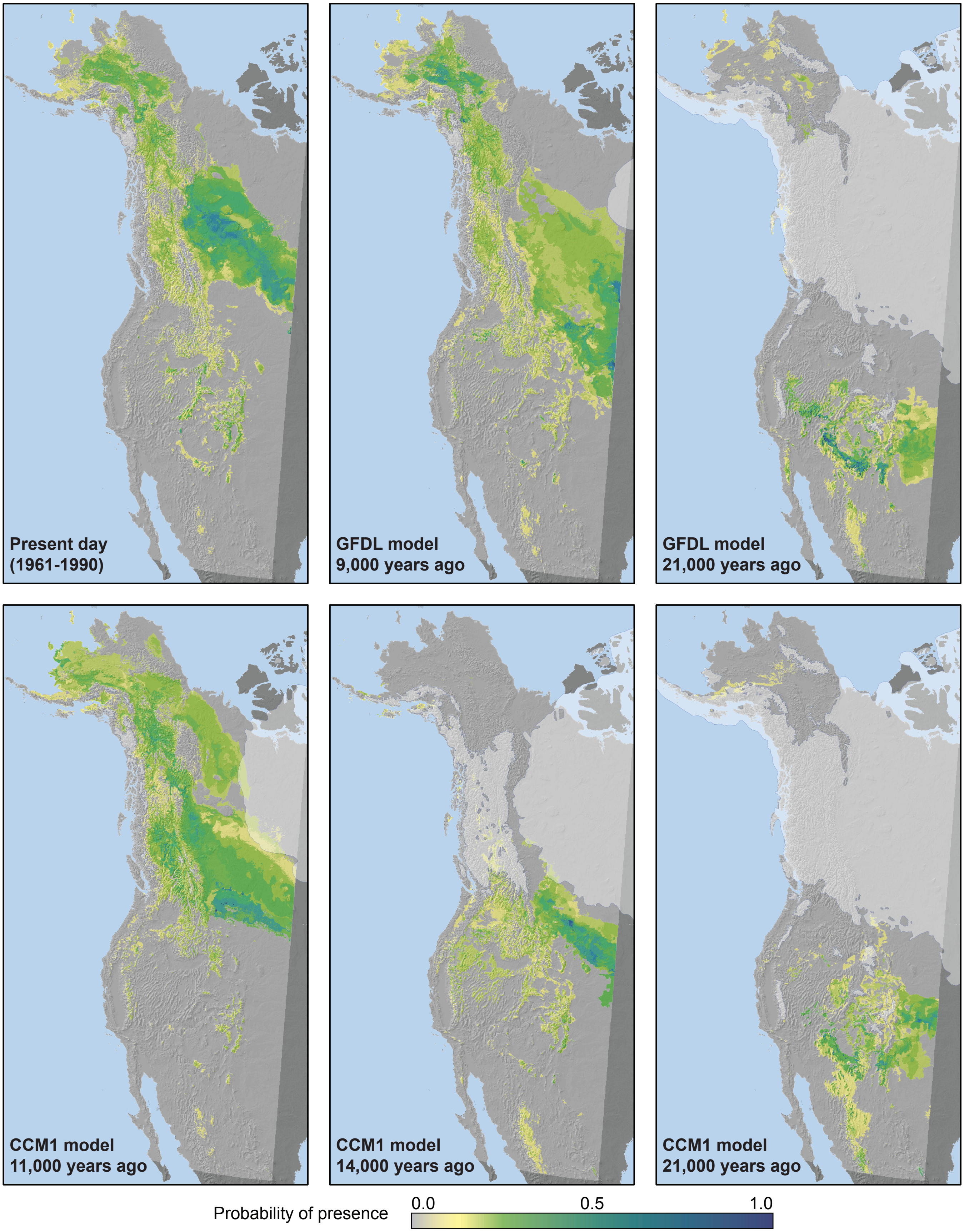

Figure S4.1S: ***Populus tremuloides*** **(trembling aspen)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

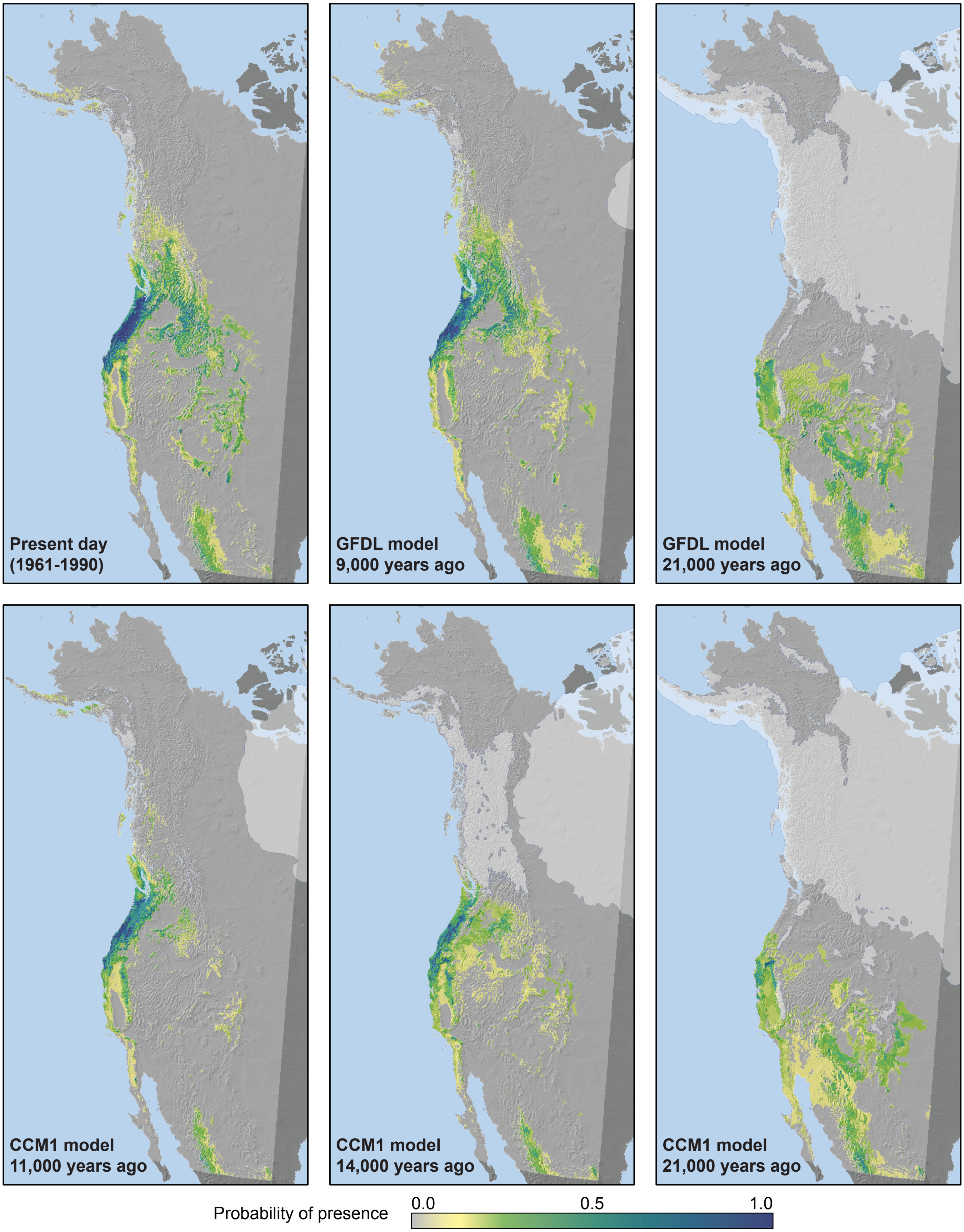

Figure S4.1T: ***Pseudotsuga menziesii*** **(Douglas-fir)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

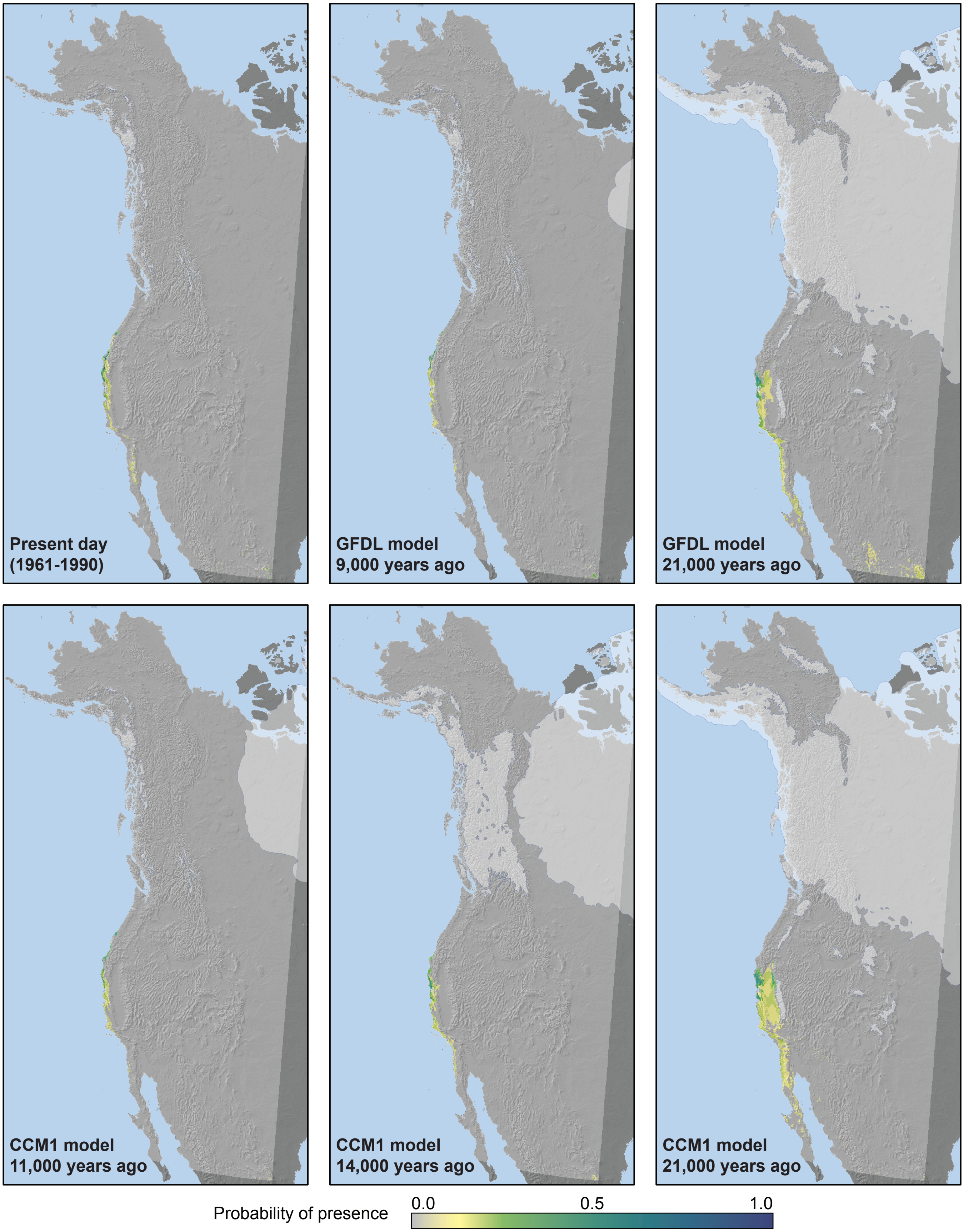

Figure S4.1U: *Sequoia sempervirens* (coast redwood) probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

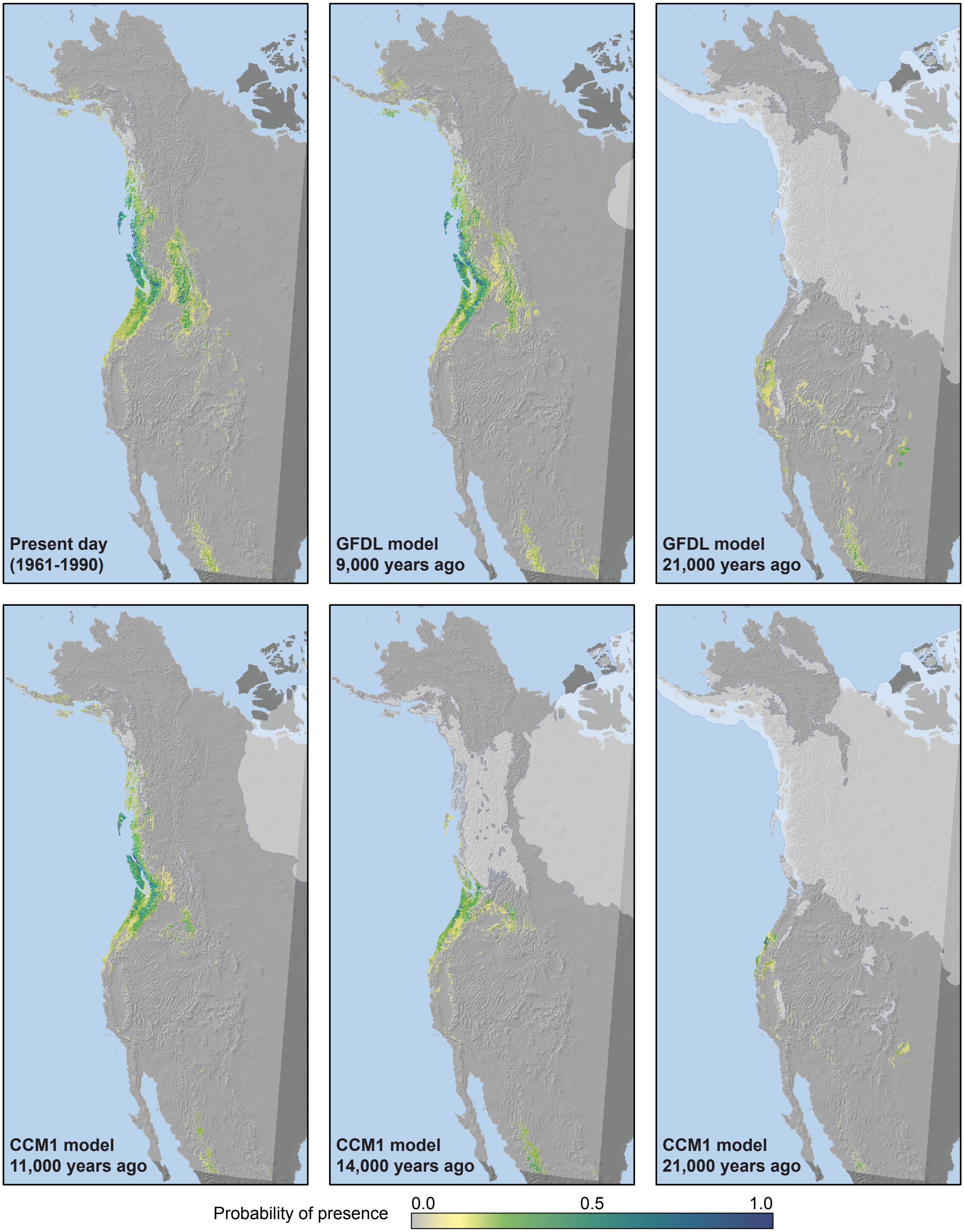

Figure S4.1V: *Thuja plicata* (western redcedar) probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

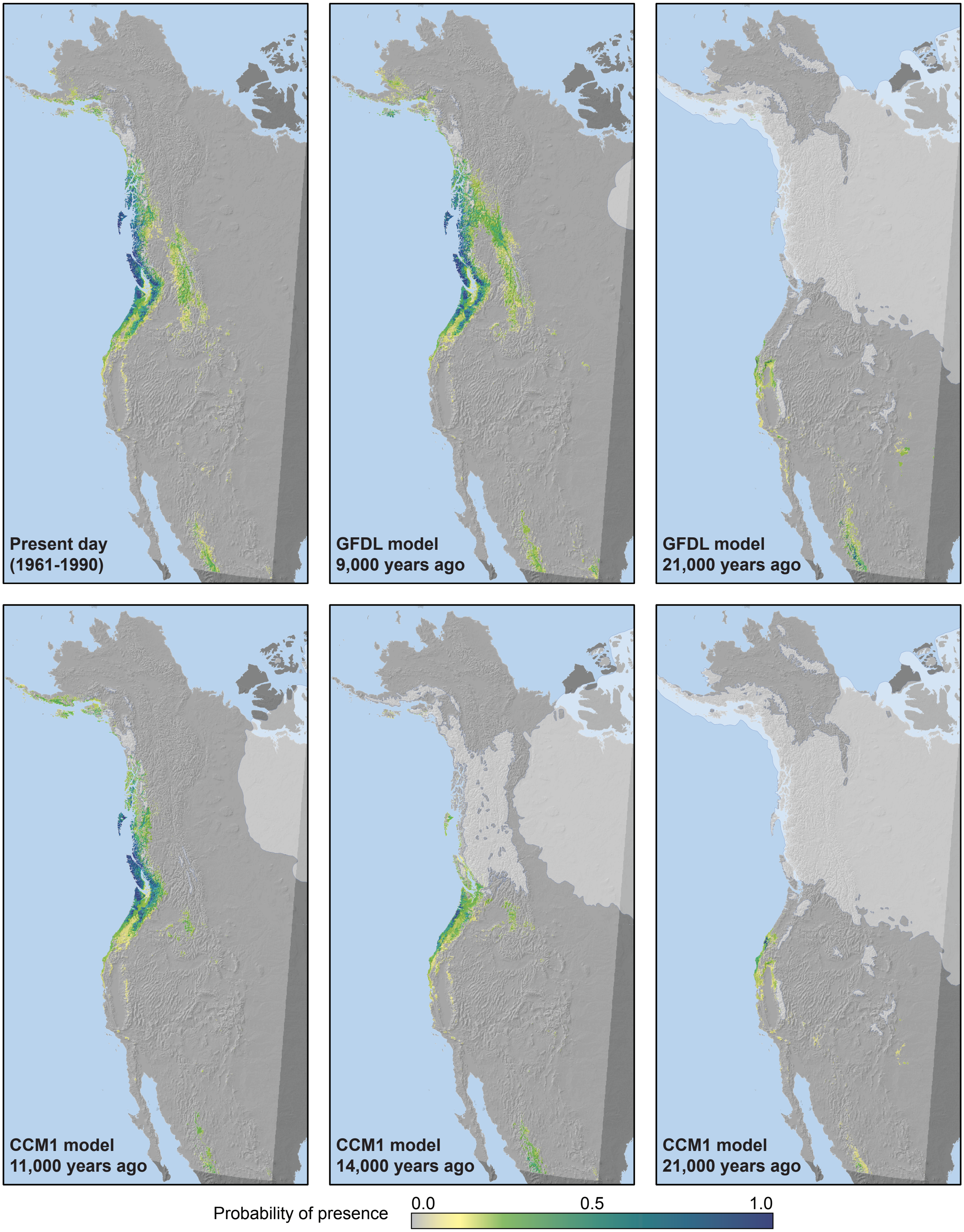

Figure S4.1W: ***Tsuga heterophylla* (western hemlock)** probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

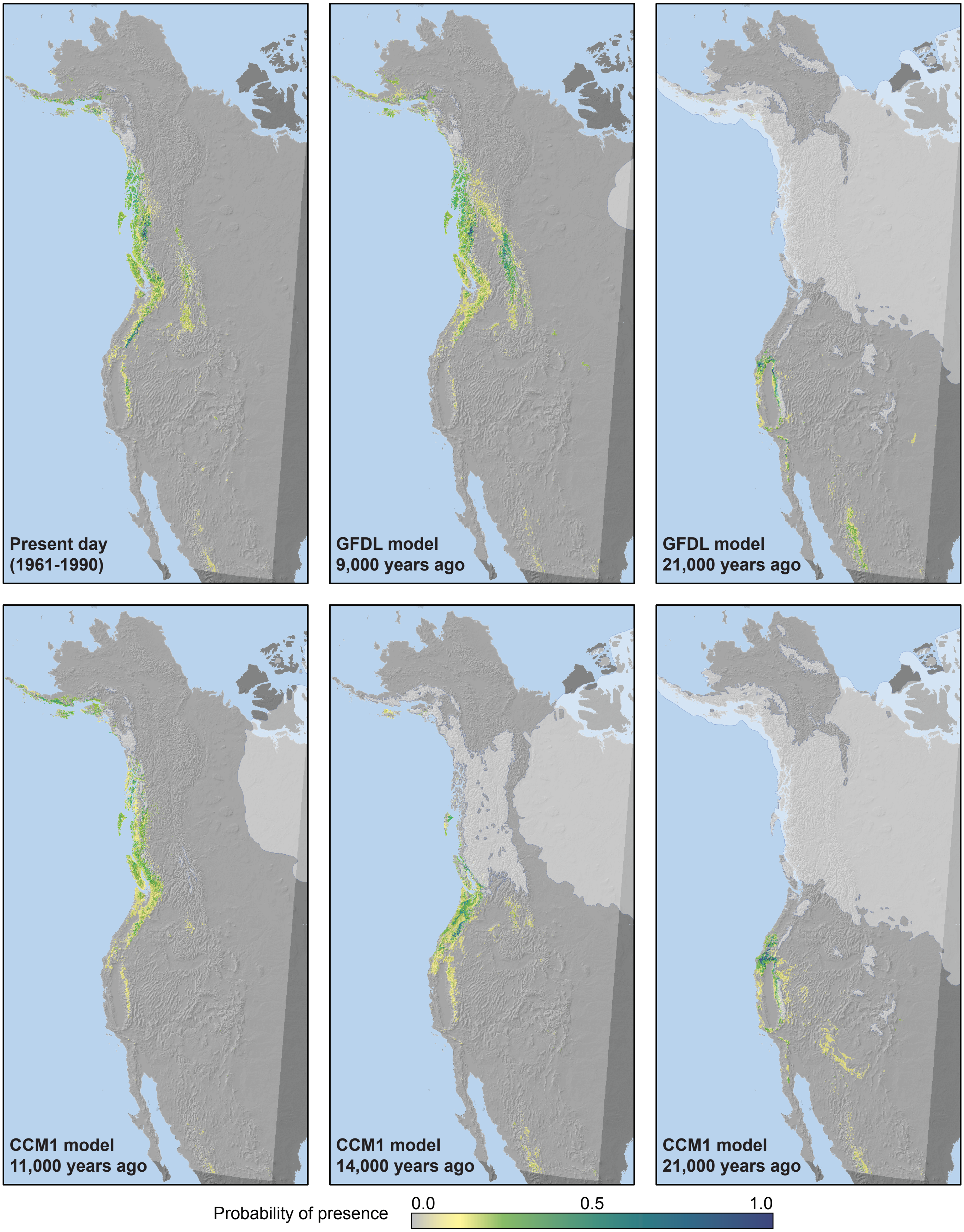

Figure S4.1X: *Tsuga mertensiana* (mountain hemlock) probability of presence as projected by the ensemble species distribution model for the observed 1961-1990 climate of the present day, for the GFDL palaeoclimate reconstruction for 9,000 and 21,000 years ago, and for the CCM1 palaeoclimate reconstructions for 11,000, 14,000, and 21,000 years ago. Continental ice is shown in transparent white.

Table S4.1: Species genetic data, compiled from various sources, listed by species and by reference. For each sampled population in each respective reference, we report the population name and verbal location, Latitude (Lat), Longitude (Long), Elevation (Elev), number of individuals sampled (N), average number of alleles per locus (A), percent polymorphic loci (P), observed heterozygosity (H_o), and expected heterozygosity (H_e). Standard errors for H_o and H_e are listed in parentheses. Where data were not reported in the literature but was calculated by the authors of this paper, I have listed the values in italics. Data that were not reported and that we were not able to calculate or estimate are noted with a hyphen. In some cases in the literature, data are presented as population averages. we have listed average values for these entries and noted them with an asterisk.

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
Abies amabilis **(Pacific silver fir)**											
(Davidson & El Kassaby, 1997)											
A	Taylor River, BC	48.30	125.37	300	8	1.64	66.7	-	(-)	0.097	(-)
B	Sebalhall Creek, BC	49.95	126.42	300	8	1.55	83.3	-	(-)	0.068	(-)
C	Maquilla Creek, BC	50.06	126.34	500	8	1.55	66.7	-	(-)	0.889	(-)
F	Fleet River, BC	48.65	124.10	710	9	1.73	66.7	-	(-)	0.156	(-)
H	Hathaway Creek, BC	50.58	127.73	212	13	1.73	83.3	-	(-)	0.124	(-)
N	Holberg Inlet, BC	50.73	128.00	215	11	1.55	66.7	-	(-)	0.109	(-)
R	Ronning Creek, BC	50.61	128.19	275	13	1.45	83.3	-	(-)	0.081	(-)
W	Mystery Creek, BC	48.80	128.15	625	17	1.82	83.3	-	(-)	0.177	(-)
Abies lasiocarpa **(subalpine fir)**											
(Ettl & Peterson, 2001)											
Mt. Dana high	Olympic Mtns., WA	47.8	123.5	1800	30	1.6	26.7	0.07	-0.042	0.079	-0.048
Klahhane Ridge high	Olympic Mtns., WA	48.0	123.4	1800	50	1.8	46.7	0.139	-0.057	0.137	-0.055
Klahhane Ridge middle	Olympic Mtns., WA	48.0	123.4	1575	50	2.1	60.0	0.138	-0.056	0.145	-0.059
Klahhane Ridge low	Olympic Mtns., WA	48.0	123.4	1350	50	2.1	60.0	0.139	-0.057	0.143	-0.058
Blue Mtn. high	Olympic Mtns., WA	47.6	123.2	1800	50	2.0	53.3	0.108	-0.054	0.115	-0.054
Blue Mtn. middle 1	Olympic Mtns., WA	47.6	123.2	1575	50	1.9	46.7	0.116	-0.054	0.117	-0.055

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
Blue Mtn. middle 2	Olympic Mtns., WA	47.6	123.2	1575	50	1.9	46.7	0.103	-0.043	0.124	-0.053
Blue Mtn. low 1	Olympic Mtns., WA	47.6	123.2	1350	50	2.0	60.0	0.12	-0.046	0.139	-0.053
Blue Mtn. low 2	Olympic Mtns., WA	47.6	123.2	1350	50	1.9	53.3	0.13	-0.059	0.143	-0.059
***Abies procera* (noble fir)** **(Yeh & Hu, 2005)**											
1 - Odell Butte	WA & OR	43.27	121.52	1950	34	1.303	41.67	-	(-)	0.297	-0.055
2 - Juniper Ridge	WA & OR	43.35	122.20	1700	34	1.250	41.67	-	(-)	0.254	-0.058
3 - Grass Mtn.	WA & OR	44.26	123.40	1060	34	1.247	37.50	-	(-)	0.231	-0.061
4 - Mary's Peak	WA & OR	44.3	123.33	1065	33	1.157	33.33	-	(-)	0.171	-0.051
5 - Laurel Mtn.	WA & OR	44.56	123.35	975	39	1.258	41.67	-	(-)	0.243	-0.055
6 - Fisher Point	WA & OR	44.33	122.02	1220	36	1.252	45.83	-	(-)	0.254	-0.054
7 - Snow Peak	WA & OR	44.39	122.35	1060	36	1.214	41.67	-	(-)	0.218	-0.054
8 - Elk Lake	WA & OR	44.49	122.06	1200	36	1.264	45.83	-	(-)	0.265	-0.061
9 - One Hundred Road	WA & OR	45.06	122.18	1130	36	1.213	33.33	-	(-)	0.210	-0.058
10 - Elk Mtn.	WA & OR	45.20	121.39	1220	33	1.245	41.67	-	(-)	0.249	-0.054
11 - Larch Mtn.	WA & OR	45.32	122.06	975	38	1.235	45.83	-	(-)	0.234	-0.055
12 - Mt. Defiance	WA & OR	45.38	121.44	1125	39	1.251	45.83	-	(-)	0.261	-0.053
13 - Larch Mtn.	WA & OR	45.43	122.17	975	35	1.231	45.83	-	(-)	0.244	-0.051
14 - Red Mtn.	WA & OR	45.56	121.50	1220	33	1.189	33.33	-	(-)	0.190	-0.056
15 - Hungry Peak	WA & OR	46.07	121.54	1280	34	1.150	37.50	-	(-)	0.153	-0.051
16 - French Butte	WA & OR	46.20	121.57	1300	35	1.216	37.50	-	(-)	0.221	-0.053
17 - Mud Lake	WA & OR	46.24	121.37	1425	37	1.226	41.67	-	(-)	0.231	-0.052
18 - McKinley Lake	WA & OR	46.35	122.08	900	33	1.157	33.33	-	(-)	0.179	-0.047
19 - Corral Pass	WA & OR	47.01	121.08	1615	38	1.258	45.83	-	(-)	0.270	-0.049

