Journal of Biogeography (*J. Biogeogr.*) (2006) **33**, 342–360

ORIGINAL ARTICLE

Classification tree and minimum-volume ellipsoid analyses of the distribution of ponderosa pine in the western USA

Jodi R. Norris[1]*, Stephen T. Jackson[1] and Julio L. Betancourt[2]

[1]*Department of Botany, University of Wyoming, Laramie, WY, USA and* [2]*Desert Laboratory, US Geological Survey, Tucson, AZ, USA*

*Correspondence: Jodi Norris, Department of Botany, University of Wyoming, Laramie, WY 82071, USA.
E-mail: jnorris@uwyo.edu

ABSTRACT

Aim Ponderosa pine (*Pinus ponderosa* Douglas ex Lawson & C. Lawson) is an economically and ecologically important conifer that has a wide geographic range in the western USA, but is mostly absent from the geographic centre of its distribution – the Great Basin and adjoining mountain ranges. Much of its modern range was achieved by migration of geographically distinct Sierra Nevada (*P. ponderosa* var. *ponderosa*) and Rocky Mountain (*P. ponderosa* var. *scopulorum*) varieties in the last 10,000 years. Previous research has confirmed genetic differences between the two varieties, and measurable genetic exchange occurs where their ranges now overlap in western Montana. A variety of approaches in bioclimatic modelling is required to explore the ecological differences between these varieties and their implications for historical biogeography and impending changes in western landscapes.

Location Western USA.

Methods We used a classification tree analysis and a minimum-volume ellipsoid as models to explain the broad patterns of distribution of ponderosa pine in modern environments using climatic and edaphic variables. Most biogeographical modelling assumes that the target group represents a single, ecologically uniform taxonomic population. Classification tree analysis does not require this assumption because it allows the creation of pathways that predict multiple positive and negative outcomes. Thus, classification tree analysis can be used to test the ecological uniformity of the species. In addition, a multidimensional ellipsoid was constructed to describe the niche of each variety of ponderosa pine, and distances from the niche were calculated and mapped on a 4-km grid for each ecological variable.

Results The resulting classification tree identified three dominant pathways predicting ponderosa pine presence. Two of these three pathways correspond roughly to the distribution of var. *ponderosa*, and the third pathway generally corresponds to the distribution of var. *scopulorum*. The classification tree and minimum-volume ellipsoid model show that both varieties have very similar temperature limitations, although var. *ponderosa* is more limited by the temperature extremes of the continental interior. The precipitation limitations of the two varieties are seasonally different, with var. *ponderosa* requiring significant winter moisture and var. *scopulorum* requiring significant summer moisture. Great Basin mountain ranges are too cold at higher elevations to support either variety of ponderosa pine, and at lower elevations are too dry in summer for var. *scopulorum* and too dry in winter for var. *ponderosa*.

Main conclusions The classification tree analysis indicates that var. *ponderosa* is ecologically as well as genetically distinct from var. *scopulorum*. Ecological

www.blackwellpublishing.com/jbi
doi:10.1111/j.1365-2699.2005.01396.x

differences may maintain genetic separation in spite of a limited zone of introgression between the two varieties in western Montana. Two hypotheses about past and future movements of ponderosa pine emerge from our analyses. The first hypothesis is that, during the last glacial period, colder and/or drier summers truncated most of the range of var. *scopulorum* in the central Rockies, but had less dramatic effects on the more maritime and winter-wet distribution of var. *ponderosa*. The second hypothesis is that, all other factors held constant, increasing summer temperatures in the future should produce changes in the distribution of var. *scopulorum* that are likely to involve range expansions in the central Rockies with the warming of mountain ranges currently too cold but sufficiently wet in summer for var. *scopulorum*. Finally, our results underscore the growing need to focus on genotypes in biogeographical modelling and ecological forecasting.

Keywords

Bioclimatic model, classification tree analysis, ecotypes, minimum-volume ellipsoid, *Pinus ponderosa*, range shifts, western USA.

INTRODUCTION

The unusual distribution of ponderosa pine (*Pinus ponderosa* Douglas ex Lawson & C. Lawson) in the western USA raises questions about the species' climatic tolerances. Ponderosa pine is a montane conifer, with suitable habitat available somewhere along the elevation gradient of most mountain ranges in western North America. It is present in all western states from California to Nebraska and from near the USA/Mexican border to more than 300 km north of the USA/Canadian border (Fig. 1). It is conspicuously absent, however, from the mountain ranges in the physiographic Great Basin of north and central Nevada and western Utah as well as western Wyoming and southern Idaho. The wide distribution of ponderosa pine encompasses at least two varieties, several suggested races, and numerous populations that exhibit genetic differences (Korstian, 1924; Weidman, 1939; Conkle & Critchfield, 1988; Sorensen *et al.*, 2001). The biogeography of ponderosa pine is a product of an evolutionary history of unknown span, geographic isolation of varieties, and substantial range expansions and contractions during glacial–interglacial cycles. The goal of this study was to determine whether the present distribution of ponderosa pine can be modelled using readily available climatic and edaphic data, and to identify the environmental factors that determine its unique distribution.

The creation of a model of ponderosa pine distribution is complicated by genetic variation within the species. Common-garden studies across the range of ponderosa pine have demonstrated genetic differences between the interior variety *P. ponderosa* var. *scopulorum* and the western variety, *P. ponderosa* var. *ponderosa*. The same studies have also confirmed differences among many populations within each variety, including variations in ecologically important characteristics such as growth rate, drought hardiness, water-use efficiency, and cold tolerance (Squillace & Silen, 1962; Wells, 1964; Conkle & Critchfield, 1988; Cregg *et al.*, 2000). Although many western conifers show clinal population variations in common-garden studies, ponderosa pine exhibits relatively abrupt differences between populations (Sorensen *et al.*, 2001). The known phylogeographic structure of ponderosa pine suggests that it may not have a single coherent niche, but rather multiple niches corresponding to multiple geographic areas as a result of some combination of migration, isolation, and adaptation to local or regional conditions. In this context, it is puzzling that no populations are adapted to conditions in the geographic centre of its distribution. The interior mountain ranges from which this species is absent have a wide range of substrates, elevations, temperatures, and precipitation regimes suitable for forests. Species that co-occur with ponderosa pine [e.g. *Juniperus scopulorum Sargent*, *Pseudotsuga menziesii (Mirbel) Franco*, *Populus tremuloides Michaux*] occur in most of these interior mountain ranges. Thus, either climate or dispersal limitations prevent ponderosa pine from growing in the centre of its range.

Although mountains in the Great Basin represent remote islands of suitable habitat, the distances between them are not likely to represent a serious dispersal barrier for the species over millennial time-scales. Ponderosa pine has a demonstrated dispersal record in the late Pleistocene and Holocene that includes northward expansion of *c.* 1500 km for var. *scopulorum*, including expansion into isolated mountain ranges and escarpments (Anderson, 1989; Betancourt *et al.*, 1990). Climate remains the factor most likely to be responsible for the absence of both varieties of ponderosa pine in the Great Basin, and may also control the western geographic extent of var. *scopulorum*. Describing how climate controls ponderosa pine distribution is difficult because of the regional distribution and range of environments involved as well as because of the possible complications of genetic

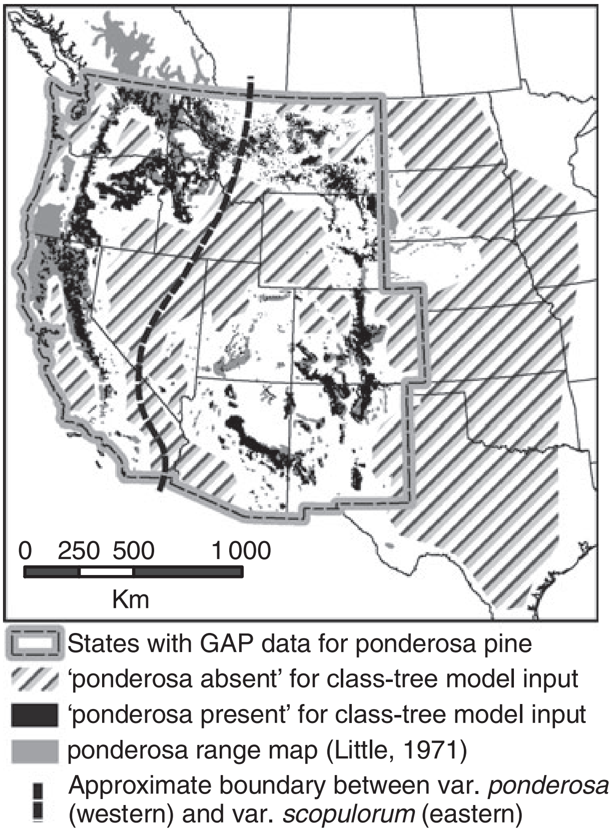

Figure 1 Distribution of ponderosa pine (*Pinus ponderosa* Douglas ex Lawson & C. Lawson) in the western United States and Canada. Little's (1971) distribution is a good overall representation of the maximum extent of ponderosa but is less useful as model input. GAP analysis data are more conservative and are used for the classification tree model creation. Also displayed is the range for ponderosa pine absence that was used for the classification tree model. All maps shown in this manuscript have been projected to Lambert Conformal Conic projection; First std Parallel: 30°; Second std parallel: 50°; Latitude of Origin: 40°; Central Meridian: −106°.

variation within the species. An objective approach to assessing climatic control of ponderosa pine distribution would be to try to model the distribution to see if it could be adequately predicted using climate variables, and then to use the model to try to explain unique features of ponderosa pine distribution.

Modelling

Species and vegetation distribution models are routinely used in ecological studies for several purposes, including to predict the response to environmental change (Franklin, 1998; Guisan *et al.*, 1999) or to understand and predict current distribution (McKenzie & Halpern, 1999; Guisan & Theurillat, 2000; Anderson *et al.*, 2003). The type of model chosen depends largely on the characteristics of the species being modelled, on the type and scale of the available distribution and predictor data, and on the goals of the modelling study. Available source data for constructing distribution models across a region are plagued by low resolution for both distribution and predictor variables, with large (*c.* 20–50%) classification errors associated with species distribution data (see Methods), and errors associated with interpolating climatic averages for points and elevations between weather stations. Any model that faithfully reproduces this imperfect distribution in spite of the imperfect predictor information is likely to be overfitted, because such a model would incorporate not only the 'signal' of the patterns in the data, but also the 'noise' associated with the large errors in the training data. Despite these shortcomings, broad niche requirements and patterns may still be evident in regional species distribution models using currently available data.

We aim to answer four questions from our distribution models for ponderosa pine.

Question 1. Can the regional distribution of ponderosa pine be explained using available distribution data and climatic and edaphic variables? We created a classification tree model (Breiman *et al.*, 1984), which can contain multiple pathways for both presence and absence outcomes and thus allows for expression of multiple genetic variants with unique environmental niches.

Question 2. Does ponderosa pine represent a single ecologically distinct taxon, or are there two or more subtaxa, each with a unique ecological niche? If there are ecologically distinct subtaxa, how do they compare with varieties previously described from morphological and genetic studies? Documented genetic differences among populations of ponderosa pine suggest that ecologically distinct variants exist that may best be modelled separately. This hypothesis was tested using a classification tree model, and compared the geographic distribution of the resulting tree 'leaves' with the distribution of the varieties of ponderosa pine.

Question 3. Is the location of the hybrid zone and boundary between the varieties *ponderosa* and *scopulorum* a product of chance, or a consequence of the ecological adaptations of each variety? This hybrid zone, located in western Montana (Latta & Mitton, 1999), is of interest because ponderosa pine forests in the intermountain region are relatively young, having been restricted to more southerly locations through the last glacial maximum. Ponderosa pine macrofossils appear in woodrat middens only within the last 6000 years at sites in northern Colorado and central Wyoming (Betancourt *et al.*, 1991; J. L. Betancourt, M. E. Lyford & S. T. Jackson, unpubl. data) and within the last 3000 years in the Bighorn Mountains of northern Wyoming and the Little Belt and Big Belt Mountains of central Montana (J. R. Norris, unpubl. data). The location of the current boundary between the varieties might be explained by dispersal and range expansion of the two varieties until they met at an arbitrary location unrelated to their environmental niches. Alternatively, the boundary location may represent a climatic boundary that corresponds to the fundamental niches of the varieties. We examine individual 'leaves' created by the classification tree model to identify the location of ecological boundaries within the species, to see whether the ecological niches represented by different leaves coincide with the geographic distribution of the varieties.

Question 4. What climatic or edaphic variables are limiting in areas where ponderosa pine is absent? There is no single answer to this question: a species may be limited by different factors at upper, lower, northern, and southern boundaries. Our goal is to display these limitations in map form, to improve understanding of the biogeography of ponderosa pine. This effort requires estimation of how limiting each variable was at a particular location, which can be done using a simple geometric model of ponderosa pine niche space in the shape of a multidimensional ellipsoid.

Using geometry as a realized niche: the minimum-volume ellipsoid

The minimum-volume ellipsoid has been used by hydrologists seeking to define hydrological regimes for use in watershed modelling (Hodge & Tasker, 1995). The approach has considerable potential in ecological applications. The mathematics used to create the ellipsoid is based on extensions from two-dimensional geometry, and the concept is relatively simple to describe and understand in familiar two- or three-dimensional terms. It is easiest to start with a simpler geometric description of a rectangle, which in two dimensions can be used as a crude model and forms the basis of an environmental limitation model called ECOSIEVE (Box, 1981). Box describes the ECOSIEVE program as creating 'rectangular hypervolumes within eight-dimensional ecoclimate hyperspace' (Box, 1981). For example, if a taxon is know to tolerate July temperatures within a particular range (X_1–X_2) and January temperatures within a particular range (Y_1–Y_2), then a crude model of where the taxon occurs can be created based on those variables (Fig. 2a). The species is not likely to occur at any geographic point whose joint values for the variables fall outside the rectangle. The position of the point relative to the rectangle would suggest, for example, that conditions are either too hot or too dry or both. This process can be repeated for any number of additional variables. In climate space, this would exist as a multidimensional box to describe what conditions the species can tolerate.

Many of these climate-related limits are likely to be interdependent for an individual species. The results of interdependence become evident when plotting the occurrence points for a species on two axes. Thompson *et al.* (1999) created a series of plots that display species occurrence points in two-dimensional climate space (Fig. 2b). From their work it is apparent that a circle or ellipse, rather than a box, might better represent the range of suitable conditions for many taxa. This geometric representation of acceptable climate space, whether as a box, circle, or ellipse, is a crude model of the taxon distribution in environmental space, and this model can be expressed using basic mathematics. Because the actual distribution of ponderosa pine in climate space resembles an ellipse rather than a rectangle (Fig. 2b), an ellipse was used as the basis for the second distribution model.

A very simple two-dimensional ellipse model can be created by simply drawing an ellipse that is as small as possible around plotted points, yet still includes all the points (Fig. 2c). This task is more complicated with multiple variables, but computing technology makes multidimensional geometric calculations relatively simple.

Hodge & Tasker (1995) have developed a FORTRAN program that estimates the dimensions of a minimum-volume ellipsoid in multidimensional space from a control data set. The advantage of this type of geometric model lies in the ability to estimate numerically why a site is not suitable for a species, because it is possible to calculate which climatic conditions are not being met. This was done in the rectangle example by calculating the distance from the rectangle in the *x*-dimension and *y*-dimension. The same calculations can be made for the

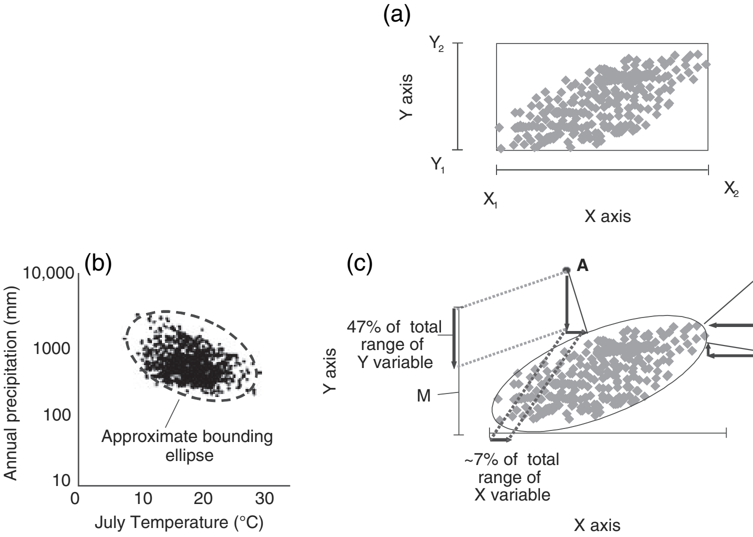

Figure 2 (a) The minimum and maximum values for climate variable *X* and climate variable *Y* form a geometric shape of a rectangle or box in two dimensions. (b) In reality, the climate space of the taxon is not likely to be box-shaped. This figure, modified from Thompson *et al.* (1999), shows the distribution of ponderosa pine (*Pinus ponderosa* Douglas ex Lawson & C. Lawson) in the United States in climate space. (c) For the new points A, B and C that fall outside the ellipse by about the same distance: point A is outside with respect to the *Y*-axis, point B is outside with respect to both the *X* and *Y* axes, and point C is outside with respect to the *X*-axis. The distance to the multidimensional ellipsoid was calculated as a percentage of the total range in that dimension (line M).

ellipse (Fig. 2c). In the ellipse example case, both distances would be calculated in degrees Celsius (°C). The use of dimensional units (e.g. °C) becomes more difficult when comparing different variables. Instead, distances can be expressed as a ratio or percentage of the overall range tolerated by the species. For example, a 2 °C difference might constitute a large or small percentage depending on whether a species has narrow or broad temperature tolerances.

METHODS

Model selection

Three characteristics of our study determined model choice: (1) uncertainty about whether ponderosa pine comprises a single coherent ecological species (Wells, 1964; Rehfeldt, 1990, 1993; Sorensen *et al.*, 2001); (2) ability to explain ponderosa pine absence in the centre of its geographic range; and (3) the nature of errors in distribution data for ponderosa pine.

The classification tree model addresses the first issue by creating a tree structure composed of a set of decision rules based on explanatory variables. These rules are formatted as 'true/false' or 'greater-than/less-than' decisions that create successive, nested binary splits of the observed data. The process is optimized to create a tree whose terminal ends or 'leaves' contain mostly homogeneous groups of either 'presence' or 'absence' values. Because of this structure, different 'presence' leaves could represent substantial differences in the taxon being modelled. For example, it would be possible to create a model using input of undifferentiated 'presence' points for two allopatric species. The resulting tree would have final groups or 'leaves' that were highly segregated, with some presence leaves consisting only of the first species and other presence leaves consisting only of the second species. By examining the geographic distribution of the points forming the 'presence' leaves this segregation would become evident. In applying the classification tree model for ponderosa pine, we examined the geographic extent of the environmental conditions described by the different leaves. Although other researchers have used classification tree models with multiple presence leaves (De'Ath & Fabricius, 2000; Larsen & Speckman, 2004), we are not aware of studies using the approach as a tool for identifying genotypic variation within a species.

The second contributing factor to our choice of models was the ability of the model to provide some explanation of why 'absence' points were not 'presence' points in terms of climatic or edaphic variables. The minimum-volume ellipsoid model allows us to calculate for an 'absence' point the distance from the ellipsoid as a vector component for each of the climate variables (Fig. 2c). Thus, maps can be created of the western United States that are colour-coded to denote the relative distance from the niche at that point.

The final consideration in the choice of models is the large potential error in the distribution data for ponderosa pine. Source distribution data for the classification tree model came from the US Geological Survey Gap Analysis Program (GAP) (http://www.gap.uidaho.edu/). Accuracy estimates for the GAP data vary from state to state, and some states had no specific accuracy estimates. Montana reported 76% absolute accuracy for ponderosa pine vegetation type with 81.5% accuracy if very similar vegetation types are included (Rocky Mountain juniper class and mixed xeric forest class) (Redmond *et al.*, 1998). The Wyoming GAP initial accuracy report indicated that 79% of all checked locations had the correct assignment (Merrill *et al.*, 1996). Colorado reported an absolute accuracy of 57.7% for all vegetation types (Schrupp *et al.*, 2000). New Mexico reported conservative, liberal, and ecotonal accuracies for ponderosa pine vegetation type of 52%, 64% and 89% respectively (Thompson *et al.*, 1996). Although we cannot estimate accuracy across all of the western states, an optimistic projection of overall accuracy would be no higher than 80%.

Errors of this magnitude will affect any model negatively. The likely effects of the classification errors on the classification tree take two forms: it should be difficult to obtain high classification accuracies for leaves that include many erroneous source points; and there should be a large number of small leaves for which decision points are influenced by chance associations. To be conservative, we chose to remove all small leaves and any large leaves with large errors.

The ellipsoid model is more sensitive than the classification tree model to the effects of outlying data points, because a single outlying point can change the resulting dimensions of the ellipsoid. In order to avoid this issue, a different data source was used for the minimum-volume ellipsoid model. A data set of point observations of ponderosa pine was compiled using multiple data sources (Table 1). The likelihood of errors in this data set is greatly reduced because each observation represents a field identification of ponderosa pine, rather than the typical methods used for GAP analysis data, which are based largely on characteristics identified from remotely sensed data. The resulting data set of point observations is unevenly distributed because some states had very large amounts of data readily available, whereas other states had very little data readily available. The unevenness in the point distribution would have affected the classification tree model, so this point data set was only used for the ellipsoid model. The uneven distribution of points should have little effect on the minimum-volume ellipsoid, because the ellipsoid created does not depend on the density of points within an area of climate space. However, the general lack of point information from Montana may have resulted in a lack of representation of that area of climate space, and may affect the model performance in that geographic area.