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
20 - Stampede Pass	WA & OR	47.14	121.22	1065	33	1.208	37.50	-	(-)	0.210	-0.055
21 - Stevens Pass	WA & OR	47.43	121.08	1000	33	1.258	37.50	-	(-)	0.253	-0.058
***Acer macrophyllum* (bigleaf maple)**											
(Iddrisu & Ritland, 2004)											
Jericho	Jericho area, BC	49.2	123.1	0	40	2.2	80	0.108	-0.029	0.102	-0.026
Fraser	Fraser area, BC	49.0	121.0	50	36	2.0	60	0.112	-0.036	0.105	-0.033
Artic	Artic area, WA	46.5	123.4	60	14	1.5	50	0.160	-0.060	0.189	-0.066
Cascadia	Cascadia area, OR	44.4	122.5	250	20	1.6	60	0.121	-0.037	0.164	-0.046
Elbe	Elbe area, WA	46.5	122.1	380	20	1.6	60	0.109	-0.035	0.172	-0.049
Helmick	Helmick area, OR	44.8	123.2	60	20	1.6	60	0.118	-0.038	0.176	-0.049
Oakville	Oakville area, WA	46.8	123.2	30	20	1.6	60	0.117	-0.033	0.148	-0.041
Siletz	Siletz area, OR	44.7	123.9	40	20	1.6	60	0.102	-0.031	0.163	-0.047
***Alnus rubra* (red alder)**											
(Xie *et al.*, 2002)											
Mainland - GVWD	Seymour, BC	49.40	122.97	200	26	1.53	31.6	0.099	-0.038	0.112	-0.043
Mainland - CHEA	Cheakmus River, BC	50.07	123.10	540	39	1.53	36.8	0.111	-0.041	0.119	-0.044
Mainland - KING	Kingcome Inlet, BC	51.50	126.13	30	52	1.53	31.6	0.109	-0.040	0.121	-0.044
Mainland - NACH	Nachelor Bay, BC	52.37	126.92	30	55	1.58	31.6	0.129	-0.048	0.130	-0.048
Mainland - SALL	Salloomt River, BC	52.43	126.92	150	56	1.53	31.6	0.116	-0.044	0.113	-0.042
Mainland - SNOW	Snow Creek, BC	54.25	129.55	10	63	1.47	31.6	0.144	-0.057	0.153	-0.056
Mainland - PRINT	Prince Rupert, BC	54.27	130.27	46	65	1.53	31.6	0.115	-0.047	0.112	-0.043
Mainland - SHAM	Shames River, BC	54.43	128.92	100	68	1.47	31.6	0.119	-0.048	0.120	-0.046
Islands - PORTR	Port Renfrew, BC	48.60	124.23	20	4	1.47	26.3	0.070	-0.032	0.071	-0.032

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
Islands - COW	Cowichan M.F., BC	48.77	123.65	150	9	1.42	26.3	0.058	-0.029	0.066	-0.031
Islands - UCLU	Ucluelet , BC	49.00	125.57	40	14	1.37	26.3	0.068	-0.029	0.090	-0.038
Islands - LOWR	Lowry Lake, BC	49.40	125.15	120	28	1.63	36.8	0.093	-0.038	0.103	-0.041
Islands - WOSS	Woss #2, BC	49.97	126.25	150	37	1.47	26.3	0.067	-0.030	0.075	-0.034
Islands - BIGT	Bigtree #2, BC	50.23	125.72	300	45	1.53	26.3	0.121	-0.054	0.123	-0.052
Islands - PORTH	Port Hardy, BC	50.62	127.25	37	49	1.53	31.6	0.077	-0.033	0.092	-0.038
Islands - NE62	Ne 62, BC	50.72	127.98	170	51	1.53	31.6	0.104	-0.046	0.102	-0.044
Islands - POOL	Poole Inlet, BC	52.35	131.35	1	53	1.58	31.6	0.070	-0.029	0.090	-0.035
Islands - COPP	Copper Bay, BC	53.12	131.67	10	57	1.47	26.3	0.079	-0.035	0.088	-0.038
Islands - MASS	Masset , BC	54.05	132.00	10	61	1.53	31.6	0.102	-0.044	0.105	-0.044
***Calocedrus decurrens* (incense cedar)**											
(Harry, 1984)											
Kilarc-LS	Sierra Nevada, CA	40.7	121.9	890	43	2.8	52	-	(-)	0.207	-0.043
Kilarc-LN	Sierra Nevada, CA	40.7	121.9	890	44	2.8	52	-	(-)	0.184	-0.039
Kilarc-HS	Sierra Nevada, CA	40.7	121.9	1390	50	3.0	60	-	(-)	0.222	-0.041
Kilarc-HN	Sierra Nevada, CA	40.7	121.9	1390	52	3.0	60	-	(-)	0.181	-0.039
Bailey-LS	Sierra Nevada, CA	38.4	120.3	1510	46	2.7	44	-	(-)	0.177	-0.042
Bailey-LN	Sierra Nevada, CA	38.4	120.3	1510	40	2.6	56	-	(-)	0.186	-0.035
Bailey-HS	Sierra Nevada, CA	38.4	120.3	2010	49	2.4	56	-	(-)	0.176	-0.039
Bailey-HN	Sierra Nevada, CA	38.4	120.3	2010	50	2.3	48	-	(-)	0.182	-0.043
Greenhorn-LS	Sierra Nevada, CA	35.7	118.3	1790	49	2.1	48	-	(-)	0.172	-0.041
Greenhorn-LN	Sierra Nevada, CA	35.7	118.3	1790	51	2.4	40	-	(-)	0.166	-0.040
Greenhorn-HS	Sierra Nevada, CA	35.7	118.3	2290	45	1.9	40	-	(-)	0.149	-0.038

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
Greenhorn-HN	Sierra Nevada, CA	35.7	118.3	2290	51	2.4	44	-	(-)	0.154	-0.040
***Chamaecyparis nootkatensis* (yellow cedar)**											
(Ritland *et al.*, 2001)											
Mt. Baker	Mt. Baker, WA	48.8	121.8	>1000	32	1.8	60	-	(-)	0.153	-0.014
Port Hardy	Port Hardy, BC	50.7	127.5	0	32	1.7	60	-	(-)	0.132	-0.011
Mt. Washington	Mt. Washington, BC	49.8	125.3	1000	35	1.5	40	-	(-)	0.090	-0.013
Tofino	Tofino, BC	49.1	125.9	0	36	1.7	50	-	(-)	0.161	-0.011
Anchorage	Anchorage, AK	61.2	149.9	0	27	1.4	30	-	(-)	0.062	-0.006
Juneau	Juneau, AK	58.4	134.2	0	30	1.7	50	-	(-)	0.198	-0.010
Petersburg	Petersburg, AK	56.8	132.9	0	33	1.8	60	-	(-)	0.197	-0.010
Ketchikan	Ketchikan, AK	52.4	131.7	1000	30	1.8	60	-	(-)	0.192	-0.011
Prince Rupert	Prince Rupert, BC	54.3	130.3	0	32	1.7	50	-	(-)	0.163	-0.011
Bella Coola	Bella Coola, BC	52.4	126.8	<1000	32	1.7	50	-	(-)	0.141	-0.008
Hurricane Ridge	Hurricane Ridge, WA	48.0	123.5	>1000	31	1.8	60	-	(-)	0.149	-0.009
Mt. Rainier	Mt. Rainier, WA	46.9	121.8	<1000	32	1.6	40	-	(-)	0.117	-0.015
Oakridge	Oakridge, OR	43.7	122.5	>1000	32	1.8	60	-	(-)	0.156	-0.010
Whiskey Peak	Whiskey Peak, OR	42.0	123.3	>1000	30	1.5	30	-	(-)	0.101	-0.011
Castlegar	Castlegar, BC	49.3	117.7	<1000	23	1.6	40	-	(-)	0.187	-0.007
Black Tusk	Black Tusk, BC	50.0	123.0	1000	32	1.8	60	-	(-)	0.183	-0.015
Queen Charlotte	Queen Charlotte, BC	53.8	132.0	400	31	1.7	50	-	(-)	0.123	-0.011
***Larix occidentalis* (western larch)**											
(Jaquish & El-Kassaby, 1998)											
1 - FLAT	Flathead, BC	49.07	114.43	1380	34	1.95	40	0.151	-0.038	0.151	-0.037

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
2 - PLUM	Plumbob, BC	49.25	115.40	1160	30	1.80	50	0.177	-0.040	0.173	-0.040
3 - CARR	Carrol Creek Road, BC	49.10	116.20	975	35	1.90	65	0.174	-0.035	0.185	-0.037
4 - SALM	Salmo, BC	49.23	117.27	800	35	1.85	40	0.151	-0.041	0.143	-0.038
5 - CHRI	Christina Lake, BC	49.22	118.15	1200	35	1.80	70	0.206	-0.045	0.219	-0.040
6 - BLUE	Blue Ridge Kaslo, BC	50.05	117.08	1375	36	1.85	65	0.204	-0.038	0.220	-0.042
7 - WILS	Wilson Lake Road, BC	50.23	117.72	1300	36	1.80	70	0.236	-0.045	0.237	-0.043
8 - BECK	Becker Lake, BC	50.25	119.17	1200	38	1.80	65	0.213	-0.042	0.204	-0.039
9 - MERR	Merritt, BC	50.28	120.93	1300	34	1.70	55	0.151	-0.036	0.156	-0.037
(Fins & Seeb, 1986)											
Eastern Washington	7 stands, eastern WA	48.6	118.4	1197.9	69	1.3	30*	0.079	-0.035	0.089	-0.039
Northern Idaho	9 stands, northern ID	47.3	116.3	1282.4	79	1.3	30*	0.049	-0.020	0.074	-0.032
Western Montana	3 stands, Western MT	46.6	114.5	1473.3	29	1.2	30*	0.054	-0.031	0.068	-0.033
Payette	One stand, Payette, ID	45.1	116.4	1646.0	10	1.3	30*	0.078	-0.048	0.089	-0.051
***Picea engelmanii* (Engelmann spruce)**											
(Ledig *et al.*, 2006)											
Kootenay River	Nelson Forest Region, BC	50.97	116.27	1830	27.5*	2.7	66.7	-	(-)	0.247	-0.047
Moyie	Nelson Forest Region, BC	49.53	116.10	1700	27.5*	2.6	70.8	-	(-)	0.247	-0.044
Bluebird Creek	Nelson Forest Region, BC	49.13	116.88	1430	27.5*	2.7	66.7	-	(-)	0.268	-0.046
Summit Lake	Payette N.F., ID	45.05	115.92	2075	27.5*	2.3	75.0	-	(-)	0.252	-0.043
Beartooth Pass	Shoshone N.F., WY	44.93	109.52	2985	27.5*	2.5	75.0	-	(-)	0.307	-0.045
Panther Creek	Salmon N.F., ID	44.88	114.30	2135	27.5*	2.5	75.0	-	(-)	0.253	-0.045
Six Bit Spring	Cache N.F., UT	41.58	111.43	2560	27.5*	2.1	54.2	-	(-)	0.227	-0.045
Highline Trailhead	Wasatch N.F., UT	40.73	110.87	3180	27.5*	2.4	58.3	-	(-)	0.228	-0.043

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
Ferron Reservoir	Manti-Lasal N.F., UT	39.13	111.45	2930	27.5*	2.2	62.5	-	(-)	0.215	-0.040
Barlow Lake	San Juan N.F., CO	37.75	107.98	2955	27.5*	2.6	70.8	-	(-)	0.296	-0.046
Navajo Lake	Dixie N.F., UT	37.53	112.77	2865	27.5*	2.2	54.2	-	(-)	0.231	-0.044
East Gavilan Canyon	Carson N.F., NM	36.73	106.30	2970	27.5*	2.5	75.0	-	(-)	0.273	-0.046
San Francisco Mtns.	Coconino N.F., AZ	35.33	111.70	2775	27.5*	2.4	66.7	-	(-)	0.269	-0.046
Sierra Blanca	Sierra Blanca, Lincoln N.F., NM	33.40	105.78	2925	27.5*	2.1	66.7	-	(-)	0.282	-0.050
Mt. Graham	Mt. Graham, Coronado N.F., AZ	32.67	109.87	3018	27.5*	2.4	66.7	-	(-)	0.261	-0.047
Flys Peak	Flys Peak, Coronado N.F., AZ	31.87	109.28	2866	27.5*	2.0	62.5	-	(-)	0.227	-0.043
***Picea glauca* (white spruce)**											
(Godt *et al.*, 2001)											
Site 1	SK	-	-	-	47.5*	1.83	44.4	0.181	-0.042	0.158	-0.053
Site 2	SK	-	-	-	47.5*	1.78	55.6	0.167	-0.046	0.158	-0.049
Site 3	SK	-	-	-	47.5*	1.72	44.4	0.153	-0.042	0.166	-0.052
Site 4	SK	-	-	-	47.5*	1.72	44.4	0.154	-0.042	0.159	-0.052
Site 5	SK	-	-	-	47.5*	1.67	44.4	0.164	-0.045	0.167	-0.050
Site 6	SK	-	-	-	47.5*	1.61	44.4	0.162	-0.041	0.157	-0.052
Site 7	SK	-	-	-	47.5*	1.72	50.0	0.156	-0.044	0.160	-0.049
(Furnier *et al.*, 1991)											
A	NF East	-	-	-	20	2.2	100	0.367	(-)	0.329	(-)
B	NF West	-	-	-	20	2.2	100	0.300	(-)	0.291	(-)
C	NB	-	-	-	20	2.5	100	0.242	(-)	0.280	(-)

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
D	MN	-	-	-	20	2.5	83.3	0.283	(-)	0.256	(-)
E	QC Central	-	-	-	19	2.0	100	0.263	(-)	0.275	(-)
F	MA	-	-	-	19	2.0	100	0.272	(-)	0.250	(-)
G	NY	-	-	-	20	2.2	100	0.333	(-)	0.301	(-)
H	QC South	-	-	-	20	2.0	100	0.275	(-)	0.293	(-)
I	ON Southeast	-	-	-	20	2.2	100	0.325	(-)	0.307	(-)
J	ON East	-	-	-	20	1.8	83.3	0.250	(-)	0.230	(-)
K	ON Northeast	-	-	-	20	2.5	100	0.283	(-)	0.295	(-)
L	MI	-	-	-	19	2.0	100	0.386	(-)	0.340	(-)
M	WI	-	-	-	20	2.2	100	0.300	(-)	0.309	(-)
N	ON West	-	-	-	19	2.3	100	0.404	(-)	0.353	(-)
O	MN	-	-	-	20	2.3	100	0.283	(-)	0.281	(-)
P	MB South	-	-	-	20	2.3	100	0.383	(-)	0.357	(-)
Q	MB North	-	-	-	20	2.3	100	0.350	(-)	0.323	(-)
R	SD	-	-	-	20	2.0	100	0.308	(-)	0.294	(-)
S	SK North	-	-	-	20	1.8	83.3	0.233	(-)	0.198	(-)
T	MT	-	-	-	5	1.8	66.7	0.367	(-)	0.319	(-)
U	BC	-	-	-	20	2.2	100	0.233	(-)	0.261	(-)
V	AK	-	-	-	20	2.0	83.3	0.283	(-)	0.239	(-)
***Picea mariana* (black spruce)** **(Rajora & Pluhar, 2003)**											
E1-FNM	Post-fire nat mature, Pine Falls, MB	50.68	95.90	274	35	2.59	65.6	0.215	-0.037	0.296	-0.045

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
E1-FNR	Post-fire nat young regen, Pine Falls, MB	50.68	95.89	274	35	2.31	62.5	0.176	-0.033	0.265	-0.045
E1-HNR	Post-harvest nat regen, Pine Falls, MB	50.68	95.90	274	35	2.53	62.5	0.222	-0.040	0.293	-0.047
E1-PLT	Plantation, Pine Falls, MB	50.67	95.91	290	35	2.53	65.6	0.198	-0.037	0.279	-0.045
E2-FNM	Post-fire nat mature, Bissett, MB	50.78	95.28	320	35	2.59	75.0	0.223	-0.035	0.317	-0.042
E2-FNR	Post-fire nat young regen, Bissett, MB	50.78	95.29	320	35	2.69	75.0	0.263	-0.041	0.335	-0.042
E2-HNR	Post-harvest nat regen, Bissett, MB	50.78	95.28	320	35	2.50	59.4	0.267	-0.047	0.322	-0.050
E2-PLT	Plantation, Bissett, MB	50.83	95.32	305	35	2.66	71.9	0.240	-0.039	0.328	-0.045
N1-FNM	Post-fire nat mature, The Pas, MB	54.29	101.39	294	35	2.34	59.4	0.201	-0.044	0.272	-0.046
N1-FNR	Post-fire nat regen, The Pas, MB	54.27	101.38	294	35	2.34	62.5	0.193	-0.042	0.280	-0.046
N1-HNR	Post-harvest nat regen, The Pas, MB	54.30	101.48	290	35	2.56	71.9	0.219	-0.043	0.335	-0.045
N1-PLT	Plantation, The Pas, MB	54.29	101.44	290	35	2.16	59.4	0.203	-0.041	0.291	-0.045
N2-FNM	Post-fire nat mature, Snow Lake, MB	54.90	99.76	290	35	2.63	68.8	0.242	-0.045	0.341	-0.048
N2-FNR	Post-fire nat young regen, Snow Lake, MB	54.72	99.98	274	35	2.47	68.8	0.169	-0.03	0.279	-0.043
N2-HNR	Post-harvest nat regen, Snow Lake, MB	54.90	99.76	290	35	2.66	71.9	0.239	-0.044	0.328	-0.047
N2-PLT	Plantation, Snow Lake, MB	54.90	99.76	290	35	2.69	68.8	0.279	-0.048	0.362	-0.047

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
(O'Reilly *et al.*, 1985)											
Sioux Narrows - Upland	Sioux Narrows, ON	49.42	94.13	-	10		53.3		(-)	0.20	(-)
Sioux Narrows - Lowland	Sioux Narrows, ON	49.42	94.13	-	10		46.7		(-)	0.22	(-)
Sioux Lookout - Upland	Sioux Lookout, ON	50.15	91.75	-	10		33.3		(-)	0.16	(-)
Sioux Lookout - Lowland	Sioux Lookout, ON	50.15	91.75	-	10		53.3		(-)	0.21	(-)
Nipigon - Upland	Nipigon, ON	48.83	88.58	-	10		64.3		(-)	0.26	(-)
Nipigon - Lowland	Nipigon, ON	48.83	88.58	-	10		73.3		(-)	0.27	(-)
Macdiarmid - Upland	Macdiarmid, ON	49.42	88.12	-	10		46.7		(-)	0.20	(-)
Macdiarmid - Lowland	Macdiarmid, ON	49.42	88.12	-	10		46.7		(-)	0.22	(-)
Cochrane - Upland	Cochrane, ON	49.08	80.92	-	10		40.0		(-)	0.23	(-)
Cochrane - Lowland	Cochrane, ON	49.08	80.92	-	10		66.7		(-)	0.25	(-)
(Isabel *et al.*, 1995)											
325	QC	50.2	74.2	-	15	2.3	76.9	0.267	-0.056	0.287	-0.063
336	QC	48.0	78.9	-	15	2.0	69.2	0.337	-0.085	0.285	-0.062
342	QC	50.4	68.4	-	15	2.5	61.5	0.305	-0.077	0.301	-0.070
345	QC	47.2	74.3	-	15	2.1	69.2	0.348	-0.083	0.311	-0.071
369	QC	48.3	64.2	-	15	2.2	69.2	0.435	-0.099	0.324	-0.065
***Picea sitchensis* (sitka spruce)**											
(Yeh & El-Kassaby, 1980)											
IUFRO-3024	Duck Cr., AK	58.37	134.58	30	76	2.00	67	-	(-)	0.16	-0.03
IUFRO-3030	Ward L., AK	55.42	131.70	15	78	1.62	50	-	(-)	0.15	-0.04
IUFRO-3040	Usk Ferry, BC	54.77	128.25	137	79	1.69	46	-	(-)	0.15	-0.04
IUFRO-3044	Inverness, BC	54.20	130.25	15	79	1.77	50	-	(-)	0.13	-0.04

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
IUFRO-3049	Link Rd., BC	53.50	132.17	90	78	1.85	50	-	(-)	0.16	-0.04
IUFRO-3058	Salmon Bay, BC	50.38	125.95	0	78	1.69	42	-	(-)	0.14	-0.04
IUFRO-3062	Big Qualicum R., BC	49.38	124.62	0	78	1.92	54	-	(-)	0.16	-0.04
IUFRO-3003	Forks, WA	48.07	124.30	137	79	1.92	46	-	(-)	0.17	-0.04
IUFRO-3008	Hoquiam, WA	47.08	124.05	7	79	1.85	50	-	(-)	0.13	-0.04
IUFRO-3012	Necanicum, OR	45.82	123.77	46	78	1.85	58	-	(-)	0.13	-0.04
***Pinus albicaulis* (whitebark pine)**											
(Bruederle *et al.*, 1998)											
1. Miller Creek	Gallatin Park, MT	45.4	110.2	-	30	1.6	42.1	0.167	-0.052	0.146	-0.045
2. Henderson Mtn.	Gallatin Park, MT	45.0	110.0	-	31	1.7	36.8	0.139	-0.043	0.164	-0.049
3. Fisher Creek	Gallatin Park, MT	45.1	109.9	-	34	1.5	42.1	0.139	-0.043	0.151	-0.047
4. Island Lake	Shoshone Park, WY	44.9	109.5	-	49	1.7	36.8	0.153	-0.047	0.151	-0.046
5. Mt. Washburn	Yellowstone N.P., WY	44.8	110.4	-	33	1.6	36.8	0.139	-0.045	0.145	-0.046
6. Union Pass	Shoshone Forest, Fremont County, WY	43.5	109.9	-	20	1.5	36.8	0.161	-0.054	0.146	-0.047
7. Togwotee Pass	Bridger-Teton Forest, Teton County, WY	43.8	110.1	-	25	1.7	36.8	0.141	-0.047	0.153	-0.046
8. Sheep Pass	Bridger-Teton Forest, Lincoln County, WY	42.5	110.8	-	30	1.6	36.8	0.152	-0.045	0.155	-0.047
9. Commissary Ridge	Bridger-Teton Forest, Lincoln County, WY	42.2	110.6	-	30	1.6	42.1	0.138	-0.042	0.161	-0.048
(Jorgensen & Hamrick, 1997)											
1. Cascade - WASH	Washington Pass, WA	48.53	120.67	-	48	2.00	25	0.092	-0.009	0.096	-0.039
2. Cascade - RAIN	Mt. Rainier N.P., WA	46.92	121.65	-	48	2.22	35	0.092	-0.009	0.109	-0.035

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
3. Cascade - ADAM	Mt. Adams, WA	46.25	121.53	-	48	2.00	25	0.083	-0.009	0.075	-0.033
4. Cascade - HOOD	Mt. Hood, OR	45.40	121.67	-	48	2.25	15	0.084	-0.009	0.089	-0.039
5. Cascade - BACH	Bachelor Peak, OR	43.98	121.68	-	48	2.00	20	0.082	-0.009	0.080	-0.037
6. Cascade - CRAT	Crater Lake N.P., OR	42.93	122.17	-	48	2.00	20	0.060	-0.008	0.070	-0.031
7. Cascade - SHAS	Mt. Shasta, CA	41.37	122.20	-	48	2.25	15	0.083	-0.009	0.091	-0.039
8. Cascade - LASS	Lassen Volcanic N.P., CA	40.48	121.48	-	48	2.00	15	0.079	-0.009	0.075	-0.035
9. Sierra Nevada - ROSE	Mt. Rose Summit, NV	39.32	119.90	-	48	2.43	20	0.119	-0.010	0.120	-0.046
10. Sierra Nevada - SONO	Sonora Pass, CA	38.33	119.65	-	48	2.43	20	0.104	-0.010	0.104	-0.042
11. Sierra Nevada - TIOG	Tioga Pass, CA	37.92	119.25	-	48	2.29	20	0.087	-0.009	0.090	-0.039
12. Sierra Nevada - MOSQ	Mosquito Lake, CA	37.43	118.73	-	48	2.00	20	0.091	-0.009	0.100	-0.041
13. Rocky Mtns. - JASP	Mt. Edith Cavell, AB	52.68	118.08	-	48	2.43	25	0.080	-0.009	0.088	-0.036
14. Rocky Mtns. - PARK	Parker Ridge, AB	52.18	117.12	-	48	2.17	25	0.076	-0.009	0.095	-0.035
15. Rocky Mtns. - CABI	Libby, MT	48.55	115.73	-	48	2.30	25	0.082	-0.009	0.113	-0.037
16. Rocky Mtns. - GLAC	Glacier N.P., MT	48.68	113.75	-	48	2.25	30	0.094	-0.009	0.117	-0.040
17. Rocky Mtns. - JEWL	Jewel Basin, MT	48.17	113.93	-	48	2.00	20	0.090	-0.009	0.088	-0.035
18. Rocky Mtns. - FRNT	Choteau, MT	47.93	112.82	-	48	2.36	25	0.108	-0.010	0.111	-0.039
19. Rocky Mtns. - NINE	Ninemile Divide, MT	47.13	114.32	-	48	2.00	25	0.083	-0.009	0.099	-0.033
20. Rocky Mtns. - WILL	Corvallis, MT	46.32	113.92	-	48	2.08	20	0.083	-0.009	0.103	-0.037
21. Rocky Mtns. - TOBA	Tobacco Roots, MT	45.52	112.00	-	48	2.09	20	0.106	-0.010	0.105	-0.038
22. Rocky Mtns. - GRAV	Gravelly Range, MT	44.82	111.87	-	48	2.25	30	0.096	-0.010	0.111	-0.036
23. Rocky Mtns. - BELT	Little Belts, MT	46.78	110.65	-	48	2.09	30	0.098	-0.010	0.105	-0.034
24. Rocky Mtns. - YELL	Dunraven Pass, WY	44.78	110.50	-	48	2.29	25	0.075	-0.009	0.081	-0.034

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
25. Rocky Mtns. - BEAR	Beartooth Plateau, WY	44.95	109.53	-	48	2.33	25	0.086	-0.009	0.106	-0.038
26. Rocky Mtns. - TOGW	Togwotee Pass, WY	43.75	110.07	-	48	2.14	25	0.079	-0.009	0.089	-0.037
27. Rocky Mtns. - CIDA	Yellow Pine, ID	45.07	115.42	-	48	2.50	20	0.098	-0.010	0.097	-0.040
28. Rocky Mtns. - SAWT	Galena Summit, ID	43.87	114.70	-	48	2.50	25	0.107	-0.010	0.110	-0.040
29. Great Basin - JARB	Jarbridge, NV	41.83	115.47	-	48	2.14	25	0.077	-0.009	0.090	-0.035
30. Great Basin - RUBY	Ruby Range, NV	40.60	115.38	-	48	2.00	20	0.056	-0.008	0.064	-0.032
***Pinus contorta* (lodgepole pine)**											
(Yeh & Layton, 1979)											
Marginal - 32	Southern YK	62.14	136.18	671	15	1.80	53	0.160	-0.044	0.157	-0.042
Marginal - 33	Southern YK	63.18	136.28	877	15	1.72	44	0.165	-0.049	0.144	-0.044
Marginal - 34	Southern YK	60.41	136.11	747	15	1.80	52	0.125	-0.032	0.138	-0.038
Marginal - 35	Northern BC	59.48	133.47	788	15	1.80	56	0.147	-0.042	0.143	-0.040
Intermediate - 30	Northern BC	59.59	128.33	640	15	1.76	52	0.149	-0.044	0.150	-0.043
Intermediate - 31	Southern YK	61.10	129.30	884	15	1.96	64	0.173	-0.039	0.178	-0.043
Central - 36	Northern BC	57.29	130.13	815	15	2.12	76	0.176	-0.036	0.184	-0.043
Central - 28	Northern BC	58.4	124.10	762	15	1.92	64	0.184	-0.043	0.178	-0.043
Central - 71	Interior BC	50.43	119.27	1524	15	2.08	68	0.189	-0.045	0.169	-0.037
(Wheeler & Guries, 1982)											
BC - 1	Ssp. latifolia	-	-	-	-	1.86	70	-	(-)	0.114	-0.026
BC - 4	Ssp. latifolia	-	-	-	-	1.69	61	-	(-)	0.125	-0.029
WA - 5	Ssp. latifolia	-	-	-	-	1.81	70	-	(-)	0.103	-0.025
AB - 6	Ssp. latifolia	-	-	-	-	1.93	73	-	(-)	0.128	-0.028

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
AB - 7	Ssp. latifolia	-	-	-	-	2.24	82	-	(-)	0.144	-0.028
AB - 8	Ssp. latifolia	-	-	-	-	1.88	73	-	(-)	0.102	-0.025
AB - 9	Ssp. latifolia	-	-	-	-	1.95	70	-	(-)	0.120	-0.027
BC - 10	Ssp. latifolia	-	-	-	-	2.04	70	-	(-)	0.115	-0.026
YK - 14	Ssp. latifolia	-	-	-	-	1.71	64	-	(-)	0.108	-0.025
BC - 15	Ssp. latifolia	-	-	-	-	1.83	70	-	(-)	0.114	-0.028
YK - 16	Ssp. latifolia	-	-	-	-	1.76	70	-	(-)	0.129	-0.029
YK - 18	Ssp. latifolia	-	-	-	-	1.76	64	-	(-)	0.109	-0.028
YK - 21	Ssp. latifolia	-	-	-	-	1.83	79	-	(-)	0.113	-0.028
BC - 22	Ssp. latifolia	-	-	-	-	1.83	67	-	(-)	0.113	-0.027
CO - 23	Ssp. latifolia	-	-	-	-	1.81	64	-	(-)	0.119	-0.028
BC - 27	Ssp. latifolia	-	-	-	-	1.97	76	-	(-)	0.111	-0.025
BC - 29	Ssp. latifolia	-	-	-	-	1.86	73	-	(-)	0.119	-0.028
WA - 30	Ssp. latifolia	-	-	-	-	1.76	61	-	(-)	0.114	-0.027
AB - 32	Ssp. latifolia	-	-	-	-	1.81	64	-	(-)	0.105	-0.025
MT - 34	Ssp. latifolia	-	-	-	-	2.09	73	-	(-)	0.116	-0.026
YK - 36	Ssp. latifolia	-	-	-	-	1.76	61	-	(-)	0.130	-0.031
BC - 37	Ssp. latifolia	-	-	-	-	1.83	70	-	(-)	0.128	-0.029
MT - 46	Ssp. latifolia	-	-	-	-	1.67	58	-	(-)	0.097	-0.026
YK - 50	Ssp. latifolia	-	-	-	-	1.88	67	-	(-)	0.144	-0.030
BC - 45	Ssp. contorta	-	-	-	-	1.88	61	-	(-)	0.130	-0.030
BC - 47	Ssp. contorta	-	-	-	-	1.88	70	-	(-)	0.123	-0.027
BC - 48	Ssp. contorta	-	-	-	-	1.18	64	-	(-)	0.125	-0.029
BC - 49	Ssp. contorta	-	-	-	-	1.57	55	-	(-)	0.114	-0.029

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
WA - 38	Ssp. contorta	-	-	-	-	1.93	73	-	(-)	0.139	-0.029
CA - 17	Ssp. murrayana	-	-	-	-	1.86	70	-	(-)	0.120	-0.026
OR - 39	Ssp. murrayana	-	-	-	-	2.02	76	-	(-)	0.128	-0.028
CA - 40	Ssp. Bolanderi	-	-	-	-	1.62	58	-	(-)	0.109	-0.280
(Dancik & Yeh, 1983)											
Hinton	Hinton, AB	53.43	117.48	1170	62	3.0	52.4	0.186	(-)	0.184	-0.045
Cypress Hills	Cypress Hills, AB	49.65	110.28	1410	30	2.5	52.4	0.165	(-)	0.179	-0.049
Coleman	Coleman, AB	49.75	114.50	1090	30	2.3	52.4	0.202	(-)	0.239	-0.063
Canmore	Canmore, AB	51.10	115.28	1630	30	2.4	47.6	0.187	(-)	0.241	-0.061
Lake Abraham	Lake Abraham, AB	52.25	116.45	1750	30	2.3	52.4	0.179	(-)	0.201	-0.054
***Pinus edulis* (pinyon pine)**											
(Premoli *et al.*, 1994)											
OC-North	Northeastern CO	-	-	1835	25	2.1	80	0.214	(-)	0.262	(-)
OC-Centre	Northeastern CO	-	-	1835	25	2.0	50	0.192	(-)	0.187	(-)
OC-East	Northeastern CO	-	-	1835	25	1.9	60	0.220	(-)	0.231	(-)
OC-West	Northeastern CO	-	-	1835	25	2.0	50	0.203	(-)	0.204	(-)
OC-South	Northeastern CO	-	-	1835	25	1.9	60	0.205	(-)	0.213	(-)
Lone Site	Northeastern CO	-	-	2050	44	2.3	50	0.164	(-)	0.219	(-)
Bonner Peak	Northeastern CO	-	-	2270	43	2.1	40	0.181	(-)	0.188	(-)
Windy Site	Northeastern CO	-	-	2100	16	1.5	40	0.150	(-)	0.179	(-)
Hewlett Gulch	Northeastern CO	-	-	1820	9	1.6	60	0.193	(-)	0.233	(-)

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
***Pinus monticola* (western white pine) (Steinhoff *et al.*, 1983)**											
Interior - 5	Beaver, BC	51.55	117.48	1145	-	1.4	42	-	(-)	0.15	-0.057
Interior - 58	Hungry Horse, MT	48.32	113.98	1130	-	1.7	67	-	(-)	0.13	-0.036
Interior - 55	Flower Creek, MT	48.38	115.57	670	-	1.8	75	-	(-)	0.16	-0.050
Interior - 45	Beaver Creek, ID	48.73	116.87	915	-	1.8	67	-	(-)	0.16	-0.049
Interior - 49	Crystal Creek, ID	47.13	116.37	915	-	1.8	83	-	(-)	0.19	-0.048
Interior - 52	Elk River, ID	46.82	116.17	915	-	1.7	67	-	(-)	0.13	-0.032
Interior - 35	Fry Meadow, OR	45.78	117.82	1525	-	1.4	42	-	(-)	0.11	-0.050
Coastal - 3	Coronation, BC	49.00	124.00	870	-	1.8	75	-	(-)	0.18	-0.054
Coastal - 4	Victoria, BC	48.42	123.62	350	-	1.6	58	-	(-)	0.15	-0.060
Coastal - 8	Forks, WA	47.93	124.37	150	-	1.2	17	-	(-)	0.06	-0.038
Coastal - 10	Humptulips, WA	47.30	123.95	150	-	1.1	8	-	(-)	0.04	-0.041
Coastal - 12	Everett, WA	48.08	122.13	30	-	1.5	50	-	(-)	0.10	-0.049
Coastal - 13	Belfair, WA	47.48	122.90	90	-	2.0	92	-	(-)	0.15	-0.039
North-Central - 16	White Pass, WA	46.57	121.37	1310	-	1.6	58	-	(-)	0.13	-0.045
North-Central - 23	Mt. Hood, OR	45.27	121.77	975	-	1.6	58	-	(-)	0.13	-0.050
North-Central - 25	Sweet Home, OR	44.42	122.17	1250	-	1.7	58	-	(-)	0.18	-0.058
South-Central - 29	La Pine, OR	43.80	121.68	1830	-	1.8	67	-	(-)	0.25	-0.060
South-Central - 27	Oakridge, OR	43.60	122.12	1370	-	1.8	67	-	(-)	0.16	-0.049
South-Central - 26	Black Butte, OR	43.53	122.92	1065	-	1.8	75	-	(-)	0.20	-0.053
South-Central - 29	Glide, OR	43.07	122.95	1160	-	1.8	75	-	(-)	0.23	-0.052
South-Central - 30	Prospect, OR	42.87	122.37	1525	-	1.9	75	-	(-)	0.21	-0.052
South-Central - 33	Pinehurst, OR	42.20	122.30	1645	-	1.8	67	-	(-)	0.21	-0.061

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
Siskiyou - 31	Chicago Creek, CA	41.97	123.67	1295	-	1.8	58	-	(-)	0.23	-0.069
Siskiyou - 32	Elliot Creek, CA	42.00	123.00	1735	-	2.0	92	-	(-)	0.24	-0.054
Warner - 34	Lakeview, OR	42.10	120.28	2286	-	2.0	83	-	(-)	0.32	-0.058
Warner - 38	Willow Mtn., CA	41.83	122.23	1980	-	2.3	92	-	(-)	0.29	-0.058
Warner - 41	Lake Audrian, CA	38.65	120.07	2225	-	2.2	83	-	(-)	0.26	-0.069
Warner - 42	Blue Lakes, CA	38.62	119.92	2530	-	2.0	75	-	(-)	0.26	-0.064
(Kim *et al.*, 2011)											
ReSt	Revelstoke, BC	51.00	118.19	500	27	1.606	87.9	-	(-)	0.194	(-)
TeXa	Texada Island, BC	48.26	113.88	1100	25	1.621	90.9	-	(-)	0.198	(-)
SwCr	Swamp Creek, WA	47.59	116.03	1345	29	1.667	87.9	-	(-)	0.208	(-)
ElCr	Elya Creek, MT	49.70	124.37	215	27	1.636	84.8	-	(-)	0.190	(-)
MoSa	Moon Saddle, ID	48.57	120.78	1190	22	1.576	87.9	-	(-)	0.187	(-)
HoLy	Holly, WA	47.57	122.92	150	17	1.636	89.4	-	(-)	0.215	(-)
SmCr	Smokey Creek, WA	46.02	121.68	1114	21	1.621	90.9	-	(-)	0.204	(-)
VeLa	Veda Lake, OR	45.26	121.76	1310	26	1.727	90.9	-	(-)	0.233	(-)
InCrb	Indian Creek, OR	44.36	118.79	1585	25	1.758	92.4	-	(-)	0.274	(-)
DeCr	Deer Creek, OR	43.25	121.86	1740	28	1.894	97.0	-	(-)	0.316	(-)
NaCr	National Creek, OR	43.00	122.38	1150	23	1.697	95.5	-	(-)	0.225	(-)
FlSp	Fly Catcher Spring, OR	42.35	124.30	800	27	1.773	89.4	-	(-)	0.275	(-)
GaQu	Gasquet, CA	41.86	123.91	435	21	1.789	93.9	-	(-)	0.264	(-)
LoSt	Lodge Pole Station, CA	41.83	122.21	1880	22	1.773	92.4	-	(-)	0.289	(-)
SpCa	Spur Canyon, CA	38.70	120.10	2500	17	1.721	81.8	-	(-)	0.254	(-)

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
***Pinus ponderosa* (ponderosa pine)**											
(O'Malley *et al.*, 1979)											
OMALLEY-1979	10 pop., WA, ID, MT	-	-	-	47	2.25	-	-	(-)	0.123	(-)
(Woods *et al.*, 1983)											
WOODS-1983	1 pop., Eastern MT	-	-	-	150	2.6	-	-	(-)	0.126	-0.037
(Hamrick *et al.*, 1979)											
HAMRICK-1979	7 pop., Eastern CO	-	-	-	-	2.0	68.4	-	(-)	0.226	(-)
(Niebling & Conkle, 1990)											
Pacific - Warner Mtns.	1 pop., Northern CA	-	-	-	65	2.39	70	-	(-)	0.150	-0.036
Pacific - Mt. Rose	1 pop., Northern CA	-	-	-	55	2.29	70	-	(-)	0.137	-0.035
North Plateau	12 pop., Pacific Northwest US	-	-	-	204	2.77	64	-	(-)	0.178	-0.041
Rocky Mtns.	20 pop., US Rocky Mtns.	-	-	-	186	2.77	75	-	(-)	0.164	-0.039
***Populus tremuloides* (trembling aspen)**											
(Cheliak & Dancik, 1982)											
1	AB	-	-	-	32	2.0	85	0.49	(-)	0.41	(-)
2	AB	-	-	-	30	2.3	88	0.46	(-)	0.40	(-)
3	AB	-	-	-	30	2.3	88	0.49	(-)	0.44	(-)
4	AB	-	-	-	30	2.3	84	0.53	(-)	0.41	(-)
5	AB	-	-	-	30	2.5	92	0.60	(-)	0.43	(-)
6	AB	-	-	-	30	2.3	85	0.52	(-)	0.42	(-)
7	AB	-	-	-	40	2.4	88	0.56	(-)	0.42	(-)

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
(Jelinski & Cheliak, 1992)											
Prairie I	Waterton Lakes N.P., AB	-	-	1300	37	2.5	87.5	0.30	(-)	0.315	-0.042
Prairie II	Waterton Lakes N.P., AB	-	-	1300	25	2.4	75.0	0.35	(-)	0.258	-0.047
Akamina	Waterton Lakes N.P., AB	-	-	2700	25	2.2	81.3	0.31	(-)	0.346	-0.046
Galwey	Waterton Lakes N.P., AB	-	-	1810	23	2.4	75.0	0.34	(-)	0.267	-0.047
Crandall	Waterton Lakes N.P., AB	-	-	1360	22	2.4	81.3	0.33	(-)	0.306	-0.056
Copper	Waterton Lakes N.P., AB	-	-	1400	24	2.5	87.5	0.28	(-)	0.246	-0.040
Prairie I	Waterton Lakes N.P., AB	-	-	1300	37	2.5	87.5	0.30	(-)	0.315	-0.042
Prairie II	Waterton Lakes N.P., AB	-	-	1300	25	2.4	75.0	0.35	(-)	0.258	-0.047
Akamina	Waterton Lakes N.P., AB	-	-	2700	25	2.2	81.3	0.31	(-)	0.346	-0.046
Galwey	Waterton Lakes N.P., AB	-	-	1810	23	2.4	75.0	0.34	(-)	0.267	-0.047
Crandall	Waterton Lakes N.P., AB	-	-	1360	22	2.4	81.3	0.33	(-)	0.306	-0.056
Copper	Waterton Lakes N.P., AB	-	-	1400	24	2.5	87.5	0.28	(-)	0.246	-0.040
(Liu & Furnier, 1993)											
LIU1993	MI, MN, WI	-	-	-	118	2.8	77	0.19	-0.05	0.32	-0.06
(Lund *et al.*, 1992)											
A	MN	-	-	-	40	2.7	90	0.205	(-)	0.206	(-)
B	MN	-	-	-	38	2.7	100	0.224	(-)	0.235	(-)
C	MN	-	-	-	40	2.6	90	0.205	(-)	0.215	(-)
D	MN	-	-	-	40	2.8	100	0.235	(-)	0.244	(-)
E	MN	-	-	-	40	2.4	80	0.230	(-)	0.234	(-)
F	MN	-	-	-	40	2.7	100	0.195	(-)	0.212	(-)
G	MN	-	-	-	40	2.9	100	0.242	(-)	0.232	(-)

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H$_o$	(SE)	H$_e$	(SE)
H	MN	-	-	-	29	2.3	80	0.217	(-)	0.210	(-)
I	MN	-	-	-	40	2.2	80	0.197	(-)	0.193	(-)
(Hyun *et al.*, 1987)											
Moosonee	ON	-	-	-	19	2.5	80.0	0.108	(-)	0.248	(-)
Kenora	ON	-	-	-	29	2.9	93.3	0.119	(-)	0.233	(-)
Cochrane	ON	-	-	-	28	2.8	66.7	0.133	(-)	0.207	(-)
Thunder Bay	ON	-	-	-	29	2.8	73.3	0.160	(-)	0.227	(-)
Hearst	ON	-	-	-	27	2.9	73.3	0.101	(-)	0.234	(-)
Sudbury	ON	-	-	-	30	2.5	86.7	0.114	(-)	0.212	(-)
Barrie	ON	-	-	-	10	2.1	73.3	0.140	(-)	0.246	(-)
Simcoe	ON	-	-	-	28	1.9	86.7	0.127	(-)	0.270	(-)
***Pseudotsuga menziesii* (douglas-fir)**											
(El-Kassaby & Ritland, 1996)											
Alberni	Port Alberni, BC										
Bowen Island	Bowen Island, BC										
Cassidy	Cassidy, BC										
Caycuse	Caycuse, BC										
Chehalis	Chehalis, BC	49.2	124.8	-	51	2.25	55	-	(-)	0.153	(-)
Chilliwack High	Chilliwack, BC	49.4	123.4	-	51	2.25	55	-	(-)	0.162	(-)
Chilliwack Low	Chilliwack, BC	49.1	123.9	-	51	2.05	45	-	(-)	0.146	(-)
Courtenay	Courtenay, BC	48.9	124.4	-	51	2.25	50	-	(-)	0.186	(-)
Darrington	Darrington, WA	46.7	123.0	-	51	2.15	50	-	(-)	0.162	(-)
Denman Island	Denman Island, BC	49.2	121.9	-	51	2.25	60	-	(-)	0.159	(-)

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
Duncan	Duncan, BC	49.2	121.9	-	51	2.20	50	-	(-)	0.158	(-)
Empress Mtn.	Empress Mtn., BC	49.7	125.0	-	51	2.20	40	-	(-)	0.144	(-)
E. Thurlow	East Thurlow, BC	48.3	121.6	-	51	2.15	40	-	(-)	0.122	(-)
Forbidden Plat.	Forbidden Plateau, BC	49.5	124.8	-	51	2.25	55	-	(-)	0.188	(-)
Franklin R.	Franklin River, BC	48.8	123.7	-	51	2.15	60	-	(-)	0.169	(-)
Gold River	Gold River, BC	48.4	123.7	-	51	2.15	50	-	(-)	0.181	(-)
Granite Falls	Granite Falls, WA	50.4	125.4	-	51	2.25	65	-	(-)	0.164	(-)
Haney	Haney, BC	49.7	125.2	-	51	2.30	50	-	(-)	0.177	(-)
Harrison Lake	Harrison Lake, BC	49.1	124.8	-	51	2.10	50	-	(-)	0.151	(-)
Hernando Is.	Hernando Island, BC	49.8	126.1	-	51	2.10	50	-	(-)	0.152	(-)
Hoh	Hoh, WA	48.1	122.0	-	51	2.10	40	-	(-)	0.139	(-)
Hope	Hope, BC	49.2	122.6	-	51	2.10	55	-	(-)	0.160	(-)
Jervis	Jervis Inlet, BC	49.3	121.8	-	51	2.05	55	-	(-)	0.166	(-)
Jeune Landing	Jeune Landing, BC	50.0	124.9	-	51	2.05	65	-	(-)	0.181	(-)
Jordan River	Jordan River, BC	47.9	123.9	-	51	2.30	70	-	(-)	0.181	(-)
Kaouk River	Kaouk River, BC	49.4	121.4	-	51	2.10	50	-	(-)	0.126	(-)
Klinaklini High	Klinaklini, BC	50.2	124.0	-	51	2.15	50	-	(-)	0.164	(-)
Klinaklini Low	Klinaklini, BC	50.4	127.5	-	51	2.15	55	-	(-)	0.185	(-)
Kelsey Bay	Kelsey Bay, BC	48.4	124.0	-	51	2.20	55	-	(-)	0.169	(-)
Kennedy Lake	Kennedy Lake, BC	50.1	127.1	-	51	2.15	55	-	(-)	0.189	(-)
Malcolm Island	Malcolm Island, BC	51.3	125.8	-	51	2.00	55	-	(-)	0.190	(-)
Mead Cr.	Mead Creek, BC	51.3	125.8	-	51	2.00	65	-	(-)	0.157	(-)
Nimpkish	Nimpkish, BC	50.4	126.0	-	51	2.10	55	-	(-)	0.160	(-)
Nanaimo L.	Nanaimo Lake, BC	49.0	125.6	-	51	1.75	45	-	(-)	0.138	(-)

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
Parksville	Parksville, BC	50.6	127.0	-	51	1.90	40	-	(-)	0.143	(-)
Pitt River	Pitt River, BC	48.8	124.1	-	51	2.35	50	-	(-)	0.181	(-)
Port Hardy	Port Hardy, BC	50.3	126.9	-	51	2.15	65	-	(-)	0.169	(-)
Powell River	Powell River, BC	49.1	124.2	-	51	2.15	45	-	(-)	0.166	(-)
Quadra Island	Quadra Island, BC	49.3	124.3	-	51	2.20	55	-	(-)	0.146	(-)
San Juan R.	San Juan River, BC	49.3	122.6	-	51	2.05	50	-	(-)	0.171	(-)
Sechelt	Sechelt, BC	50.7	127.5	-	51	2.20	55	-	(-)	0.198	(-)
Shelton	Shelton, WA	49.9	124.6	-	51	2.15	50	-	(-)	0.161	(-)
Sooke	Sooke, BC	50.1	125.3	-	51	2.05	65	-	(-)	0.167	(-)
Squamish	Squamish, BC	48.5	124.0	-	51	2.15	50	-	(-)	0.157	(-)
Stella Lake	Stella Lake, BC	49.5	123.8	-	51	2.05	50	-	(-)	0.158	(-)
Tahsis	Tahsis, BC	47.2	123.1	-	51	2.10	45	-	(-)	0.148	(-)
Texada Island	Texada Island, BC	48.4	123.7	-	51	2.20	55	-	(-)	0.161	(-)
Tinhat Mtn.	Tinhat Mtn., BC	49.7	123.2	-	51	2.20	55	-	(-)	0.169	(-)
Toba River	Toba River, BC	50.3	125.5	-	51	2.15	45	-	(-)	0.160	(-)
(Yeh & O'Malley, 1980)											
I	South Vancouver Island, BC	48.55	124.08	244	100	2.33	42.86	-	(-)	0.156	-0.041
II	South Vancouver Island, BC	48.87	123.75	457	100	2.33	57.14	-	(-)	0.156	-0.035
III	Skagit Valley, BC	49.23	121.23	930	100	2.29	52.38	-	(-)	0.168	-0.040
IV	Lower Mainland, BC	49.25	122.35	152	100	2.29	47.62	-	(-)	0.161	-0.041
V	Hope, BC	49.50	121.35	884	100	2.10	57.14	-	(-)	0.149	-0.039
VI	Halfmoon Bay, BC	49.50	123.87	175	100	2.34	47.62	-	(-)	0.139	-0.038
VII	North Vancouver Island, BC	49.87	125.83	250	100	2.14	47.62	-	(-)	0.142	-0.036
VIII	North Vancouver Island, BC	50.25	125.73	137	100	2.05	47.62	-	(-)	0.140	-0.037

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
IX	West Lillooet, BC	50.53	122.47	457	100	1.91	61.91	-	(-)	0.167	-0.041
X	West Lillooet, BC	50.57	122.53	1067	100	2.29	57.14	-	(-)	0.184	-0.042
XI	Bella Coola, BC	52.42	126.25	244	100	2.05	47.62	-	(-)	0.140	-0.041
(Hamrick *et al.*, 1979)											
YEH	11 pop., Interior BC	-	-	-	-	2.23	-	-	(-)	0.180	(-)
MORRIS	11 pop., Coastal CA	-	-	-	-	3.17	74.2	-	(-)	0.332	(-)
CONKLE	1 pop., Interior CA	-	-	-	-	1.78	100	-	(-)	0.330	(-)
HAMRICK	5 pop., Eastern CO	-	-	-	-	1.86	64	-	(-)	0.264	(-)
Sequoia sempervirens **(coast redwood)** **not included in analysis*											
(Rogers, 1994)											
High Rock River Access	Humboldt Redwoods S.P., CA	-	-	50	121	3.00	80	-	(-)	-	(-)
Kent Grove	Humboldt Redwoods S.P., CA	-	-	55	121	3.25	92	-	(-)	-	(-)
Grasshopper Hill	Humboldt Redwoods S.P., CA	-	-	213	121	3.25	83	-	(-)	-	(-)
Squaw Creek Ridge	Humboldt Redwoods S.P., CA	-	-	210	121	2.92	75	-	(-)	-	(-)
Thuja plicata **(western redcedar)**											
(Yeh, 1988)											
1 - Cowichan Lake	Interior BC	48.817	124.167	161	20	1.16	15.79	0.44	(-)	0.050	-0.031
2 - Salmo	Interior BC	49.183	117.250	610	30	1.16	15.79	0.32	(-)	0.038	-0.026
3 - Tashme	Interior BC	49.333	121.283	734	30	1.16	15.79	0.42	(-)	0.036	-0.026
4 - Kootney Lake	Interior BC	49.667	116.667	532	30	1.21	15.79	0.55	(-)	0.051	-0.031
5 - Kamloops	Interior BC	50.700	120.333	344	30	1.16	15.79	0.32	(-)	0.035	-0.026
6 - Revelstoke	Interior BC	51.033	118.250	610	30	1.11	10.53	0.30	(-)	0.036	-0.027
7 - Adams Lake	Interior BC	51.217	119.500	404	30	1.21	21.05	0.29	(-)	0.037	-0.026

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
8 - Quesnel Lake	Interior BC	52.617	120.967	697	30	1.16	15.79	0.24	(-)	0.031	-0.026
(El-Kassaby *et al.*, 1994)											
Strait of Juan De Fuca	2 pop., Van Isl & BC Coast	-	-	-	28	1	12	-	(-)	0.06	(-)
(Copes, 1981)											
COPES	5 pop., WA, OR	-	-	-	49	1	0	0	0	0	0
***Tsuga heterophylla* (western hemlock)**											
(El-Kassaby *et al.*, 2003)											
OG-TSUGHET	Campbell River, BC	49.917	125.417	795	30	1.6	38.1	0.098	(-)	0.113	(-)
(Wellman, 2004)											
HOLBERG	Holberg, BC	50.633	128.083	92	49	2	50	0.143	-0.071	0.171	-0.051
TOBA	Toba River, BC	50.500	124.167	366	49	1.9	50.0	0.149	-0.072	0.146	-0.045
UCONA	Ucona River, BC	49.667	126.000	475	49	1.9	50.0	0.131	-0.069	0.145	-0.045
FLEET	Fleet River, BC	48.633	124.067	370	50	1.6	28.6	0.117	-0.071	0.132	-0.054
SOMBRIO	Sombrio Creek, BC	48.533	124.300	365	50	1.6	42.9	0.143	-0.075	0.147	-0.052
CAMPER	Camper Creek, BC	48.567	124.500	300	50	1.4	28.6	0.119	-0.075	0.114	-0.051
SECHELT	Sechelt, BC	-	-	-	52	1.6	42.9	0.160	-0.076	0.148	-0.049
NANAIMO	Nanimo River, BC	49.000	124.167	550	50	1.9	35.7	0.140	-0.072	0.131	-0.046
UBC	UBC Res. Forest, BC	49.283	122.550	275	49	1.7	28.6	0.112	-0.073	0.137	-0.054
KAOUK	Kaouk, BC	50.083	126.983	60	50	1.6	42.9	0.149	-0.075	0.147	-0.053
UNKNOWN	Unknown natural, BC	48.833	123.750	600	49	1.5	42.9	0.161	-0.079	0.131	-0.048

Continued on next page

Table S4.1 – continued from previous page

Population	Location/Comments	Lat	Long	Elev	N	A	P	H_o	(SE)	H_e	(SE)
***Tsuga mertensia* (mountain hemlock)**											
(Ally *et al.*, 2000)											
A - Meade Crk.	Meade Crk. , BC	48.55	124.05	1067	40	1.9	36.8	-	(-)	0.092	-0.026
B - Wakeman High	Wakeman High , BC	51.10	126.25	1100	40	1.7	42.1	-	(-)	0.097	-0.028
C - Hanna Ridge	Hanna Ridge , BC	56.18	129.20	700	40	1.5	31.6	-	(-)	0.065	-0.023
D - Wakeman Low	Wakeman Low , BC	51.17	126.17	600	40	1.6	42.1	-	(-)	0.084	-0.025
E - Garbage Crk.	Garbage Crk. , BC	48.33	124.06	850	40	1.6	31.6	-	(-)	0.106	-0.040
F - Mission	Mission , BC	49.18	122.26	900	40	1.7	31.6	-	(-)	0.122	-0.042
G - Lyon Lk.	Lyon Lk. , BC	49.39	123.54	1005	40	1.4	26.3	-	(-)	0.068	-0.028
H - Parksville	Parksville , BC	49.16	124.33	824	40	1.6	31.6	-	(-)	0.102	-0.033
I - Mayo Crk.	Mayo Crk. , BC	54.47	129.02	683	40	1.7	36.8	-	(-)	0.088	-0.027
J - Zeballos	Zeballos , BC	50.10	126.47	700	40	1.3	21.1	-	(-)	0.050	-0.025
K - Kearsley Crk.	Kearsley Crk. , BC	49.19	122.22	1280	40	1.7	42.1	-	(-)	0.106	-0.036
L - Blue Ox Crk.	Blue Ox Crk. , BC	50.18	127.16	660	40	1.2	10.5	-	(-)	0.056	-0.036
M - Hkusam Mt.	Hkusam Mt. , BC	50.20	125.50	950	40	1.4	31.6	-	(-)	0.075	-0.028
N - Port Alice	Port Alice , BC	50.24	124.27	750	40	1.4	26.3	-	(-)	0.086	-0.035
O - Sale Mt.	Sale Mt. , BC	51.10	118.10	1700	40	1.2	10.5	-	(-)	0.050	-0.034
P - Hodoo Crk.	Hodoo Crk. , BC	51.20	125.32	1250	40	1.6	31.6	-	(-)	0.095	-0.030
Q - Copper Canyon	Copper Canyon , BC	48.56	124.13	1100	40	1.8	47.4	-	(-)	0.120	-0.057
R - Ashley Crk.	Ashley Crk. , BC	50.01	123.33	1000	40	1.7	36.8	-	(-)	0.109	-0.034
S - Woss Lk.	Woss Lk. , BC	50.07	126.35	900	40	1.6	36.8	-	(-)	0.082	-0.027

Chapter 5

The velocity of species migration in complex landscapes: western North American trees under past and future climate change.

5.1 Summary

Due to differences in time periods (decades vs. millennia), rates of future migration are almost certain to be orders of magnitude faster than estimated post-glacial rates. However, it is possible that future migrations will also be different in pattern and quality, as climates move into conditions not seen for millions of years in North America. The ability of species to persist through the next century may depend on their ability to keep pace, either through physical migration or gene flow, with changes in environmental habitat. For 24 North American trees, we calculate distances between similar climate habitats in the past, present, and future, with a *k*-sample nearest neighbour algorithm that measures distances to suitable climate habitats in different time periods, as a means of estimating required migration and gene flow rates. We use two measurements for future climate change: 1) measuring from the present range to future habitats, to assess how far modern populations must travel to find suitable habitat in the 2080s, and 2) measuring from projected future habitats to the present range, to determine from how far away suitably adapted populations must travel to ensure productive future forests. To place these rates in context, we also estimate post-glacial migration rates with the same nearest neighbour method. Results show that past and future migration rates by species are strongly related (r^2=0.59, $p<0.001$) suggesting that, while the magnitude may differ, the quality or patterns of migrations in the future may be similar to those through the late Quaternary. Measurements into the future suggest that many

modern populations, particularly those in high elevations and higher latitudes, may be far removed from suitable habitats in the future. Further, to be occupied by suitably-adapted populations, much future species range area may require genotypes from long distances, suggesting that assisted migration or assisted gene flow could be required for productive populations to establish and persist.

5.2 Introduction

By altering the geographic ranges of environmental conditions to which species are adapted, climate change may present threats to tree species, which are long-lived and have limited dispersal capacity (Elith & Leathwick, 2009; Thomas *et al.*, 2004). The ability of such species to survive the environmental changes of the next century may depend on their ability to keep pace, either through physical migration into new range area or though the movement of genetic material within their range (i.e. gene flow), with environmental habitat that remains within their fundamental niche space. All species in existence today have proven their ability to migrate with suitable habitat through the glacial-interglacial oscillations during the last 2.5 million years of the Quaternary, with temperatures being approximately 4-6°C colder during glacial periods. However, future changes are projected towards climate conditions that have not existed for millions of years. For example, temperatures 3-4°C warmer than the present last occurred during the Miocene, 5 to 20 million years ago (Crowley, 1990). The inference that species, because they have withstood global mean annual temperature fluctuations of 4-6°C in the past, will be able to tolerate such changes in the future, is questionable even if changes were to occur at similar rates.