Source distribution data

Ponderosa pine presence and absence data were determined separately. Presence data for the classification tree model were obtained from GAP (http://www.gap.uidaho.edu/) for 11 of the 14 western states where ponderosa pine occurs (Fig. 1). GAP vegetation classes were only included if the class

Table 1 Sources of point observation and collection data used to create the minimum-volume ellipsoid models

Source	Conditions in addition to those described in methods	Individual providing data	URL
Calflora website	Screened according to 'type of observation' category, using only incidental observation, plot species list, or secondary source of compiled records	–	http://www.calflora.org/
Colorado State University Herbarium Database	Only records identified to the nearest land survey section (*c.* 2.6 square km)	–	http://herbarium.biology.colostate.edu/database.htm
Colorado University Museum Herbarium's Specimen Database of Colorado Vascular Plants	–	–	http://cumuseum.colorado.edu/Research/Botany/Databases/search.php
Digital version of the Atlas of the Vascular Plants of Utah	–	R. Douglas Ramsey	http://www.nr.usu.edu/Geography-Department/utgeog/utvatlas/
Institute of natural resource analysis and management (INRAM) database	–	Chris Frazier	http://abies.nmsu.edu/INRAM/biodiversity/index2.php
International Tree-Ring Data Bank	–		http://www.ncdc.noaa.gov/paleo/ftp-treering.html
Oregon Plant Atlas Database	Only locations with a reported accuracy of 1 mile or less	Linda Hardison	http://www.oregonflora.org/oregonplantatlas.html
Rocky Mountain Herbarium Atlas of the Vascular Plants of Wyoming	Removed locations where specimens were planted rather than naturally occurring (R. Hartman, pers. comm.)	Ron Hartman, Joy Handley	http://www.sbs.utexas.edu/tchumley/wyomap/atlas.htm
Southwest Environmental Information Network	Removed specimen locations described as planted	–	http://seinet.asu.edu/
University of Idaho Herbarium Database	–	Pam Brunsfeld, Tyler Morrison	http://www.sci.uidaho.edu/biosci/herbarium/database.asp
Vegbank, created by the Ecological Society of America	–	Michael Lee	http://vegbank.org
A small number of personal observations	–	–	–

specifically listed ponderosa pine in the name (e.g. ponderosa pine–Rocky Mountain juniper association) or as a species whose presence defined that vegetation class (e.g. ponderosa pine forest). Vegetation types were not included if ponderosa pine was listed as one of several species that might be found in a vegetation class, such as many 'mixed conifer' classes. South Dakota, North Dakota and Nebraska were not included in the data set because digital data were not available for a ponderosa-specific vegetation type. The GAP data varied in format and resolution from state to state. All data were converted and reprojected to a grid with uniform 2.5-min cell spacing (*c.* 4 km) to match the grid resolution of the available climate data.

Ponderosa pine absence for the classification tree model was determined conservatively by including only areas of fairly certain ponderosa pine absence, by comparing a combination of distribution maps (Little, 1971) and the GAP analysis data. These absence data could theoretically extend across eastern North America, as well as to Central and South America. However, we chose to restrict the extent of ponderosa pine absence data so as to exclude areas where dispersal could be limiting. The western Great Plains were retained in the model because of dispersal potential by east-flowing rivers, extreme winds, and humans (Fig. 1).

The final data set for the classification tree model consisted of 163,544 grid cells of 2.5-min size. From this data set, a 10,000-cell stratified subset was used as input for the classification tree model. The original data set was heavily biased towards absence data, which skews the model towards predicting absence. To prevent this, the 10,000-point subset was stratified to included 4000 randomly selected cells of ponderosa pine presence and 6000 randomly selected cells of ponderosa pine absence. As reported in the model selection section, GAP data have substantial errors, and an optimistic projection of GAP accuracy could go no higher than *c.* 80%. Therefore, because the random subset of GAP data contributing to the model consists of 4000 ponderosa pine 'presence' points, the estimated minimum number of points erroneously included is 20% or 800 points.

A different presence data set was used to construct the ellipsoid model because that model is more sensitive to the effects of outlying data points. A data set of point observations of ponderosa pine was compiled using multiple data sources (Table 1). There remains potential for mislocation of the collection or observation point on a map, error in data transcription, and misidentification of the species. In order to reduce the likelihood of these errors, all points more than 20-km distant from ponderosa pine populations identified by Little (1971) were removed by using the digital version of Little's (1971) maps created by Thompson *et al.* (1999). Then, points were removed if any climate variable at the individual point was more than three standard deviations from the mean value for that variable. For precipitation-related variables, the standard deviations were calculated from the logged values of the precipitation to produce a more nearly normal distribution. This resulted in a final data set for the ellipsoid model of 514 points for var. *scopulorum* and 1222 points for var. *ponderosa* (Fig. 3).

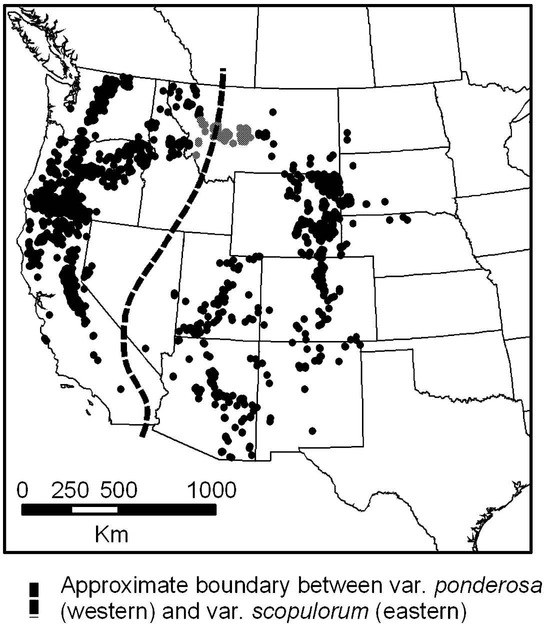

Figure 3 The distribution of point observation and collection data used for the minimum-volume ellipsoid models of the two varieties. Grey points near the hybrid zone were excluded from the models.

Independent variables

Variable selection process

Independent variable selection began with a suite of variables known to have an effect on ponderosa pine growth, survival, or recruitment (Table 2). Final variables were selected as the classification tree model was constructed. The input for the program creating the tree consists of information from the 10,000 data cells, including ponderosa pine presence or absence, and many climatic and edaphic values estimated for that particular cell. The computer program creates a tree on which branching points are created that best allow discrimination between the two possible outcomes (ponderosa-present/ponderosa-absent). These branching points are simple 'greater than' or 'less than' distinctions made for a particular variable, such as whether average January temperature is greater than 5 °C. Ten variables were represented in the final set of decision points and were used for the subsequent ellipsoid model: average January temperature, average April temperature, average July temperature, average January precipitation, average June precipitation, average July precipitation, average growing-season precipitation, average frost-free period, an estimate of topographic roughness, and the ratio of growing-season precipitation to average May–August temperature as an estimator of moisture stress (Table 2). The climatic variables are average values (1960–90) calculated from National Oceanic

Table 2 Variables used to create the classification tree model. The final tree included the variables shown in bold, which were then used to create the minimum-volume ellipsoid model

Variable	Data source
Average January temperature	PRISM data 1960–1990 (Daly *et al.*, 1997)
Average April temperature	PRISM data 1960–1990 (Daly *et al.*, 1997)
Average July temperature	PRISM data 1960–1990 (Daly *et al.*, 1997)
Average January precipitation	PRISM data 1960–1990 (Daly *et al.*, 1997)
Average June precipitation	PRISM data 1960–1990 (Daly *et al.*, 1997)
Average July precipitation	PRISM data 1960–1990 (Daly *et al.*, 1997)
Average growing-season precipitation	Calculated from PRISM data 1960–1990 (Daly *et al.*, 1997) using average monthly precipitation, first frost, and last frost data
Average daily growing season precipitation	Calculated from PRISM data 1960–1990 (Daly *et al.*, 1997) using average monthly precipitation, first frost, and last frost data
Average percent summer precipitation (April–September)	Calculated using PRISM data 1960–1990 (Daly *et al.*, 1997)
Average growing season precipitation/ average May–August temperature	Calculated from PRISM data 1960–1990 (Daly *et al.*, 1997) using average monthly temperature, precipitation, first frost, and last frost data
Average frost-free period	PRISM data 1960–1990 (Daly *et al.*, 1997)
Standard deviation of summer (June–July–August) precipitation, from 1949 to 1997	Calculated from PRISM high-resolution data set (Daly & Taylor, 1998)
Standard deviation of winter (December–January–February) precipitation, from 1949 to 1997	Calculated from PRISM high-resolution data set (Daly & Taylor, 1998)
Standard deviation of spring (April–May) precipitation, from 1949 to 1997	Calculated from PRISM high-resolution data set (Daly & Taylor, 1998)
Maximum summer (June–July–August) precipitation, from 1949 to 1997	Calculated from PRISM high-resolution data set (Daly & Taylor, 1998)
Maximum winter (December–January–February) precipitation, from 1949 to 1997	Calculated from PRISM high-resolution data set (Daly & Taylor, 1998)
Maximum spring (April–May) precipitation, from 1949 to 1997	Calculated from PRISM high-resolution data set (Daly & Taylor, 1998)
Minimum summer (June–July–August) precipitation, from 1949 to 1997	Calculated from PRISM high-resolution data set (Daly & Taylor, 1998)
Minimum winter (December–January–February) precipitation, from 1949 to 1997	Calculated from PRISM high-resolution data set (Daly & Taylor, 1998)
Minimum spring (April–May) precipitation, from 1949 to 1997	Calculated from PRISM high-resolution data set (Daly & Taylor, 1998)
Soil depth (from STATSGO data set)	USDA-United States Department of Agriculture (1993) data
Roughness – maximum range in elevation between the subject cell and its eight neighbouring cells on a 500-m digital elevation grid	DEM data from GLOBE Task Team and others (1999) calculated from 30-arc-second elevation data that had been projected to an equal area projection before performing calculations. The resolution of this grid was coarsened to match that of the climate data using a method that retained the maximum values for each new cell

and Atmospheric Administration (NOAA) weather-station data that have been gridded across the United States using topographically adjusted regression equations (PRISM data, Daly *et al.*, 1997). The roughness was calculated from a digital elevation model (GLOBE Task Team and others, 1999), by computing the maximum change in elevation between the subject cell and its eight neighbouring cells on a 500-m grid, and retaining the maximum value for each grid cell when resampling to overlay on the coarser *c.* 4-km resolution grids of other variables.