Directly measuring species migration rates, particularly for immobile species such as trees, is complicated by the many slow, stochastic, and often unobservable processes involved. Migration rates are typically estimated by measuring species location changes over time. For North American and European trees, several post-glacial migration rate estimates have been made, largely by comparing where and when species appear in the paleoecological record. Most of these estimates range around a few hundreds of metres per year, depending on the species (Davis, 1981; Delcourt & Delcourt, 1987; Huntley & Birks, 1983; King & Herstrom, 1997), but may be as low as only tens of metres per year or less (McLachlan *et al.*, 2005). However, due to the potential for stochastic events such as long-distance dispersal or confounding colonisation sources such as cryptic glacial refugia (Clark *et al.*, 1998; McLachlan *et al.*, 2005; Pearson, 2006), these estimates of late Quaternary migration rates must be interpreted with caution. Furthermore, the paucity of paleoecological data in some areas and for many species has resulted in a relatively short list of species for which palaeoecologically-based migration rates are available.

Required future rates of species migration have previously been estimated by comparing projected species range changes over time. Early calculations of required migration rates were made for North American trees by Davis (1989), Davis and Zabinski (1992), and Schwartz (1992), and ranged from about a kilometre per year to tens of kilometres per year. More recent studies have expanded the breadth of species considered and, employing newer climate projections and more sophisticated species models, largely supported these original rate estimates (Iverson & Prasad, 2002; Malcolm *et al.*, 2002; Meier *et al.*, 2012). Given projections of 2-6°C of warming over the next hundred years (IPCC, 2007), and compared to past changes of 4-6°C warming over several thousand years through the late Quaternary, it is not surprising that required future migration rates are consistently calculated to be orders of magnitude faster than reconstructed post-glacial rates. However, just as the spatial distribution of global warming will not be consistent across the earth, neither will be the required rates of species migration. Heterogeneity in future warming and precipitation changes, as well as diversity within the landscapes on which they occur, will produce great variation in rates of movements of suitable habitat for trees (Loarie *et al.*, 2009).

Different required migration rates may dictate different response by tree populations to environmental change, as outlined by Aitken et al. (2008). Organisms may remain in situ and adapt genetically to new climate conditions. Alternatively or in addition, populations may physically migrate to nearby habitats that become more suitable. Calculated rates of required future migration could account for both mechanisms. Where species' current ranges overlap with their projected future ranges, populations may require the introduction of genetic material from other distant populations that are adapted to the future climate habitats for that location. Gene flow rates between tree populations are difficult to quantify, but because pollen is easily transported by wind (tens or even hundreds of kilometres) gene flow rates in trees is generally understood to be higher than in other plants with similar mating systems (Burczyk *et al.*, 2004; Petit & Hampe, 2006).

In this study, we contribute estimates of required migration rates via seed dispersal and required gene flow rates via pollen for populations of 24 western North American tree species, based on calculated rates of climatically defined habitat movement. We also estimate migration rates since the last glacial maximum (21,000 years before present) for perspective. While we know that rates estimated for the future will be an order of magnitude faster than those required of species since the last glacial maximum, we also hypothesise that they will be qualitatively different. Depending on the current latitudinal and topographic position of a species (or its current realised climate niche) in western North America, it is far from certain that migration requirements in the past are indicative of migration requirements under future climate change. Specifically, we want to identify species and populations of species for which migration rates since the last glacial maximum, representing a warming signal of about 4°C do not correlate with migration rates estimated for an additional of 4°C of anthropogenic warming.

This study is inspired by Loarie et al. (2009) who calculated the velocity of climate envelope movements, given projected changes in temperature, across complex topographic landscapes. Here, we extend this approach to consider multiple climate variables and further extend the method to directly measure habitat change for western North American tree species. We employ a *k*-sample nearest neighbour algorithm to estimate the velocity of climate habitat movement at any given point in the landscape. We run these calculations in two directions, from current populations to potential future habitat, as well as from projected future climates back to current populations to answer two related, yet distinct questions. First, what rates of migration or gene flow are required for today's locally-adapted populations to find suitable climate habitats in the future? The resulting statistics have implication for gene conservation. Genes from populations with no future climate equivalents in the vicinity are most likely to be lost. The second question has implications for assisted migration in forestry: given the climate conditions of a future planting site, how far away is the nearest appropriately adapted population? If no matching populations exist in the vicinity, then seed transfer in reforestation programs may be required to maintain healthy and productive forests.

5.3 Methods

5.3.1 Climate data

Climate data was generated based on a 1km resolution digital elevation model, using the ClimateWNA software package (Wang *et al.*, 2012). The climate variables include mean annual temperature, mean temperature of the warmest month (MWMT), mean temperature of the coldest month (MCMT), the difference between MWMT and MCMT as a measure of continentality, mean annual precipitation, growing season precipitation from May to September, the number of growing degree days above 5°C, the number of frost-free days, and an annual and summer heat moisture index (Wang *et al.*, 2012).

We used the 1961-1990 climate normal period to represent current climate conditions, and a recent 10-year climate average from 1997-2006 to represent observed climate trends. Palaeoclimate conditions for the last 21,000 years was based on two coupled oceanic-atmospheric general circulation models (GCMs), the Community Climate Model version 1 (CCM1) of the National Center for Atmospheric Research (NCAR) for the periods 21,000, 16,000, 14,000, 11,000, and 6,000 years ago (Kutzbach *et al.*, 1998), and the Geophysical Fluid Dynamics Laboratory (GFDL) model of the National Oceanic and Atmospheric Administration (NOAA) for the periods 21,000, 16,000, 9,000, and 6,000 years ago (Bush & Philander, 1999). Future climate data was based on an ensemble average of 14 GCMs from the IPCC's Fourth Assessment Report (IPCC, 2007) for the 2071-2100 period, hereafter referred to as the 2080s. An optimistic (B1) and a pessimistic (A2) emission scenario were considered. To generate climate surfaces for past and future periods, past and future climate projections were added as interpolated anomalies to the 1961-1990 high-resolution baseline dataset, implemented by the ClimateWNA software package (Wang *et al.*, 2012).

5.3.2 *k*-sample nearest neighbour algorithm

The k-sample nearest neighbour algorithm was implemented with the yaImpute package (Crookston & Finley, 2008) for the R programming environment (R Core Team, 2009), which facilitates searches between datasets, identifying a pre-described number of data points most similar to the source point. Because the k-sample nearest neighbour algorithm is computationally intensive depending on the number of variables, we extracted two principal components from the ten climate variables described above, accounting for 81.7% of the variance in the climate dataset (Table 5.1). Thus, our measure of climate distance (Eucleadian distance based on the first two principal components), is approximately equivalent to the widely used Mahalanobis multivariate distance measure.

For our purposes, the k-sample nearest neighbour algorithm allows us to identify, for every point in a source climate data set, the k most similar points in an alternative climate data set. Once a k-sized sample of the most climatically similar points is generated, Euclidian geographic distances between the source point and all k alternative points can be calculated from the geographic x and y coordinates of the source and target locations. The shortest geographic distance to a set of reasonably similar climate equivalents was then used to determine the required migration rate. The robustness of the analysis was enhanced by choosing a k-sample size where migration rate estimates do not significantly change as a function of the k-sample size. This value is quite small ($k \approx 50$), derived from a scree plot of k-sample sizes against the subsequently calculated migration rates (Figure 5.1). At >50 points, rate estimates continue to decrease gradually, but interpretation of rate maps are made based on relative rates, which remain consistent.

5.3.3 Rates of migration and gene flow

To calculate the rate of climatically defined habitat movement, as an estimate of migration and gene flow rates, for local populations of each of the 24 western North America tree species, we constrained the k-sampling nearest neighbour analysis to projected species habitat. Past, present, and future species ranges were projected using an ensemble ecological niche modelling approach incorporating the mean output (probability of species presence) of three ecosystem-based model projections, as described in Chapters 3 and 4 of this thesis. Summaries of model validations are shown in Table 4.2, in Chapter 4.

To determine the distance between modern climates and their nearest future analogues, the k-nearest neighbour algorithm was applied to measure geographic distances from every point in the current climate data falling within the current species range, to the most similar climate points in the past or future climate data falling within the modelled past or future species range for that period. To determine the distance between future climates and their nearest modern analogue, the opposite measurement was undertaken (measuring from the future to the present). In both cases, geographic distances are converted to migration rates by dividing by the number of interim years. To differentiate required migration rates and required gene flow rates, we separately evaluated overlapping and non-overlapping habitat projections. This analysis assumes that populations are adapted to their local climate, which is generally supported in the literature (e.g. Savolainen *et al.*, 2007).

5.4 Results

Rates of species migration and gene flow estimated for climate change projections and reconstructions (but considering multiple climate variables) are shown in Figure 5.2 and Table 5.2, which includes median and 90^{th} percentile rates of tree populations within 1x1 km grid cells used for modelling but do not distinguish between gene flow and migration. Table 5.3 breaks the estimates down for rates within overlapping species ranges (representing gene flow) and non-overlapping portions of the species ranges (representing migration). Histograms of estimated post-glacial and future required gene flow or migration rates combined, representing populations within 1x1 km grid cells, are provided in Figure 5.3 and the mapped rates that form the basis of these figures and tables are shown in Supplemental Figures S5.1A-S5.1X.

Median post-glacial rates ranged from only tens of metres per year for some coastal species to more than two hundred metres year for most boreal species, with 90^{th} percentile rates at about double these medians. Future rates were roughly an order of magnitude, or more, faster than post-glacial rates, ranging from a few hundred metres per year for coastal species to more than three kilometres per year for some boreal species. Estimates of migration rates were typically higher than those for gene flow rates (Table 5.3). Coast redwood (*Sequoia sempervirens*) was notable as having the slowest rates in almost all calculations, moving only 10-30 m yr^{-1} in post-glacial calculations and only 0.11-0.23 km yr^{-1} under the future projections. In contrast, trembling aspen (*Populous tremuloides*), white spruce (*Picea glauca*), black spruce (*Picea mariana*), and paper birch (*Betula papyrifera*), all boreal species, had the highest rates of post-glacial migration (0.20-0.25 km yr^{-1}) and future migration/gene flow (2.88-5.12 km yr^{-1})

For the tree species that we considered, most projected future range was not new area gained, which would require physical migration of the species, but rather present area maintained within the current range. The proportion of future rate requiring gene flow (overlapping range) vs. migration (into new range) was 65% and 50% on average for all species in the B1 and A2 scenarios, respectively. Sub-alpine and xeric species generally required more migration (only 29% to 52% gene flow), while coastal species tended to require the least migration (58% to 75% gene flow) (Table 5.3). Notably, we find a fairly clear linear relationship (r^2=0.59, $p<0.001$) between required migration rates estimated for the past and future. The species that deviates most from the relationship is pinyon pine (*Pinus edulis*) (Figure 5.2, #23), indicating that migration requirements through the North American landscape in the future are fundamentally different than those in the past for this species.

5.5 Discussion

5.5.1 Past and future migration rates

Our overall estimates of gene flow and migration are generally similar to previous estimates for the post-glacial period (e.g. Davis, 1981; Huntley & Birks, 1983; King & Herstrom, 1997; Lawing & Polly, 2011; McLachlan *et al.*, 2005) and for the future (e.g. Davis, 1989; Davis & Zabinski, 1992; Iverson & Prasad, 2002; Lawing & Polly, 2011; Malcolm *et al.*, 2002; Meier *et al.*, 2012; Schwartz, 1992). While past and future rates are of different magnitudes, an exact interpretation of these differences is difficult because we do not know the maximum post-glacial rates of warming that occurred within the intervals for which we have data. Aside from the magnitude difference, we observed a positive correlation (r^2=0.59) between past and future estimates of required migration. Thus, species that had to migrate most in the late Quaternary must do so in the future as well. One reason appears to be that species with extensive post-glacial migrations to boreal latitudes are also those faced with a strong northern warming projection, and species in southern and coastal locations with less projected warming are those that also experienced a milder late Quaternary warming (IPCC, 2007).

We therefore reject the hypothesis that patterns of migration may be qualitatively different in the future than they have been in the past. Disregarding the rate of change, there are similar migration requirements for most species that arise from approximately 4°C increase in global mean annual temperature projected for the future, as have been the result of an approximately 4°C temperature rise since the last glacial maximum. This may at first sight appear to be a trivial result, but there are reasons for the initial hypothesis to be plausible. Species and populations that have survived many repeated cycles of glaciation during the Pleistocene may, for example, be those that can easily find habitat at high elevation during warm periods and at lower elevations during glaciations. However, additional warming might result in equivalent climate spaces at higher elevation than is realised in the landscape nearby and, therefore, may require very different migration requirements than the species has experienced in the past. Our analysis shows that this is generally not the case, perhaps with the exception of *Pinus edulis* (pinyon pine), where current southern montane distributions had close-by climate equivalents in the past, but not in the future (Figure 5.2, #23 and Figure S5.1P).

Furthermore, what applies to the median required migration rates for *Pinus edulis* (pinyon pine) at the species level, may also apply to individual populations within the species range for many other trees. While a species as a whole is not subject to unprecedented migration requirements (aside from much faster rates), uniquely adapted individual populations might be. This is evident in maps of post-glacial rates for species such as *Pseudotsuga menziesii* (Douglas-fir, Figure S5.1T) and *Pinus ponderosa* (ponderosa pine, Figure S5.1R), where notable boundaries in migration rates divide modern coastal and interior subspecies, which are adapted to very different climate habitats. This pattern is in contrast to species that expanded post-glacially from a single refugium, which results in much more uniform required migration rates, such as *Chamaecyparis nootkatensis* (yellow cedar, Figure S5.1F) and *Abies amabilis* (subalpine fir, Figure S5.1A).

5.5.2 Identifying populations at risk

This analysis allows us to quantify at what rate locally-adapted populations may have to move to find suitable climate equivalents in the future, given our modelled range projections. For some populations within the species range, similar future habitat is very close to the current location, whereas genotypes found in other populations may be at risk of extirpation due to their geographic separation from suitable climate equivalents in the future. In our measurements from the present to the future, some of the fastest required migration rates emerge for populations currently at high elevations, particularly for species with extensive modern range in the mountainous areas of the southern United States such as *Abies lasiocarpa* (subalpine fir, Figure S5.1A), *Pinus contorta* (lodgepole pine, Figure S5.1O), *P. albicaulis* (whitebark pine, Figure S5.1N), *P. edulis* (pinyon pine, Figure S5.1P), and *Populus tremuloides* (trembling aspen, Figure S5.1S). In these southern range extents, species with habitat envelopes forced by warming temperatures to continue upslope literally run out of physical habitat at mountaintops, leaving these populations several hundred or thousand kilometres from their nearest climate analogues (in cooler locations much farther north or in other disjunct mountainous areas). In contrast, low required migration rates downslope suggest that genotypes of these populations may persist through short-distance upslope migration or gene flow (visible in green or yellow areas around mountaintops in red in Figures S5.1A, S5.1O, S5.1N, S5.1P, and S5.1S).

These topographically diverse locations may have supported the same genotypes for millions of years through the Quaternary, with populations surviving millennia of warming and cooling cycles by simply shifting their ranges very short distances up and down the local mountainsides. Under climate change, these populations are now at risk of being permanently lost, as they are forced to either adapt to different climate conditions and, in the process, alter their current gene pool, or face the difficulty of long-distance migration or gene flow to equivalent future habitats. This observation may also provide a partial answer to the "Quaternary conundrum" raised by Botkin et al. (2007), which asks why so many species are projected at risk of extinction under future climate change, yet so few were lost through the numerous climatic oscillations of the Quaternary. The answer may lie in the new direction of climate change, as the warmest conditions of the past two million years are projected to quickly emerge on the North American landscape. We show that, while the "Quaternary conundrum" may not be solved with respect to species extinctions, without nearby climate equivalents in the future, unique genotypes within species could certainly be lost.

5.5.3 Priorities for assisted migration or assisted gene flow

Our second implementation of migration distance measurements, from future habitat back to current populations, has applications in forest management, namely for assisted migration prescriptions in reforestation programs to address climate change. The distance from future habitat back to modern equivalents can be interpreted as a relative difficulty for populations to adapt to new climate conditions supported by gene flow from matching populations. Because gene flow is generally high in temperate trees, populations with nearby climate equivalents may already contain appropriate genotypes, allowing for rapid natural adaptation to new climate conditions. In contrast, projected habitats without nearby current climate equivalents may require human intervention to maintain forest health and productivity. Such assisted migration prescriptions could be implemented through existing reforestation programs, and this analysis can serve to prioritize assisted migration interventions. For commercially important species such as *Picea engelmanii* (Engelmann spruce) and *Pinus contorta* (lodgepole pine), required rates for future forest adaptation tend to be high in places such as the British Columbia interior, where both species are widely harvested (Figure S5.1J and Figure S5.1O, respectively). Because of the lack of suitably-adapted genotypes nearby, reforestation in these areas may benefit from non-local seed sourcing, which has also been suggested based on practical trials (Ukrainetz *et al.*, 2011).

While the focus of our study is on future migration rates to the end of the century, the high warming rates observed between the 1961-1990 and 1997-2006 periods (Wang *et al.*, 2012) suggest that many locally adapted populations may already be outside the climate habitats to which they are well-adapted. Recent population declines of some species, which are often observed not just at the southern and low elevation trailing edges of species ranges, may reflect the importance of considering within-species genetic variation (Chen *et al.*, 2010; van Mantgem *et al.*, 2009). Although there are conceptual reasons why species may persist in habitat outside their realised niches, we would generally expect species declines in locations in which species distribution models project habitat losses in the future, such as in the lower elevations of the mountains and highlands in the southwest United States, as can be seen in species such as *Picea engelmanii* (Engelmann spruce) (Figure S5.1J). However, tree species declines in the central or core areas may also occur due to the emergence of climatic conditions for which well-adapted genotypes are not closely available. An example of locally-adapted populations that may already be at the limits of their climatic tolerances may be *Chamaecyparis nootkatensis* (yellow cedar). Documented declines of this species (Hennon *et al.*, 2012, 2005) coincide with populations that were identified in our analysis as having no nearby climate equivalents (Figure S5.1H), underscoring the importance of considering assisted migration prescriptions even within portions of species ranges that are not projected to become unsuitable habitat for the species as a whole.

5.6 References

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T. L., & Curtis-McLane, S. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**(1), 95–111.

Botkin, D. B., Saxe, H., Arajo, 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. 2007. Forecasting the effects of global warming on biodiversity. *Bioscience*, **57**(3), 227–236.

Burczyk, J., DiFazio, S. P., & Adams, W. T. 2004. Gene flow in forest trees: how far do genes really travel? *Forest Genetics*, **11**(3-4), 179–192.

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**(D20), 24509–24525.

Chen, P. Y., Welsh, C., & Hamann, A. 2010. Geographic variation in growth response of Douglas-fir to interannual climate variability and projected climate change. *Global Change Biology*, **16**(12), 3374–3385.

Clark, J. S., Fastie, C., Hurtt, G., Jackson, S. T., Johnson, C., King, G. A., Lewis, M., Lynch, J., Pacala, S., Prentice, C., Schupp, E. W., Webb, T., & Wyckoff, P. 1998. Reid's paradox of rapid plant migration - Dispersal theory and interpretation of paleoecological records. *Bioscience*, **48**(1), 13–24.

Crookston, N. L., & Finley, A. O. 2008. yaImpute: An R package for kNN imputation. *Journal of Statistical Software*, **23**(10).

Crowley, T. J. 1990. Are there any satisfactory geologic analogs for a future greenhouse warming? *Journal of Climate*, **3**(11), 1282–1292.

Davis, M. B. 1981. Quaternary history and the stability of forest communities. *Pages 132–153 of:* West, D. C., Shugart, H. H., & Botkin, D. B. (eds), *Forest Succession*. New York: Springer-Verlag.

Davis, M. B. 1989. Lags in Vegetation Response to Greenhouse Warming. *Climatic Change*, **15**(1-2), 75–82.

Davis, M. B., & Zabinski, C. 1992. Changes in geographical range resulting from greenhouse warming: effects on biodiversity in forests. *Pages 297–308 of:* Peters, R. L., & Lovejoy, T. E. (eds), *Global Warming and Biological Diversity*. New Haven, CT: Yale University Press.

Delcourt, P. A., & Delcourt, H. R. 1987. *Long-term forest dynamics of the temperate zone: A case study of Late-Quaternary forests in eastern North America*. New York: Springer-Verlag.

Elith, J., & Leathwick, J. R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution and Systematics*, **40**, 677–697.

Hennon, P. E., D'Amore, D. V., Zeglen, S., & Grainger, M. 2005. *Yellow-Cedar decline in the north coast forest district of British Columbia, Research Note PNW-RN-549.* Portland, Oregon: United States Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Hennon, P. E., D'Amore, D. V., Schaberg, P. G., Wittwer, D. T., & Shanley, C. S. 2012. Shifting Climate, Altered Niche, and a Dynamic Conservation Strategy for Yellow-Cedar in the North Pacific Coastal Rainforest. *Bioscience*, **62**(2), 147–158.

Huntley, B., & Birks, H. J. B. 1983. *An atlas of past and present pollen maps for Europe: 0-13000 years ago.* Cambridge: Cambridge University Press.

IPCC. 2007. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.*

Iverson, L. R., & Prasad, A. M. 2002. Potential redistribution of tree species habitat under five climate change scenarios in the eastern US. *Forest Ecology and Management*, **155**(1-3), 205–222.

King, G. A., & Herstrom, A. A. 1997. *Holocene tree migration rates objectively determined from follis pollen data.* New York: Springer-Verlag. Pages 91–101.

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**(6-7), 473–506.

Lawing, A. M., & Polly, P. D. 2011. Pleistocene Climate, Phylogeny, and Climate Envelope Models: An Integrative Approach to Better Understand Species' Response to Climate Change. *Plos One*, **6**(12).

Loarie, S. R., Duffy, P. B., Hamilton, H., Asner, G. P., Field, C. B., & Ackerly, D. D. 2009. The velocity of climate change. *Nature*, **462**(7276), 1052–U111.

Malcolm, J. R., Markham, A., Neilson, R. P., & Garaci, M. 2002. Estimated migration rates under scenarios of global climate change. *Journal of Biogeography*, **29**(7), 835–849.

McLachlan, J. S., Clark, J. S., & Manos, P. S. 2005. Molecular indicators of tree migration capacity under rapid climate change. *Ecology*, **86**(8), 2088–2098.

Meier, E. S., Lischke, H., Schmatz, D. R., & Zimmermann, N. E. 2012. Climate, competition and connectivity affect future migration and ranges of European trees. *Global Ecology and Biogeography*, **21**(2), 164–178.

Pearson, R. G. 2006. Climate change and the migration capacity of species. *Trends in Ecology & Evolution*, **21**(3), 111–113.

Petit, R. J., & Hampe, A. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution and Systematics*, **37**, 187–214.

R Core Team. 2009. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Savolainen, O., Pyhajarvi, T., & Knurr, T. 2007. Gene flow and local adaptation in trees. *Annual Review of Ecology Evolution and Systematics*, **38**, 595–619.

Schwartz, M. W. 1992. Modeling Effects of Habitat Fragmentation on the Ability of Trees to Respond to Climatic Warming. *Biodiversity and Conservation*, **2**(1), 51–61.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., de Siqueira, M. F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A. S., Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Peterson, A. T., Phillips, O. L., & Williams, S. E. 2004. Extinction risk from climate change. *Nature*, **427**(6970), 145–148.

Ukrainetz, N. K., O'Neill, G. A., & Jaquish, B. 2011. Comparison of fixed and focal point seed transfer systems for reforestation and assisted migration: a case study for interior spruce in British Columbia. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **41**(7), 1452–1464.

van Mantgem, P. J., Stephenson, N. L., Byrne, J. C., Daniels, L. D., Franklin, J. F., Fule, P. Z., Harmon, M. E., Larson, A. J., Smith, J. M., Taylor, A. H., & Veblen, T. T. 2009. Widespread Increase of Tree Mortality Rates in the Western United States. *Science*, **323**(5913), 521–524.

Wang, T. L., Hamann, A., Spittlehouse, D. L., & Murdock, T. Q. 2012. ClimateWNA-High-Resolution Spatial Climate Data for Western North America. *Journal of Applied Meteorology and Climatology*, **51**(1), 16–29.

Table 5.1: Loadings for the first 4 principal components (PC1 to PC4) for each climate variable, based on an analysis of the combined climate data for all time periods and all models. The proportion of the total variance explained by each component (PropVar) and the cumulative variance explained at each component (CumVar) are also noted. A list of climate variables is provided below. For a complete explanation of all climate variables, see Wang et al. (2012).

	Principal component			
Variable	PC1	PC2	PC3	PC4
Mean annual temperature	-0.388	-0.198	-	0.117
Mean temperature of the warmest month (MTWM)	-0.357	-	-0.427	0.24
Mean temperature of the coldest month (MTCM)	-0.37	-0.249	0.18	-
Difference between MWMT and MCMT	0.141	0.278	-0.807	0.14
Mean annual precipitation (log transformed)	0.112	-0.602	-	0.289
Mean summer precipitation (log transformed)	0.139	-0.509	-0.341	-0.629
Annual heat moisture index	0.321	-0.329	-	0.536
Summer heat moisture index	0.363	-0.23	-	-0.2
Degree days above 5°C	-0.389	-	-	-0.302
Number of frost-free days	-0.383	-0.159	-	-
Proportion of variance explained	*0.585*	*0.232*	*0.107*	*0.036*
Cumulative variance explained	*0.585*	*0.817*	*0.923*	*0.960*

Table 5.2: Calculated required migration rates for past, present, and future. Median migration rates (with 90th percentiles in brackets) of climate habitat for each species (km yr^{-1}) are given for the periods of fastest climate change in the past (from 16,000 to 14,000 years ago in the CCM1 reconstruction and from 16,000 to 9,000 years ago in the GFDL reconstructions), the observed present (from weather station data from the 1961-1990 normals to the 1997-2006 period), and for the future (from the 1961-1990 normals to the 2080s period) as projected in an ensemble GCM projection for the B1 and A2 scenarios. Present and future rates represent the rates required for properly-adapted genotypes to reach suitable habitat.

		Past				Present		Future 2080s			
		GFDL		CCM1		97-06		B1		A2	
Species name	Common name	Med	p90	Med	p90	Med	p90	Med	p90	Med	p90
Mesic: coastal only											
1. Abies amabilis	Pacific silver fir	0.13	0.21	0.08	0.18	0.63	3.46	0.40	1.74	0.63	2.47
2. Abies procera	noble fir	0.09	0.21	0.04	0.14	0.34	3.49	0.38	2.00	0.78	3.93
3. Acer macrophyllum	bigleaf maple	0.07	0.15	0.02	0.13	0.44	2.33	0.40	1.42	0.66	2.08
4. Alnus rubra	red alder	0.12	0.21	0.03	0.14	0.46	2.35	0.41	1.30	0.63	1.78
5. Calocedrus decurrens	incense cedar	0.05	0.11	0.02	0.08	0.31	1.85	0.34	1.42	0.54	1.88
6. Chamaecyparis nootkatensis	yellow cedar	0.19	0.35	0.08	0.17	0.69	3.11	0.44	1.53	0.66	2.15
7. Picea sitchensis	Sitka spruce	0.15	0.23	0.06	0.19	0.63	3.10	0.46	1.61	0.75	2.63
8. Sequoia sempervirens	coast redwood	0.03	0.09	0.01	0.04	0.22	1.54	0.11	0.71	0.23	1.10
Group average		*0.1*	*0.2*	*0.04*	*0.13*	*0.47*	*2.65*	*0.37*	*1.47*	*0.61*	*2.25*
Mesic: coastal & interior											
9. Pinus monticola	western white pine	0.14	0.24	0.05	0.22	0.83	5.45	0.68	2.65	1.13	4.65
10. Pseudotsuga menziesii	Douglas-fir	0.12	0.24	0.08	0.25	1.38	7.26	0.97	3.10	1.48	5.15
11. Thuja plicata	western redcedar	0.12	0.22	0.11	0.25	0.94	5.08	0.65	2.13	1.06	3.29
12. Tsuga heterophylla	western hemlock	0.16	0.27	0.10	0.22	0.90	4.59	0.63	2.11	1.01	3.19
13. Tsuga mertensiana	mountain hemlock	0.16	0.27	0.07	0.21	0.83	4.32	0.56	2.28	0.91	3.31

Continued on next page

Table 5.2 – continued from previous page

		Past				Present		Future 2080s			
		GFDL		CCM1		97-06		B1		A2	
Species name	Common name	Med	p90	Med	p90	Med	p90	Med	p90	Med	p90
Group average		*0.14*	*0.25*	*0.08*	*0.23*	*0.98*	*5.34*	*0.70*	*2.45*	*1.12*	*3.92*
Boreal & sub-boreal											
14. Betula papyrifera	paper birch	0.22	0.60	0.20	0.30	4.47	13.94	2.88	6.62	4.95	12.33
15. Picea glauca	White spruce	0.23	0.60	0.20	0.29	4.35	14.08	3.01	6.69	5.04	10.89
16. Picea mariana	black spruce	0.25	0.60	0.20	0.29	4.47	14.47	3.02	6.50	5.03	10.80
17. Pinus contorta	lodgepole pine	0.15	0.3	0.13	0.26	2.19	9.43	1.84	6.42	3.61	9.49
18. Populus tremuloides	trembling aspen	0.24	0.53	0.20	0.29	4.41	13.85	2.99	6.94	5.12	12.41
Group average		*0.22*	*0.53*	*0.19*	*0.29*	*3.98*	*13.15*	*2.75*	*6.63*	*4.75*	*11.18*
Sub-alpine											
19. Abies lasiocarpa	subalpine fir	0.15	0.28	0.15	0.27	2.09	8.47	1.60	4.96	2.82	9.07
20. Picea engelmannii	Engelmann spruce	0.22	0.33	0.13	0.31	2.01	8.50	1.38	4.66	2.56	8.42
21. Pinus albicaulis	whitebark pine	0.13	0.26	0.14	0.31	2.01	9.56	1.33	5.30	2.42	8.87
Group average		*0.17*	*0.29*	*0.14*	*0.3*	*2.04*	*8.84*	*1.44*	*4.97*	*2.6*	*8.79*
Xeric sub-xeric											
22. Larix occidentalis	western larch	0.14	0.27	0.11	0.28	1.72	8.19	1.21	3.70	2.14	5.54
23. Pinus edulis	pinyon pine	0.03	0.18	0.09	0.33	2.02	8.31	1.35	3.12	1.80	3.75
24. Pinus ponderosa	ponderosa pine	0.10	0.24	0.10	0.35	1.75	8.5	1.37	3.97	2.28	6.44
Group average		*0.09*	*0.24*	*0.10*	*0.30*	*1.81*	*7.90*	*1.30*	*3.51*	*2.03*	*5.28*
Average for all species		*0.14*	*0.29*	*0.10*	*0.23*	*1.67*	*6.88*	*1.18*	*3.45*	*2.01*	*5.65*

Table 5.3: Required rates of gene flow and migration (in km yr^{-1}) for habitats to be filled in the 2080s with suitably-adapted populations, listed as median (Med) and 90^{th} percentile (p90) rates. The percent gene flow is noted (% GF).

		2080s B1					2080s A2				
			Gene flow		Migration			Gene flow		Migration	
Species name	Common name	%GF	Med	p90	Med	p90	%GF	Med	p90	Med	p90
Mesic: coastal only											
1. Abies amabilis	Pacific silver fir	79.1	0.34	2.48	1.20	5.19	70.5	0.59	4.15	2.33	13.19
2. Abies procera	noble fir	49.6	0.23	2.06	1.39	5.53	36.3	0.44	3.69	2.51	17.34
3. Acer macrophyllum	big leaf maple	73.3	0.37	2.73	0.77	4.29	60.3	0.73	4.13	1.61	9.61
4. Alnus rubra	red alder	81.6	0.46	3.27	0.63	5.27	68.5	0.95	4.78	1.57	11.83
5. Calocedrus decurrens	incense cedar	66.0	0.31	1.86	0.86	3.40	56.1	0.55	3.44	1.64	4.63
6. Chamaecyparis nootkatensis	yellow cedar	82.4	0.41	2.62	1.24	5.43	77.8	0.66	4.25	2.07	11.71
7. Picea sitchensis	Sitka spruce	77.1	0.49	2.91	1.23	5.04	67.2	0.98	5.38	1.94	7.66
8. Sequoia sempervirens	coast redwood	44.7	0.07	0.34	0.41	2.70	28.3	0.15	0.51	1.87	6.83
Group average		69.2	0.34	2.28	0.97	4.61	58.1	0.63	3.79	1.94	10.35
Mesic: coastal & interior											
9. Pinus monticola	western white pine	65.2	0.54	3.33	1.47	4.52	56.8	0.95	4.34	2.48	6.19
10. Pseudotsuga menziesii	Douglas-fir	69.4	0.90	3.28	1.65	5.38	51.9	1.44	4.53	2.80	14.86
11. Thuja plicata	western redcedar	77.1	0.73	3.27	1.31	8.30	62.2	1.46	4.72	2.44	15.60
12. Tsuga heterophylla	western hemlock	82.0	0.66	3.16	1.35	4.60	71.9	1.21	4.58	2.11	6.43
13. Tsuga mertensiana	mountain hemlock	79.0	0.49	3.06	1.39	4.47	70.8	0.88	4.27	2.00	8.15
Group average		74.5	0.66	3.22	1.43	5.45	62.7	1.19	4.49	2.37	10.25

Continued on next page

Table 5.3 – continued from previous page

		2080s B1					2080s A2				
			Gene flow		Migration			Gene flow		Migration	
Species name	Common name	%GF	Med	p90	Med	p90	%GF	Med	p90	Med	p90
Boreal & sub-boreal											
14. Betula papyrifera	paper birch	68.0	3.13	8.70	2.85	5.77	47.1	4.05	10.39	4.09	8.07
15. Picea glauca	White spruce	68.6	2.92	7.78	3.02	5.76	49.5	3.92	8.50	4.33	7.80
16. Picea mariana	black spruce	67.0	3.06	7.83	3.00	5.66	47.4	3.91	8.32	4.31	7.69
17. Pinus contorta	lodgepole pine	46.6	1.44	3.90	1.81	8.75	24.4	2.50	5.46	3.69	17.35
18. Populus tremuloides	trembling aspen	76.5	3.55	8.78	3.03	7.71	66.9	5.96	11.18	4.48	9.54
Group average		65.3	2.82	7.40	2.74	6.73	47.1	4.07	8.77	4.18	10.09
Sub-alpine											
19. Abies lasiocarpa	subalpine fir	59.3	1.30	3.64	1.64	5.37	33.8	1.88	4.66	2.51	10.26
20. Picea engelmannii	Engelmann spruce	58.1	1.05	3.58	1.81	5.35	38.5	1.73	4.88	2.78	12.25
21. Pinus albicaulis	whitebark pine	39.1	0.97	4.18	1.82	5.16	21.4	1.51	5.21	2.31	6.33
Group average		52.2	1.11	3.80	1.76	5.29	31.2	1.71	4.92	2.53	9.61
Xeric & sub-xeric											
22. Larix occidentalis	western larch	43.0	0.75	2.62	1.59	4.81	27.1	1.17	4.14	2.76	14.67
23. Pinus edulis	pinyon pine	47.5	1.26	3.60	1.99	5.69	25.5	1.71	4.89	3.82	8.75
24. Pinus ponderosa	ponderosa pine	52.7	0.95	3.34	1.76	4.73	34.5	1.24	4.28	2.83	10.04
Group average		47.7	0.99	3.19	1.78	5.08	29.0	1.37	4.44	3.14	11.15
Average for all species		64.7	1.10	3.80	1.63	5.40	49.8	1.69	5.20	2.72	10.30

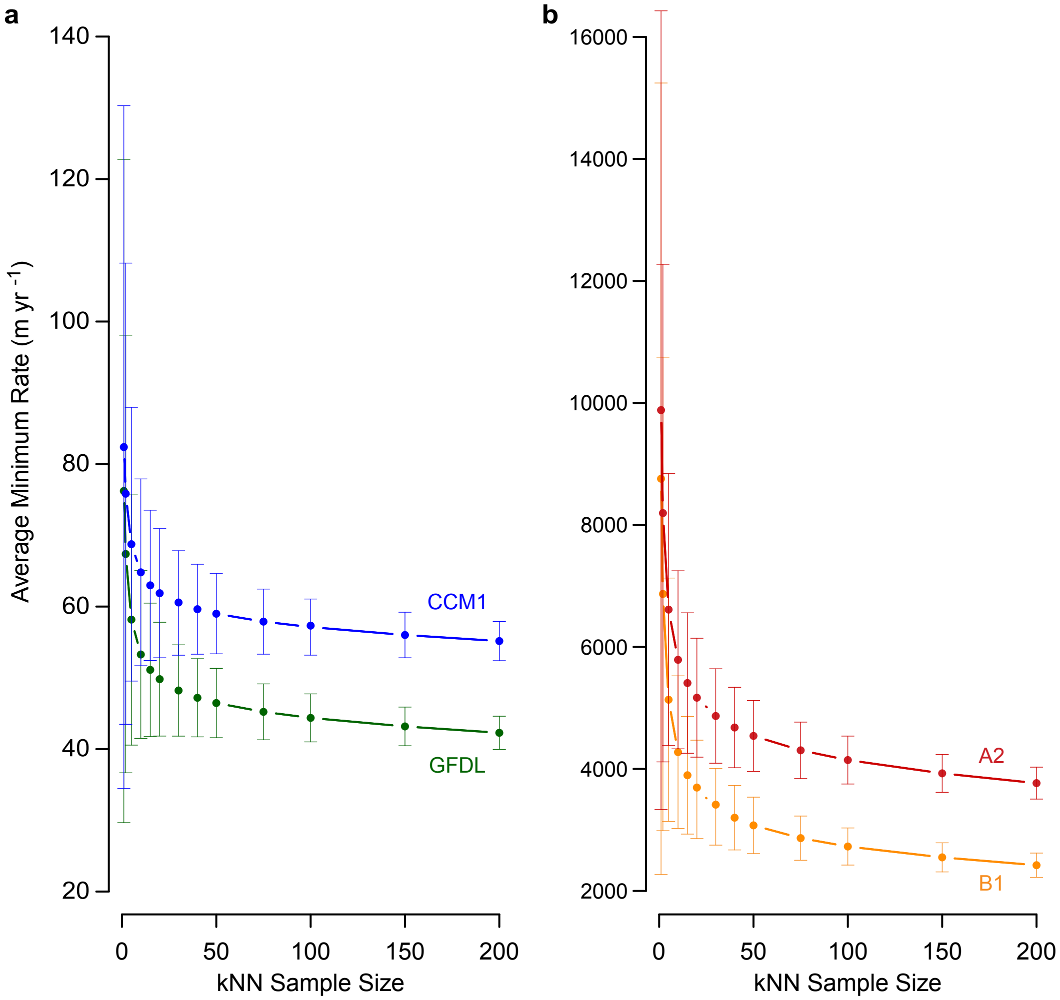

Figure 5.1: Minimum climate migration rates, averaged across the study area, for **(a)** calculations between the 1961-90 period and the 21,000 years ago period for the CCM1 and GFDL climate reconstructions, and **(b)** calculations between the 1961-90 period and the 2080s future period for the A2 and B1 climate projections, plotted as a function of the nearest neighbour sample size used in the *yaImpute* routine in R. Bars represent standard error.

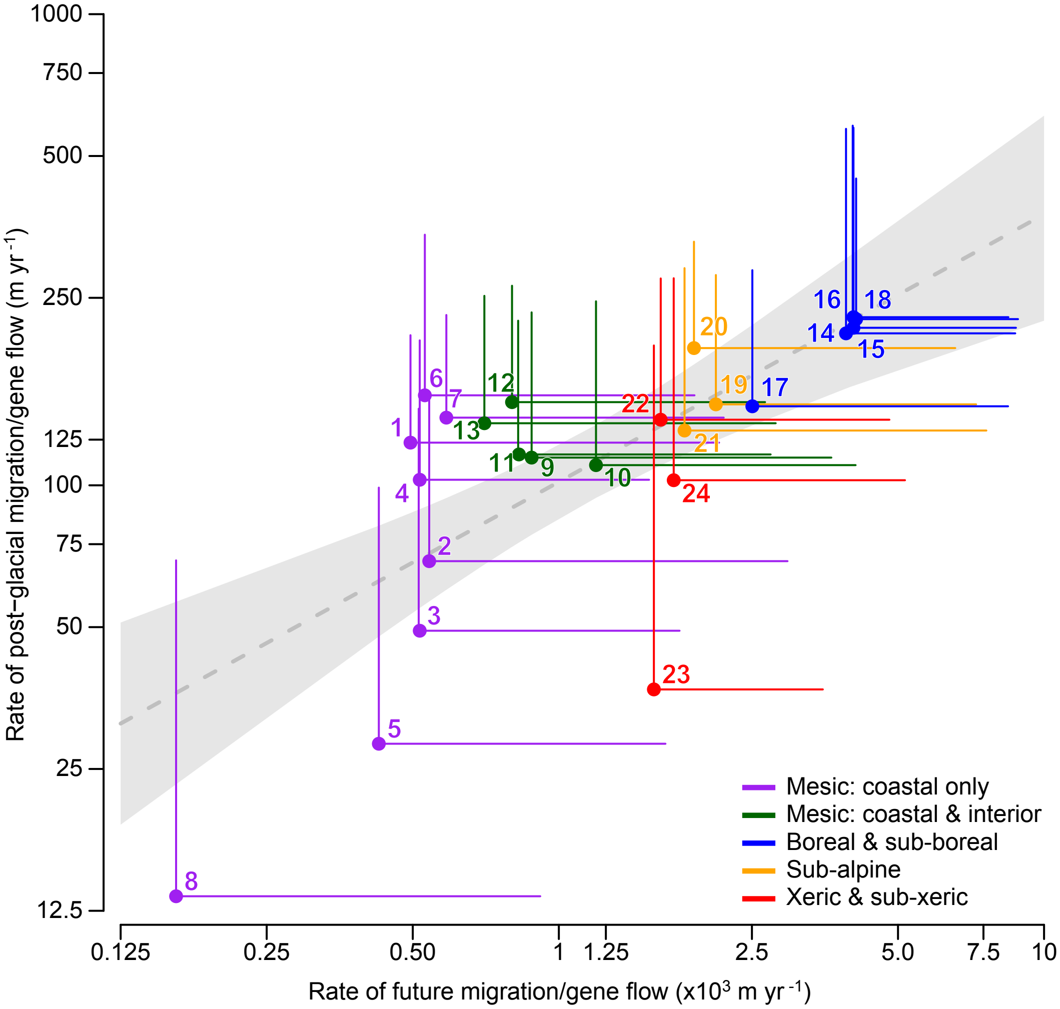

Figure 5.2: Regression plot showing species' median post-glacial migration rates relative to the projected median rate required for modern populations to find suitable habitat in the 2080s. Post-glacial rates are an average median rate for the CCM1 and GFDL reconstructions. The grey line shows the least-squares regression of the data ($r^2 = 0.59$, $p < 0.001$) and the grey shaded area represents the 90% confidence interval of the regression. Future rates are based on an average of median rate for both the A2 and B1 projection. Lines extend to the 90^{th} percentile for both post-glacial and future rates, and species are coloured based on their biogeographic grouping. Species numbers, biogeographic groups, and median/90^{th} percentile rates for individual projections are listed in Table RATES.

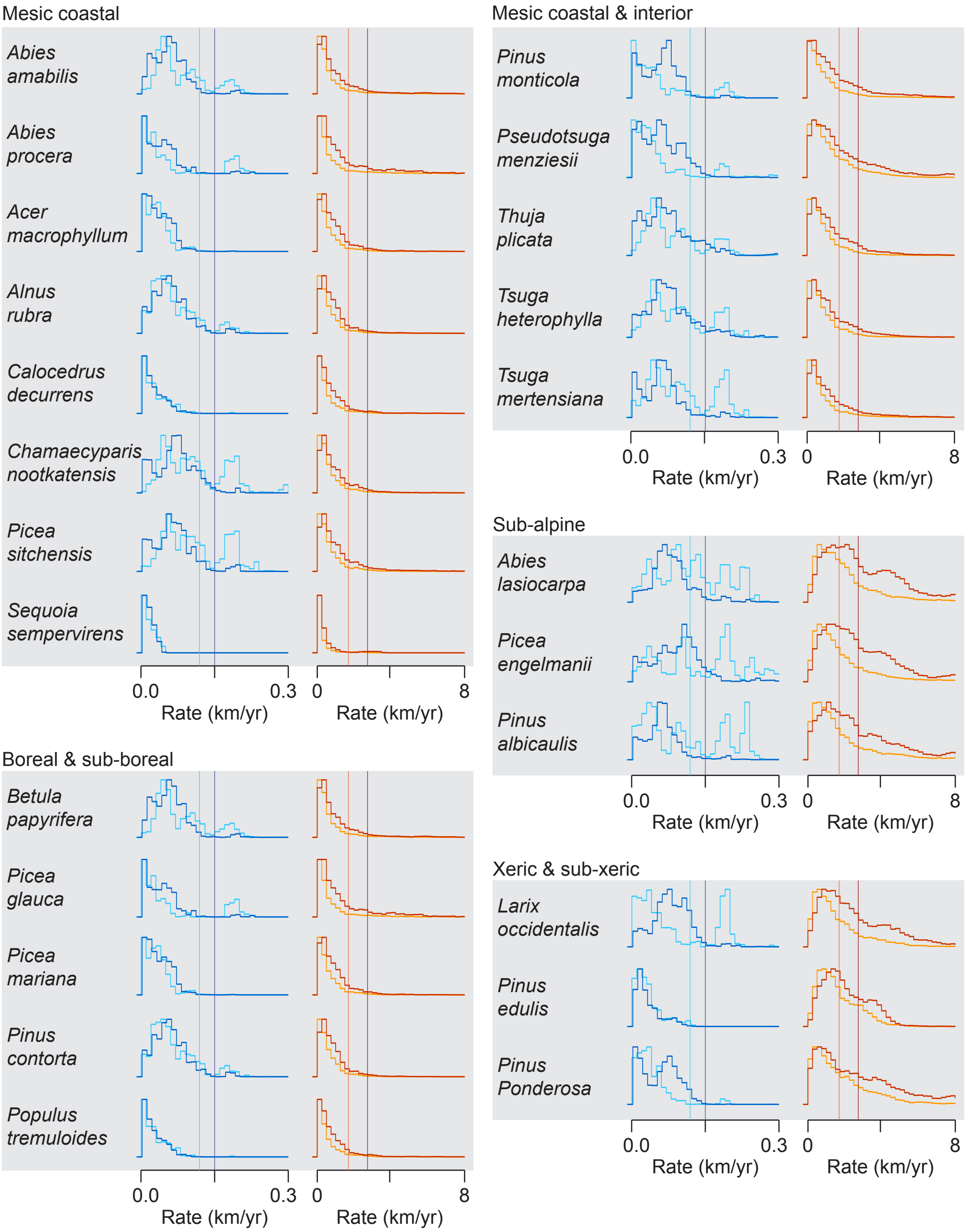

Figure 5.3: Distributions of post-glacial migration/gene flow rates for the two paleoclimate reconstructions (CCM1 in light blue and GFDL in dark blue) are compared to future rates required for present day populations to reach suitable climate habitat by the 2080s, based on two emissions scenarios (the optimistic B1 in orange and pessimistic A2 in red). To facilitate comparison between species, vertical lines are shown representing the median rate for each respective model, averaged across all species.

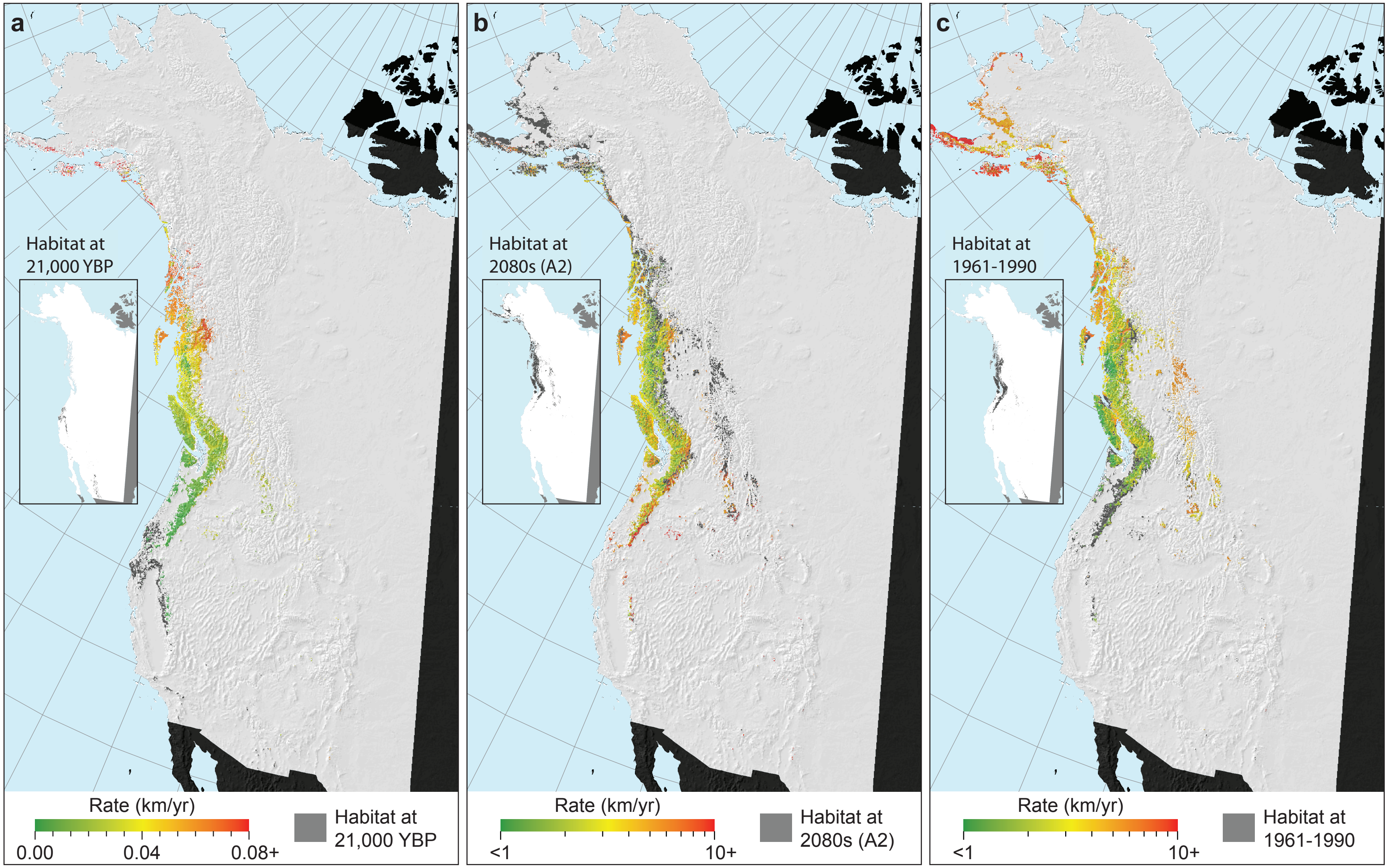

Figure S5.1A: ***Abies amabilis*** **(Pacific silver fir)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

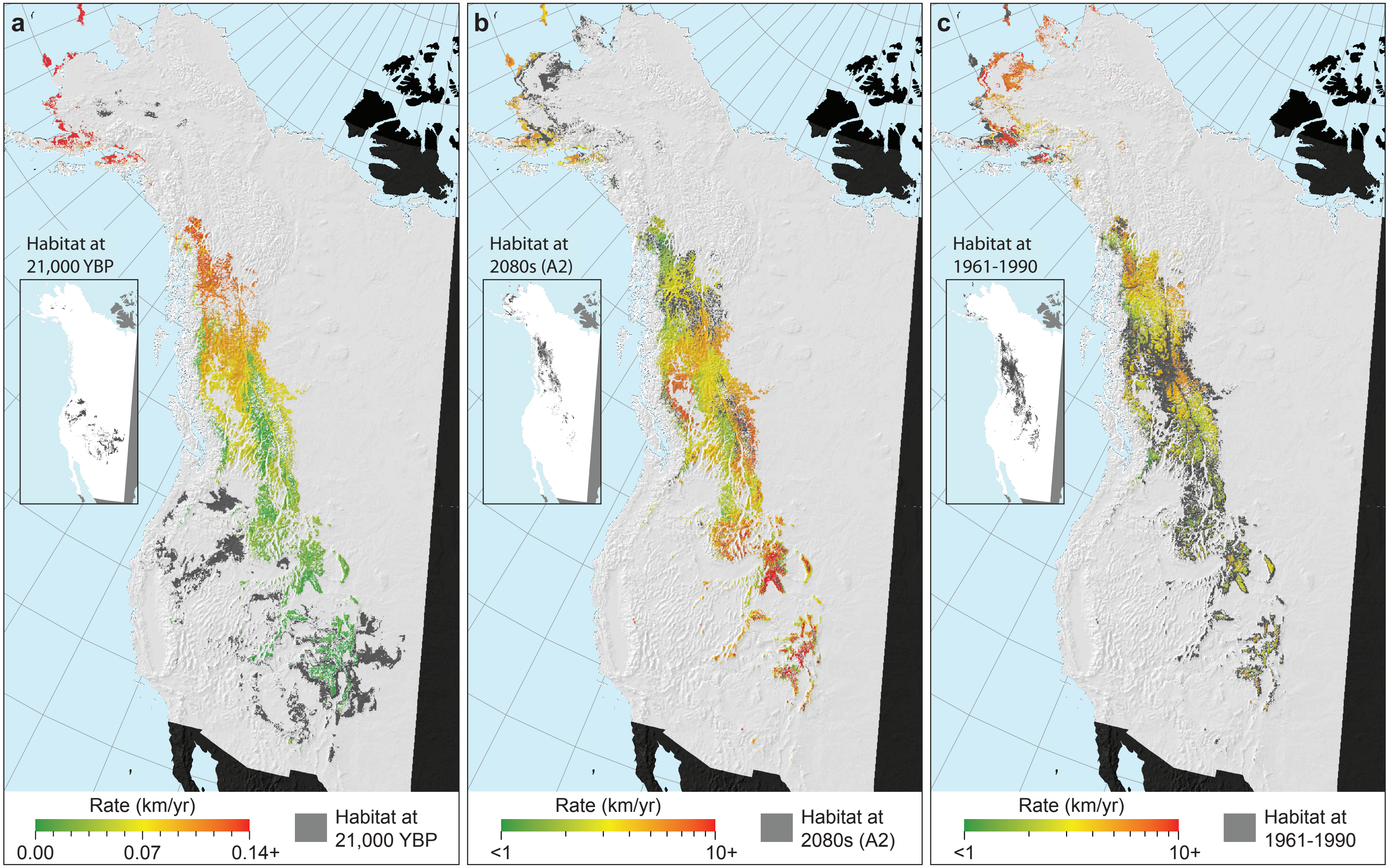

Figure S5.1B: ***Abies lasiocarpa*** **(subalpine fir)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

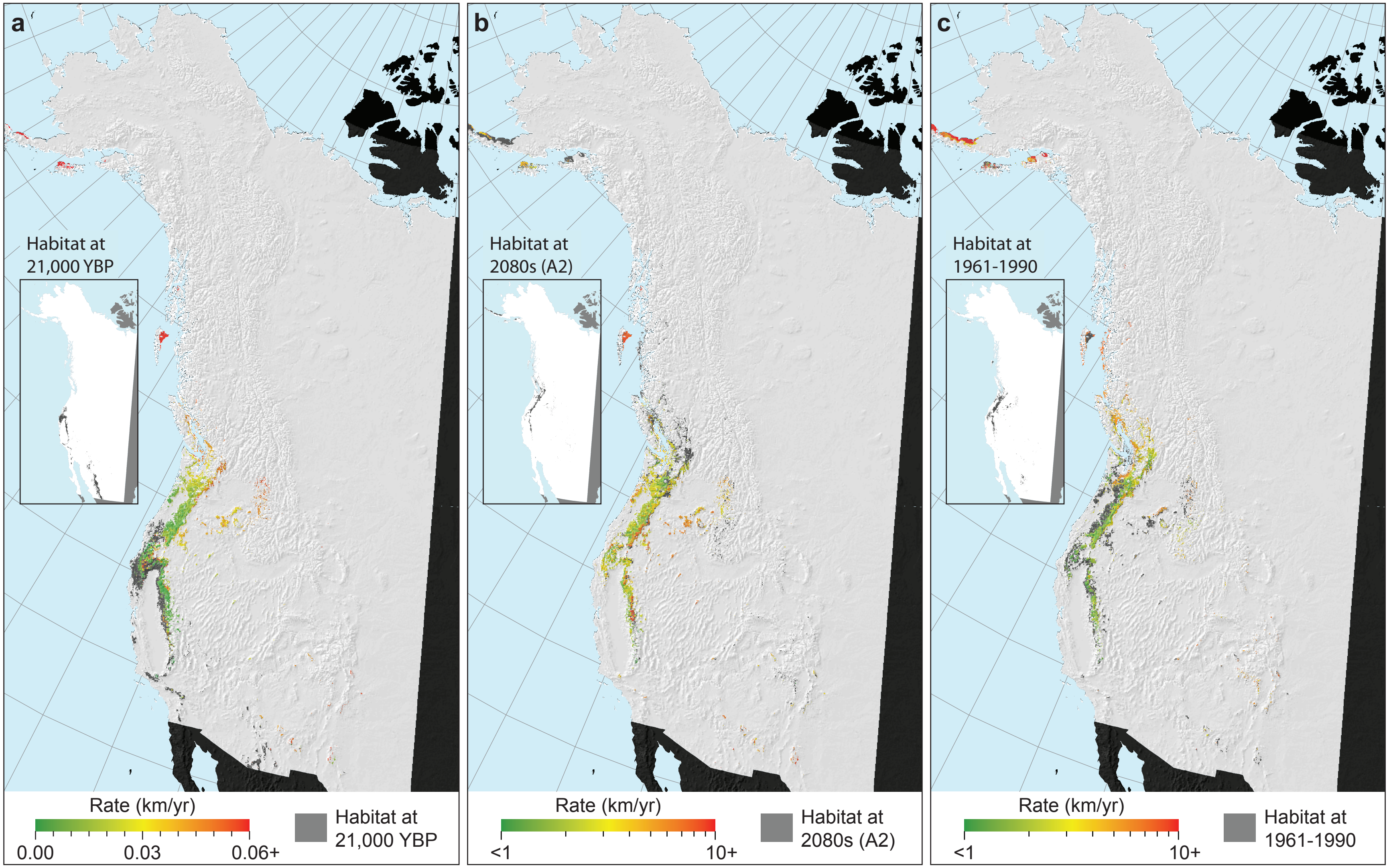

Figure S5.1C: ***Abies procera*** **(noble fir)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