The roughness variable is not expected directly to control ponderosa pine distribution. We used it in the absence of a nationwide soils coverage as a proxy for topographic diversity and bedrock outcrops. Areas of topographic diversity (high relief) often have local sites where soil texture, slope, and aspect combine to yield high effective moisture suitable for tree growth. Coarse, stony soils near bedrock outcrops allow percolation, thus increasing effective moisture. As a result, sites with coarse-textured substrate can effectively be wetter than sites with fine-textured substrate (inverse texture effect) (Noy-Meir, 1973; Knight, 1994). In addition, topographic roughness and rock outcrops increase snow drifting, and in areas of high topographic relief the variation in aspect increases the variation in effective moisture, with north-facing slopes being cooler and effectively wetter and south-facing ones warmer and drier. Rock outcrops are also sites where there is less competition from grasses, and the topographic relief can provide natural firebreaks. This may reduce the frequency of grassland fire to give a long-enough interval to allow some seedlings to survive through their early, fire-susceptible years. Substrate conspicuously affects ponderosa pine distribution along the transition between the Rocky Mountains and the Great Plains in eastern Colorado, Wyoming and Montana. In these areas, ponderosa pine occurs on outcrops or areas of increased topographic relief that are often no higher than nearby prairie-covered plains. Specific examples include the Missouri River corridor in central Montana and the Hartville structural uplift in southeastern Wyoming, which is up to 500 m lower than adjacent shortgrass prairie to the south. We expected that soil depth might account for this effect, but it was absent from the final classification tree.

Rejected variables

There were several variables that were included in the modelling process but were not ultimately included as nodes in the classification tree, such as depth to bedrock (USDA-United States Department of Agriculture, 1993) and frequency-based variables such as the standard deviation of precipitation and maximum precipitation in a 50-year period (Table 2). The frequency-based variables were considered because of the ecology of ponderosa pine. Although ponderosa pine occurs in relatively dry areas when compared with most other western conifers, it reproduces best in the southwestern USA during unusually wet years (Pearson, 1950; Savage *et al.*, 1996). To allow for this, variables were included representing variability in precipitation that might be necessary for reproduction. To do this, precipitation estimates were used for individual months of every individual year from 1948 to 1997 (Daly & Taylor, 1998). These estimates were used to calculate the standard deviation of summer (June + July + August) precipitation as a percentage of average summer precipitation, and the maximum summer precipitation in that 49-year period. Although PRISM estimates of individual months are available from 1896 to 1997, pre-1948 data were not used because the period of record of the network of weather station data from which the PRISM data were derived was described in metadata as 'generally 1948 to present'. None of these variables was identified as a decision point in the classification tree model.

Collinearity and spatial autocorrelation

Biogeographical models are complicated by multicollinear climatic variables, spatial autocorrelation, and the coarse scale and inaccuracies of the source data. Collinearity does not alter the accuracy of the predictions of these models, but can make it difficult or impossible to determine the separate effect of a variable from other collinear variables (Belsley *et al.*, 1980). In some model types such as regression models, even the direction of effect of a multicollinear variable may not be reliably determined. A common test for collinearity is to calculate the variance inflation factor (VIF) for the variable matrix (Belsley *et al.*, 1980), with a typical cutoff for collinearity being a VIF > 10. Collinearity was tested for by calculating the VIF for the predictor variables. Successive calculations of the VIF with different combinations of variables indicated collinearity between growing season precipitation and the ratio of summer precipitation to temperature, and between frost-free period and April and July temperatures. Because of this collinearity, the effect represented by one variable may be indistinguishable from the effect of any variable collinear with it.

Climate data are inherently spatially autocorrelated, and violate the assumption of independence used to calculate error estimates for a model. A resulting model may track spatial patterns in addition to the interaction of dependent and independent variables. A common index of spatial autocorrelation is Moran's I statistic, for which 0 indicates no spatial autocorrelation, -1 indicates strong negative autocorrelation, and $+1$ indicates strong positive spatial autocorrelation. Moran's I statistic was calculated for the binary residuals of the classification tree model using distance-weighting with multiple distance thresholds. It had a maximum value of 0.348, indicating that spatial autocorrelation of the residuals was present but not severe. The classification tree model input was a 10,000-point stratified subsample from a starting data set of 163,544 points, or *c.* 1 in every 16 points. There is stronger spatial autocorrelation for many of the variables in the full 163,544-point data set. As a result we cannot assume that model error will be geographically uniform.

Classification tree methods

The classification tree was created using the statistical package S-Plus 2000 Release 3 (Insightful Corporation, Seattle, WA, USA). The tree separates input points using true/false tests that can be either greater than/less than or categorical. If allowed, the tree continues to be expanded until the terminal 'leaves' achieve 100% accuracy. To avoid overfitting, the tree was pruned to 25 leaves (Fig. 4) using a minimum-deviance pruning technique (Breiman *et al.*, 1984). Twelve of the leaves predicted ponderosa pine presence and thirteen predicted ponderosa pine absence. The tree correctly classified 94.9% of the 10,000 input points. Of the 4000 input 'ponderosa-present' points, only 171 were misclassified as ponderosa-absent, but, according to GAP accuracy statements mentioned earlier, at least 20% or 800 of the input points representing ponderosa presence are likely to be in error. The errors in ponderosa pine absence data are believed to be small, and so some ponderosa-present leaves may have very small errors. The effects of large errors in ponderosa pine presence data on a classification tree take two forms: they lower the accuracy (percentage correct assignment) for leaves that include many misclassified points, making it difficult to identify whether such leaves represent true patterns in nature or are artefacts of the high error rate in the source data, and numerous small leaves may be created for which decision points are influenced by chance associations. To be conservative, all small leaves were removed as well as any

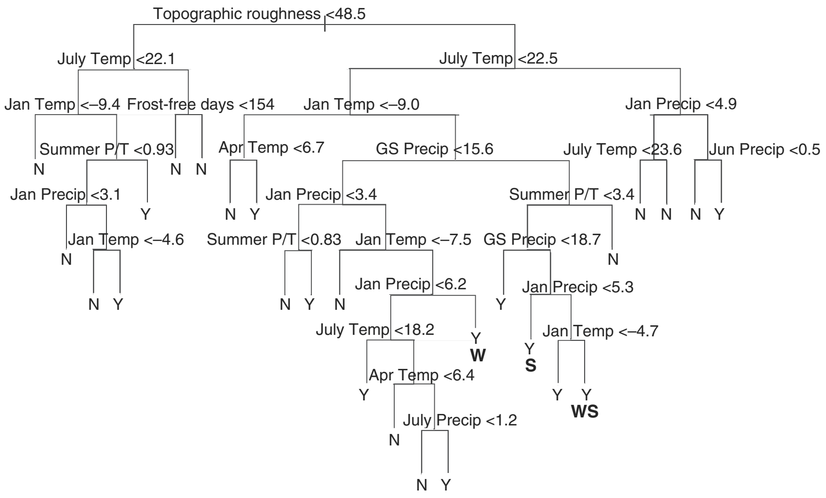

Figure 4 Classification tree for ponderosa pine (*Pinus ponderosa* Douglas ex Lawson & C. Lawson), where each branch is defined by a 'less than' condition. If the less-than statement is true, follow the left branch; if false, follow the right branch. 'Y'/'N' indicate that ponderosa presence/absence is predicted. All precipitation units are cm; temperature units are °C; topographic roughness is in metres and is described further in the section on methods. The leaves W, S, and WS met size and accuracy thresholds. The three leaves are similar except that leaf W is defined by high winter and low summer moisture, leaf S is defined by high summer and low winter moisture, and leaf WS is defined by high winter and summer moisture.

Table 3 Statistics for each leaf of the classification tree model. Leaves match those shown in Fig. 4 and are numbered from left to right. Leaves shown in bold met size and accuracy standards and are discussed in the text

Number	Terminal node	Number of values	Predicted value	Per cent correct
1	January temperature < −9.4	522	No	99.4
2	Jan precipitation < 3.1	233	No	99.1
3	January temperature < −4.6	25	No	84
4	January temperature > −4.6	124	Yes	86.3
5	Summer precipitation/temperature > 0.93	192	Yes	87.5
6	Frost-free days < 154	211	No	93.8
7	Frost-free days > 154	3319	No	99.8
8	April temperature < 6.7	213	No	93
9	April temperature > 6.7	19	Yes	68.4
10	Summer precipitation/temperature < 0.83	337	No	90.2
11	Summer precipitation/temperature > 0.83	140	Yes	58.6
12	January temperature < −7.5	70	No	84.3
13	July temperature < 18.2	239	Yes	87
14	April temperature < 6.4	85	No	75.3
15	July precipitation < 1.2	38	No	71.1
16	July precipitation > 1.2	95	Yes	94.7
17	**Jan precipitation > 6.2**	**771**	**Yes (W)**	**94**
18	GS precipitation < 18.7	565	Yes	86.7
19	**Jan precipitation < 5.3**	**1205**	**Yes (S)**	**98.2**
20	January temperature < −4.7	119	Yes	76.5
21	**January temperature > −4.7**	**564**	**Yes (WS)**	**95.9**
22	Summer precipitation/temperature > 3.4	96	No	79.2
23	July temperature < 23.6	129	No	83
24	July temperature > 23.6	528	No	98.7
25	June precipitation < 0.5	25	No	88
26	June precipitation > 0.5	136	Yes	96.3

large leaves with errors of greater than 10% (Table 3). The seven smallest leaves, which accounted for 817 of 4000 points, were removed. Of the remaining five leaves, two were removed that incorrectly classified more than 10% of the input points.

Minimum-volume ellipsoid methods

The minimum-volume ellipsoid describes an area in multidimensional climate and edaphic space in which each variety of ponderosa pine is able to grow. The source data points consisted only of ponderosa-present points. This model is resistant to spatial distributional bias, because the shape of the ellipsoid does not depend on the density of points at a particular location. Two separate ellipsoids were created, one for each variety, using a data set of 514 points for var. *scopulorum* and of 1222 points for var. *ponderosa*. The same ten variables were used, but the five precipitation-related variables were log-transformed and all the variables were then rescaled to a common 0–1000 scale. The matrix describing the approximate minimum-volume ellipsoid was calculated using a FORTRAN program modified from Hodge & Tasker (1995), which is based on an algorithm described in Weisburg (1980). To determine if a point fell within the ellipsoid required evaluation of the equation of the ellipsoid:

$$c = (\mathbf{x} - \mathbf{m})\mathbf{M}^{-1}(\mathbf{x} - \mathbf{m})^{\mathrm{T}}.$$

In the case of the 10-dimensional ellipsoid, $\mathbf{x}$ was the row vector of the 10 coordinates of the data point to be tested, $\mathbf{M}$ was a symmetric positive-definite matrix of dimension 10 × 10, $\mathbf{m}$ was a row vector with 10 values, and c was a constant. If evaluation of the right side of the equation results in a value greater (smaller) than c, then the data point falls outside (inside) the ellipsoid. To mitigate the effects of inaccuracies in the source data, the ellipsoid was restricted by reducing the outer boundary of the ellipsoid in order to better reflect the central portion of the niche. The reduction of the ellipsoid was accomplished by reducing the size of the constant c, which behaves similarly to the radius of a circle, in that increasing (reducing) the size of c will increase (reduce) the size of the ellipsoid. The amount of the reduction was arbitrary, and was done to optimize visually the match between the predicted distribution and the actual distribution. Points across the western USA were then tested to determine if they fell within the ellipsoid. If a point fell outside, the distance to the ellipsoid was calculated in each dimension, and reported as percentages of the total range of values for each variable (Fig. 2c).