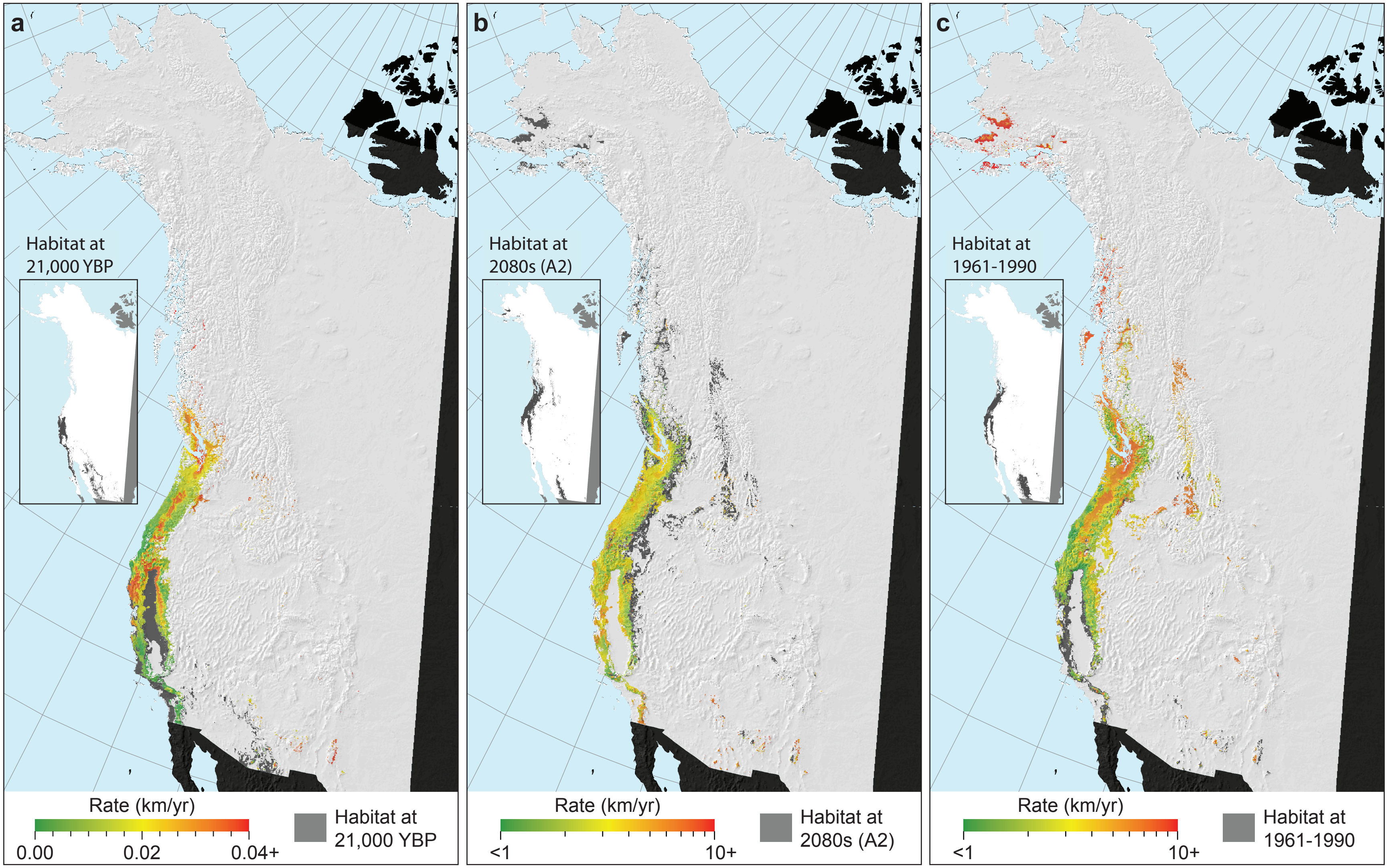

Figure S5.1D: ***Acer macrophyllum*** **(bigleaf maple)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

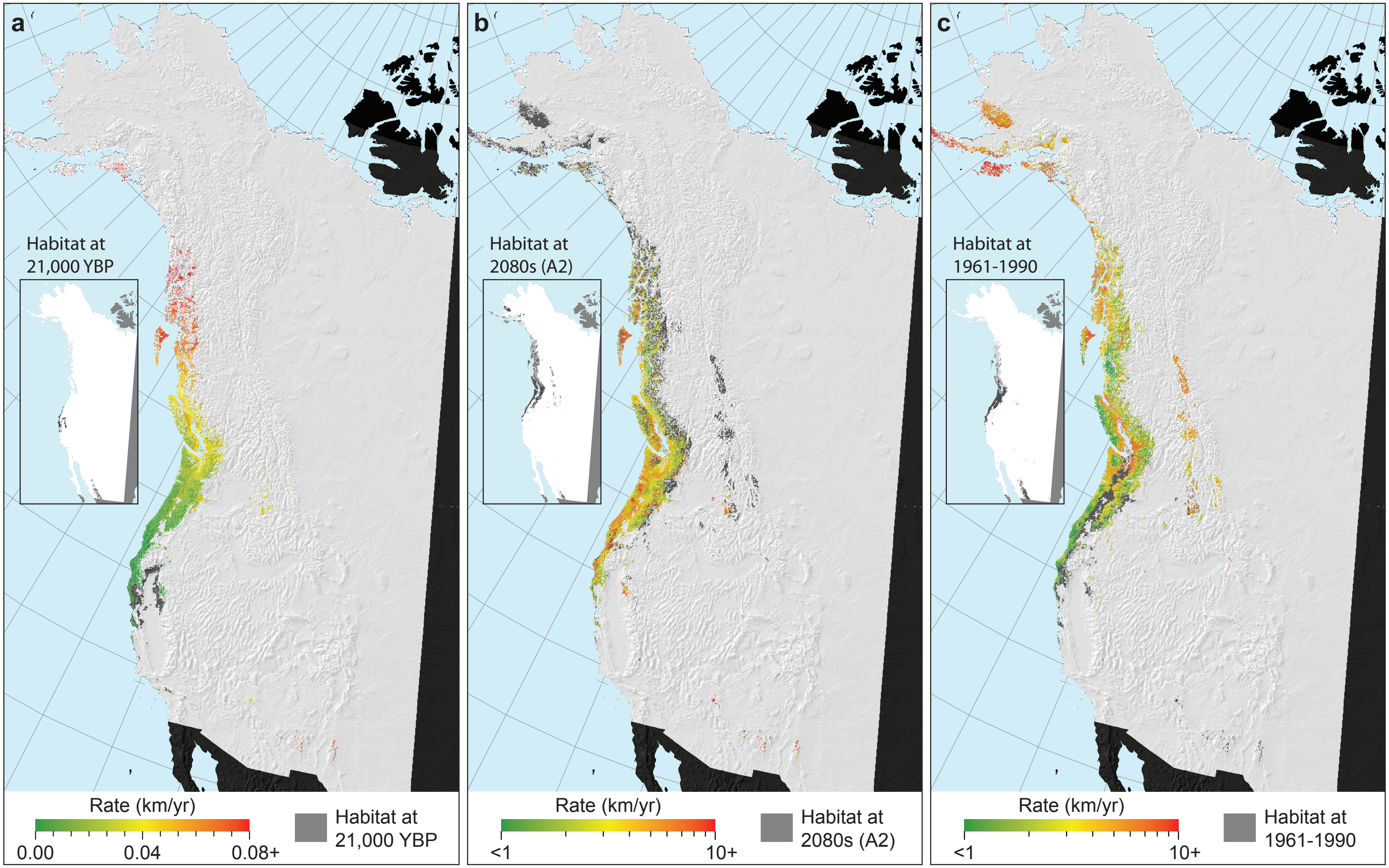

Figure S5.1E: ***Alnus rubra*** **(red alder)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

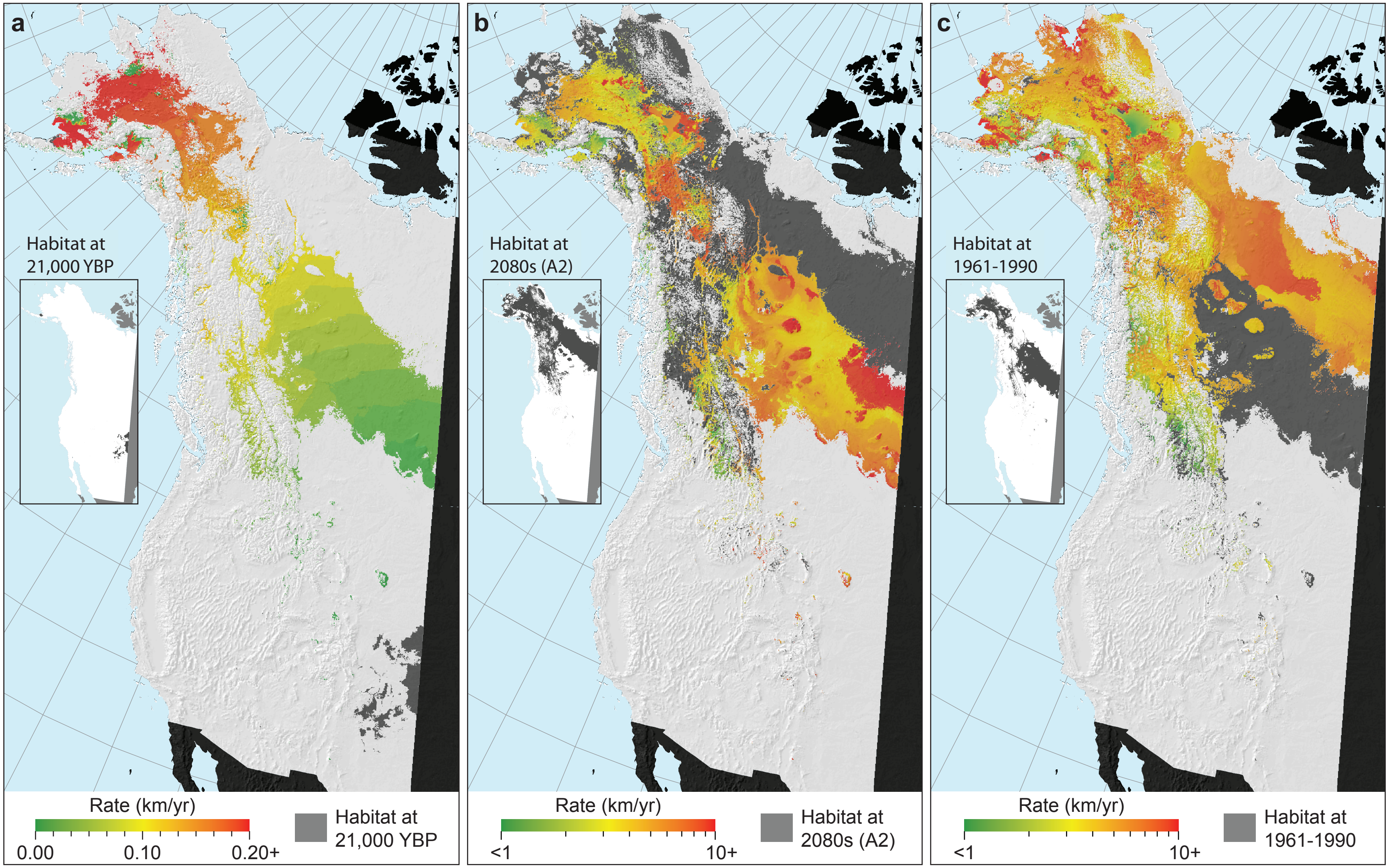

Figure S5.1F: ***Betula papyrifera*** **(paper birch)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

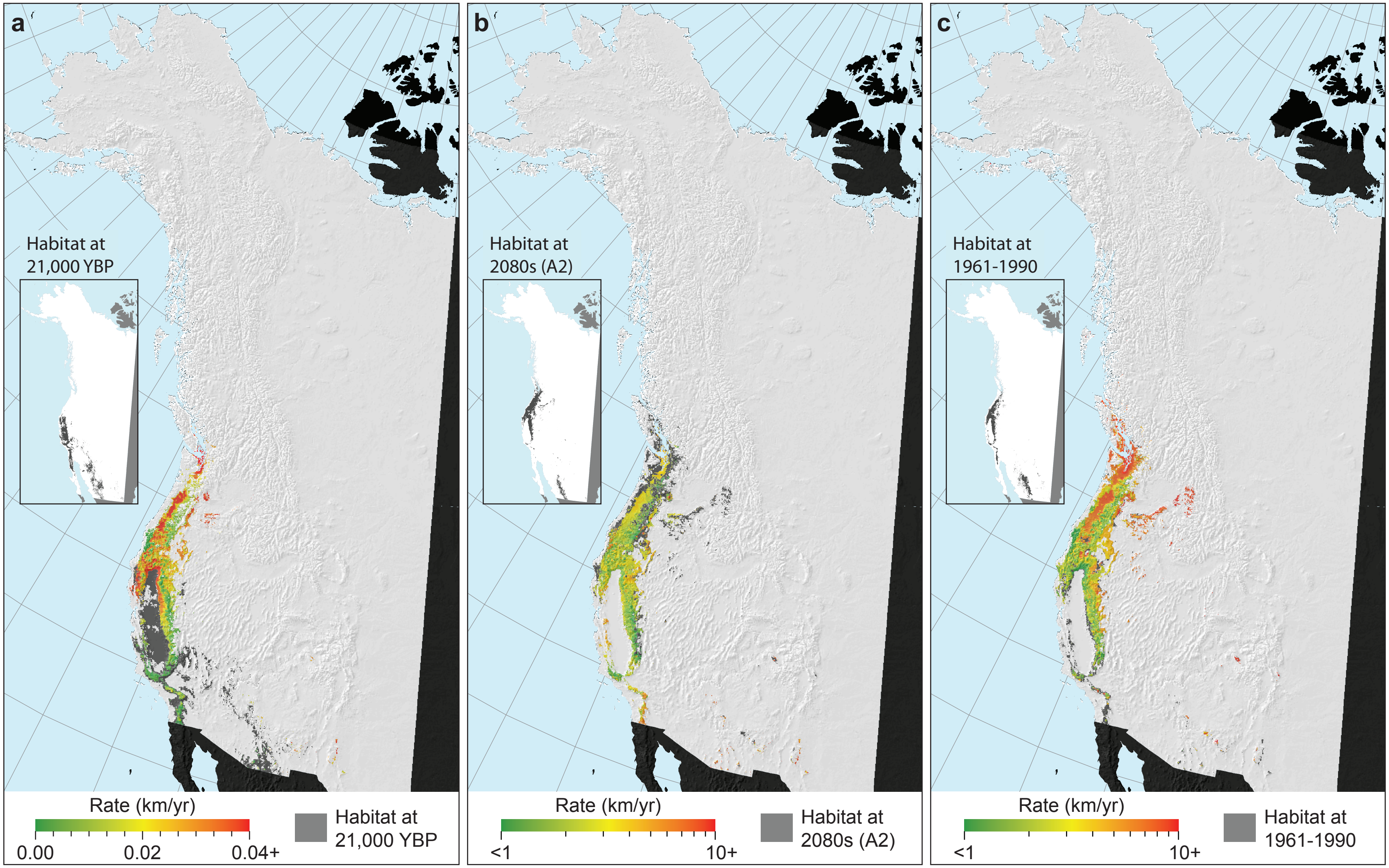

Figure S5.1G: *Calocedrus decurrens* (incense cedar) Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

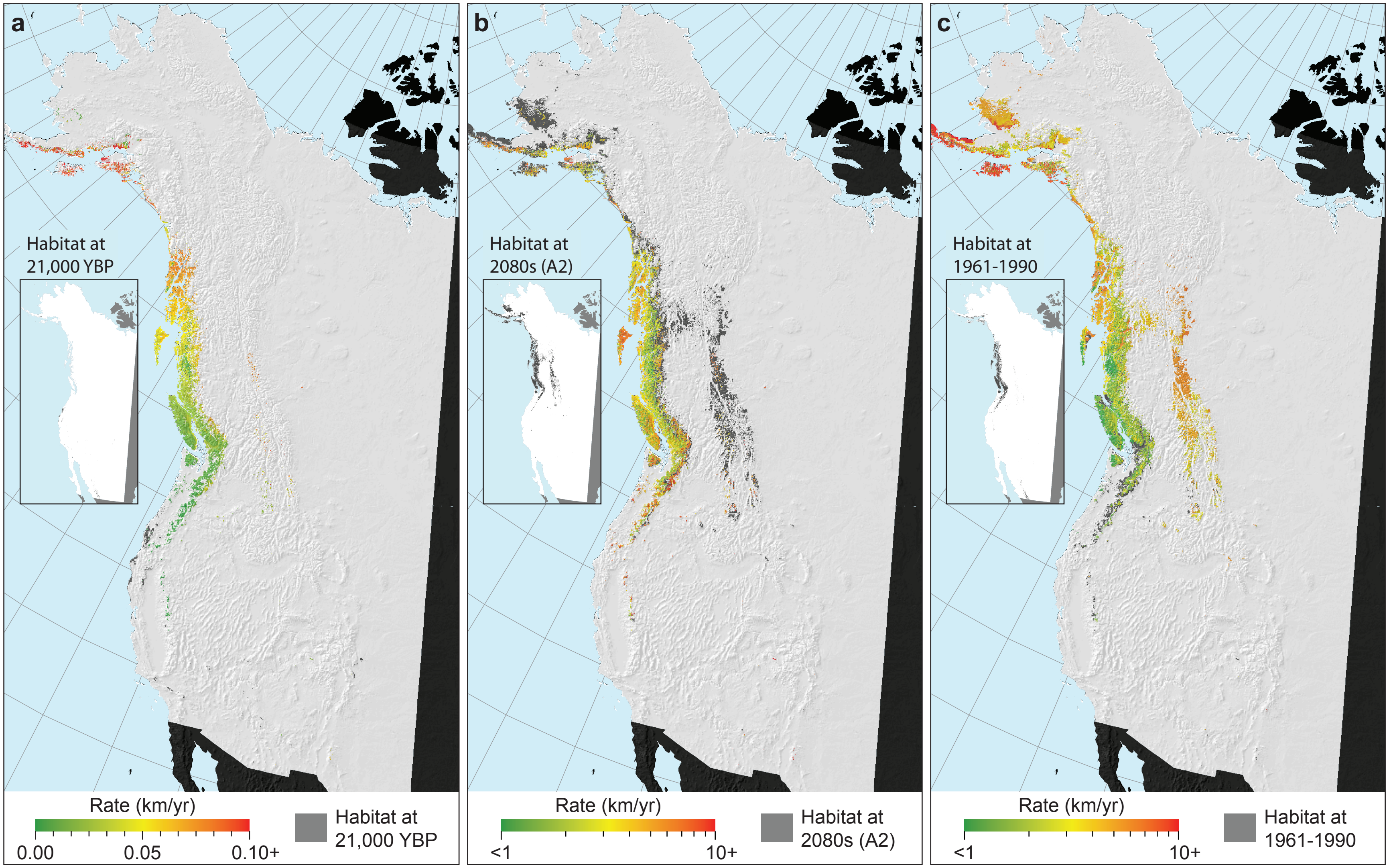

Figure S5.1H: ***Chamaecyparis nootkatensis*** **(yellow cedar)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

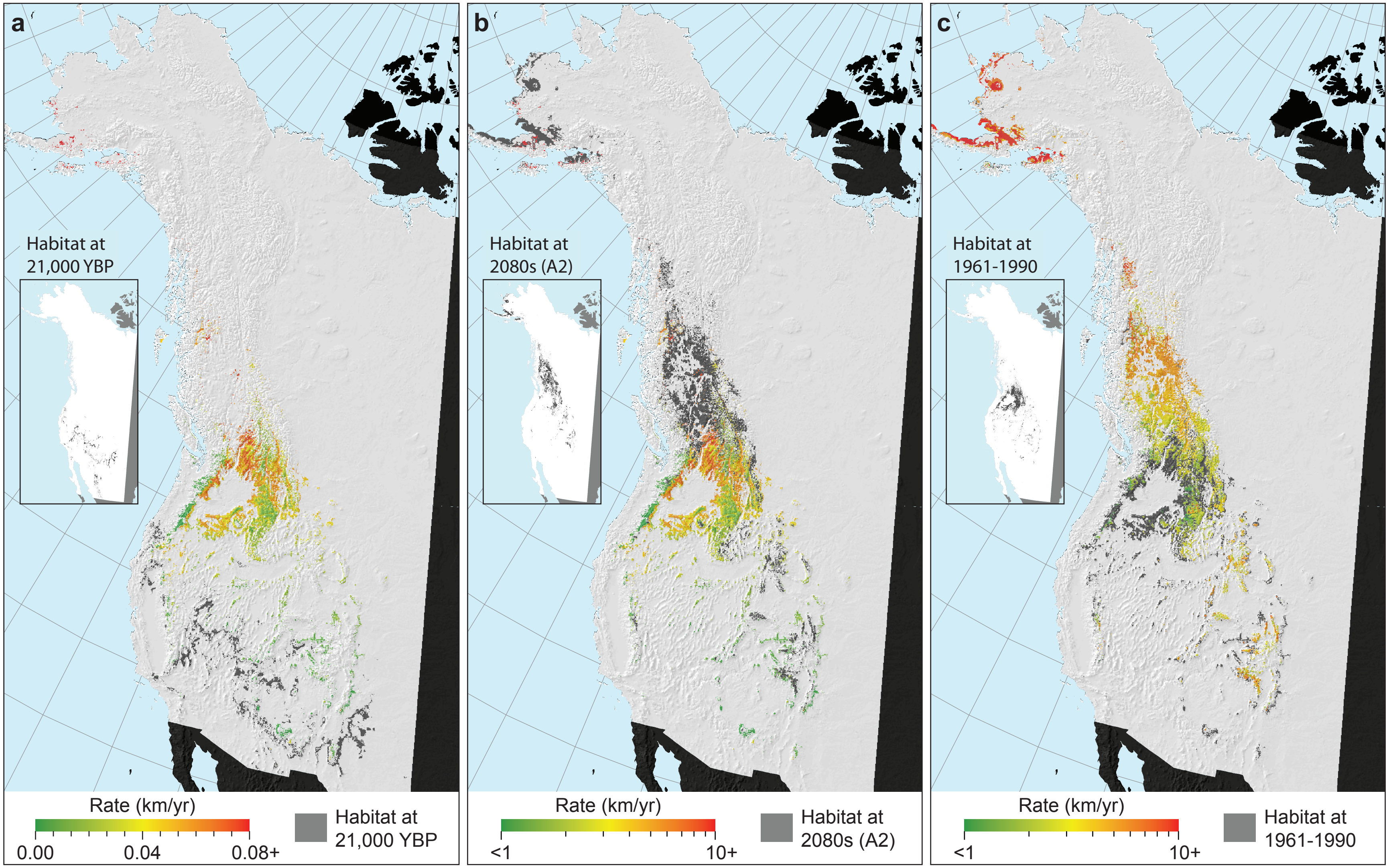

Figure S5.1I: ***Larix occidentalis*** **(western larch)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

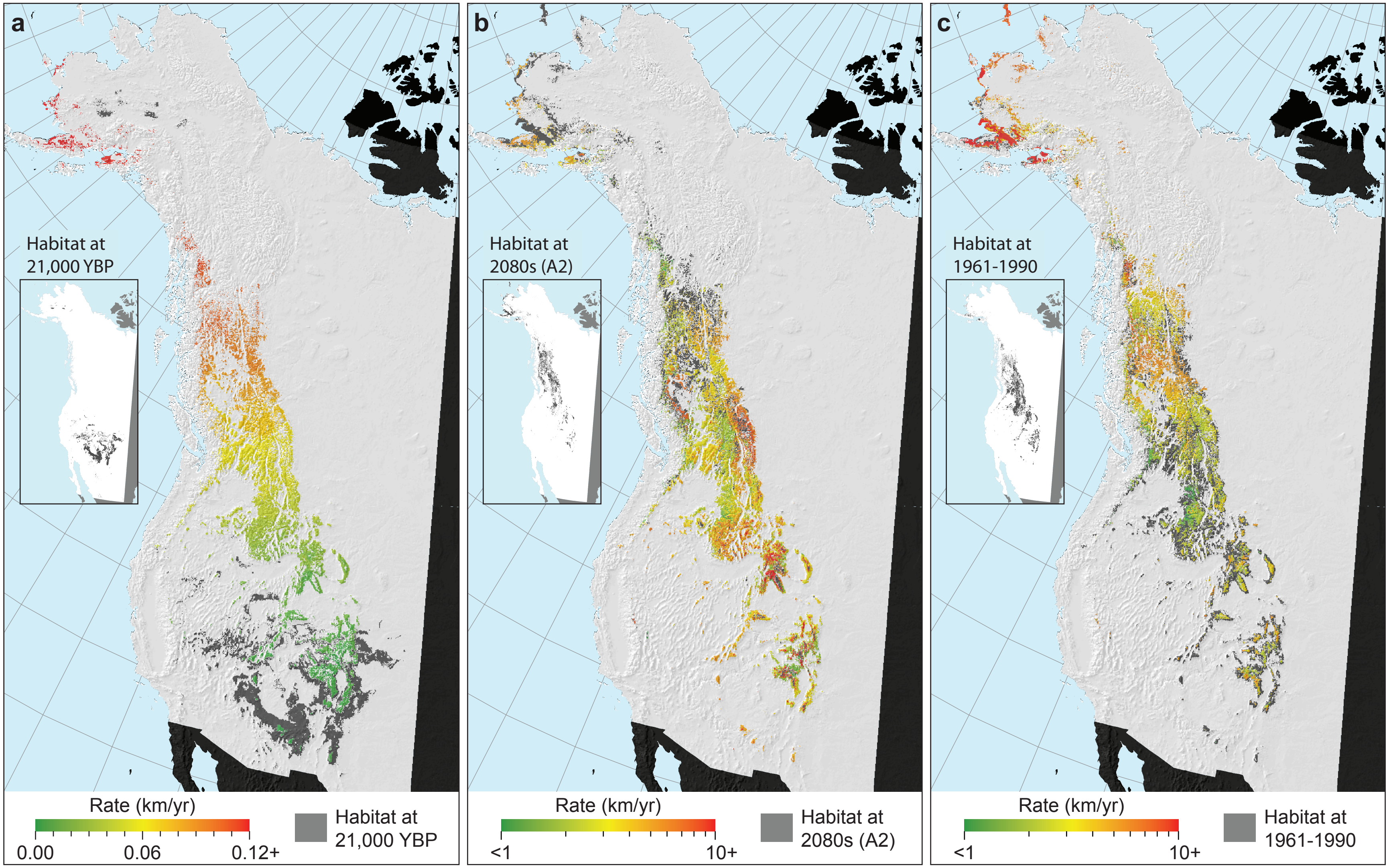

Figure S5.1J: ***Picea engelmannii*** **(Engelmann spruce)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

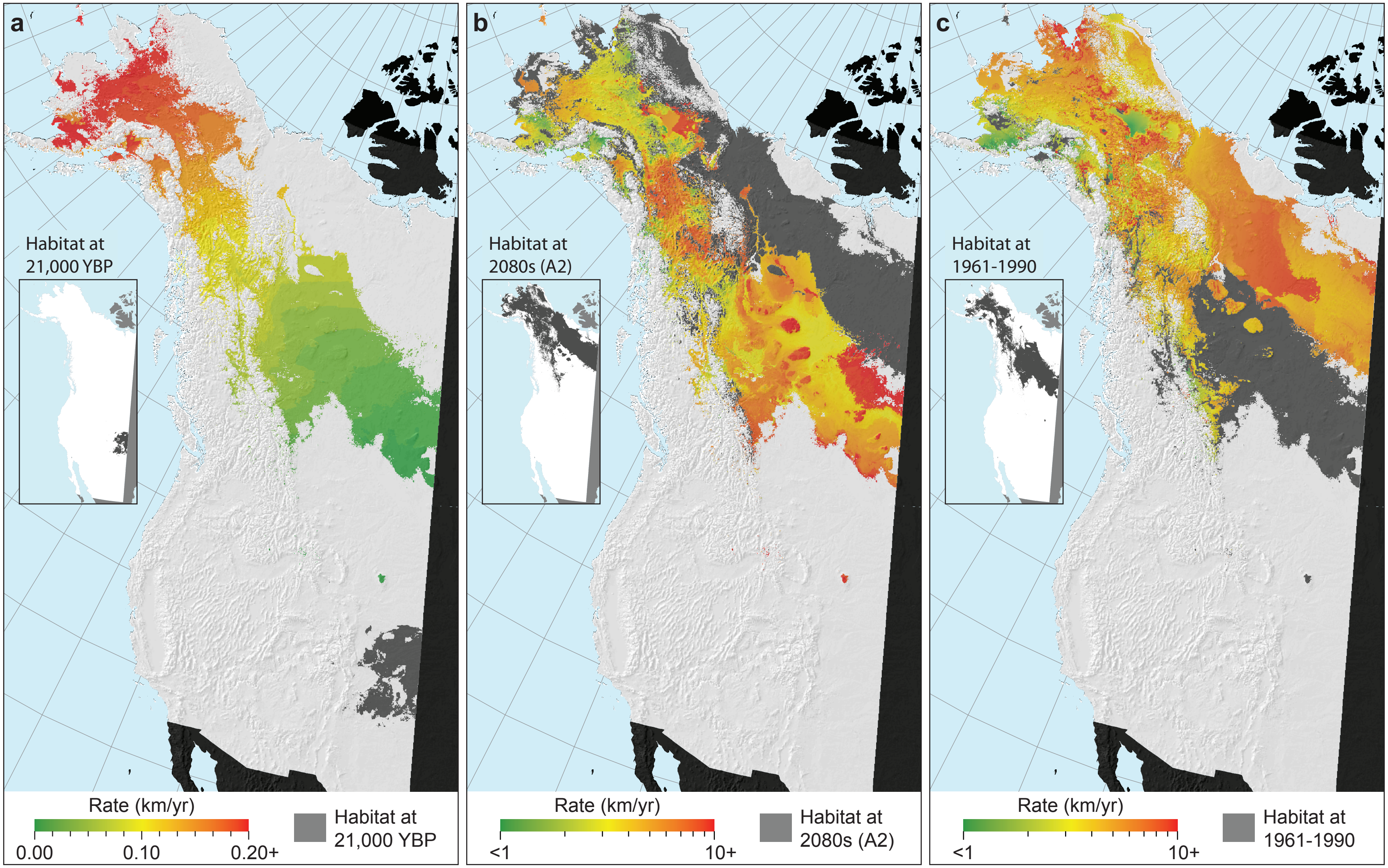

Figure S5.1K: *Picea glauca* (white spruce) Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

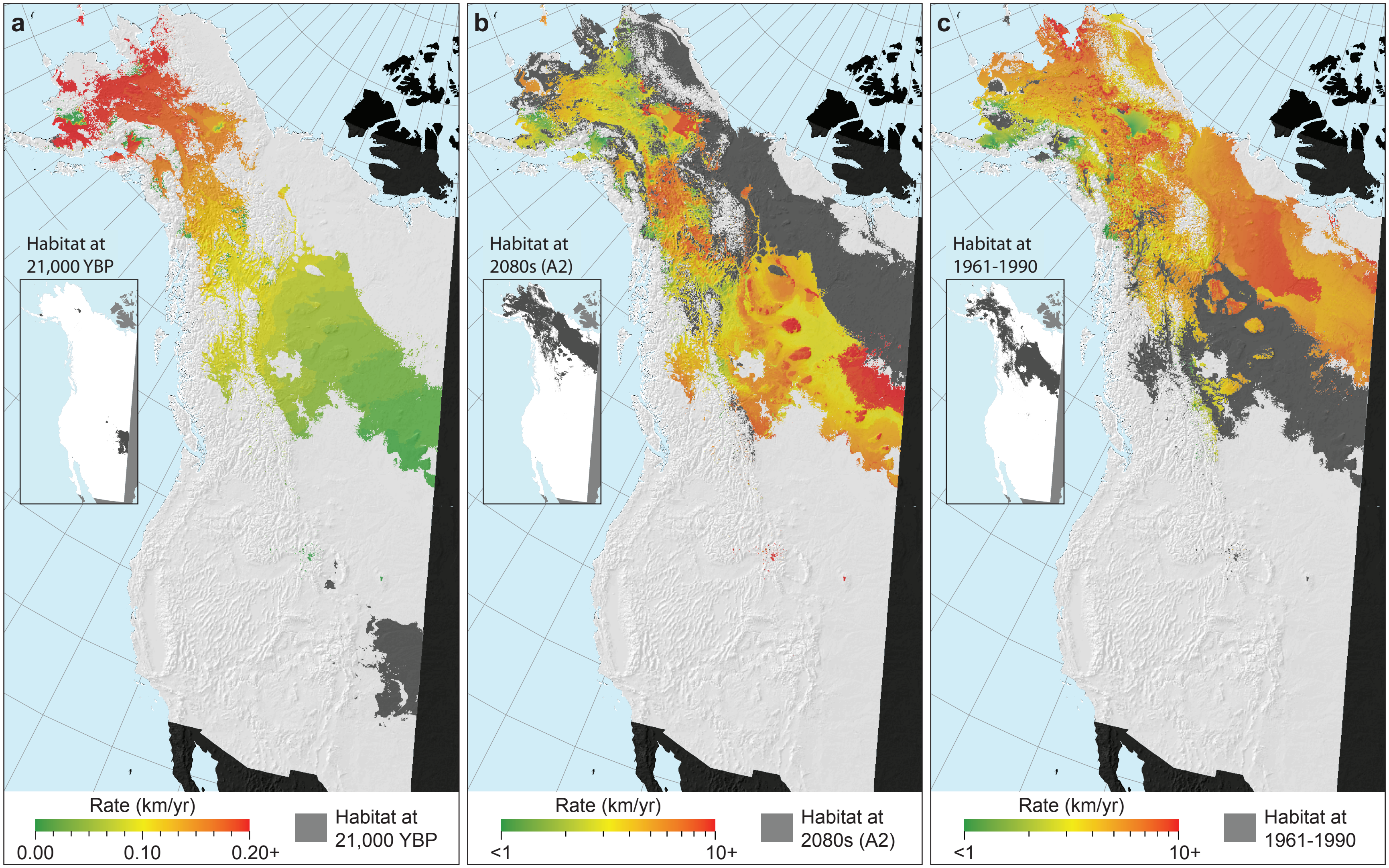

Figure S5.1L: ***Picea mariana*** **(black spruce)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

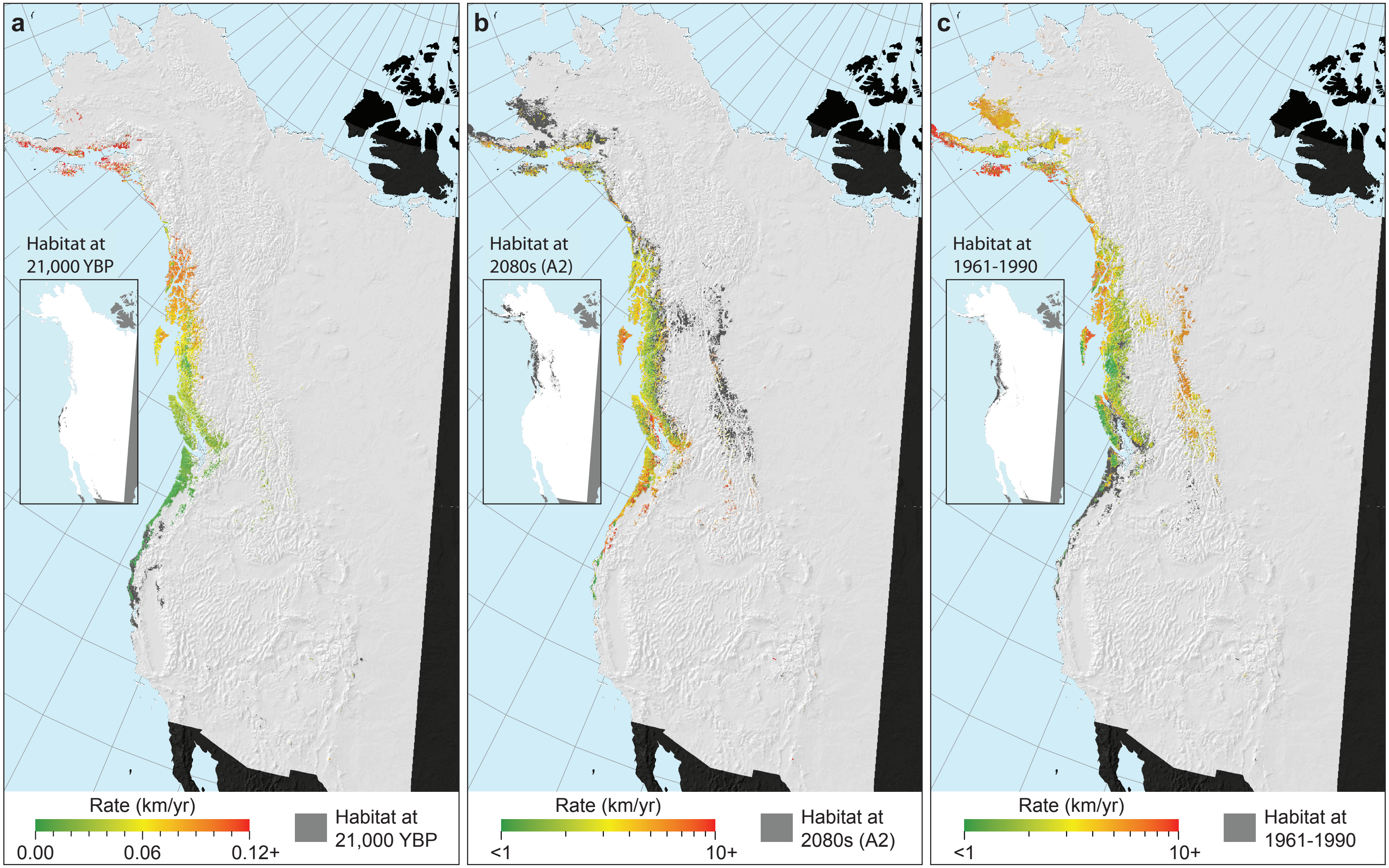

Figure S5.1M: ***Picea sitchensis*** **(sitka spruce)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

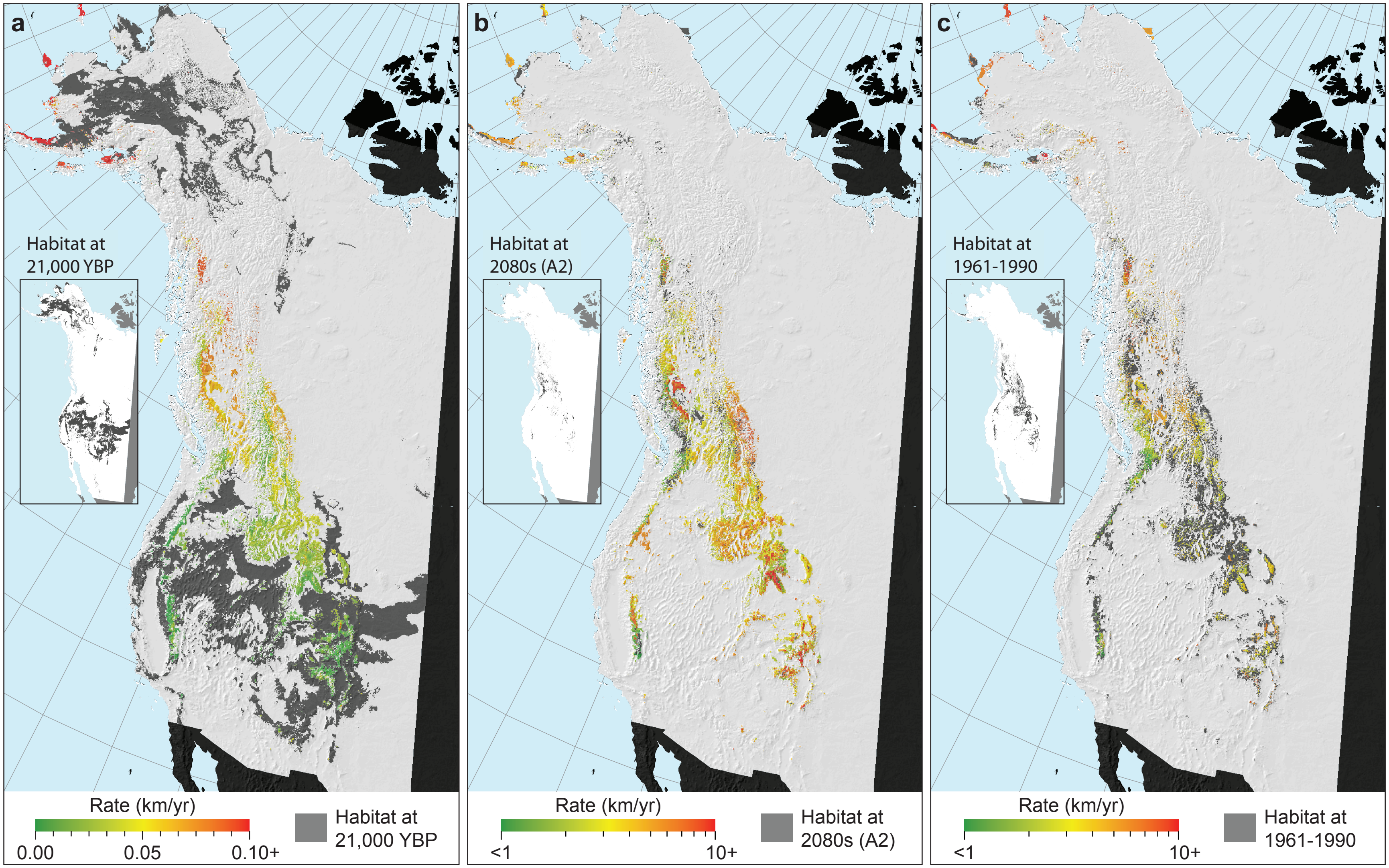

Figure S5.1N: ***Pinus albicaulis*** **(whitebark pine)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

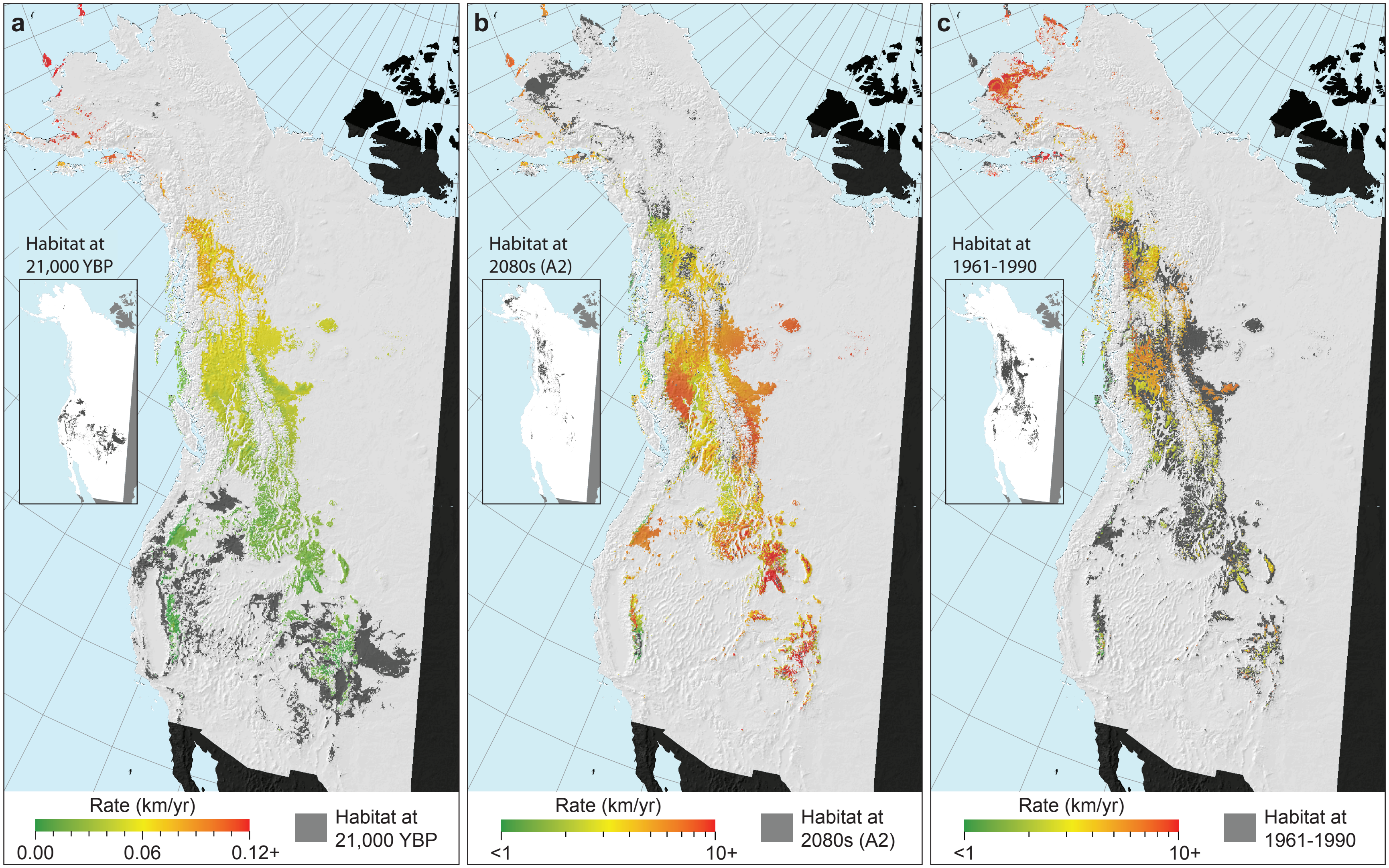

Figure S5.1O: ***Pinus contorta*** **(lodgepole pine)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

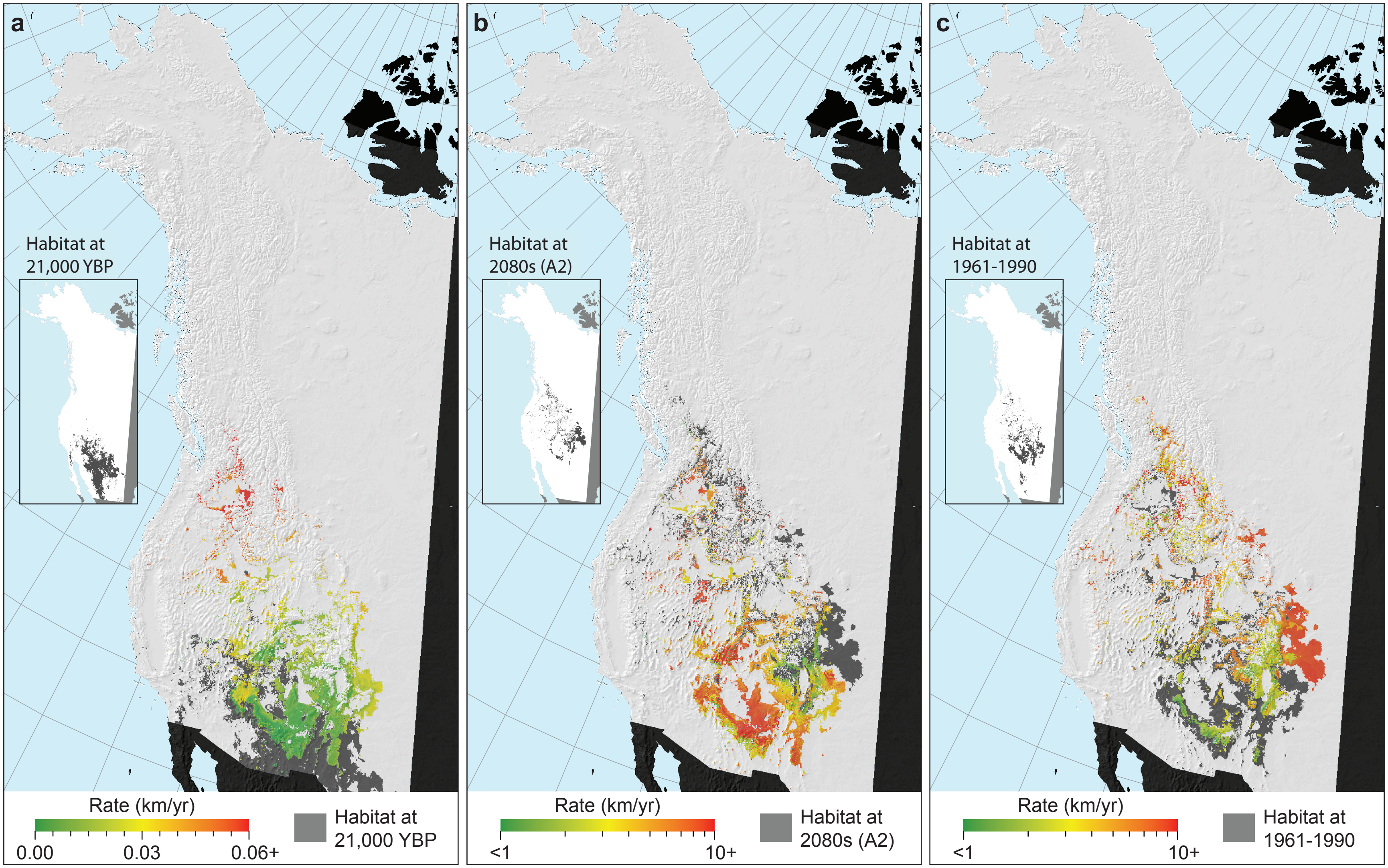

Figure S5.1P: ***Pinus edulis*** **(pinyon pine)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

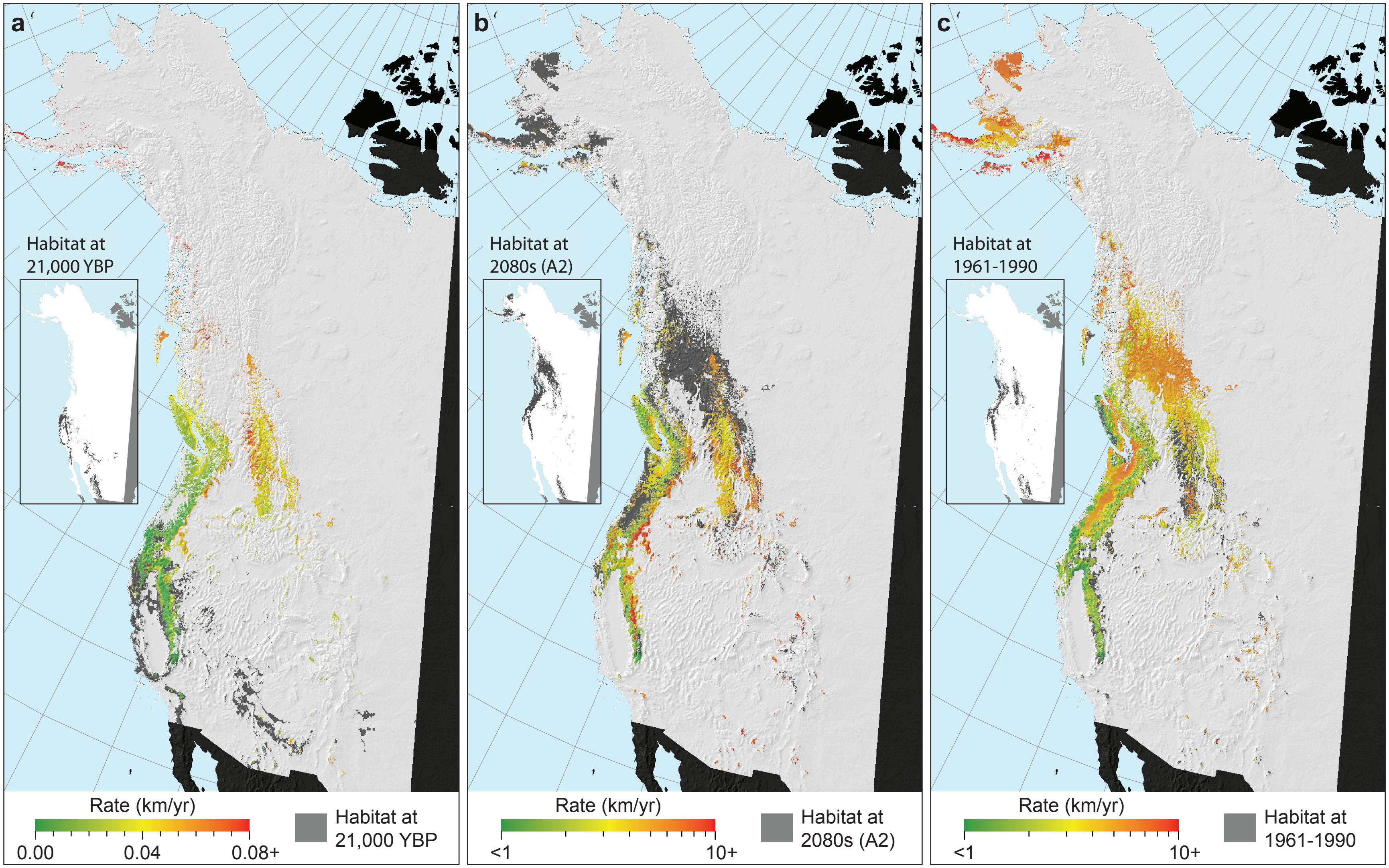

Figure S5.1Q: ***Pinus monticola*** **(western white pine)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

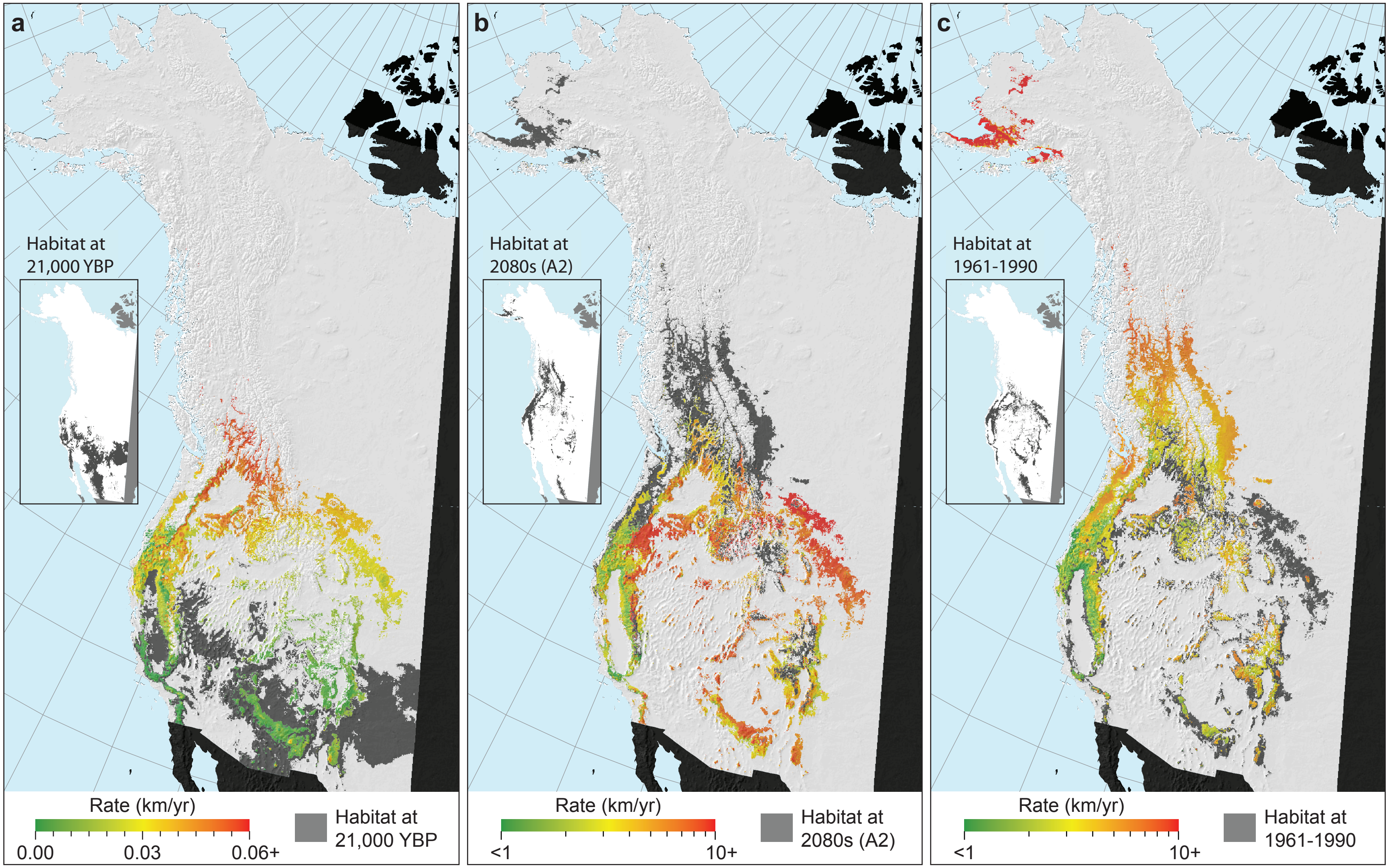

Figure S5.1R: ***Pinus ponderosa*** **(ponderosa pine)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

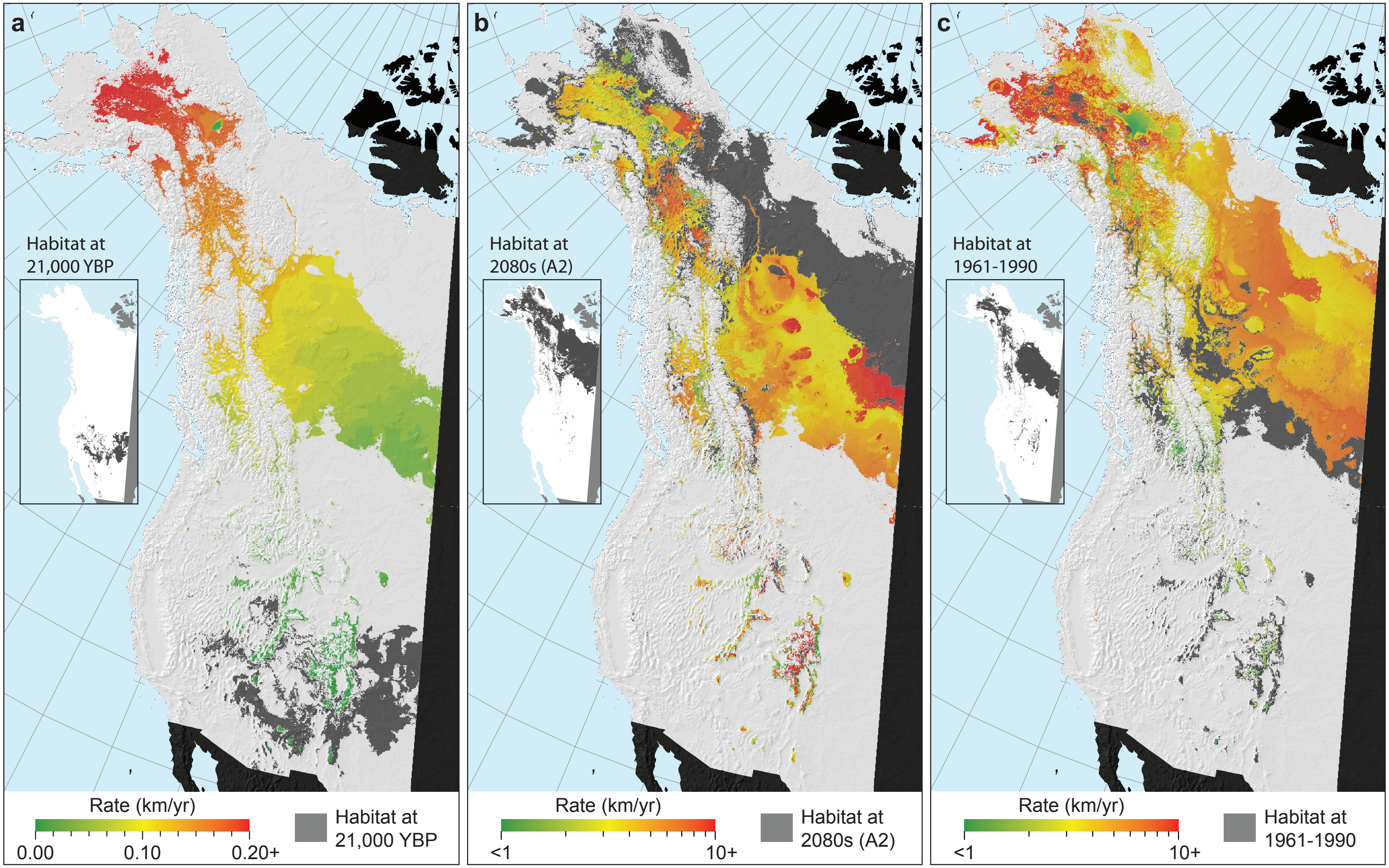

Figure S5.1S: *Populus tremuloides* (trembling aspen) Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