RESULTS

Classification tree

The input to the classification tree model consisted of ponderosa presence/absence data without regard to variety. The resulting classification tree contained twelve leaves, or pathways, of predicted presence of the species (Fig. 4). Because of the potential errors in the source data (see Methods: Model Selection), seven of the smallest leaves were discarded as well as two larger leaves with low prediction accuracy. The remaining three leaves were tested against the entire 163,544-point data set, and had errors of commission (false positives) of 23% (leaf W) 10% (leaf S) and 17% (leaf WS). One leaf (S) corresponds to the distribution of var. *scopulorum* (compare Figs 1 & 5). The other two leaves correspond to the distribution of var. *ponderosa*, although conditions matching leaf WS also occur in northern Arizona and western Colorado, within the range of var. *scopulorum* (compare Figs 1 & 5).

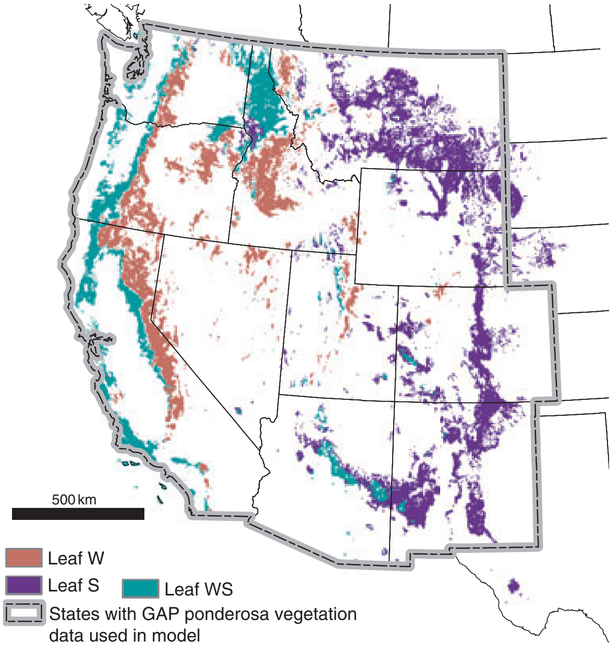

Figure 5 Distribution of habitat of classification tree leaves W, S, and WS (see Fig. 4 for definition of leaves).

The three pathways match a large area of ponderosa pine distribution with a simple tree structure (Figs 4 & 5). All three pathways share decision points for topographic roughness, July temperature (< 22.5 °C), and January temperature (> −9 °C). The leaves vary substantially in the minimum or maximum amount of summer or winter precipitation, and have been designated to indicate predominant precipitation seasonality (S for summer; W for winter). The *scopulorum* leaf (leaf S) is defined by high growing-season precipitation and low January precipitation (Fig. 4). The two var. *ponderosa* leaves (W and WS) have high January precipitation. However, leaves WS and W differ in expression of growing-season precipitation: leaf WS is characterized by higher growing-season precipitation (> 18.7 cm) than leaf W (< 15.6 cm). The distribution of leaf WS is spread across three areas, one west of the crest of the Sierra Nevada and Cascades, a second in northern Idaho and adjacent Washington, and a third on the Mogollon Rim in Arizona (Fig. 5). Three climatic patterns explain this distribution. The first area (coastal ranges, west slopes of the Cascades and Sierra Nevada) receives little June precipitation and is extremely dry in July and August. However, heavy March–April–May precipitation, coinciding with the start of the growing season, results in high growing-season precipitation. The northern Washington and Idaho areas receive significant summer (June–July–August) precipitation, although the precipitation regime is still winter-dominated. In contrast, the Mogollon rim area differs from other WS areas in receiving substantial July and August precipitation from monsoonal circulation and generally experiencing dry conditions during the early summer (May/June). Like the W and other WS regions, however, it also receives high winter precipitation.

Ellipsoid

The climatic controls on ponderosa pine distribution are apparent from the mapped distances from the ellipsoid niche (Fig. 6). The differences between the maps for var. *ponderosa* and var. *scopulorum* are revealing. The temperature maps are similar, with the patterns nearly matching in most areas. High mountain ranges are too cold for either variety of ponderosa pine. July temperatures exceed the realized niches of both varieties in the low-elevation deserts of the southwestern USA and across much of the Great Plains. The realized niche of var. *scopulorum* may be less restricted by extreme temperatures in both summer and winter. The frost-free-period maps are also similar for the two varieties, probably as a result of the wide range of frost-free periods tolerated by the two varieties.

In contrast, the precipitation maps show large differences between the varieties, with var. *ponderosa* limited by low January precipitation across most areas occupied by var. *scopulorum*, and var. *scopulorum* limited by low growing-season precipitation across most of the area occupied by var. *ponderosa* (Fig. 6). These winter dry–summer dry conditions overlap across much of the Great Basin and interior mountain ranges in Utah, Idaho, Wyoming, and parts of Colorado, all of which lie in the rain shadows of the Sierra Nevada (to the west) and the eastern ranges of the Rocky Mountains (to the east). Thus, both summer and winter precipitation is deficient in these areas. For var. *scopulorum*, however, the growing-season precipitation deficit can also reflect a shortened growing season rather than a lack of June and July precipitation. Such is the case for the Yellowstone Plateau, the Tetons, and the mountain ranges of north-central Colorado (Fig. 6a). However, in much of the Great Basin and nearby Wind River Mountains in central Wyoming, July precipitation is insufficient at low elevations with long growing seasons.

DISCUSSION

The classification tree model created leaves whose geographic distributions generally match the respective distributions of var. *ponderosa* and var. *scopulorum*, supporting the hypothesis that the varieties have distinct environmental niches (Fig. 5). Both the classification tree and ellipsoid model results suggest that the primary difference involves seasonality of precipitation. The results of the classification tree model corroborate the morpho-taxonomic distinctions noted in early studies of ponderosa pine (Sudworth, 1908; Korstian, 1924; Weidman, 1939). The differences between varieties were noted initially in naturally occurring trees that had different needle lengths and cone sizes (Sudworth, 1917; Korstian, 1924). Common-garden studies established the genetic basis for these differences, as well as for differences in pollen-cone colour, hypoderm thickness, frost tolerance, drought tolerance of seedlings and seeds, and morphological features such as growth rate and the age of formation of terminal buds (Korstian, 1924; Squillace & Silen, 1962; Conkle & Critchfield, 1988). Molecular genetic research has also demonstrated differences between the varieties of ponderosa pine. Pinaceae exhibit paternal inheritance of chloroplast DNA (cpDNA) and maternal inheritance of mitochondrial DNA (mtDNA), making it possible to consider the effects of dispersal by both pollen and seeds. Different haplotypes have been found both within and between varieties for cpDNA and mtDNA (Latta & Mitton, 1999). Johansen & Latta (2003) found a very steep cline (< 10 km) in mtDNA haplotypes between var. *ponderosa* and var. *scopulorum* in Montana.

Before our study, the location of the boundary between the varieties, including the limited zone of introgression in western Montana, might have been attributed to competitive exclusion. The coincidence between the modelled distribution of the varieties and their actual distribution suggests that the boundary is a consequence of the geographic expression of the niches of the respective varieties. The importance of precipitation seasonality is corroborated by the steep gradient from strongly summer-dominated to strongly winter-dominated precipitation in west-central Montana (Fig. 7). Depending on the mechanism of genetic–ecological selection, these ecological differences may slow dispersal and gene flow between the varieties near the hybrid zone. The two varieties may have coincidentally migrated to the current boundary at the same time, but our results indicate that the boundary is controlled by the respective responses of the two varieties to the prevailing precipitation regime.

The minimum-volume ellipsoid model demonstrated that the ecological niches of the varieties are similar in temperature, but differ in seasonality of precipitation. Variety *ponderosa* requires substantial winter precipitation, whereas var. *scopulorum* requires substantial summer precipitation. In the centre of the species' distribution, both types of precipitation are deficient. Across much of the western interior, this precipitation dependence itself rests on the larger controls of topography, because the precipitation–temperature–elevation relationship changes depending on the position of mountains or valleys relative to latitude, to oceanic moisture sources, and to other mountains that create rain shadows.

The precipitation–temperature–elevation dependence was explored further by graphing the relationship between average summer temperature and average summer precipitation for several interior mountain ranges and the Mogollon Rim of Arizona (Fig. 8). In general, on any single mountain range, canyon, or ridge, summer temperature and summer precipitation are inversely related. However, individual mountain ranges clearly differ in the slope, y-intercept, and scatter of the temperature/precipitation relationship (Fig. 8). In particular, systematic differences in the y-intercept (i.e. the position of the point-cloud relative to the y-axis) occur among ranges. For example, mountain ranges traversed by moist air masses in summer (e.g. Laramie Range, Mogollon Rim) receive precipitation after the air rises and cools only slightly. In contrast, air must rise and cool substantially before precipitation falls on mountain ranges exposed to air masses already depleted in