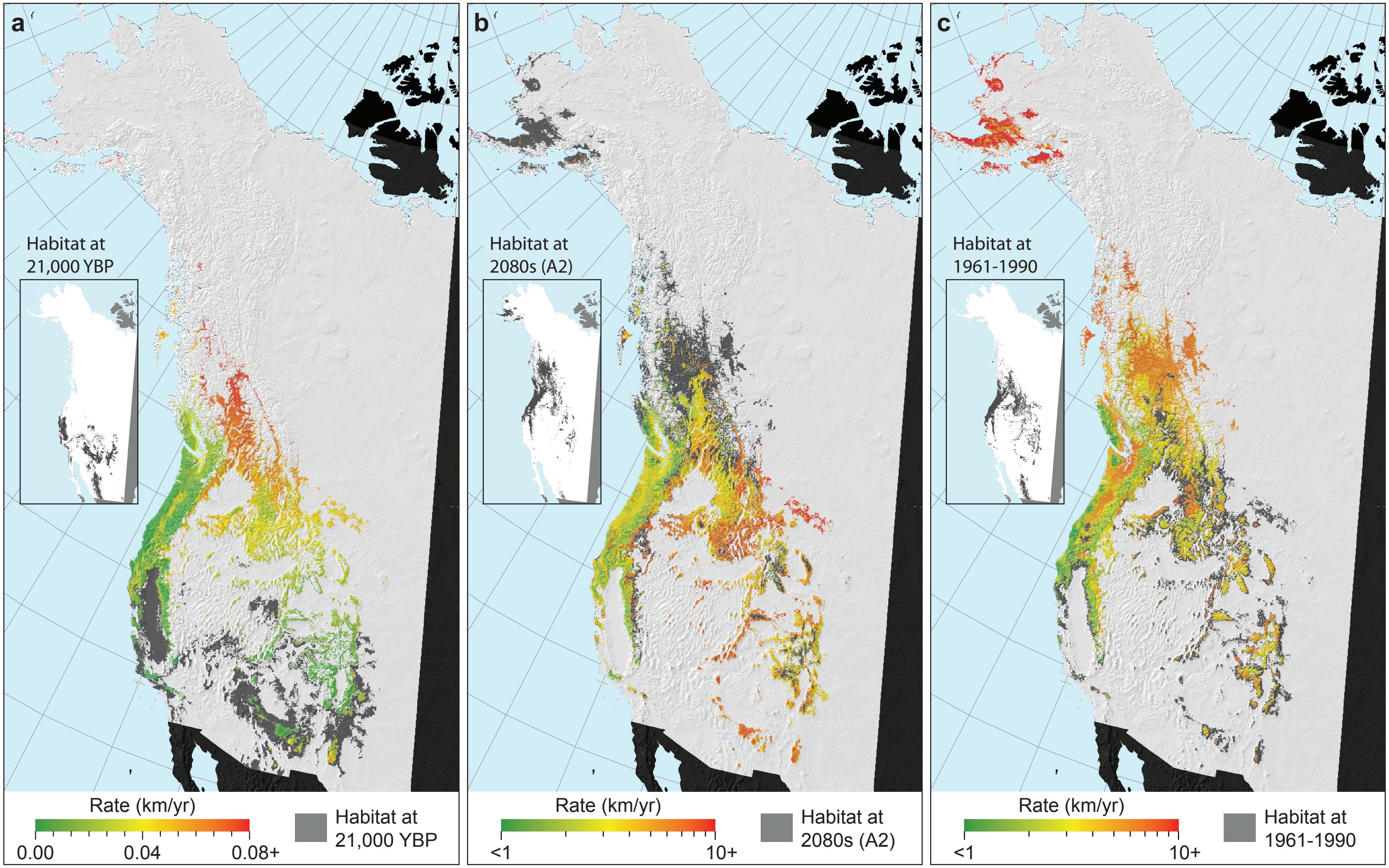

Figure S5.1T: ***Pseudotsuga menziesii*** **(Douglas-fir)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

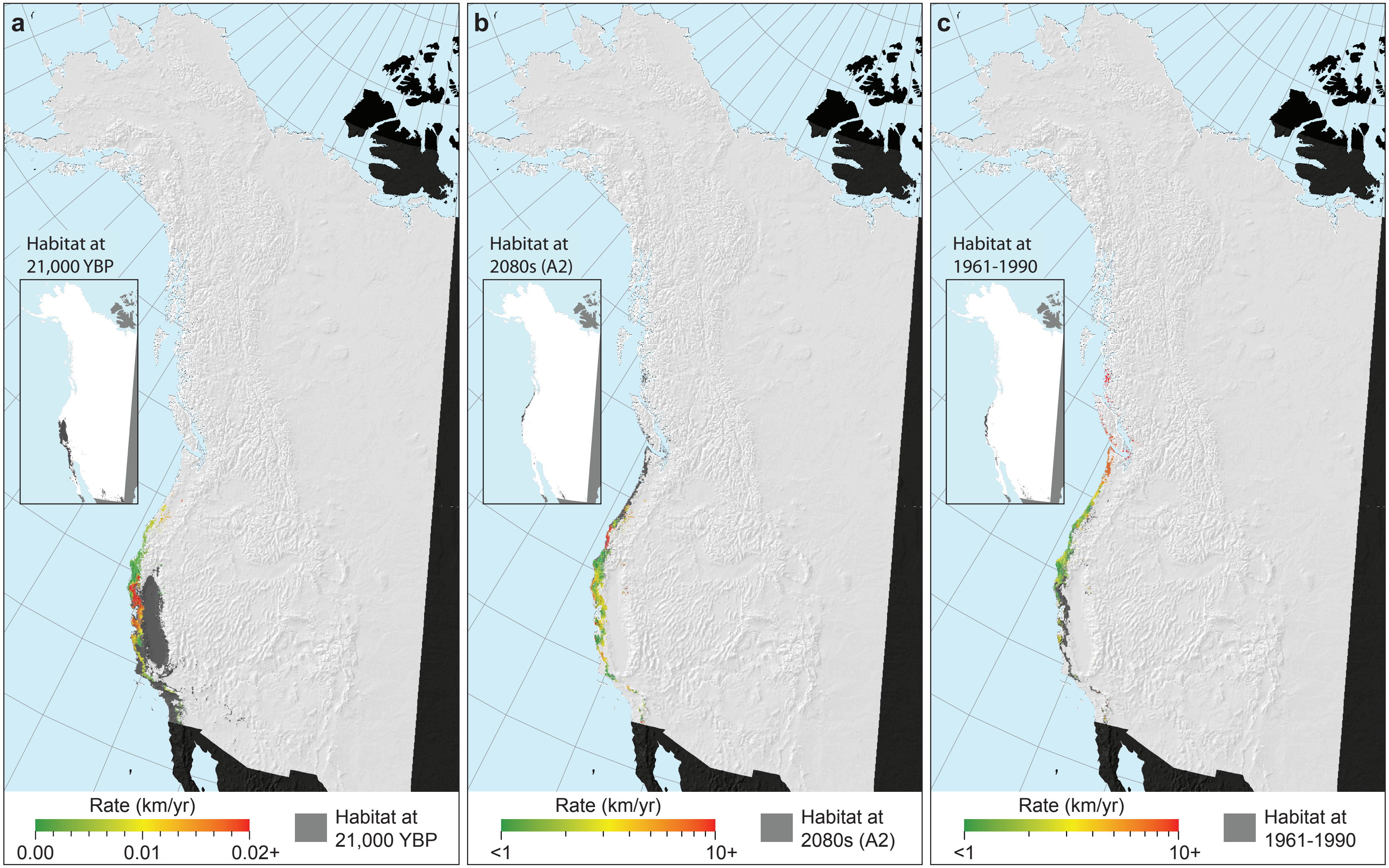

Figure S5.1U: ***Sequoia sempervirens*** **(coast redwood)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

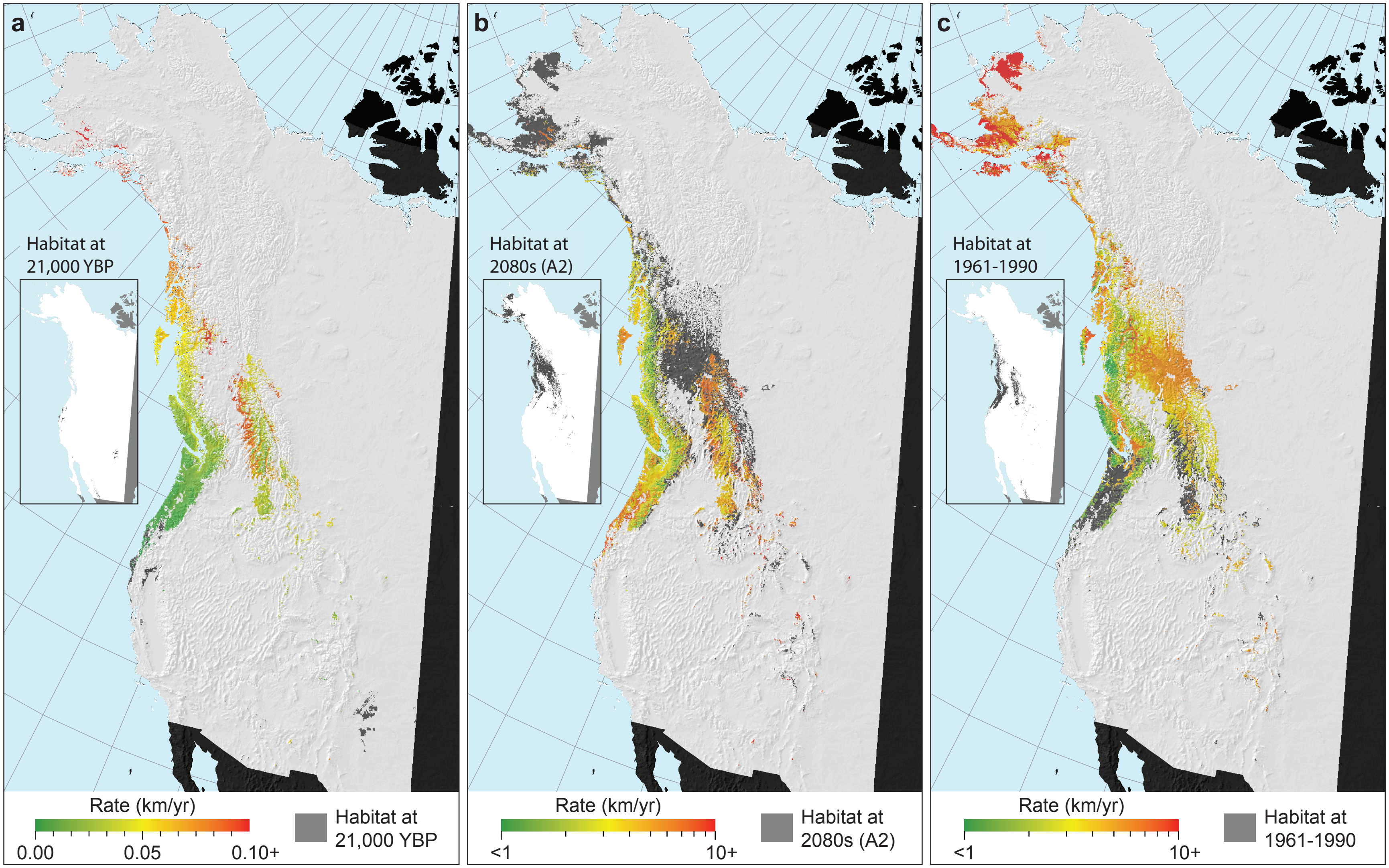

Figure S5.1V: ***Thuja plicata*** **(western redcedar)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

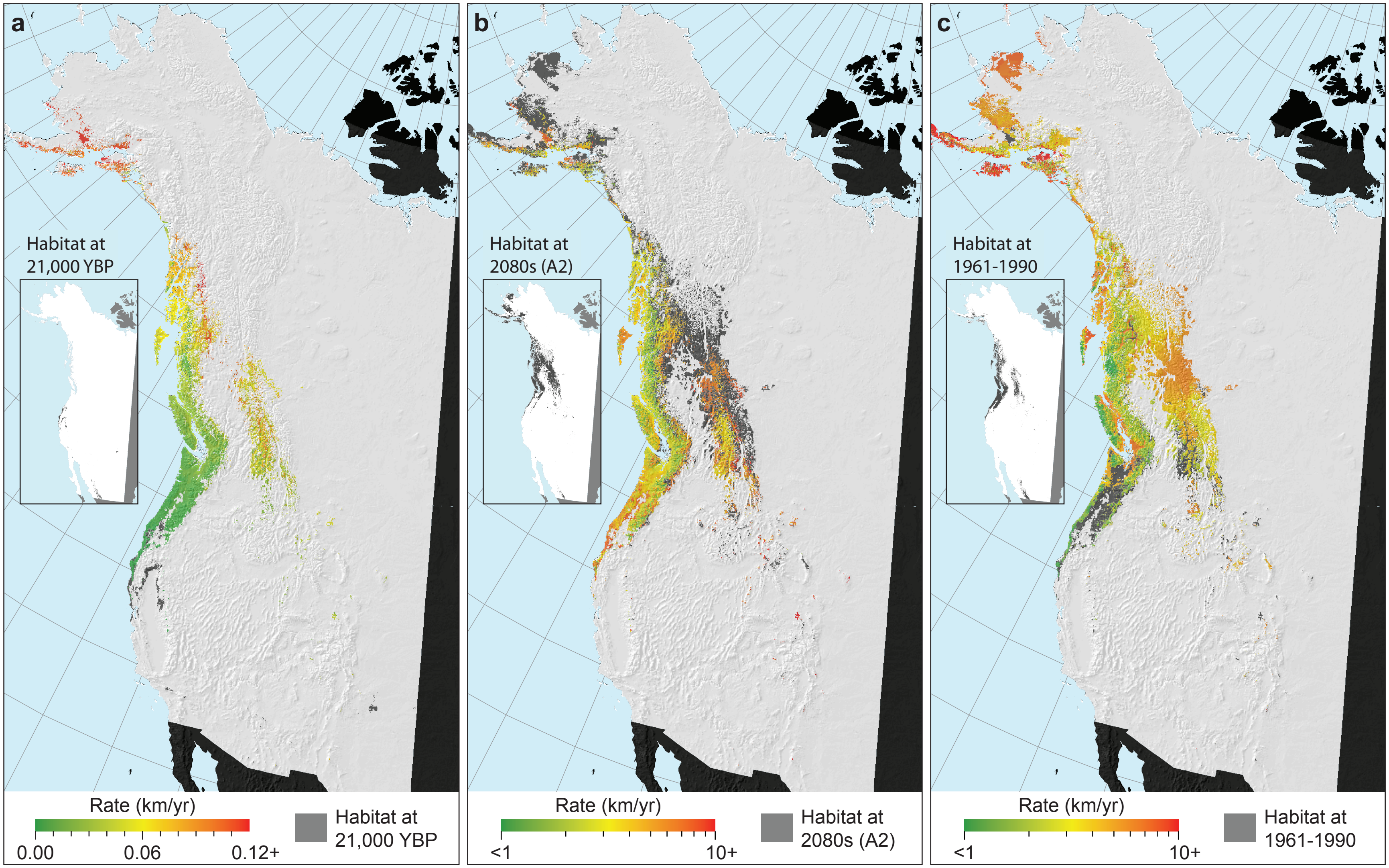

Figure S5.1W: *Tsuga heterophylla* (western hemlock) Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

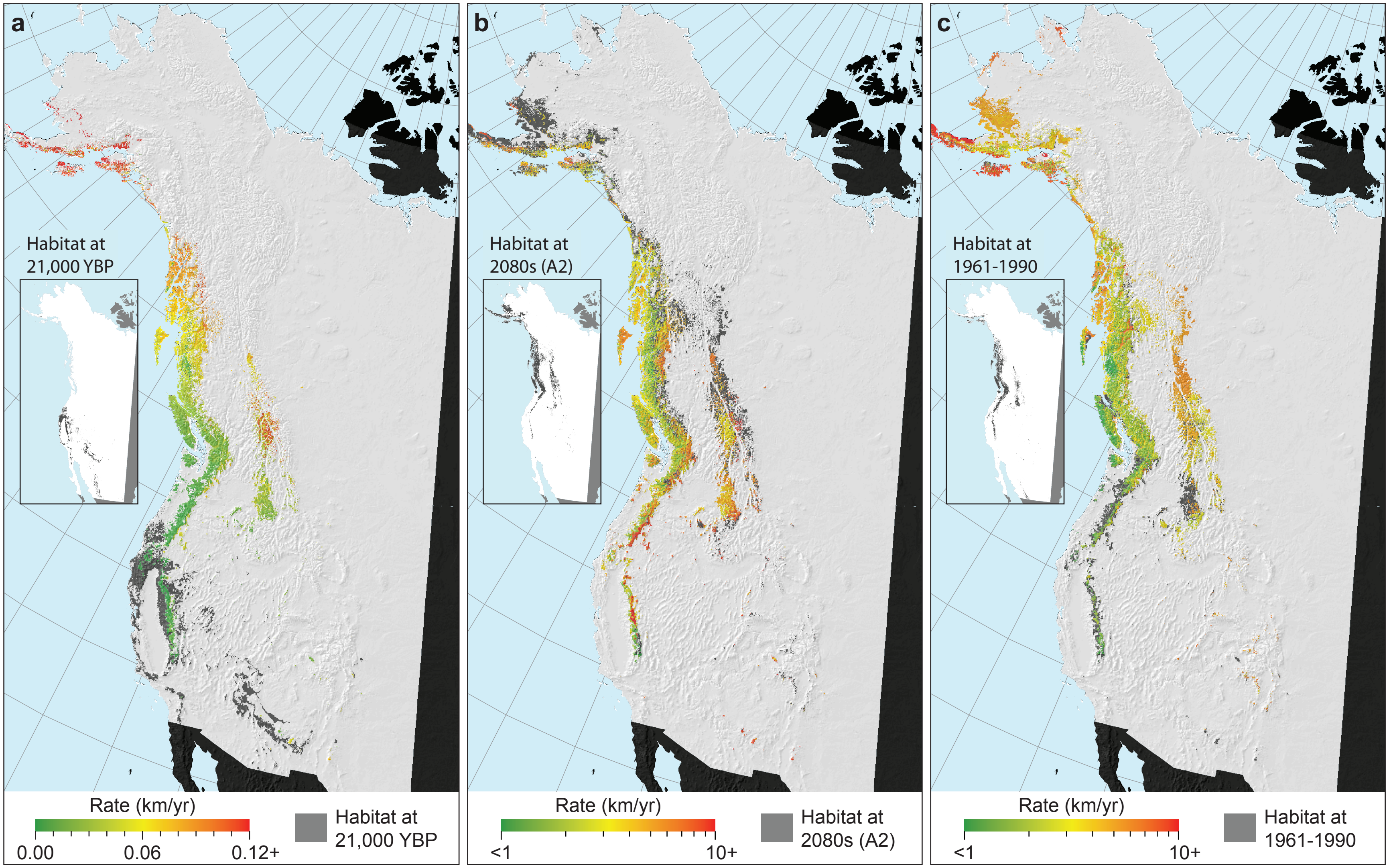

Figure S5.1X: ***Tsuga mertensiana*** **(mountain hemlock)** Required migration or gene flow rates as measured **(a)** between the GFDL climate reconstruction for 21,000 years ago and the 1961-1990 observed climate (representing post-glacial rates), **(b)** from the 1961-1990 observed climate to the climate of the A2 scenario for the 2080s period (representing how fast modern local populations must travel to suitable future habitats), and **(c)** from the climate of the A2 scenario of the 2080s period to the 1961-1990 observed climate (representing, for future habitats, the required rate for suitably-adapted present-day populations to arrive).

Chapter 6

Conclusions

This thesis makes contributions to the fields of species distribution modeling, biogeography, and genetics of western North American tree species. I further developed methodology that allows some of the results to assist in the development of better climate change adaptation strategies for important forest trees.

6.1 Developing effective modelling methods

Regarding species distribution modeling, I developed a new methodological approach, and tested a number of established techniques for their accuracy. Because the widely-used cross-validation approaches have recently been questioned, I used truly independent approaches to model validation: spatial extrapolation beyond data coverage and temporal extrapolation and validation against fossil and pollen data back to the last glacial maximum. Such validations are rarely done, and my results were interesting and unexpected in that they found strong correlations among semi-independent cross validations and truly independent tests of model accuracy. Thus, widely-used cross validation techniques may be over-optimistic, but they remain adequate model selection tools. Thus, I broadly affirmed the validity of using semi-independent data-splitting approaches for model selection. I also confirmed previous studies showing that while certain individual modelling methods consistently outperform others, ensembles that incorporate a number of individual methods performed best. A novel finding in this respect was that excluding the worst performing models in my validations did not increase the accuracy of ensemble projections.

By including both species- and ecosystem-based methods, I also assessed the difference in accuracy and robustness between these approaches. While species-based models were generally higher in measures of overall accuracy, ecosystem-based models tended to be much higher in specificity, generating species range maps that lacked virtually any tendency to over-predict species ranges. When converting probability of presence predictions to presence-absence binary maps, ecosystem-based methods were less sensitive to the selected threshold of probability. Also, presence-absence maps that considered all probability of presence values above zero appeared to be plausible species range maps. In contrast, individual species models had to be trimmed using thresholds for probability of presence predictions, and the resulting maps were quite sensitive to this level. Furthermore, individual species models lacked robustness when extrapolating beyond data coverage. Models trained with data from Canada typically over-predicted species ranges in the United States, and hind-casts toward the last glacial maximum often resulted in quite high probability of presence projections in habitat that should be unsuitable, such as areas under the continental ice sheet.

I also investigated potential reasons for the limitation of species distribution models, in particular their behavior when projecting into climate conditions that lack modern analogues. This is a widely stated note of caution against applying species distribution models under future climate scenarios. Such climate projections may contain climate arrangements or extremes without equivalent today, and therefore correlative approaches may fail in reliably predicting future habitat. In the second research chapter, I tested model accuracies through past time periods, using palaeoecological data for validation, in comparison with multivariate measurements of climate similarity to the present day. Results showed that, at measurements of high climate dissimilarity with the training data, the accuracy of ecosystem-based species distribution models may indeed be compromised, while at lower levels of climate dissimilarity, models showed better performance. However, while no-analogue climate conditions at the last glacial maximum were common, the majority of these non-analogous climates appeared under the continental ice in northern North America where vegetation would not be present. Last glacial maximum climate dissimilarities for much of the southern portion of the North American continent and Beringia, as well as those for projected climate conditions of the coming century, were much lower. Comparatively lower dissimilarity measurements in the future climate projections also suggested that future no-analogue conditions may not be widespread through the study area. This study allowed the no-analogue climate issue to be put into a paleoecological perspective, and I concluded that habitat future climate projections for the 21^{st} century are likely within model capabilities.

6.2 Historical biogeography and genetic diversity

Using ensemble projections of three ecosystem-based modeling techniques, I reconstructed tree species habitat back to the last glacial maximum with the goal of investigating how modern genetic diversity and the structure of species into subspecies and genetic varieties may have been shaped by species refugial and biogeographical histories. Interpretations of species habitat hindcasts from the last glacial maximum identified a number of geographic areas which held notably high levels of tree diversity at the last glacial maximum, appearing mostly through California and Oregon, as well as through the highlands of the southern United States interior. Outside of these areas and south of the continental ice, habitat reconstructions showed a scarcity of tree species with tundra, desert, and savannah dominating the landscape. Areas along the Pacific coast of British Columbia and Alaska often exhibited suitable habitat, particularly for temperate rainforest or boreal species. Through the post-glacial warming of the Holocene, as continental ice receded, habitat for tree species gradually emerged in more northern locations. While many species likely expanded post-glacially from a single refugial region located south of the continental ice, other species show notably more complex patterns, including those with possible refugia in Beringia (Alaska) or along the Pacific coast west of the continental ice.

Examination of range reconstructions also suggested differences in historical biogeographies for species with different patterns of genetic diversity. For example, species with recognized variants or sub-species (typically a wetter coastal variant and drier interior variant), such as lodgepole pine, ponderosa pine, and Douglas-fir, showed extensive and widespread habitat at the last glacial maximum, largely maintaining geographic separation between coastal and interior regions. By comparison, hindcast refugia for temperate rainforest species with less genetic diversity, even those with modern disjunct interior populations, tended to be restricted to small coastal areas (e.g. hemlock, redcedar, and alder). These patterns suggest parallels between species historical biogeographies, including their hindcasted glacial ranges and post-glacial migration routes and patterns of modern genetic diversity.

Relationships between species' genetic diversities and landscape configuration metrics of their glacial ranges suggested that modern genetic diversity is largely determined by species' historical biogeographies, particularly in the case of allelic richness. Generally, this result supports genetic theory suggesting strong effects of bottlenecks on genetic diversity. It also suggests that genetic diversity, particularly as measured by the presence of rare alleles, may increase very slowly with range expansion. This has implications for conservation and management in a changing climate where ranges have potential to alter dramatically for many species. If genetic diversity is lost due to range contractions or extirpated populations, the recovery of rare alleles can be expected to be an extremely slow process.

6.3 Past and future rates of habitat migration

Comparing calculated past and future rates of habitat migration suggested that, while magnitudes of future rates may far outpace past rates, the general patterns of migrations remain similar. General trends in the movement of similar climate habitats were northward at higher rates across more uniform prairie landscapes and upslope at slower rates through topographic terrain with more climate variability. Calculated future habitat migration rates for local populations indicated that, while some modern populations may find suitable habitat nearby, others may be forced to travel long distances. Further, some future habitats may require genetic material from distant sources to establish well-adapted forest communities. This suggests that, while species as a whole may not be at risk for extinction (as there will likely be locations within the species range in which future suitable habitat maintains in situ or very close to the current location), populations in certain localities may be at risk of maladaptation due to the inability of necessary gene flow to keep pace with the migration of climate habitats.

In light of these findings, genetic conservation strategies for present day populations should be considered, particularly in areas in which long-distance gene flow or migration is likely to be required for the maintenance of modern genotypes. Increasing the potential for modern genotypes to find suitable habitat in warming climates, without direct management intervention at the population level, should involve the conservation of contiguous terrain, particularly that with topographic heterogeneity to maximise climatic habitat diversity. In my rate calculations, topographic terrain served to limit habitat migration rates for downslope locations which could move short distances uphill, but dramatically increase habitat migrations rates for already upslope populations which would be required to make long-distance jumps to suitable habitat in distant areas. For this reason, preservation of contiguous habitat space in and around locations such as the Rocky Mountains, Sierra Nevada, and Cascade Ranges, would serve to maximise ratios of land area to habitat heterogeneity, potentially conserving suitable habitats in closer locations, even in topographically-driven microsites.

At the trailing edge of modern species distributions, such as at low elevations in southern locations, populations existing at the tolerance limits of the species may experience local range loss under climate change. Recently observed species declines in these areas, linked to changes in local climate, would suggest that climate change is already affecting marginal populations. The observed rates of temperature change for the past 30 years as calculated in my analysis, which aligned with those projected for the end of the century, coincides with these species declines. However, species declines in mid-range areas far from the margins of species tolerances may be driven by the lack of suitably-adapted genotypes in the nearby vicinity, particularly if my calculated rates of gene flow suggested a lack of similar climate habitat nearby. In these cases, management strategies that consider seed sourcing from non-local areas may be appropriate. Implementations of assisted migration or assisted gene flow strategies have been suggested (Aitken & Whitlock, 2013; Gray *et al.*, 2011) and tested (Ukrainetz *et al.*, 2011) for regions within western North America, and may represent reasonable strategies to mitigate maladaptation due to climate change.

6.4 References

Aitken, S. N., & Whitlock, M. C. 2013. Assisted gene flow to facilitate local adaptation to climate change. *Annual Review of Ecology, Evolution and Systematics.*

Gray, L. K., Gylander, T., Mbogga, M. S., Chen, P. Y., & Hamann, A. 2011. Assisted migration to address climate change: recommendations for aspen reforestation in western Canada. *Ecological Applications*, **21**(5), 1591–1603.

Ukrainetz, N. K., O'Neill, G. A., & Jaquish, B. 2011. Comparison of fixed and focal point seed transfer systems for reforestation and assisted migration: a case study for interior spruce in British Columbia. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere*, **41**(7), 1452–1464.