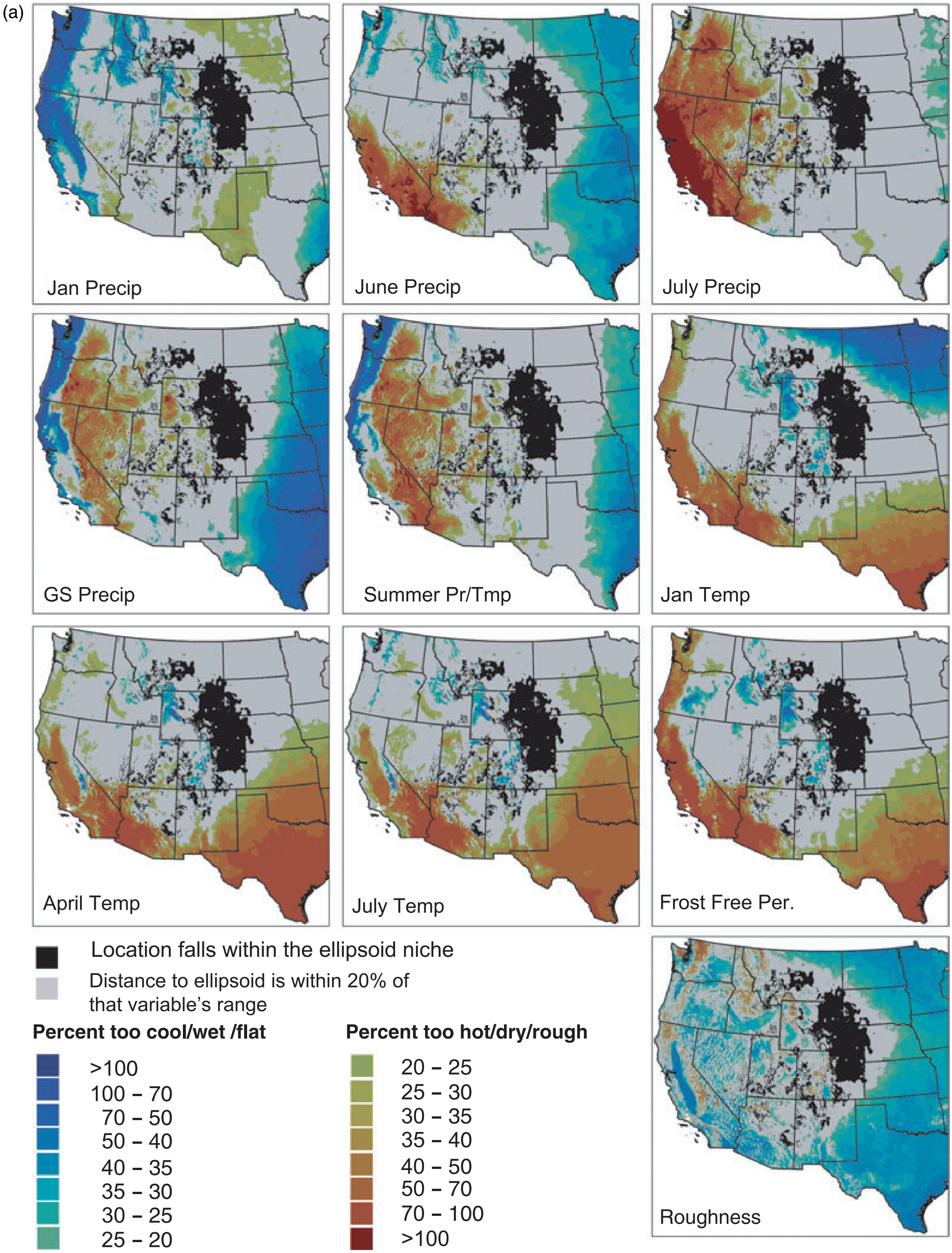

Figure 6 Calculated distances to the ellipsoid niche of var. *scopulorum* (a) and var. *ponderosa* (b) expressed as a percentage of the total range for each of 10 variables (see Fig. 2). The varieties have very similar maps for temperature variables, but show seasonal differences for precipitation variables.

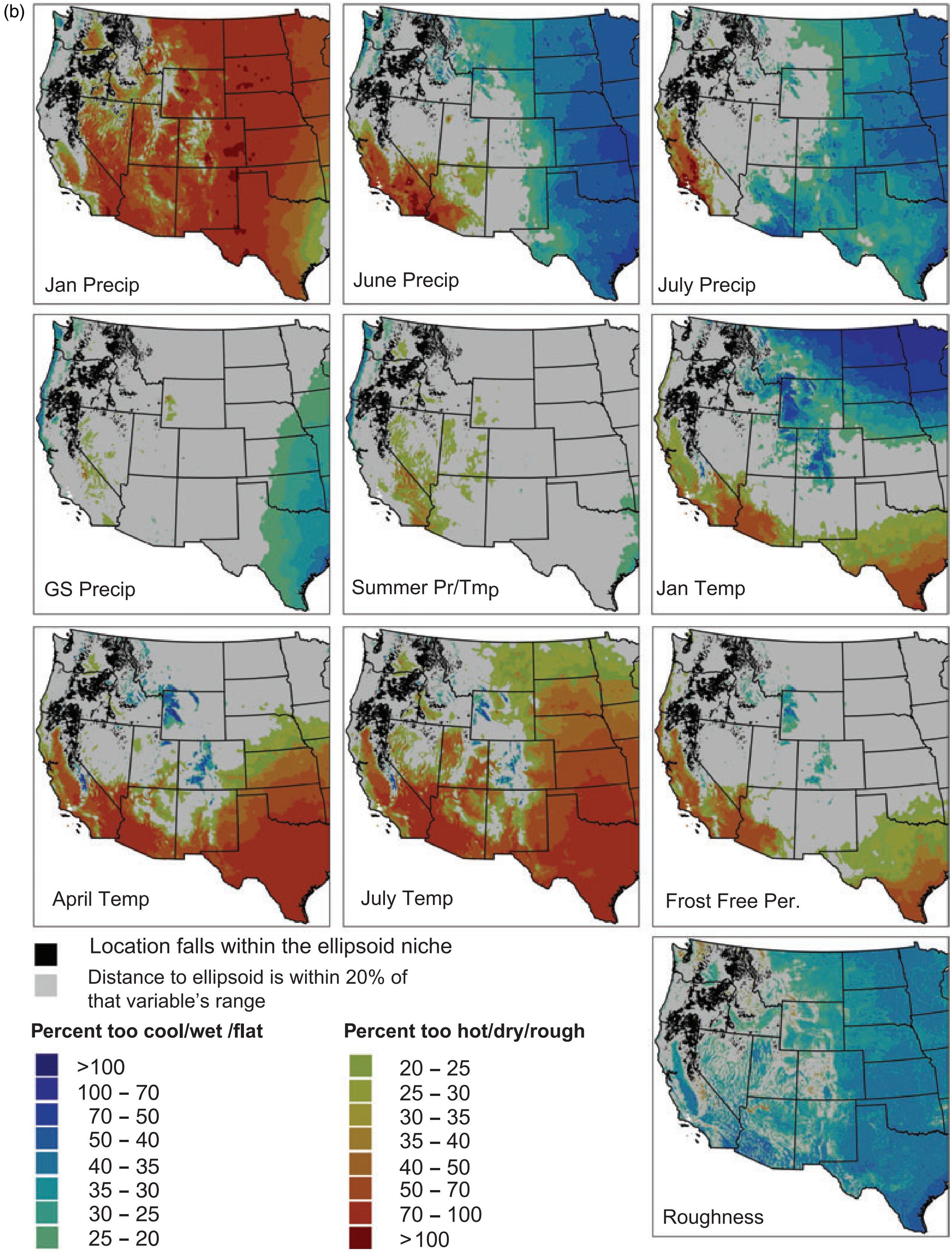

Figure 6 *continued*

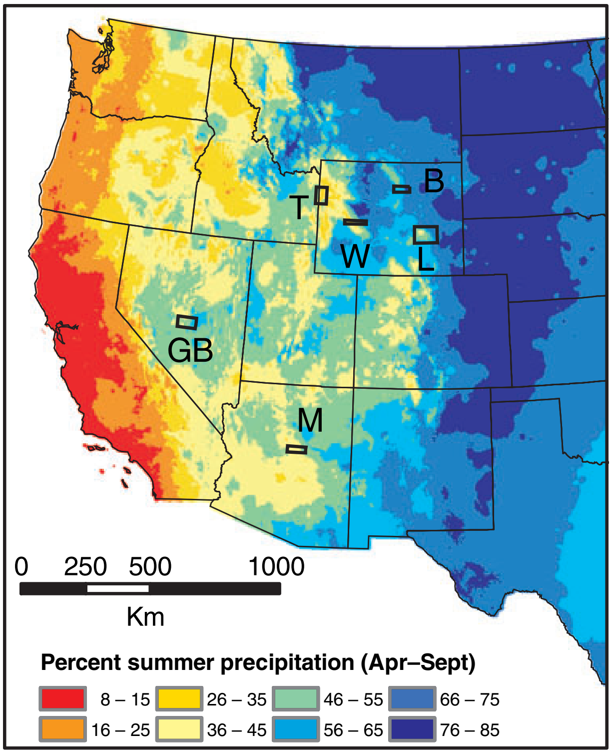

Figure 7 Average percentage summer precipitation (calculated as April–September) in the western United States using *c.* 4-km resolution PRISM climate data (Daly *et al.*, 1997). Variety *scopulorum* (*ponderosa*) grows largely in areas dominated by summer(winter) precipitation. Boxes delineate areas graphed in climate space in Fig. 8. Abbreviations are: B, Bighorn Mountains; T, Teton Mountains; W, Wind River Range; L, Laramie Range; M, Mogollon Rim; GB, Great Basin Ranges.

moisture (e.g. Great Basin ranges). Thus, specific precipitation values are associated with colder temperatures in the Great Basin than in the Laramie Range, and conversely warmer temperatures are associated with higher precipitation in the Laramie Range than in the Great Basin (Fig. 8). When graphed, the depletion of moisture from an air mass results in the downward shift in point clouds, and hence a lower temperature for a given amount of precipitation. When mountain ranges that support var. *scopulorum* are graphed against mountain ranges that do not support var. *scopulorum*, the latter are generally shifted downwards, indicating that less moisture is available at suitable temperatures than is available in mountain ranges that support var. *scopulorum*. Exceptions include the Tetons and Yellowstone Plateau, which receive plenty of growing-season precipitation but cool temperatures and a short growing season exclude ponderosa pine – these ranges lack low-elevation sites with warm temperatures (Figs 6a & 8). It is possible that the Teton and Yellowstone regions could support ponderosa pine forests if temperatures increase. In most other areas, such as the Great Basin, an increase in temperature alone would be unlikely to allow ponderosa growth unless there were also an increase in the moisture content of the air masses reaching those regions. This relationship between temperature and moisture is only a part of the overall niche, but it may play a prominent role in explaining the absence of var. *scopulorum* across many interior mountain ranges.

Ponderosa distributions in past climates

Biogeographical studies of the palaeodistribution and migration of ponderosa pine may benefit from an understanding

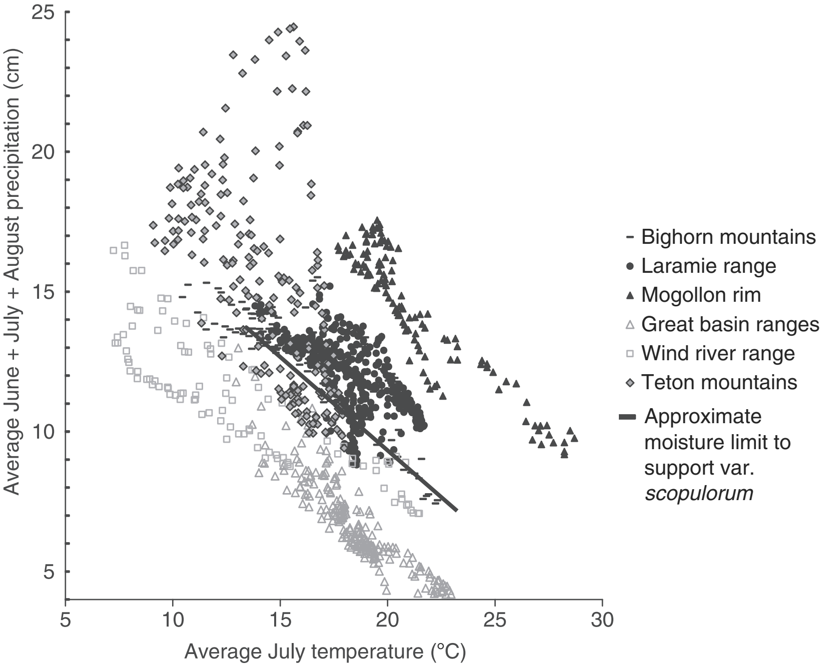

Figure 8 Relationship of temperature to moisture using *c.* 4-km resolution PRISM climate data (Daly *et al.*, 1997) for selected areas in the western United States; see Fig. 7 for locations. Solid black symbols are used for ranges that have ponderosa pine present, grey symbols are used for ranges where ponderosa pine is absent. The Teton Mountain Range (black outline, grey fill) is conspicuous because, although ponderosa is absent, the mountain range receives substantial summer moisture. The lack of low elevation–warm temperature montane sites may exclude ponderosa pine from the Tetons. A line is shown for the approximate temperature–precipitation threshold that distinguishes mountain ranges that support var. *scopulorum* from those that do not.

of the modern realized niche. In particular, ponderosa pine distribution during the last glacial maximum (*c.* 25,000–15,000 yr BP) may be explored in the context of the modelled niche space of var. *ponderosa* and var. *scopulorum*. Because ponderosa pine pollen is difficult to distinguish reliably from that of other members of *Pinus* subgenus *Pinus* (e.g. *P. contorta*), most of the evidence of past distribution is limited to plant macrofossils. The richest source of glacial macrofossils is the woodrat midden record from semi-arid, rocky areas that would have been wetter during the glacial period. Interestingly, the midden record indicates that neither variety of ponderosa pine occupied the Great Basin in either the past glacial or the present interglacial, and that most of the current distribution of var. *ponderosa* in the central Rockies was achieved by northward migration during the Holocene.

Variety ponderosa*, Late Pleistocene–Holocene glacial biogeography*

Records of var. *ponderosa* are scarce for the last glacial period. Most fossil records that extend to the last glacial maximum are not likely to contain var. *ponderosa* macrofossils because of their location in wetlands, large lake basins (Owens Lake, Pyramid Lake), or altitudes that would have been subalpine or alpine during the last glacial. In reviewing the literature, we found only two full-glacial macrofossil records with possible var. *ponderosa* macrofossils. A seed identified as either ponderosa or lodgepole pine (*P. contorta*) was reported from glacial-age sediments of Battle Ground Lake in Washington (Barnosky, 1983; North American Plant Macrofossil Database, 2004). Needles identified as ponderosa pine/Jeffrey pine complex (*Pinus ponderosa/jeffreyi*) were reported from a glacial-age woodrat midden from Kings Canyon, California (Cole, 1983). The search for more glacial evidence of var. *ponderosa* could be narrowed substantially by focusing on macrofossil-yielding deposits (small lakes or woodrat middens) at elevations that would have been climatically suitable. The minimum-volume ellipsoid and classification tree models suggest that var. *ponderosa* establishment and survival are dependent on winter moisture, and on the absence of extremely cold temperatures. Vegetation reconstructions and some climate models for the last glacial maximum in the Pacific Northwest suggest conditions colder and drier than today (Thompson *et al.*, 1993; Bartlein *et al.*, 1998). Thompson *et al.* (1993) have suggested that the apparent decrease in winter precipitation at *c.* 21,000 yr BP may have resulted from a glacial anticyclone causing a pronounced weakening of westerlies or even seasonal reversal in wind and weather tracks in the Pacific Northwest. Palaeowind evidence in the form of sand and loess deposits does not support the existence of a glacial anticyclone in the Great Plains (Muhs & Bettis, 2000). However, Sweeney *et al.* (2004) found evidence supporting the presence of a glacial anticyclone that weakened southwesterly flow in the Pacific Northwest, creating drier conditions there between 35,000 and 15,000 yr BP. Depending on the seasonal and geographic extent of this weakening, much of the northwestern USA could have been unsuitable for var. *ponderosa* owing to a combination of decreased temperature and decreased precipitation. Therefore the most promising locations for glacial-age macrofossils of var. *ponderosa* may be along the coastal margins of western North America, where temperatures are moderated and even weak westerlies might supply moisture. Although conventional lake sites may be scarce in the coastal ranges, ponderosa pine macrofossils may be preserved in small basins or in alluvial and debris flow deposits such as those now being investigated in several western states (Meyer & Pierce, 2003). The success of such macrofossil studies may be enhanced by advances in wood identification, including extraction and sequencing of DNA from ray parenchyma (Dumolin-Lapègue *et al.*, 1999; Deguilloux *et al.*, 2002).

Variety scopulorum*, Late Pleistocene–Holocene glacial biogeography*

The glacial record of var. *scopulorum* is much better defined than that of var. *ponderosa*. Extensive glacial-age macrofossil records exist for low- and mid-elevation sites across much of New Mexico, Arizona, Utah, Nevada, and parts of Colorado (Betancourt, 1990; North American Packrat Midden Database, 2004; North American Plant Macrofossil Database, 2004). Ponderosa pine needles have only been identified from glacial-age middens in the San Andres Mountains of southern New Mexico (33.18° N, 106.60° W, 1700 m) and the Santa Catalina Mountains of southern Arizona. (32.35° N, 110.88° W, 1463 m). Ponderosa pine is absent from glacial-age records north of *c.* 33.5° latitude, despite the presence in those records of many species associated with ponderosa pine today (e.g. Douglas fir, *Pseudotsuga menziesii*, and Rocky Mountain juniper, *J. scopulorum*) (Anderson, 1989; Betancourt, 1990; North American Packrat Midden Database, 2004).

Two possible reasons for this restriction are suggested by the modelled niche of var. *scopulorum*. Variety *scopulorum* today requires significant growing-season precipitation, but may also be restricted by low temperatures from mountain ranges that effectively have no lower montane zone with mild temperatures. During the glacial period, the influence of the summer monsoon was apparently restricted to < 1800 m in southern Arizona/southern New Mexico and points south, based on the relative importance of C_4 species in plant macrofossil assemblages (Betancourt, 1990; Holmgren *et al.*, 2003) and on carbon isotope ratios from soil carbonates and megaherbivore teeth (Connin *et al.*, 1998). In addition to summer dryness, cooler temperatures might have shifted temperature belts far enough downslope to eliminate the lower montane zone in the central Rockies of Colorado, Utah, Wyoming, and Montana (Jackson *et al.*, 2005). Despite having greater effective moisture than today (Thompson *et al.*, 1993), glacial climates were clearly unfavourable for var. *ponderosa* north of 36° N.

CONCLUSION

The classification tree model demonstrates that *Pinus ponderosa* var. *ponderosa* and *P. ponderosa* var. *scopulorum* are ecologically as well as genetically distinct entities. The model differentiation is based largely on seasonal differences in the timing of precipitation associated with the modern range of each variety. A closer examination of the niche of each variety using the minimum-volume ellipsoid model highlights the importance of precipitation seasonality. The location of the current hybrid zone between var. *ponderosa* and var. *scopulorum* is controlled by the geographic shift from summer-dominant to winter-dominant precipitation.

The restriction of var. *scopulorum* to southern New Mexico and Arizona during the last glacial period suggests that much of the central and southern Rockies and the Colorado Plateau were either too cold or had insufficient growing-season precipitation to support var. *scopulorum* populations. In the central Rockies, a relatively small decrease in temperature during the last glacial could have shifted the foothills of the central Rockies into the cooler temperature regimes today experienced by the Tetons and Yellowstone Plateau, forcing ponderosa pine completely off the mountain flanks. In the southern Rockies and the Colorado Plateau, however, a more substantial temperature shift would be required. Decreased summer precipitation rather than cooler temperatures may be responsible for the glacial-age absence of ponderosa pine. Variety *ponderosa* has a very limited palaeorecord, but was likely to be less severely restricted by climate than var. *scopulorum* during the last glacial, with potential for a moderating influence of maritime climates as well as significant moisture in coastal areas (Bartlein *et al.*, 1998). Determining the extent of var. *ponderosa* during the last glacial will require additional macrofossil records, especially from coastal areas.

A number of issues remain to be resolved. If adaptations to precipitation seasonality occurred in response to selection, then the Quaternary and perhaps pre-Quaternary history of precipitation seasonality of western North America may have contributed to the development of the varieties. Did ancestral ponderosa pine evolve in habitats that were both summer and winter wet, and then split into lineages adapted to either summer dry or winter dry conditions? Where, when and in what order did these adaptations occur? Most of the Quaternary has not been spent in interglacial conditions: did ponderosa pine differentiation occur during glacial or interglacial times, or did it occur before the Quaternary?

If moisture transport patterns are held constant, then increasing summer temperatures in the future should produce changes in the distribution of var. *scopulorum*. All else held constant, increased temperatures should lead to range expansions in the central and northern Rockies as the foothills in wet mountain ranges that are currently too cold for ponderosa pine become warmer. The future of ponderosa pine distribution will depend on the response of each variety to changing climatic conditions. In order to predict accurately the response of ponderosa pine and of other species it may be necessary to consider the responses of ecological varieties separately.

ACKNOWLEDGEMENTS

This study and manuscript benefited from the comments and suggestions of William Baker, Bart Geertz, Brian Huntley, Jon Mason, and an anonymous referee. This work was supported by grants to the University of Wyoming from the National Science Foundation and the Wyoming Agricultural Experiment Station (McIntyre-Stennis funds).

REFERENCES

Anderson, R.S. (1989) *Development of the southwestern ponderosa pine forests: what do we really know? Multiresource management of ponderosa pine forests*. General Technical Report RM-185. USDA Forest Service, Fort Collins, CO.

Anderson, R.P., Lew, D. & Peterson, A.T. (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*, **162,** 211–232.

Barnosky, C.W. (1983) Late-Quaternary vegetational and climatic history of southwestern Washington. PhD dissertation, University of Washington, Seattle, WA.

Bartlein, P.J., Anderson, K.H., Anderson, P.M., Edwards, M.E., Mock, C.J., Thompson, R.S., Webb, R.S., Webb, T., III & Whitlock, C. (1998) Paleoclimate simulations for North America over the past 21,000 years: features of the simulated climate and comparisons with paleoenvironmental data. *Quaternary Science Reviews*, **17,** 549–585.

Belsley, D.A., Kuh, E. & Welsch, R.E. (1980) *Regression diagnostics: identifying influential data and sources of collinearity*. Wiley, New York.

Betancourt, J.L. (1990) Late Quaternary biogeography of the Colorado Plateau. *Packrat middens: the last 40,000 years of biotic change* (ed. by J.L. Betancourt, T.R. Van Devender and P.S. Martin), pp. 259–292. University of Arizona Press, Tucson, AZ.

Betancourt, J.L., Van Devender, T.R. & Martin, P.S. (1990) Synthesis and prospectus. *Packrat middens: the last 40,000 years of biotic change* (ed. by J.L. Betancourt, T.R. Van Devender and P.S. Martin), pp. 435–447. University of Arizona Press, Tucson, AZ.

Betancourt, J.L., Schuster, W.S., Mitton, J.B. & Anderson, R.S. (1991) Fossil and genetic history of a pinyon pine (*Pinus edulis*) isolate. *Ecology*, **72,** 1685–1697.

Box, E.O. (1981) The ecological model: life form limitation, cover and dominance. *Macroclimate and plant forms: an introduction to predictive modelling in phytogeography*, pp. 33–46. Dr W. Junk Publishers, The Hague, The Netherlands.

Breiman, L., Friedman, R.A., Olshen, R.A. & Stone, C.J. (1984) *Classification and regression trees*. The Wadsworth statistics/probability series 26, Wadsworth International Group, Belmont, CA.

Cole, K.L. (1983) Late Pleistocene vegetation of Kings Canyon, Sierra Nevada, California. *Quaternary Research*, **19,** 117–129.

Conkle, M.T. & Critchfield, W.B. (1988) Genetic variation and hybridization of ponderosa pine. *Ponderosa pine: the species and its management* (ed. by D.M. Baumgartner and J.E. Lotan), pp. 27–43. Washington State University Cooperative Extension, Pullman, WA.

Connin, S.L., Betancourt, J.L. & Quade, J. (1998) Late Pleistocene C_4 plant dominance and summer rainfall in the southwestern U.S.A. from isotopic study of herbivore teeth. *Quaternary Research*, **50,** 179–193.

Cregg, B.M., Olivas-García, J.M. & Hennessey, T.C. (2000) Provenance variation in carbon isotope discrimination of mature ponderosa pine trees at two locations in the Great Plains. *Canadian Journal of Forest Research*, **30,** 428–439.

Daly, C. & Taylor, G.H. (1998) *102-year high-resolution climate data set for the conterminous United States*. Oregon Climate Service, Oregon State University, Corvallis, OR.

Daly, C., Taylor, G.H. & Gibson, W.P. (1997) The PRISM approach to mapping precipitation and temperature. *Proceedings of the 10th Conference on Applied Climatology*, pp. 10–12. American Meteorological Society, Boston, MA.

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

Deguilloux, M., Pemonge, M. & Petit, R.J. (2002) Novel perspectives in wood certification and forensics: dry wood as a source of DNA. *Proceedings of the Royal Society of London Series B, Biological Sciences*, **269,** 1039–1046.

Dumolin-Lapègue, S., Pemonge, M.H., Gielly, L., Taberlet, P. & Petit, R.J. (1999) Amplification of oak DNA from ancient and modern wood. *Molecular Ecology*, **8,** 2137–2140.

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

GLOBE Task Team and others (1999) *The Global Land One-kilometer Base Elevation (GLOBE) digital elevation model, Version 1.0* (ed. by D.A. Hastings, P.K. Dunbar, G.M. Elphingstone, M. Bootz, H. Murakami, H. Maruyama, H. Masaharu, P. Holland, J. Payne, N.A. Bryant, T.L. Logan, J.P. Muller, G. Schreier and J.S. MacDonald), National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, CO. Digital database on the World Wide Web (URL: http://www.ngdc.noaa.gov/seg/topo/globe.shtml).

Guisan, A. & Theurillat, J.P. (2000) Equilibrium modelling of alpine plant distribution and climate change: how far can we go? *Phytocoenologia*, **30,** 353–384.

Guisan, A., Weiss, S.B. & Weiss, A.D. (1999) GLM versus CCA spatial modelling of plant species distribution. *Plant Ecology*, **143,** 107–122.

Hodge, S.A. & Tasker, G.D. (1995) *Magnitude and frequency of floods in Arkansas*. US Geological Survey Water Resources Investigations Report, 95-4224, Little Rock, AR.

Holmgren, C.A., Peñalba, M.C., Aasen Rylander, K. & Betancourt, J.L. (2003) A 16,000 ^{14}C yr BP packrat midden series from the U.S.A.-Mexico borderlands. *Quaternary Research*, **60,** 319–329.

Jackson, S.T., Betancourt, J.L., Lyford, M.E. & Aasen Rylander, K. (2005) A 40,000-year woodrat-midden record of vegetational and biogeographical dynamics in north-eastern Utah, USA. *Journal of Biogeography*, **32,** 1085–1106.

Johansen, A.D. & Latta, R.G. (2003) Mitochondrial haplotype distribution, seed dispersal and patterns of postglacial expansion of ponderosa pine. *Molecular Ecology*, **12,** 293–298.

Knight, D.H. (1994) *Mountains and plains: the ecology of Wyoming landscapes*. Yale University Press, New Haven, CT.

Korstian, C.F. (1924) A silvical comparison of the Pacific Coast and Rocky Mountain forms of Western Yellow Pine. *American Journal of Botany*, **11,** 318–324.

Larsen, D.R. & Speckman, P.L. (2004) Multivariate regression trees for analysis of abundance data. *Biometrics*, **60,** 543–549.

Latta, R.G. & Mitton, J.B. (1999) Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. *Evolution*, **53,** 769–776.

Little, E.L., Jr (1971) *Atlas of United States trees, volume 1, conifers and important hardwoods*. U.S. Department of Agriculture Miscellaneous Publication 1146, Washington, DC.

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

Merrill, E.H., Kohley, T.W., Herdendorf, M.E., Reiners, W.A., Driese, K.L., Marrs, R.W. & Anderson, S.H. (1996) *The Wyoming Gap analysis project final report*. Wyoming Cooperative Fish and Wildlife Research Unit, University of Wyoming, Laramie, WY.

Meyer, G.A. & Pierce, J.L. (2003) Climatic controls on fire-induced sediment pulses in Yellowstone National Park and Central Idaho: a long-term perspective. *Forest Ecology and Management*, **178,** 89–104.

Muhs, D.R. & Bettis, E.A. (2000) Geochemical variations in Peoria Loess of western Iowa indicate paleowinds of mid-continental North America during last glaciation. *Quaternary Research*, **53,** 49–61.

North American Packrat Midden Database (2004) Accessed: 15 August 2004. http://climchange.cr.usgs.gov/data/midden/.

North American Plant Macrofossil Database (2004) Accessed: 15 August 2004. http://uwadmnweb.uwyo.edu/Botany/NAPMD/.

Noy-Meir, I. (1973) Desert ecosystems: environment and producers. *Annual Review of Ecology and Systematics*, **4,** 25–51.

Pearson, G.A. (1950) Management of ponderosa pine in the Southwest. *US Department of Agriculture Monograph*, **6,** Washington, DC.

Redmond, R.L., Hart, M.M., Winne, J.C., Williams, W.A., Thornton, P.C., Ma, Z., Tobalske, C.M., Thornton, M.M., McLaughlin, K.P., Tady, T.P., Fisher, F.B. & Running, S.W. (1998) *The Montana Gap analysis project: final report*. Montana Cooperative Wildlife Research Unit, The University of Montana, Missoula, MT.

Rehfeldt, G.E. (1990) Genetic differentiation among populations of *Pinus ponderosa* from the upper Colorado River Basin, U.S.A. *Botanical Gazette*, **151,** 125–137.

Rehfeldt, G.E. (1993) Genetic variation in the Ponderosae of the Southwest. *American Journal of Botany*, **80,** 330–343.

Savage, M., Brown, P.M. & Feddema, J. (1996) The role of climate in a pine forest regeneration pulse in the southwestern United States. *Ecoscience*, **3,** 310–318.

Schrupp, D.L., Reiners, W.A., Thompson, T.G., O'Brien, L.E., Kindler, J.A., Wunder, M.B., Lowsky, J.F., Buoy, J.C., Satcowitz, L., Cade, A.L., Stark, J.D., Driese, K.L., Owens, T.W., Russo, S.J. & D'Erchia, F. (2000) *Colorado Gap analysis program: a geographic approach to planning for biological diversity - final report*. USGS Biological Resources Division, Gap Analysis Program and Colorado Division of Wildlife, Denver, CO.

Sorensen, F.C., Mandel, N.L. & Aagaard, J.E. (2001) Role of selection versus historical isolation in racial differentiation of ponderosa pine in southern Oregon: an investigation of alternative hypotheses. *Canadian Journal of Forest Research*, **31,** 1127–1139.

Squillace, A.E. & Silen, R.R. (1962) Racial variation in ponderosa pine. *Forest Science Monograph, 2*, American Society of Foresters, Washington, DC.

Sudworth, G.B. (1908) *Forest trees of the Pacific Slope*. US Department of Agriculture, Government Printing Office, Washington, DC.

Sudworth, G.B. (1917) *The pine trees of the Rocky Mountain region*. United States Department of Agriculture Bulletin, 460, Washington, DC.

Sweeney, M.R., Busacca, A.J., Richardson, C.A., Blinnikov, M. & McDonald, E.V. (2004) Glacial anticyclone recorded in Palouse loess of northwestern United States. *Geology*, **32,** 705–708.

Thompson, R.S., Whitlock, C., Bartlein, P.J., Harrison, S.P. & Spaulding, W.G. (1993) Climatic changes in the Western United States since 18,000 yr B.P. *Global climates since the last glacial maximum* (ed. by H.E. Wright, Jr, J.E. Kutzbach, T.I. Webb, W.F. Ruddiman, F.A. Street-Perrott and P.J. Bartlein), pp. 468–513. University of Minnesota Press, Minneapolis, MN.

Thompson, B.C., Crist, P.J., Prior-Magee, J.S., Deitner, R.A., Garber, D.L. & Hughes, M.A. (1996) *Gap analysis of biological diversity conservation in New Mexico using geographic information systems*. New Mexico Cooperative Fish and Wildlife Research Unit, Las Cruces, NM.

Thompson, R.S., Anderson, K.H. & Bartlein, P.J. (1999) *Atlas of relations between climatic parameters and distributions of important trees and shrubs in North America*. US Geological Survey Professional Paper 1650 A & B, Denver, CO.

USDA-United States Department of Agriculture (1993) *State Soil Geographic Data Base (STATSGO) – data user's guide*. Soil Conservation Service Miscellaneous Publication, 1492, Washington, DC.

Weidman, R.H. (1939) Evidence of racial influences in a 25-year test of ponderosa pine. *Journal of Agriculture Research*, **59,** 855–888.

Weisburg, S. (1980) *Applied linear regression, Appendix 9A.1*. John Wiley and Sons, New York, NY.

Wells, O.O. (1964) Geographical variation in ponderosa pine. *Silvae Genetics*, **13,** 89–103.

BIOSKETCHES

Jodi Norris is a PhD candidate at the University of Wyoming. She is interested in Quaternary and modern interactions between climate and vegetation at regional scales.

Stephen Jackson is a professor of Botany at the University of Wyoming, where he also directs the Program in Ecology. His research focuses on responses of populations and communities to environmental change across a range of temporal and spatial scales.

Julio Betancourt is a senior scientist with the US Geological Survey. He is interested in how climate variability affects physical and biological systems on time-scales from years to millennia. He has employed a battery of morphological, geochemical and genetic analyses of fossil plant and animal material to test long-standing hypotheses about ecophysiological and evolutionary responses to changes in atmospheric CO_2 levels and climate.

Editor: Brett Riddle